

Marten Hofstede

Het zoekboek voor het Web

Zoeken en vinden op internet en World Wide Web

[versie 0.92]

© Marten Hofstede

De Creative Commons Naamsvermelding-NietCommercieel-GelijkDelen Licentie is van toepassing op dit werk. Ga naar <http://creativecommons.org/licenses/by-nc-sa/2.5/nl/> of stuur een brief naar Creative Commons, 559 Nathan Abbott Way, Stanford, Californië 94305, VS om deze licentie te bekijken.

Inhoud

1. Het World Wide Web: waar zoek ik in?	3
1.1 Omvang van het Web	3
1.2 Het Web in domeinen	6
1.3 Actualiteit en veroudering	8
1.4 Hypertekst als bindmiddel	9
1.5 Typen resources	11
1.6 Internet als communicatiemiddel	11
1.7 Het onzichtbare Web	13
2. Van webwijzers tot AskAnExpert diensten: waar zoek ik mee?	15
2.1. Inleiding	15
2.2. Adressen en links	15
2.3. Directories, (verzamelingen van) webgidsen en startpagina's	16
2.4. WebRings	21
2.5. Zoekmachines, springplankzoekmachines en metazoekmachines	21
2.5.1. Metazoekmachines	..26
2.5.2. Springplankzoekmachines	27
2.6. Discussiefora: nieuwsgroepen, mailing lists, forums en e-groups	28
2.6.1. Nieuwsgroepen	28
2.6.2. FAQs	32
2.6.3. Mailing lists	32
2.6.4. Nieuwsbrieven en e-zines	35
2.7. Bloggers en 'experts'	36
2.7.1. Bloggers	37
2.7.2. 'Experts'	38
3. Het hoe van het zoeken	43
3.1. De keuze van het juiste zoekinstrument	43
3.2. Zoeken in directories	46
3.3. Zoeken met zoekmachines en metazoekmachines	48
3.3.1. Booleaanse operatoren	49
3.3.1.1. Nesting	51
3.3.1.2. Nabijheidsoperatoren (proximity operators)	51
3.3.1.3. Phrases	52
3.3.2. Best match en andere niet-Booleaanse zoeksystemen	53
3.3.3. Unitaire operatoren	55
3.3.3.1. Term weighting	57
3.3.4. Goochelen met zoektermen: trunceren, maskeren, word stemming, fonetisch en fuzzy zoeken; regular expressions, ranges	58
3.3.4.1. Google's ~ (tilde)-operator	60
3.3.5. Hoofdletters/kleine letters	60
3.3.6. Alfabetten en bijzondere tekens (diacrieten)	61
3.3.7. Leestekens en andere niet-lettertekens	61
3.3.8. Extra's	..62
3.3.8.1. Heel precies zoeken: Title en URL	..67
3.3.8.2. Meer van hetzelfde: link en anchor, related e.d.	67
3.3.8.3. Een zoekresultaat beperken met speciale criteria	68
3.3.8.4. Sortering	70
3.3.8.5. Shortcuts en rekenfuncties	70
3.3.9. Standaard Search en Advanced Search	70

3.3.9.1. Manipulatie van de URL	71
3.3.10. Het zoekresultaat	73
3.3.11. Customizing	74
3.3.12. Samenvatting	75
4. De afzonderlijke zoekinstrumenten	76
4.1. Directories, (verzamelingen van) webgidsen en startpagina's	77
4.1.1. Algemene, mondiale directories	77
4.1.1.1. Yahoo!	77
4.1.1.2. The Open Directory Project (ODP)	81
4.1.2. Directories Plus	84
4.1.2.1. About.com - The Human Internet	85
4.1.2.2. Communities	88
4.1.3. Als directories georganiseerde verzamelingen van webgidsen	89
4.1.3.1. Usenet FAQ's by Category	89
4.1.3.2. The Argus Clearinghouse en de Internet Public Library	90
4.1.3.3. WWW Virtual Library en Resource Discovery Network	91
4.1.4. De kleinere subject directories	91
4.1.4.1. Librarians' Internet Index	92
4.1.5. Nederlandse directories	94
4.1.5.1. De dekking van Nederlandse directories en startpagina's	94
4.1.5.2. Open Directory Project Nederlands	94
4.1.5.3. Startnederland.nl	94
4.1.6. Web directories van andere landen	96
4.2. Zoekmachines	97
4.2.1. Wat maakt een zoekmachine tot een goede zoekmachine?	98
4.2.1.1. Overlap	99
4.2.2. Google	100
4.2.2.1. In kort bestek	100
4.2.2.2. Basic Search	115
4.2.2.3. Language Tools (Taalhulpmiddelen)	116
4.2.2.4. Preferences (Voorkeuren)	116
4.2.2.5. Advanced Search	116
4.2.2.6. Google's resultatenlijst	119
4.2.2.7. Bespreking Google	123
4.2.2.8. Wat met Google kan dat met Google niet kan: Google Alert, Soople, Findforward, Google Total	128
4.2.2.9. Google's speciale zoekmachines: Google Images, Google Scholar Beta, Google Book Search, Google Labs	130
4.2.3. Yahoo!	152
4.2.3.1. In kort bestek	152
4.2.3.2. Basic Search	164
4.2.3.3. Language Tools (Taalhulpmiddelen)	165
4.2.3.4. Preferences (Voorkeuren)	166
4.2.3.5. Advanced Search	166
4.2.3.6. Yahoo!'s resultatenlijst	169
4.2.3.7. Bespreking Yahoo!	170
4.2.3.8. Y!Q	172
4.2.3.9. My Yahoo!, MyYahoo! Search en MyWeb!	173
4.2.4. MSN Search	177
4.2.4.1. In kort bestek	177
4.2.4.2. Zoekscherm	186

4.2.4.3. Search builder	187
4.2.4.4. Settings	189
4.2.4.5. Het resultatenschermb	190
4.2.4.6. Commandoversie	191
4.2.4.7. Bespreking MSN Search	191
4.2.5. Ask	200
4.2.5.1. In kort bestek	200
4.2.5.2. Basic Search	206
4.2.5.3. Preferencesen Settings	206
4.2.5.4. Advanced Search	207
4.2.5.5. Het resultatenschermb	209
4.2.5.6. Bespreking Ask	210
4.2.6. Andere zoekmachines	213
4.2.6.1. A9	213
4.2.6.2. Exalead	215
4.3. Metazoekmachines	219
4.4. Speciale zoekmachines	221
4.4.1. Nationale zoekmachines: Nederland	221
4.4.2. Nationale en regionale zoekmachines van andere landen	223
4.4.3. Onderwerpsspecifieke zoekmachines	224
4.4.4. Een eigen zoekmachine	231
5. Op zoek naar het Invisible Web	233
5.1. Beleidsmatige beperkingen aan het door zoekmachines geïndexeerde Web	234
5.2. Technische beperkingen	235
5.3. Databases opsporen	236
5.3.1. Speciale instrumenten om databases te vinden	236
5.3.2. Met Google en webgidsen naar databases zoeken	238
5.3.3. Databases zoeken op de plaatsen waar ze te verwachten zijn	240
5.3.4. Databases zoeken via treffers in zoekmachines	245
5.4. Zoeken in databases: een stoomcursus	245
5.4.1. Structuurverschillen	246
5.4.2. Interfaces	247
5.4.3. Records, velden, indexen, zoek sleutels	251
5.4.4. Termen en operatoren	252
5.4.5. Trunceren en maskeren	252
5.4.6. Nabijheidsoperatoren	253
5.4.7. Indexen en thesaurussen	253
5.4.8. 'Explode'-functie in thesaurussen en classificatiesystemen	260
5.4.9. De presentatie van zoekresultaten	262
5.5. Tot slot	263
6. Browser buttons en toolbars	264
6.1. Browser buttons (bookmarklets, favelets)	264
6.2. Toolbars	266
6.2.1. Toolbars speciaal voor Firefox	266
6.2.2. Andere search toolbars	268
7. Bijblijven	273
7.1. Alerts	274
7.2. Tracking services	275
7.3. Discussiefora: nieuwsgroepen, message boards en mailing lists	276

7.4. Weblogs en RSS	277
7.4.1. Blogpulse	278
7.4.2. Technorati	282
7.4.3. IceRocket	283
7.4.4. Feedster	284
7.4.5. Google Blog Search	287
7.4.6. Tot slot	288
8. Het zoekproces	289
8.1. Known-item zoeken	289
8.2. Re-finding	291
8.3. Exploratory	293
8.4. Don't know what you need to know	295
8.5. Visible en Invisible Web	295
Appendix A: Zoekmogelijkheden van drie grote mondiale zoekmachines	296
Appendix B: Verklaring van termen	314

Inleiding

Deze gids 'Zoeken en vinden op het World Wide Web' heeft tweeërlei doel:

- u de basiskennis bij te brengen om effectief en efficiënt op het Web te leren zoeken;
- u kennis te laten maken met de belangrijkste instrumenten die zoeken op het Web mogelijk maken.

Aan deze twee doeleinden is vrijwel het hele boek gewijd. Daarom is niet of nauwelijks aandacht geschonken aan de geschiedenis van het World Wide Web en internet, waarover op het Web en in papieren publicaties een hele berg aan literatuur verschenen is, maar bij voorbeeld wel aan HTML, waarvan een elementaire kennis onontbeerlijk is om goed te kunnen zoeken.

Ik ga ervan uit dat de lezers van dit boek toegang hebben tot een computer met een internet-aansluiting, dat ze wel eens geprobeerd hebben iets op het Web te zoeken: misschien via Ilse, MSN of Startpagina, misschien via Yahoo! of Google en dat ze moeite willen doen om goed te leren zoeken. Als u meermalen de ervaring hebt gehad dat u iets niet kon vinden waarvan u vermoedde dat het er wel was, en als u bereid bent om op het Web te leren zoeken, dan is dit boek voor u.

Goed zoeken, of het nu op het Web is of ergens anders, kan alleen als je weet of kunt beoordelen:

- wat je zoekt
- waar je het moet zoeken
- hoe je het moet zoeken
- of je gevonden hebt wat je zoekt

Aan al die aspecten van zoeken zal ik aandacht besteden, waarbij ik niet alleen zal ingaan op het hoe maar ook op het waarom. Zoeken doe je met verstand, dat geldt voor alle stadia van het zoekproces, van de formulering van de vraag tot aan de beoordeling van het zoekresultaat.

Bij het schrijven van dit boek heb ik erg veel gehad aan resources die op het World Wide Web zelf te vinden waren. In een boek dat over zoeken en vinden op het Web gaat leek het mij een aardige gedachte om in kadertjes wat losse tips te geven of wat dieper op een onderwerp in te gaan. Hier en daar verwijs ik ook naar andere literatuur, vrijwel in alle gevallen op het Web.

Net als de eerste editie van dit zoekboek, die in 2001 bij uitgeverij Elmar verscheen, is dit boek geschreven alsof het in gedrukte vorm zou worden uitgegeven. Dat was ook oorspronkelijk mijn bedoeling. Maar twee overwegingen hebben mij uiteindelijk, in overleg met Elmar, anders doen besluiten. In de eerste plaats heeft het enkele jaren geduurd voordat de eerste uitgave was uitverkocht. Dat betekent dat kopers die het boek in 2003 of 2004 hebben gekocht, voor een deel met verouderde informatie werden geconfronteerd. Het feit dat dit bij een boek over zoeken op het Web onvermijdelijk is, maakt de gedachte niet minder onaangenaam. Een uitgave als PDF-bestand ligt alleen al daarom meer voor de hand omdat een herziene uitgave in deze vorm economischer en gemakkelijker te realiseren is. Hoewel ik niet kan beloven dat er regelmatig updates zullen komen, zal de drempel om die uit te brengen in elk geval een stuk lager zijn dan nu het geval was.

Daarnaast speelt nog een andere overweging: uitgave van het boek als PDF-bestand heeft het grote voordeel dat links direct op het scherm kunnen worden aangeklikt en gevolgd; en wie toch behoefte mocht hebben aan een papieren versie, kan het boek zelf uitprinten.

Nu dit boek niet meer door een uitgever wordt gepubliceerd, rijst de vraag hoe het met het auteursrecht moet. Zoals u op de copyrightpagina kunt zien is dat geregeld via een van de "Creative Commons"-licenties. In dit geval komt die erop neer dat het boek vrij

kan worden gelezen, gedownload en geprint, miets niet voor commerciële doeleinden, dat de tekst ook door anderen kan worden bewerkt, mits onder dezelfde voorwaarden waaronder dit boek ter beschikking wordt gesteld, maar dat ik het op prijs zou stellen als daarbij of bij letterlijke overname van langere gedeelten bronvermelding plaatsvindt. In deze herziene uitgave is ten opzichte van de eerste editie van 2001 heel veel gewijzigd. Aangezien gebleken is dat de eerste editie hier en daar als facultatieve lectuur in opleidingen in gebruik is geweest, heb ik de opzet ervan zoveel mogelijk intact gelaten. Wel doe ik in deze editie het onderscheid tussen zoeken en blijven sterker uitkomen, omdat het laatste tegenwoordig een aparte behandeling vereist. De hoofdstukken 2, 4, 5 en 6 zijn geheel herschreven, 1 en 3 zijn aangepast en geactualiseerd. Een hoofdstuk over blijven is toegevoegd. Bij de vorige editie had ik beter moeten weten dan daarbij een website te willen onderhouden. Bij deze pak ik het wat minder ambitieus aan. Sinds het begin van 2005 heb ik van het vakblad Informatie Professional ruimte gekregen voor een weblog. Daarin zal ik regelmatig verslag doen van interessante nieuwe ontwikkelingen over het onderwerp dat u en mij zo boeit.

Leiden, maart 2006

1. Het World Wide Web: waar zoek ik in?

Dit hoofdstuk geeft een korte schets van wat het Web is, zijn omvang, indeling en organisatie. Het doel is u een idee te geven van de onafzienbare en onvoorstelbare ruimte waarin het zoeken zich afspeelt. Het is niet bedoeld om u af te schrikken, maar om u het gevoel te geven dat u, elke keer dat u het Web op gaat om iets te zoeken, een verre ontdekkingsreis gaat ondernemen.

In dit hoofdstuk zullen al een aantal termen worden gebruikt die veel lezers wel en andere lezers niet vertrouwd zijn. Om praktische redenen worden die niet allemaal ter plekke uitgelegd. Voor de belangrijkste termen wordt de lezer naar de Lijst van termen achterin het boek verwezen.

1.1. De omvang van het Web

Het *World Wide Web* is een belangrijk deel van *internet*, en internet is een mondiaal netwerk van computernetwerken. Dat laatste betekent dat het bij internet om twee zaken draait:

- informatie, zoals die in computerbestanden is opgeslagen, en
- communicatie, in zoverre als die informatie via telefoonlijnen en satellietverbindingen tussen computers ('in netwerken') kan worden uitgewisseld. In die uitwisseling spelen twee soorten computers een rol: de aanbieders, *hosts* of *servers*, en de afnemers van de informatie, *clients* of *terminals*.

Hoewel hosts en servers enerzijds en clients en terminals anderzijds niet hetzelfde zijn, worden de termen hier door elkaar gebruikt. Het gaat ons hier om het onderscheid tussen 'leverende' en 'afnemende' computers.

Internet verbindt allerlei soorten computers met elkaar, waarop heel uiteenlopende besturingssoftware draait. Om die computers toch met elkaar te laten communiceren is een gemeenschappelijke taal nodig die door alle computers in een netwerk wordt gesproken en begrepen. Zo'n taal heet een *protocol*. Op internet is dat protocol (eigenlijk gaat het om een combinatie van verschillende protocollen, met verschillende functies) het zogeheten *TCP/IP* ('Transmission Control Protocol/Internet Protocol'). Dit biedt u als gebruiker van een PC of werkstation (de 'terminal') de mogelijkheid via een *client*programma (bij voorbeeld de browser op uw PC) opdrachten naar een hostcomputer te sturen die door de *server*-software op die computer begrepen en uitgevoerd worden. Daarbij maakt TCP/IP gebruik van zogeheten *IP-adressen*. Alle computers in het netwerk, niet alleen de hosts, maar ook de terminals, hebben zo'n IP-adres, dat is samengesteld uit vier reeksen getallen van elk 1-255, gescheiden door punten: bij voorbeeld 237.16.8.167. Dat geldt ook voor de computer waarop u werkt: de uitwisseling van gegevens tussen uw terminal en een hostcomputer is dus de gegevensoverdracht tussen twee IP-adressen.

Een *browser* is het programma op de PC waarmee u het Web op gaat en dat u pagina's op het Web laat zien. Drie bekende browsers zijn Internet Explorer van Microsoft, Firefox/Mozilla van het Mozilla Project en Opera van Opera Software.

Zoals gezegd vormt het World Wide Web een belangrijk deel van internet. Binnen dit deel wordt tussen host en terminal een speciaal protocol gesproken, het zogeheten 'Hypertext Transfer Protocol' of *HTTP*. Dit protocol regelt de communicatie tussen momenteel (februari 2006) ruim 76 miljoen webservers enerzijds, en tegen de een miljard PC's en werkstations, maar ook mobiele telefoons en andere apparaten anderzijds.

Bron: http://news.netcraft.com/archives/web_server_survey.html

Vraag: waar vind je de betekenis van termen die met internet te maken hebben?
In een openbare bibliotheek zou je daarvoor op zoek gaan naar een internetwoordenboek of -encyclopedie. Zulke naslagwerken heb je ook op het Web zelf. Twee van zulke woordenboeken zijn:

- Webopedia (<http://www.webopedia.com/>)
- Netlingo (<http://www.netlingo.com/>)

Zulke woordenboeken vind je door in een directory zoals [Yahoo!](#) onder Computers & Internet op zoek te gaan naar dictionaries; in de directory van [Google](#): Reference – Dictionaries – By Subject – Computers.

Wat al die clients in het World Wide Web bij al die servers opvragen, zijn merendeels webpagina's. Dit zijn bestandjes zoals die welke u op uw eigen computer opslaat, maar dan niet met bij voorbeeld de *extensie* .doc (voor bestanden die u met de tekstverwerker MS Word maakt), maar met de extensie .htm of .html. Extensies zeggen iets over het *formaat* of type van bestanden. In hun simpelste vorm bestaan .html-documenten uit – betrekkelijk ongestructureerde – tekst, maar deze tekst bevat doorgaans allerlei elementen die daaraan grafische effecten, een zekere structuur, directe verwijzingen (*hyperlinks*) naar andere documenten (op dezelfde of op andere hostcomputers), stilstaande of bewegende beelden, geluid, en zelfs allerlei, soms interactieve, computerprogramma's, kunnen toevoegen. De tekst op deze webpagina's is dan ook geen gewone tekst, maar tekst die is opgemaakt in zogenaamde *HTML*-code. 'HTML' staat voor Hypertext Markup Language. Deze code maakt het Web tot wat het is: een verzameling van samenhangende informatie die over miljarden bestanden op miljoenen computers verspreid is. Op deze *hypertekst* komen we later nog meer uitgebreid terug. Behalve HTML-bestanden kunnen ook bestanden in heel andere formaten tussen servers en clients worden uitgewisseld. Dat gebeurt dan via andere protocollen dan HTTP die onder TCP/IP vallen, bij voorbeeld *FTP* en *Telnet*. Als u bij voorbeeld een .exe of een .zip bestand 'downloadt', dat wil zeggen naar uw eigen computer kopieert, gebruikt u het FTP-protocol. Een enkele uitzondering daargelaten zijn ze voor het onderwerp van dit boek niet van direct belang.

Wat is een hyperlink? Een hyperlink is een stukje tekst of een plaatje in een document op een computer waarin het adres van een bestand verstopt zit. Een hyperlink herkent u doordat een stukje tekst (een of meer woorden) in een document dat uw browser u laat zien, een afwijkende kleur heeft en (meestal) onderstreept is. Als uw muisaanwijzer op die tekst staat, of op een plaatje met een hyperlink, verandert het pijltje in een wijzend handje. Als u op een hyperlink klikt, maakt uw computer verbinding met de computer waarop het geadresseerde bestand staat. Het document wordt opgevraagd en naar het interne geheugen van uw computer gekopieerd en door de browser op uw beeldscherm getoond. Desgewenst kan de kopie daarna op uw eigen computer worden opgeslagen.

Een vraag die veel gesteld wordt is die naar de omvang van het Web en internet. De afgelopen jaren is het steeds moeilijker geworden op die vraag een zinnig antwoord te geven. Daarvoor zou het in de eerste plaats nodig zijn precies te definiëren wat er tot het Web behoort en wat tot de rest van internet, en in de tweede plaats een precies onderscheid te maken tussen het zogeheten 'zichtbare' en het zogeheten 'onzichtbare Web'. Bij dat laatste onderscheid spelen zoekmachines een grote rol: het verschil tussen zichtbaar en onzichtbaar is het verschil tussen wat de grote mondiale zoekmachines wel en niet indexerend. Tot voor een jaar of vijf indexeerden de zoekmachines alleen statische webpagina's die als een HTML-bestand op een server stonden. Samen vormden die pagina's het 'zichtbare Web'. Het 'onzichtbare Web' bestond uit documenten in niet openbaar toegankelijke 'intranetten', documenten die de grote

mondiale zoekmachines niet indexeerden, en – voor het grootste deel – uit documenten of documentbeschrijvingen (bijvoorbeeld catalogustitels) in databases die als zodanig wel op het Web te vinden waren, maar waarvan de inhoud voor de zoekmachines verstopt bleef. Die documenten en documentbeschrijvingen kunnen vanuit een database wel in HTML worden geproduceerd, en behoren als zodanig tot het Web, maar ze bestaan enkel zolang als ze door een client worden opgevraagd, en zijn dus niet statisch, maar dynamisch. En de laatste jaren zijn de grote zoekmachines, Google voorop, steeds meer van dit soort dynamische pagina's gaan indexeren en hebben ze in hun databases opgenomen. Anders gezegd: wat gisteren nog tot het onzichtbare Web behoorde, maakt nu deel uit van het zichtbare Web, en deze ontwikkeling zet zich nog steeds voort. Als u dit in gedachten wilt houden, geef ik u toch wat cijfers. Vijf jaar geleden telde Google, de grootste onder de mondiale webzoekmachines, nog 1,6 miljard pagina's; tegenwoordig zijn dat er ergens tussen de 10 en 20 miljard, inclusief dynamische pagina's. Het aantal statische pagina's op het Web wordt op meer dan 11,5 miljard geschat, voor het aantal documenten op internet kom je schattingen van tot 40 miljard tegen. In het sinds 1996 opgebouwde Internet Archive (waarover later meer) zouden meer dan 40 miljard documenten te vinden zijn, en de omvang van het onzichtbare Web is wel eens geschat op 400 à 550 maal de omvang van het zichtbare Web. Schattingen van de dagelijkse groei van het Web variëren van 3 tot 10 miljoen documenten, afhankelijk van waar de schatting betrekking op heeft: het openbare of het hele Web.



Het Internet Archive is een digitaal archief van het openbare deel van het Web sinds 1996. Het beslaat ca. 40 miljard webpagina's, te zamen meer dan 400 terabyte (400.000 gigabyte) aan opslagruimte, met een maandelijkse groei van meer dan 10 terabyte. Het archief is voor iedereen toegankelijk en men kan er dus (via de zogeheten Wayback Machine) de staat van het Web op een bepaalde datum bekijken.

Het doorsnee webdocument heeft een omvang van 10 à 20 kilobyte, dat is ca. 1000-2000 woorden. Een deel daarvan wordt in beslag genomen door HTML-codes, maar ook als we daarmee rekening houden, zal een doorsnee webpagina nog zo'n 600-1200 woorden tekst bevatten. Dat brengt de totale hoeveelheid doorzoekbare tekst in alle openbaar toegankelijke pagina's op het zichtbare Web (bij een conservatieve schatting van 10 miljard) op de verbijsterende hoeveelheid van zo'n 6-12 triljoen woorden, met een aanwas van – voorzichtig geschat – 1,8 à 3,6 miljard woorden per dag. Als miljarden webpagina's en triljoenen woorden u weinig zeggen. doen boeken en kasten dat misschien wel. Als we alle informatie die thans op het openbare, zichtbare

Web beschikbaar is in een encyclopedie zouden onderbrengen, zou die ongeveer 200.000 maal zo groot zijn als de Encyclopaedia Britannica. Als we die encyclopedie in de kast zouden willen zetten, zouden we daar 500 km kastruimte voor nodig hebben; en elke dag komen daar ruim 3600 delen of 150 m encyclopedie bij.

Zie: How much information? 2000 (<http://www.sims.berkeley.edu/research/projects/how-much-info/internet.html>); How much information? 2003 (<http://www.sims.berkeley.edu/research/projects/how-much-info-2003/internet.htm>); voor de omvang van het Web, zie: The Indexable Web is more than 11.5 billion pages (<http://www.cs.uiowa.edu/~assignori/web-size/>).

Nu we enige voorstelling hebben van de reusachtige omvang van het Web kunnen we de conclusie trekken dat er weinig onderwerpen zullen zijn waarover op het Web niets te vinden is; en dat goed zoeken op het Web niet alleen inhoudt: alles vinden wat je zoekt, maar vooral ook: alles mijden wat je niet zoekt.

1.2. Het Web in domeinen

Op het moment dat u achter uw computer gaat zitten en via uw modem, intranet of draadloos het Web op gaat, strekt de hele wijde wereld van verbindingen en computers zich achter uw beeldscherm uit. Hoe desoriënterend die wereld van miljoenen computers en miljarden computerbestanden aanvankelijk ook mag lijken, helemaal vormeloos en ongestructureerd is die ruimte niet.

Als u in de adresregel van uw browser een adres intypt en vervolgens op Go klikt of Enter geeft, probeert uw computer (via het netwerk dat de toegang verschaft) in de eerste plaats verbinding te maken met een server die een deel van het door u opgegeven adres vertaalt naar een zogeheten IP-adres. Het adres dat u opgeeft heet een *URL*, de afkorting voor Uniform Resource Locator. Deze adressering geldt voor het hele internet, niet alleen voor de servers die HTTP spreken, maar ook bij voorbeeld voor Gopher-, WAIS- en FTP-servers, die wel een IP-protocol, maar geen HTTP met elkaar spreken. Het bestaat in zijn officiële vorm uit het protocol (meestal, maar niet uitsluitend http:), gevolgd door twee slashes, het adres van de host, met evt. een poort, gevolgd door het *pad*: een of meer subdirectories of folders (tussen slashes) en ten slotte een bestandsnaam met een extensie (meestal, maar niet uitsluitend .htm of .html). Het laatste, de bestandsnaam met de extensie kan overigens ook ontbreken. In dat geval eindigt de URL met een slash.

```
http://www.address.edu:1234/path/subdir/file.ext
|
|service|
|
|__ domain __|
|_____ host _____|
|
|port|
|
| file and
|_ resource details _|
```

Bron: <http://www.utoronto.ca/webdocs/HTMLdocs/NewHTML/url.html>

Het belangrijkste deel van de URL is het hostadres zelf, het gedeelte van de URL tussen de dubbele punt en de eerste enkele slash. Dit hostadres bevat de zogeheten *domain name*, die is opgebouwd uit (van achteren naar voren) de zogeheten *top level domain name* (TLD) en een of meer andere elementen ('second level', 'third level' etc. domain) die doorgaans de naam of een andere aanduiding van de eigenaar van de domain name bevatten. Deze adreselementen samen geven in veel gevallen al een redelijk goede indruk van de locatie en het karakter van de gezochte informatie.

De 'top level domain name' is ofwel een tweeletterige aanduiding voor het land waar de computer zich bevindt (een zogeheten 'country code Top Level Domain' of ccTLD, bijvoorbeeld .nl voor Nederland), ofwel een drie- of vierletterige generieke aanduiding van het soort dienstverlener. De al langer bestaande 'generic Top Level Domains' (of gTLD's) zijn:

- .com (internationaal, oorspronkelijk voor bedrijven, nu algemeen)
- .edu (Amerikaans, hoger onderwijs)
- .gov (Amerikaanse overheid)
- .int (internationale organisatie)
- .mil (Amerikaans militair)
- .net (internationaal, oorspronkelijk netwerken, nu algemeen) en
- .org (internationaal, oorspronkelijk [overwegend non-profit] organisaties, nu algemeen).

Nieuwere generieke domeinnamen zijn .aero (internationaal, luchtvaart), .biz (internationaal, businesses, zakelijke dienstverlening), .coop (internationaal, coöperaties), .info (internationaal, voor algemeen gebruik), .museum (internationaal, musea en andere instellingen met een bewaarfunctie), .name (internationaal, personen), .pro (Verenigde Staten, Canada en Groot-Brittannië, accountants, advocaten, artsen). In de loop van 2005 zijn daar nog .travel, .jobs, .mobi, cat en .xxx bijgekomen; eveneens in 2005 is voor het eerst het ccTLD .eu (voor de Europese Unie) officieel in gebruik genomen.

Bij de beoordeling van informatie op grond van een generieke of country code TLD moet u wel rekening houden met twee belangrijke kanttekeningen. Wat de generieke TLD's betreft bestaat er een onderscheid tussen 'restricted' en 'unrestricted' TLD's. De laatste kunnen door jan-en-alleman worden geregistreerd, wat natuurlijk consequenties heeft voor de betrouwbaarheid van de informatie die op webpagina's met een unrestricted TLD geboden wordt. Bekende voorbeelden van misleidende URLs zijn www.whitehouse.org (naast het officiële adres www.whitehouse.gov) en www.gatt.org (naast www.wto.org). Voor de fijnproevers: er wordt ook nog onderscheid gemaakt tussen 'sponsored' en 'unsponsored' TLD's (sTLD's en uTLD's), maar voor de zoeker heeft dat onderscheid bij mijn weten geen praktische consequenties.

De volgende generieke TLD's zijn 'unrestricted': .com, .net, .org, .en, .info. De overige zijn 'restricted' en kunnen dus niet door iedereen worden geregistreerd.

Een tweede kanttekening betreft de 'country code' TLD's. Hoewel het in zijn algemeenheid waar is dat die het land aanduiden waar de computer zich bevindt waarop een website is opgeslagen, zijn er landen die een aardige bron van inkomsten hebben uit de verkoop van hun ccTLD aan buitenlandse belanghebbenden. Bekende voorbeelden zijn Tuvalu (.tv), Tonga (.to), Vanuatu (.vu) en Niue (.nu). Wie dus over een domeinnaam (bijvoorbeeld "[women](http://www.women.tuvalu)") wil beschikken waarvan alle generieke TLD's al zijn uitgegeven en ook die van het eigen land niet meer beschikbaar is, kan naar zo'n ccTLD uitwijken.

In sommige landen worden naast de country code als 'top level domain' ook algemene 'subdomains' gebruikt om domeinbezitters te typeren. Veel gebruikte subdomains zijn .co voor bedrijven en .ac voor wetenschappelijke instellingen, maar ook .edu, .org en .net komen als 'second level domain' voor. Voorbeelden zijn <http://www.kuleuven.ac.be> voor de Katholieke Universiteit Leuven in België, en <http://www.adelaide.edu.au> voor de University of Adelaide in Australië.

Zegt het deel van een URL vóór de eerste enkele slash veel over de organisatie achter de hostcomputer (en dus de informatie daarop) of de plaats waar die computer zich bevindt, ook het gedeelte van een URL na de slash is vaak informatief. Dat geldt natuurlijk in de eerste plaats voor de bestandsnaam aan het einde van de URL, die – hoe

kort ook – dikwijls iets over de inhoud van het bestand zegt. Maar ook het pad (de route van subdirectories) naar het bestand toe kan een waardevolle informatiebron zijn. Zo zit in dat pad vaak de naam van een auteur van een bestand verscholen, aangezien die naam vaak een gemakkelijke manier is de herkomst van een document in een organisatie aan te duiden. Voorbeeld: <http://www.fhk.eur.nl/personal/berkers/research.html>, de URL van een pagina die gewijd is aan het onderzoek van 'Ph.D. candidate' Pauwke Berkers van de Faculteit der Historische en Kunstwetenschappen van de Erasmus Universiteit van Rotterdam. Als bekend is dat iemand een of meer (persoonlijke) pagina's op het Web onderhoudt, is die persoon vaak via zijn of haar naam in de URL op te sporen. Zo'n naam hoeft overigens niet altijd in het pad of de bestandsnaam voor te komen, het aantal websites dat de naam van een persoon in het domain heeft (met .com, .net, .org etc.), is de laatste jaren explosief gestegen.

Denk er bij het intypen van URL's om, dat deze *case-sensitive* zijn, d.w.z. dat het verschil maakt of u hoofdletters of kleine letters gebruikt.

URL's kunnen u niet alleen veel informatie verschaffen over organisaties, afdelingen, onderwerpen, personen en bestanden, de elementen ervan spelen ook een belangrijke rol in de 'ranking' van zoekmachines (zie blz. 54 e.v.) en worden door diezelfde zoekmachines apart geïndexeerd. Het gevolg is dat u een webdocument niet alleen op woorden in de tekst van dat document kunt vinden, maar ook op woorden in het adres. Voor ccTLD's geldt hetzelfde: `site:be` bij voorbeeld beperkt een zoekactie tot websites en -pagina's met de domain name be.

Meer over URLs: URLs (<http://www.hennigweb.com/projects/webclass/urls.html>); een lijst van ccTLD's met subdomains vindt u bij Alldomains.com – Domain name extensions (<http://www.zianewmedia.com/alldomains/alltlds.html>); voor algemene informatie over domain names zie ook: ICANN | FAQs (<http://www.icann.org/faq/>) en ICANN | Registry Listing (<http://www.icann.org/registries/listing.html>)

1.3. Actualiteit en veroudering

Hierboven is al opgemerkt dat het Web een verre van statische informatieverzameling vormt: er komen voortdurend grote hoeveelheden documenten bij en er vallen er ook af. Het gevolg daarvan is een zekere mate van vormeloosheid en ongrijpbaarheid, die ertoe heeft geleid dat de structuur van het Web wel eens met die van een kwal is vergeleken. De documenten die wel langere tijd blijven bestaan zijn evenmin statisch. Als publicatiemedium biedt het Web ongekende mogelijkheden tot het actualiseren van informatie. Het cursusaanbod van universiteiten, aanbiedingen en prijslijsten van bedrijven, tentoonstellingen van musea, zelfs nieuws in de krant – aan de actualiteit van al die informatie mogen op het Web andere eisen worden gesteld dan op papier. Dat geldt ook en speciaal voor die informatie waarnaar documenten extern verwijzen. De doorsnee webpagina van 10 kilobyte bevat 7 externe links, een feit dat zowel een van de grote successen als een van de ergernissen van het Web vormt. Het laatste als van die 7 links er na een jaar twee verdwenen en een verhuisd blijken te zijn. Het gevolg is dan: 'broken' of 'dead' links, links die alleen een foutmelding opleveren.

Zo'n foutmelding hoeft overigens niet altijd einde verhaal te zijn. Servers zijn wel eens uit de lucht, voor kortere of zelfs langere tijd, en bovendien kan het ook voorkomen dat een document op een server verhuisd is. Twee dingen die u kunt doen als u de melding tegenkomt dat ofwel de server ofwel het document niet gevonden wordt als u een (correct!) adres hebt proberen te bereiken, zijn:

- verbinding met de server proberen te maken door pad en bestandsnaam uit het adres te verwijderen, en dan via een zoekfaciliteit of een sitemap (overzicht van het aanbod van de site) alsnog het gezochte bestand proberen te vinden (u gaat er dan van uit dat het bestand op de server zelf verhuisd is.)

- aan de hand van de titel, als u die kent, met een zoekmachine het document zoeken. Zoals we nog zullen zien is van heel veel tamelijk recente documenten nog een kopie beschikbaar bij (onder meer) de zoekmachine Google, en zijn oudere versies van documenten, en documenten die van het Web verdwenen zijn, vaak nog via het Internet Archive te bekijken.

Meer over foutmeldingen: History of 404 (<http://www.plinko.net/404/history.asp>)

Als u naar informatie op het Web zoekt, kan de actualiteit van een gevonden document – in vergelijking met die van andere gevonden, soortgelijke documenten – van doorslaggevende betekenis blijken te zijn.

3. **William of Ockham - Ockham's Razor**

Supplies a broad overview of William's principle of parsimony. Includes a link to further resources.

URL: <http://www2.strw.leidenuniv.nl/~vdmeulen/deeper/OBIG/spoilers.html>

Datum: 1997/01/31 - **Bestandsgrootte:** 2k - **Domein:** www2.strw.leidenuniv.nl

open in nieuw venster | [toon alleen resultaten van deze site](#) | [stuur naar een bekende](#)

Zoekmachines – een soort telefoongidsen van het Web – bevatten de titels van webpagina's, de URL, een korte inhoudsbeschrijving of *snippet* en nog een aantal gegevens. Daaronder kan ook de datum zijn waarop een document op de server is gezet of het laatst is gewijzigd: de zogeheten *Last Modified Date*, in dit geval 31 januari 1997.

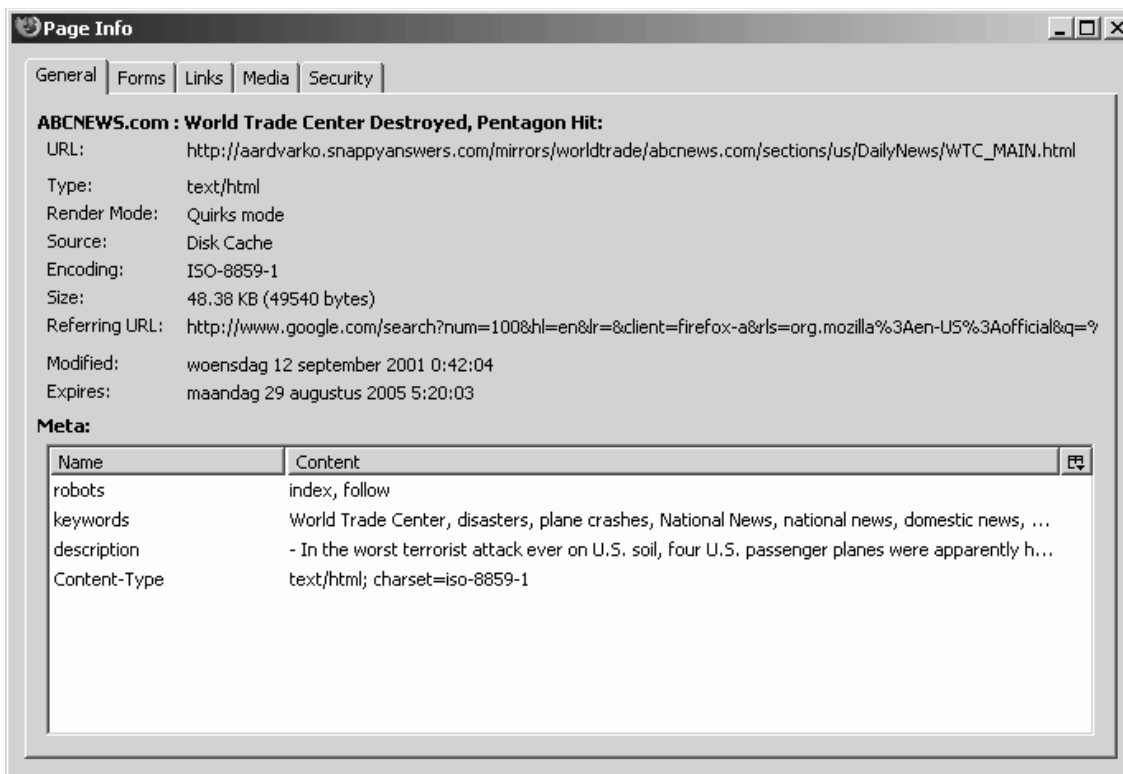
Op goede webpagina's wordt door de auteurs bijgehouden wanneer zij de geboden informatie het laatst hebben gecontroleerd en/of bijgewerkt, zodat dat voor iedere gebruiker zichtbaar is. Helaas gebeurt dat lang niet altijd. Maar daarnaast wordt de datum waarop veranderingen in een document worden aangebracht bij veel hosts ook automatisch bijgehouden: de 'Last Modified'-datum. Het browserprogramma Firefox en veel zoekmachines kunnen die datum tonen, en bij de laatste kan men er ook gebruik van maken bij het zoeken (zie volgende bladzijde).

Zie ook: 'What's in a date?', blz. 128.

1.4. **Hypertekst als bindmiddel**

Wat maakt het World Wide Web tot het wereldwijde web? Dat zijn niet de meer dan 10 miljard webpagina's die het Web omvat, maar de meer dan 70 miljard hyperlinks die deze miljarden documenten met elkaar verbinden. Net zoals de mens (in het algemeen) een sociaal wezen is, met een kring van familieleden, vrienden en kennissen, zo is ook een webpagina (in het algemeen) een sociaal document, waaromheen zich andere documenten groeperen. Ze kunnen dat doen omdat overal op aarde mensen als nijvere termieten hyperlinks tussen documenten aanbrengen. Daardoor krijgt het Web een soort sociale organisatie, met een kern van documenten die onderling naar elkaar linken, en verschillende buitengebieden van documenten waarnaar vanuit die kern niet gelinkt wordt en die zelf wel of niet naar documenten in de kern linken. Volgens een studie die tijdens een WWW-conferentie in 2000 gepresenteerd werd, bedraagt de omvang van de kern 39% van het Web, verwijst 21% van de webpagina's wel naar die kern, maar trekt zelf geen links, trekt 19% wel links vanuit de kern, maar verwijst daar niet heen terug en staat 14% helemaal op zichzelf.

Zie: Andrei Broder e.a., *Graph structure in the web* (<http://www9.org/w9cdrom/160/160.html>)



Met Firefox als browser kunt u (via de menukeuzes Tools – Page Info, of via de rechter muisknop) serverinformatie over het document dat u op het scherm hebt staan opvragen. Onder tab 'General' vindt u onder meer de Last Modified-datum.

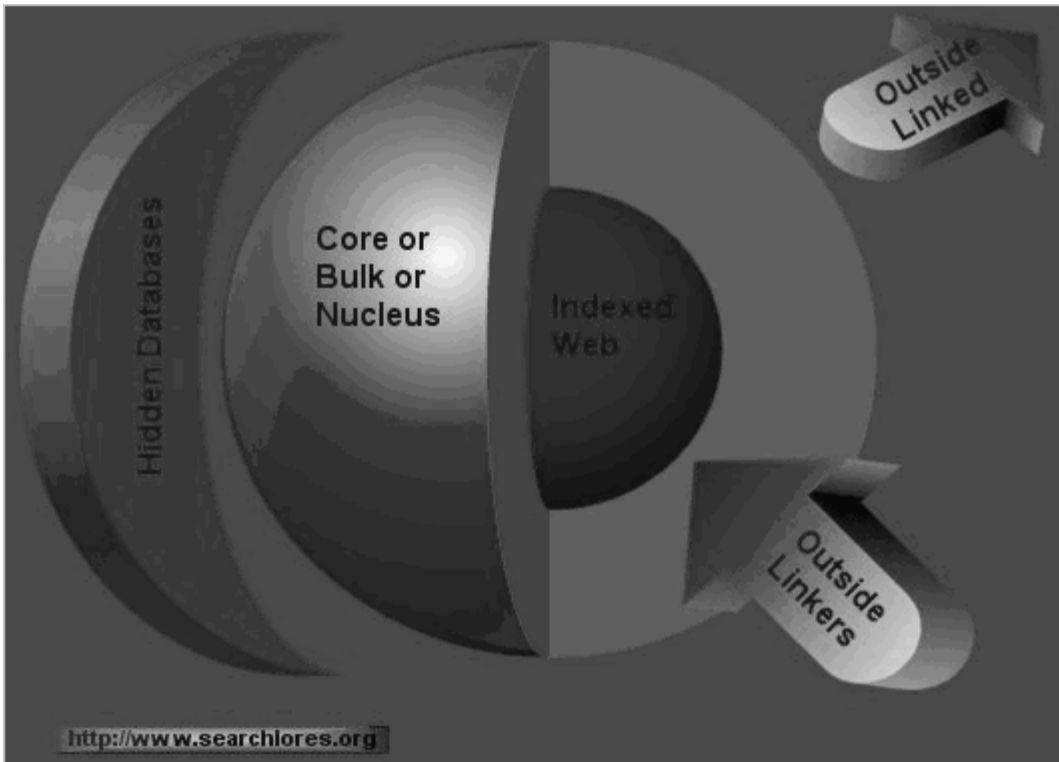
Links stellen ons in staat om over het Web te 'surfen' en door het volgen van associaties de informatie te vinden die we zoeken. Daarvoor is het niet nodig miljarden links te volgen, integendeel, het gemiddelde aantal links dat twee willekeurige pagina's van elkaar scheidt bedraagt slechts 19; net zoals de *six degrees of separation*-theorie (ook wel het *Small World-fenomeen* genoemd) stelt dat de afstand tussen twee mensen die elkaar niet kennen niet groter is dan ca. zes personen die zij en die elkaar wel kennen.

Zie: R. Albert e.a., Diameter of the World Wide Web

([http://www.nd.edu/~networks/Publication%20Categories/03%20Journal%20Articles/Computer/Diameter_Nature%20401.%20130-131%20\(1999\).pdf](http://www.nd.edu/~networks/Publication%20Categories/03%20Journal%20Articles/Computer/Diameter_Nature%20401.%20130-131%20(1999).pdf)).

Links maken het ons mogelijk het Web op een effectieve manier te doorzoeken. Er zijn telefoonboeken (zoekmachines) en 'Gouden Gidsen' (directories en startpagina's) van het Web waarmee we kunnen zoeken. Maar zonder links zouden er geen zoekmachines zijn en ook als we zoekmachines en directories gebruiken, zijn het nog altijd vaak links, verwijzingen dus, die ons bij precies die informatie brengen waarnaar we op zoek zijn.

De belangrijkste zoekinstrumenten op het Web zijn directories of webgidsen, startpagina's en zoekmachines. *Directories* en *startpagina's* zijn de Gouden Gidsen van het Web. Ze worden door mensen samengesteld. Redacteuren bekijken (al dan niet door de auteurs aangemelde) webpagina's en brengen die met een korte beschrijving onder in een onderwerpscatalogus met een boomstructuur (van algemeen naar bijzonder). De gebruiker van een directory of startpagina kan de links met de beschrijvingen dan terugvinden door in de juiste rubriek te kijken of, bij specifiekere onderwerpen, de boomstructuur te volgen. *Zoekmachines* zijn computerprogramma's die de bestanden op het Web (geautomatiseerd) catalogiseren. Ze halen een kopie op van elke webpagina, indexerend die en vormen zo een database waarin u en ik kunnen zoeken.



De belangrijkste segmenten van het Web. In het midden een kern van documenten die naar elkaar verwijzen. Daarvan is een deel door zoekmachines geïndexeerd (zie blz. 233 e.v.). 'Outside linked' zijn documenten, waarnaar vanuit de kern verwezen wordt maar die zelf geen hyperlinks naar de kern bevatten, 'Outside linkers' zijn documenten die naar documenten in de kern verwijzen maar die zelf geen verwijzingen vanuit de kern 'trekken'. Een apart segment vormt het zogeheten 'invisible' of 'hidden' Web (de 'Hidden Databases').

Bron: How to search the Web (<http://searchlores.org/paris/paris1.htm>). Met dank aan Fjalar Ravia.

1.5. Typen resources

Zoals eerder opgemerkt omvat het Web webpagina's, pagina's dus die zijn opgemaakt in HTML. Deze pagina's zelf kunnen echter zeer uiteenlopend van karakter zijn en toegang geven tot de meest uiteenlopende *resources*. Onder resources worden in het algemeen alle informatiebronnen verstaan waartoe het internet toegang biedt. Om een idee te geven van die verscheidenheid noem ik er alvast enkele, die bij de directory van Yahoo! worden onderscheiden: chats and forums, conferences, education, employment, events, institutes, journals, mailing lists, news and media, organizations, Web directories. U kunt u voorstellen dat zulke typen resources een apart aspect vormen bij heel veel onderwerpen en dat het voor een zoeker een groot voordeel is zulke resources onderscheiden te zien. Hoewel maar weinig zoekinstrumenten zoeken op het Web naar typen resources mogelijk maken, zal ik er daar waar dat nodig is op attenderen.

1.6. Internet als communicatiemiddel

Het Web, en meer in het algemeen internet, omvat niet alleen een wereld van informatie in de vorm van bestanden die op computers staan, maar ook een wereld van communicatie in de vorm van discussiefora: nieuwsgroepen, mailing lists, forums, e-groups enz. En waar communicatie plaatsvindt, zijn mensen aanwezig die communiceren, en die dat misschien ook met u willen doen.

Hiervan zijn *nieuwsgroepen* een soort elektronische prikborden, naar onderwerp georganiseerd, waarop iedereen een bericht of een vraag kan 'posten' (vastprikken), of op andere berichten kan reageren. Een bericht of vraag en de reacties die daarop komen

(plus de reacties op reacties) heten samen een *thread*. Je kunt je op nieuwsgroepen abonneren, maar dat betekent alleen dat je, als je je newsgroup reader opent (meestal een extra voorziening van je browser, die vanzelf geactiveerd wordt als je contact maakt met een newsserver door in de adresregel van je browser `news://` in te typen), de laatste berichten van alleen die newsgroups waarop je geabonneerd hebt, direct kunt lezen. Berichten van andere nieuwsgroepen kun je ook lezen, zelfs zonder geabonneerd te zijn, maar dan moet je de betreffende nieuwsgroep eerst in een lijst opzoeken en vervolgens openen. Dat kan overigens alleen als je provider op de eigen server een toegang tot de betreffende nieuwsgroep verschaft. Sommige providers doen dat uit ruimte- of andere overwegingen maar tot een beperkt gedeelte. Sinds december 2001 zijn de archieven van nieuwsgroepen ook te raadplegen via een van de databases van de zoekmachine Google, Google Groups. Abonneren op en berichten versturen naar nieuwsgroepen kan daar inmiddels ook worden gedaan.

Louis Couperus

Ton van Til Jan 27 2001, 6:13 am [show options](#)

Ik zit in het eindexamenjaar VWO en ik moet een scriptie schrijven over Louis **Couperus**.
Het thema wat ik wil behandelen is: "De invloed van Nederlands-Indie en Den Haag op het werk van **Couperus**"
Heeft iemand informatie over dit thema of weet iemand goede, uitgebreide sites over **Couperus**?

Bij voorbaat dank,

Sander van Til

Gérard Oomens Jan 27 2001, 6:33 am [show options](#)

icu02...@euronet.nl (Ton van Til) wrote in
<3a72d52e.23531...@news.euro.net>:

>Ik zit in het eindexamenjaar VWO en ik moet een scriptie schrijven
>over Louis **Couperus**.
>Het thema wat ik wil behandelen is: "De invloed van Nederlands-Indie
>en Den Haag op het werk van **Couperus**"
>Heeft iemand informatie over dit thema of weet iemand goede,
>uitgebreide sites over **Couperus**?

>Bij voorbaat dank,

>Sander van Til

<http://www.louiscouperus.nl/> ;dit lijkt mij een goed begin!

--

Vingeroefeningen
www.xs4all.nl/~gomens

Een voorbeeld van enkele berichten in het archief van de nieuwsgroep nl.kunst.literatuur.

Mailing lists zijn anders georganiseerd. Het zijn geen prikborden, maar georganiseerde gespreksgroepen, met aan het hoofd een 'moderator' (lijstbeheerder), die soms ook toeziet op de kwaliteit en toepasselijkheid van de ingestuurde berichten. (Ook nieuwsgroepen kunnen trouwens zo'n moderator hebben.) Hier *moet* je je op abonneren om de berichten überhaupt te kunnen zien, je ontvangt die namelijk in je mailbox. Mailing lists, in Nederland ook wel *rondzendlijsten* genoemd, zijn niet altijd interactief, d.w.z. lezers kunnen niet altijd reageren. De term 'mailing lists' omvat zowel de interactieve lijsten (*discussielijsten*) als aankondigingen, bij voorbeeld in de vorm van newsletters, e-zines e.d., die meer eenrichtingsverkeer vormen. Verder zijn lang niet alle mailing lists openbaar, het overgrote merendeel heeft een besloten karakter. Forums en e-groups zijn alternatieve benamingen voor soortgelijke diensten als newsgroups en mailing lists.

Groepen en lijsten zijn in de eerste plaats een middel om op de hoogte te blijven. Maar van een deel ervan zijn en worden de berichten ook gearchiveerd, wat betekent dat in die archieven kan worden gezocht. Dat betekent een niet onaanzienlijke uitbreiding van de omvang van de ruimte die we te doorzoeken hebben: alleen al de door de zoekmachine Google doorzoekbaar gemaakte archieven van de nieuwsgroepen sinds 1981 tellen niet minder dan een miljard berichten; en hoewel de archieven van mailing lists niet centraal doorzoekbaar zijn, hebben ook veel mailing lists wel zo'n – openbaar – archief. Daarbij komt dat de informatie die in de nieuwsgroepen en discussielijsten omgaat vaak een speciaal karakter heeft. Berichten die in nieuwsgroepen en discussielijsten worden uitgewisseld, hebben vaak de vorm van aankondigingen, vragen, antwoorden en reacties, en de informatie die ze te bieden hebben, komt daarmee overeen. Zo kan in de archieven van de nieuwsgroepen waar over wiskunde wordt gecommuniceerd de hele discussie in de wiskundige wereld over de oplossing van het laatste theorema van Fermat door Andrew Wiles (23 juni 1993) worden gevolgd. De kwaliteit van de informatie hangt natuurlijk sterk af van de expertise van de deelnemers aan de groepen en lijsten. Die varieert van pover op de talrijke virtuele babbelboxen tot zeer deskundig in de echt gespecialiseerde lijsten en groepen. Zoeken op internet betekent ook zoeken naar groepen en lijsten. Via directories als Yahoo! en zoekmachines als Google zijn zulke groepen en lijsten vaak wel te vinden, maar ze hebben ook hun eigen zoekinstrumenten. In het volgende hoofdstuk zullen we zien welke dat zijn.

Ten slotte bestaat het Web natuurlijk ook nog eens uit tientallen miljoenen nijvere mensen die, ieder vanuit hun eigen ervaring en kennis, buiten groepen en lijsten om aan de verdere uitbouw van dit merkwaardige monument van menselijke activiteit werken. Mensen vinden op het Web die wellicht het antwoord op uw vragen weten, is een van de vaardigheden die dit boek u wil bijbrengen.

1.7. Het onzichtbare Web

Als u op het Web aan het zoeken bent, kunt u daarop websites van bibliotheken of andere organisaties tegenkomen die hun catalogi en databases openbaar toegankelijk maken. Voorbeelden zijn de Franse [Bibliothèque Nationale](#) met een catalogus van meer dan 15 miljoen titelbeschrijvingen van boeken, en de [Nederlandse overheid](#) met een verzameling van Nederlandse wetten, die fulltext beschikbaar zijn maar niet geïndexeerd door de grote mondiale zoekmachines. De inhoud van zulke databases vormt een belangrijk deel van wat wel genoemd wordt het *Invisible Web*, dat wil zeggen een deel van het Web dat er wel is, en waartoe men ook wel toegang kan krijgen, maar waarvan de inhoud door de normale instrumenten om het Web te doorzoeken niet ontsloten wordt. De inhoud van zulke databases zit verscholen achter een webpagina die er toegang toe geeft, maar die de wereld achter die toegang niet via links maar via een

zoekmogelijkheid in een database ontsluit. Webpagina's die worden gegenereerd als u in zo'n database aan het zoeken bent, worden doorgaans niet door zoekinstrumenten 'gezien', en zijn dus onzichtbaar.

De omvang van het Invisible Web wordt geschat op enige honderden malen die van het wel zichtbare Web. In zekere zin is het dus maar gelukkig dat zoekinstrumenten deze reusachtige hoeveelheid informatie niet ontsluiten. In plaats daarvan moeten wijzelf proberen die databases op te sporen en erin te leren zoeken. Gelukkig zijn er op het Web ook plaatsen waar mensen die taak van ons overnemen en ons naar de juiste databases verwijzen.

Overigens zijn de grenzen tussen het zichtbare en het onzichtbare Web niet meer zo scherp gedefinieerd als ze een jaar of vijf geleden nog waren. Dat komt doordat grote mondiale zoekmachines als Google steeds meer in databases doordringen en de inhoud daarvan ontsluiten. Zo ontsluiten de zoekmachines van [Google](#) en [Yahoo!](#) tegenwoordig ook miljoenen titelbeschrijvingen van boeken uit WorldCat, een internationale bibliotheekcatalogus van meer dan 40 miljoen titelbeschrijvingen, en is PubMed, een grote artikelendatabase op het gebied van de geneeskunde, tegenwoordig voor een deel ook met de nieuwe wetenschappelijke zoekmachine [Google Scholar](#) (zie blz. 132 e.v.) doorzoekbaar. De internetboekhandel Amazon heeft al meer dan 120.000 boeken van ruim 200 uitgeverij via zijn zoekmachine [A9](#) fulltext doorzoekbaar gemaakt, en ook [Google](#) timmert aan de weg met een project om tezamen met uitgeverij en een aantal belangrijke Amerikaanse bibliotheken 15 miljoen boeken te digitaliseren. Ook de concurrentie, verenigd in de [Open Content Alliance \(OCA\)](#), zit niet stil.

Tot slot van dit hoofdstuk is het goed erop te wijzen dat we het tot nu toe alleen nog maar hebben gehad over het *openbare* deel van het Web. Op de computers die in dit reusachtige netwerk van netwerken zijn opgenomen staan nog terabytes (1 terabyte = 10^{12} byte of 1000 gigabyte) aan informatie die alleen voor bevoegde personen toegankelijk is. Ontelbare intranetten en duizenden elektronische tijdschriften waartoe alleen abonnees toegang krijgen, vormen daarvan slechts twee voorbeelden.

Over Internet en WWW is natuurlijk veel meer te vertellen dan in dit boek wordt gedaan. Wie informatie over het onderwerp op een ordelijke manier gepresenteerd wil raadplegen, zij verwezen naar The Living Internet (<http://livinginternet.com/>).

2. Van Webwijzers tot AskAnExpert-diensten: waar zoek ik mee?

Het Web biedt een heel scala aan mogelijkheden om informatie te vinden. In dit hoofdstuk worden die onderscheiden en beschreven, waarbij natuurlijk ook aandacht wordt besteed aan de vraag wanneer u gebruik moet maken van welke mogelijkheid of mogelijkheden. Lees dit hoofdstuk als een terreinverkenning, een eerste kennismaking met de bekendste zoekinstrumenten – directories, startpagina's en zoekmachines – en een inleiding tot een aantal wat minder gebruikte mogelijkheden, die niet tot het centrale thema van dit boek behoren.

2.1. Inleiding

Om met succes op het Web te kunnen zoeken, is het nodig een globaal idee te hebben van de diverse mogelijkheden die het Web biedt om informatie te vinden. Veel mensen komen in dat opzicht nauwelijks verder dan de mogelijkheden die de homepage van hun internetprovider of hun browser hun biedt. Maar bij elke vraag hoort een zoekinstrument dat het meest geschikt is om op die vraag een antwoord te geven. En als u meer mogelijkheden kent, zult u die misschien ook leren gebruiken.

De populairste manieren van zoeken op het Web zijn het volgen van links, het gebruik van *directories* en dat van *zoekmachines* (search engines). Directories en zoekmachines samen worden ook wel aangeduid als *zoekinstrumenten* (search tools). Naast deze manieren van zoeken bieden ook de archieven van nieuwsgroepen en discussielijsten in sommige gevallen geschikte mogelijkheden om antwoorden op vragen te vinden. Vragen kun je ook stellen. In de eerste plaats natuurlijk bij nieuwsgroepen en rondzendlijsten, maar daarnaast ook bij *experts*, personen op het Web die bereid zijn vragen te beantwoorden (of ze expert zijn in de zin waarin het woord in het normale spraakgebruik wordt gebezigd, kunnen we meestal niet beoordelen.). Zoeken doet u doorgaans vanuit een situatie waarin u een informatietekort hebt of een informatieachterstand hebt opgelopen. Maar het Web is niet alleen maar een grote vergaarbak van bestaande informatie, maar ook een nieuwsmedium dat u kunt gebruiken om bij te blijven en van de kennis en inzichten van anderen te profiteren. Rondzendlijsten, nieuwsgroepen en fora, waar u vragen kunt stellen en waarvan u de archieven kunt doorzoeken, zijn voor dit doel natuurlijk de meest voor de hand liggende bronnen, maar het is ook mogelijk u door *bloggers* op de hoogte te laten houden van wat er binnen uw interessegebied allemaal te koop is. Bloggers zijn mensen die een *weblog*, een soort dagboek op het web, bijhouden en publiceren. Hun functie is drieërlei: ze verspreiden nieuws, geven meningen en oordelen en signaleren trends. Ten slotte zijn er diensten die u een zoekprofiel laten opstellen en u dan regelmatig op de hoogte houden van nieuw materiaal op het Web dat aan dat zoekprofiel voldoet, en attenderingsdiensten die u van veranderingen op webpagina's op de hoogte houden.

2.2. Adressen en links

Voor wie weet waar hij op het Web moet zijn is het *adres* (de URL) van de betreffende webpagina natuurlijk de gemakkelijkste manier om die pagina te vinden. URLs zijn weliswaar moeilijk te onthouden, maar ze hebben wel een redelijk overzichtelijke structuur, en bovendien, als je weet waar je moet zijn, ben je daar waarschijnlijk al eens eerder geweest en kun je het adres in je bookmarks of favorieten hebben gezet.

De allersimpelste manier om de website van een bedrijf of organisatie te vinden is om de naam daarvan in de adresregel van uw browser te typen. Zowel Firefox als Internet Explorer schakelt dan een zoekmachine in en komt met een resultaat alsof u die zoekmachine zelf gebruikt had. Firefox gebruikt daarbij de zoekmachine Google en toont daarvan meteen de eerste treffer, Internet Explorer gebruikt de zoekmachine MSN Search en laat het hele zoekresultaat zien. De mogelijkheden van Firefox gaan zelfs nog wat verder: je kunt via de adresregel met de shortcuts van Google (zie blz. 70, 110)

zoeken. Met *quote PHG* vraagt u de (Amerikaanse) beurskoers van Philips op, met *Chicago movies / restaurants / driving directions / shopping / concerts / careers / real estate / news / weather / sports* kunt u via de adresregel van Firefox informatie over films / restaurants / 'de weg' / winkelen / concerten / banen / onroerend goed / nieuws / het weer / sportnieuws in Chicago opvragen. Sommige van deze 'shortcuts werken zelfs internationaal, bijvoorbeeld: *weather Prague*.

Links (in feite adressen in een document waarmee door middel van een muisklik een directe telecommunicatieverbinding tot stand wordt gebracht) zijn voor het zoeken voortreffelijk te gebruiken als men ze als bookmark heeft opgeslagen, ze per e-mail toegestuurd krijgt of als ze als relevante verwijzing in een ander webdocument voorkomen. In de praktijk van het zoeken komt men vaak op drie manieren op een al bekend webadres terecht:

- doordat men de URL heeft onthouden;
- doordat men de route naar de betreffende URL kent door het gebruik van een ander adres (met andere woorden URL en titel van het gezochte document zijn niet precies bekend, maar de route erheen is door langdurig gebruik 'ingesleten');
- de titel van het gezochte document is bekend, maar de URL moet met een zoekmachine worden opgezocht.

Maar daarnaast is het volgen van links – het zogeheten 'websurfen' – ook een geliefd tijdverdrijf, dat de ontdekkingslust van veel mensen prikkelt. Hierbij begint men met een al bekend document en volgt via de hyperlinks in dat document en volgende documenten een route door het Web.

2.3. Directories, (verzamelingen van) webgidsen en startpagina's

Directories vertegenwoordigen een poging een representatief gedeelte van het Web via een *thesaurus* of een *classificatiesysteem* – een zoekboom – op onderwerp vindbaar te maken. Ze doen dat door de webpagina's die ze in hun 'Gouden Gids' opnemen, onder te brengen in een systeem van rubrieken (of categorieën) en subrubrieken (subcategorieën). Moderne thesaurussen en classificatiesystemen stammen uit de wereld van bibliotheek en documentatie. Aan beide ligt de gedachte ten grondslag dat de wereld beschreven kan worden in een (weliswaar voortdurend veranderend) hiërarchisch systeem dat van algemeen naar bijzonder gaat en waarin elk onderwerp een hogere of diepere plaats krijgt toebedeeld. Het (praktische) verschil tussen beide is dat een classificatiesysteem alles overkoepelend is en tot in alle uithoeken uitgewerkt, terwijl een thesaurus lossier georganiseerd is en uit clusters van begrippen bestaat. Het merendeel van de onderwerpen (fiets) in een thesaurus heeft een 'broader term' (tweewieler), een of meer 'related terms' (motorfiets, bromfiets) en een of meer 'narrower terms' (bakfiets, mountain bike). In een classificatiesysteem wordt ditzelfde vaak uitgedrukt in een notatie:

3.6. tweewieler

3.6.1 fiets

3.6.1.1 vouwfiets

3.6.1.2 bakfiets

3.6.1.3 mountainbike

3.6.2. motorfiets

3.6.3. bromfiets

Hoewel er tal van pogingen zijn en worden ondernomen om een representatief gedeelte van het Web in een van de bestaande thesaurussen of classificatiesystemen (bij voorbeeld Dewey, UDC, Library of Congress Subject Headings) onder te brengen, zijn de populairste directories – Yahoo!, Open Directory Project – gebaseerd op een thesaurus van eigen ontwerp.

Voor een overzicht van pogingen om het Web in een bestaand classificatiesysteem te categoriseren, zie: Beyond Bookmarks: Schemes for Organizing the Web (<http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm>).

Directories bieden hun gebruikers over heel veel onderwerpen een overzichtelijke hoeveelheid websites die niet alleen netjes ingedeeld zijn, maar ook voorzien van een korte beschrijving, waarin extra informatie wordt gegeven over het karakter van de webpagina of website.

De termen *webpagina* en *website* worden in dit boek los door elkaar gebruikt. Eigenlijk is een website een samenhangende verzameling van webpagina's, en is een webpagina wat u ziet totdat u naar een volgend adres doorsurft. Maar vaak is een webpagina de toegangspoort tot een website, en zijn websites weer onderdeel van grotere websites. Een van de verschillen tussen directories en zoekmachines is dat directories eerder websites ontsluiten, terwijl zoekmachines over het algemeen webpagina's verzamelen en indexeren.

Directories als Yahoo! zijn commerciële ondernemingen die het moeten hebben van reclame-inkomsten. De in de directory opgenomen websites worden door redacteurs geselecteerd, op basis van eigen onderzoek en/of van aanmelding door auteurs. Die twee omstandigheden bepalen de beperkingen van dit soort zoekinstrumenten. Als het al nodig en technisch mogelijk zou zijn de aanwas van ettelijke miljoenen webpagina's per dag in de directories onder te brengen, zou dat economisch niet haalbaar zijn. Zelfs het Open Directory Project, dat met enige tienduizenden (en een nog steeds groeiend aantal) vrijwillige redacteurs werkt, groeit met niet meer dan ca. 1 miljoen pagina's per jaar. Een goede directory biedt een representatief overzicht van websites in alle rubrieken. Per rubriek is dat aantal niet te groot, anders wordt zo'n rubriek onoverzichtelijk. Hetzelfde geldt voor het aantal rubrieken op hetzelfde niveau in de 'boom'. Een en ander heeft tot gevolg dat een grote directory met enkele miljoenen webpagina's heel wat rubrieken moet tellen en ook een tamelijk uitgewerkte structuur moet hebben. De grootste directory, het Open Directory Project, telt momenteel (augustus 2005) ruim 590.000 rubrieken en gaat op sommige plekken tot wel negen niveaus diep. Overigens zitten nogal wat rubrieken op verschillende plaatsen in de boom (bij voorbeeld History of Science zit onder Science, maar ook onder History).

Een extraatje van directories is nog dat ze sommige sites een kwaliteitsaanduiding meegeven, die ze onderscheidt van de andere opgenomen sites.

Directories zijn populair, en terecht. Als je niet precies weet, of niet precies kunt formuleren, waarnaar je op zoek bent en waar je het moet zoeken, kunnen ze je daarbij helpen. Een groot voordeel van directories boven zoekmachines is dat je er taal- en termonafhankelijk in kunt zoeken: je hoeft niet precies te weten hoe iets in een document heet (of het bij voorbeeld ouderdomsdementie of Alzheimer wordt genoemd) en in welke taal het document is geschreven om dat document toch in de juiste categorie van een directory te kunnen vinden. Hoewel ze maar een klein deel van het Web dekken, profiteren ze natuurlijk zoals alle zoekfaciliteiten op het Web van het bestaan van hyperlinks: zelfs al vind je er niet precies in wat je zoekt, de links in wat je wel vindt kunnen je hopelijk verder helpen. Ook kunnen, zoals we nog zullen zien, de pagina's die je met een directory vindt, het uitgangspunt vormen voor een verdere zoekactie met een zoekmachine (zie blz. 67 e.v.). Door hun boomstructuur dwingen directories tot een ordelijke manier van denken over het onderwerp waarover u vragen heeft; en als werken via de boomstructuur niet lukt kunt u nog altijd gebruik maken van de toegevoegde zoekmachine die de hele directory, of een gedeelte ervan, doorzoekt.



[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#) [Local](#) [more »](#)

Search Directory

[Preferences](#)
[Directory Help](#)

The web organized by topic into categories.

[Arts](#)

[Movies](#), [Music](#), [Television](#),...

[Home](#)

[Consumers](#), [Homeowners](#), [Family](#),...

[Regional](#)

[Asia](#), [Europe](#), [North America](#),...

[Business](#)

[Industries](#), [Finance](#), [Jobs](#),...

[Kids and Teens](#)

[Computers](#), [Entertainment](#), [School](#),...

[Science](#)

[Biology](#), [Psychology](#), [Physics](#),...

[Computers](#)

[Hardware](#), [Internet](#), [Software](#),...

[News](#)

[Media](#), [Newspapers](#), [Current Events](#),...

[Shopping](#)

[Autos](#), [Clothing](#), [Gifts](#),...

[Games](#)

[Board](#), [Roleplaying](#), [Video](#),...

[Recreation](#)

[Food](#), [Outdoors](#), [Travel](#),...

[Society](#)

[Issues](#), [People](#), [Religion](#),...

[Health](#)

[Alternative](#), [Fitness](#), [Medicine](#),...

[Reference](#)

[Education](#), [Libraries](#), [Maps](#),...

[Sports](#)

[Basketball](#), [Football](#), [Soccer](#),...

[World](#)

[Deutsch](#), [Español](#), [Français](#), [Italiano](#), [Japanese](#), [Korean](#), [Nederlands](#), [Polska](#), [Svenska](#), ...

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2006 Google

Help build the largest human-edited directory on the web.

[Submit a Site](#) - [Open Directory Project](#) - [Become an Editor](#)

Voorbeeld van een directory, in dit geval die van Google (gebaseerd op het Open Directory Project). Op deze pagina ziet u de hoofdrubrieken, met daaronder (in een kleinere letter) een aantal van de subrubrieken die onder elk van die hoofdrubrieken vallen. Boven de hoofdrubrieken ziet u een zoekregel waarin zoektermen kunnen worden getypt. Zoeken in de directory leidt u direct naar de websites in de directory waar de door u opgegeven zoekterm in de titel of de beschrijving van de site in de directory, of – in de versie van Google – in de tekst van webpagina zelf, voorkomt. Bij elke site staat verder vermeld in welke rubriek die gevonden is.

Overigens gebruiken sommige van de algemene, mondiale directories daarvoor een grote mondiale zoekmachine die behalve de directory ook een veel grotere database van webpagina's doorzoekt (zie onder zoekmachines). Daarbij presenteert zo'n zoekmachine dan eerst de in de directory gevonden sites of pagina's en vervolgens die welke hij in de grotere zoekdatabase heeft gevonden. De bekendste voorbeelden van algemene mondiale directories zijn die van Yahoo! en het Open Directory Project (ODP). De laatste, die door vrijwilligers wordt samengesteld en op tal van sites onder verschillende namen beschikbaar is, wordt onder meer door Google gebruikt.

Hoewel de vergelijking van World Wide Web met een boek in het algemeen niet opgaat, biedt een boek, bij voorbeeld een handboek, wat de zoekmogelijkheden betreft wel aardig vergelijkingsmateriaal. Zo'n boek heeft een inhoudsopgave, een register en (meestal in de noten en de literatuurlijst) interne en externe verwijzingen naar andere plaatsen waar informatie over hetzelfde of een verwant onderwerp te vinden is. Bij een handboek gebruikt u elk van die drie zoekmogelijkheden voor verschillende doeleinden. Zo is het ook met het Web: directories zijn als het ware inhoudsopgaven van het Web, zoekmachines zijn registers en links zijn verwijzingen.

Verwant aan directories zijn *verzamelingen van webgidsen*, die eveneens door mensen worden samengesteld. Ze zijn losser georganiseerd dan directories, behandelen bredere onderwerpen met grotere lacunes ertussen dan deze, maar gaan daar vaak vollediger en dieper op in dan de algemene mondiale directories. Elk onderwerp of vak heeft zijn eigen redacteur(en), vaak iemand die het vakgebied kent of er zelfs een specialist op is. Sommige zijn gecombineerd met discussielijsten (fora) waarop de redacteur regelmatig artikelen publiceert om lezers op speciale sites, technieken of ontwikkelingen te attenderen en waarop lezers vragen kunnen stellen en informatie kunnen uitwisselen. Andere zijn verzamelingen van complete overzichten van hele vakgebieden. Enkele voorbeelden hiervan zijn [About.com](#), [The Internet Public Library](#) en de [WWW Virtual Library](#). Als representatief overzicht van het hele Web zijn ze minder bruikbaar dan directories, maar wie er een gids over het onderwerp van zijn interesse vindt, hoeft in veel gevallen niet veel verder te zoeken.

Behalve verzamelingen van webgidsen, waarin men naar gidsen over een onderwerp kan zoeken (door via de zoekboom te gaan of door de interne zoekmachine te gebruiken) en daarna de gids van zijn keuze kan raadplegen, bestaan er natuurlijk ook afzonderlijke onderwerpsgidsen, die een uitgebreid overzicht over de beschikbare resources over een onderwerp bieden. Deze losse onderwerpsgidsen kunnen op verschillende manieren worden opgespoord. Zo bevat Yahoo! een categorie *Web directories* over alle mogelijke onderwerpen, waarnaar via de zoekmachine van Yahoo! gezocht kan worden. Deze Web directories omvatten zowel echte Web directories, dus gespecialiseerde Yahoo!'s in het klein, als onderwerpsgidsen, die resources op het Web over een onderwerp in relatie tot elkaar beschrijven (zoals de onderwerpsgidsen van The WWW Virtual Library, zie blz. 92).

Algemene, mondiale directories kunnen u naar websites leiden over alle mogelijke onderwerpen, van oldtimers tot aan Chinese pentekeningen. Maar het Web is een anglocentrische wereld, en dat weerspiegelt zich ook sterk in dit soort directories. Als u iets over een onderwerp wilt weten dat dicht bij huis – of in elk geval ver van de Verenigde Staten – speelt – ons koninklijk huis, vakanties in België of Duitsland, in Nederland verkrijgbare medicijnen – kunt u beter met een Nederlandse (of andere nationale) directory of startpagina zoeken.

[« Complete Directory Results](#)

TOP 100 RELATED DIRECTORY CATEGORIES

out of 799.

1. [News and Media Web Directories](#)
2. [Entertainment > Web Directories](#)
3. [Regional > Web Directories](#)
4. [Phone Numbers and Addresses > Web Directories](#)
5. [Web Directories of Dictionaries](#)
6. [Newspaper Web Directories](#)
7. [Actors > Web Directories](#)
8. [Education Web Directories](#)
9. [Society and Culture Web Directories](#)
10. [Arts Web Directories](#)
11. [Health > Web Directories](#)
12. [Computers and Internet > Web Directories](#)
13. [College and University Web Directories](#)
14. [Government > Web Directories](#)
15. [Science > Web Directories](#)
16. [Social Science > Web Directories](#)
17. [Real Estate > Web Directories](#)
18. [Art History > Web Directories](#)
19. [Design Arts > Web Directories](#)
20. [Humanities > Web Directories](#)
21. [\(5131\) Arts Institutes > Web Directories](#)
22. [Art Museums, Galleries, and Centers > Web Directories](#)
23. [Performing Arts > Web Directories](#)

Yahoo! is een algemene directory, die over alle mogelijke onderwerpen wel een of meer sites te bieden heeft. Maar het is tegelijk (mede) een verzameling van directories op allerlei meer gespecialiseerde terreinen, die zo'n terrein verder onderverdelen en een veel uitgebreidere database van sites hebben. Je vindt die door in Yahoo! met de term 'Web directories' te zoeken.

Startpagina's zijn tot dusver nog niet ter sprake gekomen. Ze zijn enigszins te vergelijken met directories in zoverre als webpagina's waarnaar ze verwijzen door redacteuren van zo'n startpagina worden verzameld dan wel door de auteurs bij de startpagina worden aangemeld. Ook delen ze die webpagina's in onderwerpen in, en is er zelfs sprake van een zekere hiërarchische structuur in die onderwerpen. Maar terwijl directories op de homepage alleen tussen de tien en twintig zeer algemene onderwerpsgebieden – Kunst en cultuur, Wetenschap en onderwijs, Computers en Internet etc. – bieden, nog zonder links naar webpagina's, starten startpagina's direct al op de homepage met drie of vier kolommen van blokjes hyperlinks over onderwerpen waarvoor veel belangstelling is. Zoveel mogelijk onderwerpen worden daarbij op de homepage ondergebracht, tot wel 50 of 60 toe. Daartegenover staat een zeer platte structuur: directories gaan soms tot 9 niveaus diep, startpagina's zelden tot meer dan 3. Een ander verschil is dat directories de links vaak vergezeld doen gaan van een korte beschrijving. Startpagina's doen dat niet, alles is erop gericht zoveel mogelijk links over zoveel mogelijk actuele onderwerpen (banen, reizen, sport, radio en TV, bekende Nederlanders enz.) bij elkaar te zetten. Daarbij bieden de grootste en bekendste startpagina's geïnteresseerden de mogelijkheid zich met eigen gespecialiseerde 'dochterpagina's' bij hen aan te sluiten, en er zijn er bij die toegang geven tot duizenden van zulke dochters. Hoewel de kwaliteit en volledigheid daarvan aanzienlijk kan verschillen, bieden zulke dochters vaak complete, actuele en goed ingedeelde overzichten over gespecialiseerde onderwerpen binnen een nationaal domein. Een voorbeeld van een Nederlandse directory is het [Nederlandse ODP](#). De meest uitgebreide startpagina's, met veel dochters, zijn [Startpagina.nl](#), [Startkabel.nl](#) en

BooGo!. Startnederland.nl is een startpagina die zelf geen resources verzamelt, maar uitsluitend toegang geeft tot de dochters van andere startpagina's.

BOOGO! PiXUM De beste keuze voor uw digitale foto's! Nu testen!

Actueel	Beurs & Beleggen	Feest & Uitgaan	» als startpagina instellen
WK2006	KoersAlarm.nl	Festivalinfo.nl	» onderwerpen A-Z
Informatie Vogelgriep	behr.nl	Filmfocus	» uw eigen onderwerp?
meldnr dierenbescherming	Alex	Concertkaartjes en reizen	» suggesties
weerbericht	Koersverloop AEX	Team Uitstapje	» instellingen
Goedkoop tanken	Belegger.NL	Zoek een partycentrum	» help & info
Nieuwe zorgverzekering	DeBeurs.NL	Dagje weg	
NOS teletekst	Quote Nieuwslog	Feestagenda	
Laatste virus nieuws	» meer financiën	Belbios (tip)	
» meer nieuws & media		Kroegpagina.nl	
Actuele BooGO! Dochters	Bankzaken	Leuke dingen doen	Zoeken (a/t-z)
vogelgriep	sns bank	» meer feest & uitgaan	Vinden.nl
Cursus	ING bank		
Opleidingen	Online verzekering afsluiten	Gratis	
Hypotheken	Postbank	DEuro.nl	Nieuws
» meer nieuws & media	Abn Amro	Allesgratis.nl	Geen aanwijzingen vergiftiging
	Rabobank	gratis actiecoupons	Milosevic
	» meer financiën	attractieparken	Kapotte vriezer wordt
Educatie & Opleiding	Financiën	Prijz.nl prijsvragen	ingevroren echtbaar fataal
Schoolweb (tip)	Beurs gorilla	Gratis bellen via internet	Amsterdam krijgt
Kennisset.nl	De Beurs.NL	Spybot: Gratis spyware	Uitvaartmuseum
Studieinfo.nl	Brief op maat	verwijderen	
schooltool	professor fiscus	Gratis TV gids online	'Fatah niet in regering met
Studieboekenmarkt	Online Kasboek	Gratis gokkasten	Hamas'
Omroep / Wetenschap	Koersalarm	Gratis TV kijken via internet	
Informatie Beheer Groep	PayPal (online betalen)	Gratisdingen.nl	bron: Nu.nl/Nieuws
Leerlingen.com	Wellowell (vergelijken)	Gratis gokkasten	Persoonlijke links
» meer educatie & opleiding	BooGO! Eurocalculator	Alles Gratis	» links toevoegen
	» meer financiën	» meer gratis	Financiële markten
Gezondheid & Voeding	Communicatie	GSM, SMS & RINGTONES	Weer en verkeer
Ziekenhuis.nl	MSN Messenger	Ringtones en Logo's (tip)	
Hollandse Pot (recepten)	Datingservice	SMS Taal	
Calorie wijzer	Chatplaza	gsmweb	
gratis recepten	Sms via internet (1 per dag)	Brambo	
Calorie tabel	Kletscafe	Dacloglo	
ReceptenWeb	SchoolBANK.nl (reünie)	Hot sms	
Donor of niet?	Wegen.Net	GSMLinks.nl	
Voedingscentrum	GratisKaart.nl	Online WAP browser	
Stoppen met roken	SMScity	SMSme.nl (SMS kanalen)	
Smulweb	» meer communicatie	» meer gsm, sms en ringtones	
Sport je fit	Computers	Handig	
» meer gezondheid & voeding	Download.com	Routenet.nl	
Kinderen & Ouders	Virushelp	De Telefoongids	
Kid City		Tips om te besparen	
kleurplaten van sinterklaas			

Boogo! is een voorbeeld van een Nederlandse startpagina. Gericht op een jong publiek zet het zoveel mogelijk rubrieken op de startpagina die actueel zijn of om andere redenen veel bezoekers trekken. Onderin elk vak is een verwijzing naar Meer ..., waar je dan, tot twee of drie niveaus diep, een verdere onderverdeling van een rubriek vindt.

2.4 WebRings

Niet speciaal voor het zoeken bedoeld, maar wel een middel om informatie over een onderwerp op te sporen zijn de zogenaamde *WebRings*. Dit zijn kringen van mensen die in een bepaald onderwerp geïnteresseerd zijn en die daarover ieder voor zich een Website onderhouden. Vindt men zo'n Website, dan is daarop een verwijzing naar de Websites van de andere deelnemers in de WebRing te vinden. Zo trachten WebRings gezamenlijk bezoekers te trekken. Een voorbeeld is een WebRing van 21 sites die aan Napoleon gewijd zijn.

Meer over WebRings, plus een directory van tienduizenden van zulke kringen vindt u op de site WebRing Directory and Online Community (<http://dir.webring.com/rw>).

2.5. Zoekmachines, metzoekmachines en springplankzoekmachines

De meest gebruikte zoekinstrumenten naast directories zijn *zoekmachines* (search engines). Dit is de wat simplistische naam voor een hoop computer hardware en software die het u uiteindelijk mogelijk maken een of meer zoektermen in een zoekregel

in te typen en dan webpagina's te vinden die die zoekterm(en) bevatten. Zoekmachines werken dus met zoektermen, d.w.z. termen waarin de vraag van een gebruiker vervat is en die (hopelijk) ook in documenten op het Web voorkomen. Als de termen in een vraag – in de vraag 'Wat is de hoofdstad van Nederland?' bij voorbeeld 'hoofdstad' en 'Nederland' – overeenkomen met die in een webdocument dat de zoekmachine kent, vindt de zoekmachine dat webdocument.

ADVANCED SEARCH OPTIONS
➔ back to home page

FIND PAGES	<input type="text" value="containing"/> <input type="text" value="the words"/> <input type="text"/> <input type="text"/>
LANGUAGE	<input type="text" value="Only"/> search pages with <input type="text" value="any language"/>
COUNTRY	<input type="text" value="Only"/> search for websites in <input type="text" value="any country"/>
FILE FORMAT	<input type="text" value="Only"/> search for files of type: <input type="text" value="any type"/>
DOMAIN OR SITE	<input type="text" value="Only"/> search for pages from site or domain: <input type="text"/> <small>(e.g., "exalead.com" or just "com")</small>
LINKS	Search for pages that link to domain or page: <input type="text"/> <small>(e.g., "exalead.com" or "www.exalead.com/help.html")</small>
SEARCH METHOD	<input type="text" value="Exact search"/>
DOCUMENT SORTING	<input type="text" value="By relevance"/>
MODIFICATION DATE	Only keep documents modified after: <input type="text"/> <small>(e.g., "20/04/1994")</small>

Een van de invulformulieren waarmee u bij de zoekmachine Exalead kunt zoeken. Omdat zoekmachines naar zoektermen in documenten zoeken, bieden ze een heleboel mogelijkheden om aan te geven hoe die termen met elkaar moeten worden verbonden en in wat voor soort documenten ze waar moeten voorkomen.

Op alle zoekmogelijkheden die zoekmachines bieden ga ik later in. Hier wil ik alvast benadrukken dat zoekmachines op een heel andere manier moeten worden bediend dan directories. Een directory gebruik je om je erdoor naar documenten over een onderwerp te laten leiden. In een directory volg je een van tevoren uitgestippelde route, van algemeen naar bijzonder, die je naar het onderwerp van je keuze leidt, en vindt daar een titel en een korte beschrijving van documenten die door redacteurs van de directory bij dat onderwerp (een rubriek of categorie) zijn ingedeeld. In een directory kun je ook wel zoeken, maar alleen met termen die in de titel of de beschrijving ervan voorkomen, of die in een rubrieksaanduiding voorkomen. Zoekmachines brengen je in direct contact met de inhoud van de documenten zelf. U kunt die documenten alleen vinden als u de zoekmachine een of meer termen opgeeft die in de documenten zelf voorkomen.

Hoe zoekmachines werken

Wie een zoekmachine gebruikt, verwacht misschien dat de zoekmachine even het hele Web doorzoekt naar de zoektermen die hij heeft opgegeven. Zo is het natuurlijk niet. In

werkelijkheid zoek je bij een zoekmachine in een database van webpagina's die ofwel door gebruikers bij de zoekmachine zijn aangemeld ofwel door de zoekmachine zelf al eerder verzameld zijn. Die database bevat kopieën van de pagina's op het Web. Voor het aanleggen van die database maakt de zoekmachine gebruik van zogeheten *spiders* of *crawlers*, ook wel *gatherers* genoemd, programma's die webpagina's en websites in kopie naar de eigen computers van de zoekmachine overhalen. Ze doen dat door uitgaande van een centrale webpagina (of een database van webpagina's zoals die van Yahoo!) een lijst te maken van alle andere pagina's en sites waarnaar die webpagina of webpagina's door middel van hyperlinks verwijzen, en door al die pagina's op te vragen (net zoals uw browser een (kopie van een) pagina ophaalt als u op een link klikt). Hetzelfde gebeurt met de pagina's waarnaar de eerste set verwijst, enzovoorts. Van al die pagina's wordt niet alleen de inhoud, maar ook onder meer het adres (de URL), de bestandsgrootte en de laatste wijzigingsdatum opgenomen (voor een overzicht van alle informatie die spiders over een pagina verzamelen moet u eens kijken naar de Page Info die de Firefox browser verschaft als u een pagina op hebt staan en op de rechter muisknop klikt (zie boven, blz. 10).

Spiders zijn altijd actief en sporen niet alleen nieuwe pagina's en sites op, maar ook nieuwe informatie op pagina's en sites die ze al eerder gevonden hebben. Zij zijn het die bepalen welk deel van het Web door een zoekmachine wordt gedekt. De programmeur van een spider kan in zo'n programma namelijk beperkingen hebben aangebracht hoe ver of hoe diep links moeten worden gevolgd. Redenen voor zulke beperkingen kunnen zijn de gedachte dat hoe dieper webpagina's op een hostcomputer verscholen zit, hoe geringer de kans is dat het een belangrijk document betreft; of, ook wel, de wens een door de spider bezochte website niet al te zeer te belasten. De frequentie waarmee spiders websites bezoeken, verschilt per zoekmachine en per site.

De opgehaalde webpagina's worden vervolgens door een *indexer* geïndexeerd. Zo ontstaat een database van webdocumenten met een index waardoor die database snel doorzocht kan worden. Zo'n index is in wezen niets anders dan een lijst van woorden uit de geïndexeerde pagina's, met een opgave van hun plaats in de pagina en met een verwijzing naar de databaserecords (de in de database opgenomen pagina's) waarin die woorden voorkomen.

Tevens worden de pagina's geanalyseerd en samengevat door een *summarizer*, een programma dat de belangrijkste termen en concepten in de pagina's identificeert. De *summarizer* bepaalt welke informatie de zoekmachine over een pagina verschaft, maakt ingewikkelder zoekacties mogelijk en bepaalt vaak mede de plaats die een webpagina in een zoekresultaat krijgt (als derde of als driehonderdste document bij een bepaalde vraag); dit laatste heet 'ranken'. Wat dit aspect betreft wordt dus ook wel eens gesproken van een *ranker*. Ook identificeert de *summarizer* objecten op een webpagina.

De samenvatting van de inhoud van een pagina die een zoekmachine in de resultatenlijst bij een vraag laat zien, wordt *summary*, of ook wel *snippet*, genoemd.

Ten slotte zorgt de *broker* (ook wel *query server* genoemd) ervoor dat u als gebruiker een gebruikersvriendelijke zoekpagina met allerlei zoekmogelijkheden wordt gepresenteerd. Zo'n broker omvat dus enerzijds de functionele zoek- ('retrieval')mogelijkheden en anderzijds de presentatie daarvan (de gebruikersinterface).

Zie ook: How they work

(http://www.search-engines-seo.com/engine/engine/How_Search_Engines_Work/)

Voor een uitgebreidere uiteenzetting, met een wat andere indeling van functies, zie E. Liddy, How a search engine works. Searcher 9:5 (2001) (<http://www.infotoday.com/searcher/may01/liddy.htm>)

Net als bij directories zijn de bedrijven die zoekmachines in de lucht houden, commerciële ondernemingen. Ze leven van reclames en de verkoop van hun software voor intranetten, dat wil zeggen de interne computernetwerken van bedrijven en

instellingen. Er zijn er ook bij die zoekresultaten verkopen: bedrijven met een commerciële website kunnen de plaats kopen waarop hun website in het zoekresultaat verschijnt als de gebruiker van de zoekmachine met bepaalde termen zoekt. Hun voornaamste doel is het trekken van zoveel mogelijk bezoekers, niet de ontsluiting van het hele Web; dat komt pas op de tweede plaats. Vandaar dat zoekmachines wel graag prat gaan op een grote database van webdocumenten (tegenwoordig minimaal drie miljard) – want dat trekt bezoekers, en dus ook adverteerders --, maar dat er niet een bij is die het hele Web dekt. Ongeveer de helft is vermoedelijk momenteel het maximaal haalbare. Daarbij dekken verschillende zoekmachines verschillende, zij het wel tot op zekere hoogte, overlappende gedeelten van het Web. Dat betekent dus dat je er in het algemeen gesproken meer dan een moet gebruiken.

Voor de meeste zoekmachines geldt dat ze zich meer concentreren op een redelijk volledige en actuele collectie sites dan op de afzonderlijke documenten in die sites. Die laatste zijn vooral van belang als ze speciaal zijn aangemeld of wanneer er van elders veel verwijzingen naar zijn. Dan worden ze immers ook zeker door een spider gevonden. De minste kans om door een spider gevonden te worden hebben documenten waarnaar helemaal geen links bestaan. Die vormen tezamen een door de zoekmachines onontgonnen gebied van het Web. Zoekmachines vinden webpagina's immers door hyperlinks te volgen.

Het aantal grote, mondiale zoekmachines die ambiëren een representatief deel van het Web te ontsluiten is niet zo groot. Des te groter is het aantal dat zich tot een bepaald geografisch domein (Nederland, Duitsland, Spanje) of een onderwerp (bij voorbeeld celebrities) of vakgebied (bij voorbeeld filosofie, internationale betrekkingen) beperkt. *Nationale* of *regionale zoekmachines* kunt u gebruiken om dicht bij huis te zoeken, maar dat ligt minder voor de hand dan het gebruik van een nationale directory of startpagina in plaats van een internationale directory. Dat komt doordat de mondiale zoekmachines (onder meer door middel van de country code top level domains, zie blz. 7 e.v.) ook nationaal kunnen zoeken, en omdat nationale zoekmachines niet per se een groter deel van het nationale deel van het Web ontsluiten dan de mondiale. Nederlandse voorbeelden van nationale zoekmachines zijn Ilse en Vindex.

Het zoekscherf van Vindex, momenteel de grootste nationale Nederlandse zoekmachine

Nationale en regionale zoekmachines zijn in zekere zin speciale zoekmachines, want ze concentreren zich op een apart, geografisch gedeelte van het Web. Er zijn er ook die zich op een speciaal onderwerp of vakgebied concentreren: de (in strikte zin) *speciale zoekmachines*. In hun zuiverste vorm zijn dit zoekmachines die niet het hele Web of een

geografisch beperkt gedeelte daarvan indexeren, maar alleen bestanden op bepaalde servers waarvan bekend is dat daarop (overwegend) materiaal over een bepaald vakgebied of onderwerp is opgeslagen. Dit zijn de zogeheten *LASEs* of *Limited Area Search Engines*. Andere speciale zoekmachines zijn meer directories of databases van links met hun beschrijvingen, waarin men met een interne zoekmachine kan zoeken. Dan zijn niet de webpagina's zelf geïndexeerd en daarmee *full-text* (dat wil zeggen woord voor woord) doorzoekbaar gemaakt, maar alleen de links met de bijbehorende korte beschrijvingen.

Speciale zoekmachines kunnen om verschillende redenen de voorkeur verdienen boven de grote mondiale zoekmachines. Zo kunnen zoektermen in een beperkt, vakspecifiek domein een veel beter afgebakende betekenis hebben dan in de grote vergaarbak van de database van een algemene zoekmachine (de term 'vervreemding' in een juridische context betekent iets anders dan dezelfde term in een sociologische context), kan de kwaliteit van de gevonden documenten beter gewaarborgd zijn door de zorgvuldige selectie van de opgenomen sites en kan, in het geval van LASEs, de spider van een speciale zoekmachine soms dieper en verder doordringen dan die van een algemene zoekmachine.

Voorbeeld van een zogeheten LASE, een Limited Area Search Engine. De hier afgebeelde, ISN, indexeert alleen bestanden op servers met veel materiaal over internationale veiligheid en betrekkingen.

Tot de categorie speciale zoekmachines in bredere zin behoren ook zoekmachines die gespecialiseerd zijn in bepaalde mediatypes, bijvoorbeeld beeld- (foto, video) en geluidsmateriaal, een bepaald type informatie (bijvoorbeeld nieuws en blogs (zie blz. 277 e.v.)) of een bepaalde functie (bijvoorbeeld 'shopping search engines' of 'prijsvergelijkingzoekmachines', zoekmachines die de voordeligste aanbieders bij een bepaald product voor u zoeken). Grote mondiale zoekmachines zoals Google en Yahoo! hebben naast hun grote database van webdocumenten nog speciale databases voor deze doelen; Google bijvoorbeeld heeft Google Images, Google Video (Beta), Google News en Froogle (shopping), en Oristus, een audiozoekmachine van Google, komt eraan. Dat neemt niet weg dat voor allerlei speciale doeleinden en materialen zoekmachines bestaan die leemtes vullen die door de bedrijven achter de grote zoekmachines worden opengelaten. Als voorbeelden noem ik zoekmachines die P2P-netwerken doorzoeken

(P2P = peer-to-peer, bestandsuitwisseling tussen mensen die met behulp van speciaal daarvoor bestemde software de bestanden op een deel van de harde schijf van hun computer in een netwerk met elkaar delen) en zoekmachines om chatrooms mee op te sporen.

2.5.1. Metazoekmachines

Een bijzondere categorie van zoekmachines zijn de zogeheten *metazoekmachines* (meta-search engines). Dit zijn zoekmachines zonder een eigen database en index van websites of webpagina's, hoewel ze soms wel een (niet erg omvangrijke) directory hebben. Wat doen deze metazoekmachines? Zoals we hebben gezien dekken de grote zoekmachines zoals Google en Yahoo! ieder maar een deel van het Web. Er is wel overlap, maar hoe groot die is, is moeilijk te zeggen en bovendien op het niveau van de individuele vraag niet zo van belang. Daarbij komt dat deze zoekmachines ieder apart en in veel gevallen met verschillende vraagformuleringen moeten worden bediend. Metazoekmachines bieden hiervoor een oplossing: ze vertalen de vraag die u intypt naar de zoektaal van de verschillende zoekmachines en sturen de vraag gelijktijdig naar die zoekmachines op. Sommige doen dat naar alle zoekmachines die ze aan het werk zetten tegelijk, andere doen het in fases: eerst vijf zoekmachines, daarna, als u het antwoord op uw vraag nog niet bent tegengekomen, nog eens vijf, enz. Een gemak dat veel metazoekmachines ook bieden is dat ze de resultaten van de verschillende zoekmachines ontdebelen, d.w.z. pagina's die door meer dan één zoekmachine worden gevonden, maar één keer tonen (hoewel soms wel met de verschillende snippets). Een belangrijk voordeel is ten slotte dat ze de gebruiker minder afhankelijk maken van de eigenaardigheden (met name het rankingmechanisme, de volgorde waarin een zoekmachine zijn resultaten presenteert) van iedere zoekmachine apart. Vaak hebben ze een eigen rankingmechanisme, dat die treffers hoog in de resultatenlijst zet die bij meer zoekmachines hoog in de ranking voorkomen.

Metazoekmachines zijn goed voor simpele vragen, waarop ze snelle antwoorden bieden. Ook voor vragen die heel weinig treffers opleveren, zijn ze goed te gebruiken.



jux2™
search engines differ more than you think

Compare Google, Yahoo, and MSN with One Search

Jux2 is een van de weinige metazoekmachines die Google, Yahoo! en MSN in het pakket van bevroegde zoekmachines hebben zitten.

Tegenover de hierboven genoemde voordelen staan echter ook aanzienlijke nadelen. Het belangrijkste daarvan is dat de meeste metazoekmachines Google niet in hun pakket van zoekmachines hebben zitten waar ze hun resultaten vandaan halen. Google en Yahoo! ontbreken vaak, en worden dan vervangen door zoekmachines van het tweede garnituur, zoals Netscape en AOL, die weliswaar Google's database gebruiken, maar slechts een deel daarvan. Of ze worden vervangen door zoekmachines zoals Overture, die hun zoekresultaten doorspekken met treffers voor de 'ranking' waarvan ze zich laten betalen.

Een ander nadeel is dat metazoekmachines nogal eens gebruik maken van zoekmachines die, met hetzelfde rankingmechanisme, uit dezelfde database van webdocumenten putten. Een voorbeeld vormen AltaVista, Yahoo! en AlltheWeb, die uit één database van Yahoo! putten, maar toch vaak samen in het pakket van matazoekmachines zitten en zo bij de eigen ranking van die metazoekmachines drie keer tellen in plaats van één.

Daarbij komt dan ook nog een ander bezwaar, dat inherent is aan het begrip metazoekmachine: zoekmachines als Google, Yahoo! en MSN Search bieden uiteenlopende mogelijkheden, en waar ze dezelfde mogelijkheden bieden, moeten vragen waarbij daarvan gebruik wordt gemaakt vaak op uiteenlopende manieren worden geformuleerd. Een metazoekmachine heeft maar één zoekregel, waarin een vraag maar op één manier kan worden geformuleerd. Het gevolg kan zijn dat een metazoeker de geavanceerde mogelijkheden van de zoekmachines die hij bevraagt niet kan gebruiken. Al deze bezwaren zijn voor sommige experts reden om [het gebruik van metazoekmachines af te raden](#). Of dat een goede raad is hangt af van wat men ervoor in de plaats stelt: een zoekactie met de verderop in dit boek besproken grote mondiale zoekmachines Google, Yahoo!, MSN Search, Ask, en mogelijk ook Gigablast en Exalead, met gebruikmaking van de specifieke zoekmogelijkheden en de 'taal' van elk van die zoekmachines is te prefereren boven een met welke metazoeker ook, maar vergt natuurlijk ook meer.

2.5.2. Springplankzoekmachines

Veel van de bezwaren tegen metzoekers worden ondervangen door springplankzoekmachines. Deze bieden één zoekregel, waarmee niet gelijktijdig, maar na elkaar diverse zoekmachines met een identieke, maar per zoekmachine aan het idioom van die zoekmachine aangepaste zoekvraag kunnen worden bevraagd.

turboscout
access all search engines without retyping

[Like TurboScout? Please Help...](#)

Web Images Reference News Products Blogs Audio/Video

Enter keywords above, then click a search provider below

[Google](#) • [Yahoo](#) • [MSN](#) • [Ask](#)
[A9](#) • [Alexa](#) • [Alltheweb](#) • [Clusty](#) • [DogPile](#) • [EntireWeb](#) • [Gigablast](#) • [ithaki](#) • [ixquick](#)
[Looksmart](#) • [Mamma](#) • [Metacrawler](#) • [ODP](#) • [Search.com](#) • [Snap](#) • [Teoma](#) • [Wisenut](#)

[Add TurboScout to Your Firefox's Search Box :: Make TurboScout Your Homepage](#)

TurboScout, een voorbeeld van een springplankzoekmachine.

Springplankzoekmachines doen maar twee dingen, maar die doen ze goed: ze bieden ons een simpele toegang tot een uitgelezen pakket van zoekmachines, waaruit we zelf kunnen kiezen; en ze bieden ons één zoekregel waarin we maar één keer een vraag hoeven in te typen die we vervolgens – met eventuele aanpassingen – door meer zoekmachines kunnen laten uitvoeren. Voorbeelden van dit soort zoekmachines zijn [Fagan Finder](#), dat bovendien een schat aan informatie over zoeken en zoekmachines biedt, en ook over speciale zoekmachines, [Proteus Internet Search](#) en [TurboScout](#).

2.6. Discussiefora: nieuwsgroepen, message boards en mailing lists

Als bron van informatie kunnen discussiefora voor drie doeleinden worden gebruikt:

- om vragen te stellen
- om bij te blijven
- om antwoorden te vinden

Deze laatste functie komt in deze sectie van dit hoofdstuk ter sprake, de andere twee worden elders in dit boek behandeld.

Voor welk type vragen zijn discussiefora geschikt?

Het type vragen waarvoor deze lijsten bij uitstek geschikt zijn, is:

- hoe doe ik...?
- waar vind ik...?
- heeft iemand ervaring met...?
- wat vindt men van...?
- hoe werkt...?
- wat is een goed...?

Maar denk erom: raadpleeg om te beginnen de FAQ! (zie blz. 32)

Om het antwoord op een vraag te krijgen kunt u zich natuurlijk op een discussieforum abonneren en vervolgens de vraag aan anderen stellen. Maar het is heel goed mogelijk dat uw vraag al eens eerder door anderen gesteld en beantwoord is, en om zo'n antwoord dan te vinden gebruikt u de archieven van de discussiefora, die in veel gevallen op internet aanwezig en te doorzoeken zijn.

De antwoorden die u in de archieven van discussiefora zult vinden, hebben natuurlijk veel te maken met het soort communicatie dat zich in die fora afspeelt. Opinies, tips en discussies vormen de inhoud van die communicatie, en voor vragen die daarop betrekking hebben, moet u op internet bij de archieven van zulke fora zijn. En daarnaast vormen ze natuurlijk een uitstekende omgeving om deskundigen op een bepaald vakgebied (van militaire vliegtuigen tot evolutiebiologie) of gewoon mensen met dezelfde belangstelling te vinden.

2.6.1. Nieuwsgroepen

Het eenvoudigst gaat dat bij de archieven van Usenet *nieuwsgroepen* vanaf 1981, die door Google Groups (<http://groups.google.com/>) gezamenlijk en per nieuwsgroep ontsloten zijn. De database dekt de archieven van ruim 50.000 van de naar schatting meer dan 120.000 nieuwsgroepen. Daarbij maakt Google Groups het zowel mogelijk nieuwsgroepen te vinden in de archieven waarvan een antwoord op uw vraag misschien verborgen ligt, als om direct in de berichten ('posts' of 'postings') zelf te zoeken, dwars door alle archieven heen. Voor het eerste gebruikt u de directory van nieuwsgroepen die op de homepage van Google Groups beschikbaar is. Net zoals bij een directory als het ODP (zie blz. 81 e.v.) kunt u daarin van algemeen naar bijzonder afdalen, maar eenmaal in de directory kunt u er ook in zoeken, op de namen en beschrijvingen van de groepen. Een andere mogelijkheid om interessante nieuwsgroepen te vinden is natuurlijk om een paar relevante vragen op het hele archief los te laten (vanaf de homepage), en dan te

kijken bij welke nieuwsgroepen de antwoorden op die vragen vandaan komen. Direct zoeken in het hele archief doet u natuurlijk ook als u niet op zoek wilt gaan naar interessante nieuwsgroepen, maar naar het antwoord op een concrete vraag. Meer over de mogelijkheden daarvan en van het gebruik van de Advanced Search, vindt u op blz. 112 e.v.

[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#) [Local](#) [more »](#)

 [Advanced Groups Search](#)
[Preferences](#)
[Groups Help](#)

Learn more about [Google Groups](#).

Create, search, and browse groups to discuss and share ideas.

Members: Sign in	▪ Quickly access your favorite groups	▪ Start new discussions
New users: Join	▪ Follow your bookmarked (★) topics	▪ Create new groups

Arts and Entertainment	Business and Finance
Computers	Health
Home	News
Recreation	Regions and Places
Science and Technology	Society and Humanities
	Browse all of Usenet...

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2006 Google

De homepage van Google Groups, een combinatie van een directory met een zoekmachine.

Google Groups is een enorm archief van vragen en antwoorden, opinies, adviezen en discussies, waarin u een antwoord op vragen vindt die met een webzoekmachine niet zo gemakkelijk te beantwoorden zijn. Een voorbeeld vormt een (misschien niet hét) antwoord op de vraag naar de herkomst van de uitdrukking "ouwe jongens, krentenbrood" (zie volgende blz.).

Naast Google Groups bestaat er nog een andere manier om nieuwsgroepen op te sporen. Een (meestal incompleet) overzicht van nieuwsgroepen kunt u krijgen als uw internet provider over een nieuwsserver beschikt en uw browser of newsreader daarmee contact kan maken. (Een *newsreader* is een speciaal softwareprogramma waarmee u de berichten van nieuwsgroepen kunt lezen. Op Windows-computers kunt u daarvoor Outlook Express gebruiken als uw browser geen ingebouwde newsreader heeft, of u kunt op free- en sharewaresites een andere newsreader downloaden.) Sommige van zulke news readers maken het mogelijk in deze lijst te zoeken, maar Google Groups werkt waarschijnlijk handiger.

Martien Sep 27 2003, 6:20 am [show options](#)

Maikel Olde Hartingsveldt <hartingsve...@weghalen-hotmail.com> deed de wereld kond van het feit dat

>Al jaaaaaren ben ik op zoek naar de **herkomst** van dit gezegde, maar tot
>nu toe heb ik nog geen - bevredigende - verklaring gevonden. Ik ken er
>nog 1 variant op: "toffe jongens, krentenbrood". Maar volgens mij
>heeft "Ouwe jongens..." oudere papieren....

>Wie kent de oplossing?

Volgens Marc de Coster:

...De **uitdrukking** kwam begin jaren 70 in zwang in de vlotte omgangstaal, maar werd al veel eerder gebruikt onder soldaten (Salleveldt 1978 vermeldt de variant 'ouwe jongens, goeie soep'. Daarin verklaard als een kreet waarmee men een meerdere (kapitein, majoor enz) karakteriseert die ten koste van alles populair wil zijn. In zijn twee jaar nadien verschenen "Woordenboek van Jan Soldaat in Indonesie krijgt de **uitdrukking** een andere uitleg: "Alles in orde")....

--
Raffinement 'n gave

End of messages

☆ [watch this topic](#)

Vraag en antwoord over de herkomst van 'Ouwe jongens, krentenbrood'.

Namen van Usenet nieuwsgroepen hebben een structuur die een beetje lijkt op die URLs: ze lopen af van algemeen naar bijzonder. Enkele voorbeelden zijn:

- humanities.lit.authors.shakespeare
- soc.genealogy.jewish
- soc.history.war.world-war-ii
- talk.politics.european-union
- rec.autos.makers.ford.mustang.

De hoofdgroepen zijn:

- biz. – zakelijke producten, diensten, productbesprekingen
- comp. – computers (hardware, software, consumenteninfo), automatisering
- humanities. – kunst, literatuur, filosofie etc.
- news. – nieuws (over Usenet)
- rec. – recreatie: spel, hobbies, sport
- sci. – wetenschap
- soc. – samenleving, cultuur
- talk. – discussie; bij voorbeeld over politiek, religie of milieu
- misc. – banen, gezondheid en andere onderwerpen
- alt. – alle eerder genoemde onderwerpen en nog talloze andere

Naast deze mondiale nieuwsgroepen zijn er ook die actief zijn binnen elk nationaal domein. Dan begint de naam met de tweeletterafkorting die we al eerder in URLs tegenkwamen: nl.* (Nederland), be.* (België), de.* (Duitsland), uk.* (Groot Brittannië), fr.* (Frankrijk). (De asterisk staat hier als joker voor alles na de tweeletterafkorting kan volgen.) Een voorbeeld van een Nederlandse nieuwsgroep: nl.gezondheid.alt-genezen.

Ten slotte is van belang het verschil tussen 'geleide' (moderated) en ongeleide nieuwsgroepen. In 'gemodereerde' nieuwsgroepen wordt over het algemeen strakker de hand gehouden aan het onderwerp waar de nieuwsgroep over gaat.

De meeste van de hierboven genoemde nieuwsgroepen behoren tot Usenet, een netwerk van computersystemen met een verzameling van duizenden elektronische prikborden die al van vóór de tijd van internet dateert. Daarnaast bestaan er nog een zeer groot aantal bulletin, discussion of message boards (ook wel aangeduid als Web forums of Web conferences) die door bedrijven en organisaties worden gerund en die niet tot de officiële Usenet nieuwsgroepenhierarchie behoren. Twee grote providers voor zulke forums zijn ezboard (<http://www.ezboard.com/>) en Delphi Forums (<http://www.delphiforums.com/>), die u resp. op het niveau van de forums zelf en van de

IMDb NOW PLAYING MOVIE / TV NEWS MY MOVIES DVD / VIDEO MESSAGE BOARDS SHOWTIMES & TICKETS GAME BASE **FREE TRIAL!** **IMDbpro**

Earth's Biggest Movie Database™ [Home](#) | [Top Movies](#) | [Photos](#) | [Independent Film](#) | [Browse](#) | [Help](#) [Login](#) | [Register to post](#)

IMDb SEARCH All [More](#) | [Tips](#) WEB SEARCH Powered by

[Main Boards](#) | [Help](#) | [Terms & Conditions](#)
[New User](#) | [Log In](#)
[IMDbPro Message Boards](#)

Message Boards

NEW! Visit the [IMDbPro.com Message Boards](#) and discuss a wide variety of entertainment industry topics with other members of the IMDbPro.com community. Recent message board postings include: [DIRE ENDURANCE](#) [DODGEBALL](#), [Help](#), [We need: Hot Actor , 33-43, for political thriller](#) [More »](#)

Board: [Être et avoir \(2002\)](#)

This board is reserved for discussions about [Être et avoir \(2002\)](#)

Inappropriate discussions may be deleted or moved

[Start New Topic](#)

[Add to Favorites](#)

Subject	Started by	Replies	Latest Posting
childhood	makro120	0	Sat Mar 26 2005 06:24
i can not stand this movie	hemza lasri	4	Tue Mar 22 2005 03:03
Is Mr. Lopez still teaching?	Vagabond5	0	Tue Jan 25 2005 08:44
DVD RELEASE	filmloverforlife	3	Thu Jul 1 2004 14:40
Funniest scene ever?	eglaisher	11	5 days ago
Explaining the title	MrTeal	10	Sun Nov 21 2004 09:03
Not in England	Hannywoo	7	Mon Oct 25 2004 10:36

Een 'message board' bij de bekende Internet Movie Database, in dit geval bij de Franse film *Être et avoir*. Bij iedere film in de database biedt zo'n prikbord bezoekers ruimte om te reageren, andere bezoekers vragen te stellen etc. Voor de zoeker kan het overzicht van dit soort prikborden informatie opleveren die elders niet of moeilijk te vinden is.

afzonderlijke berichten laten zoeken. Een full-text doorzoekbaar bestand van het berichtenverkeer op zulke message boards biedt Lycos Discussions Web Search (<http://discussion.lycos.com/>); let hier ook op de links naar Related Organizations, bovenaan de resultatenpagina's. Gebruik Lycos Discussions Web Search als u geïnteresseerd bent in vragen als wat anderen goede HTML-editors vinden of hoe je in Parijs goedkoop vegetarisch kunt eten, of als u onlangs ergens een interessant bericht hebt gelezen en niet meer weet waar. Er zijn meer van dit soort zoekmachines, al geeft die van Lycos verreweg de meeste resultaten; voor een overzicht zie: [Data Mining: Message Board Search](#).

Een simpele manier om een Nederlandstalig forum of prikbord te vinden waar u anderen met dezelfde interesse treft, is via de dochters van startpagina's. Zowel Startpagina.nl (<http://www.startpagina.nl>) als Startkabel (<http://www.startkabel.nl>) heeft een groot aantal dochters die over zo'n prikbord of forum beschikken. Die van Startkabel staan netjes bij elkaar op de pagina Forums (<http://forums.startkabel.nl>), die van Startpagina zijn te vinden door in een pagina naar een term als prikbord, berichtenbord, prikpagina of forum te zoeken. Alle prikborden van dochters van Startpagina hebben het adres www.prikkpagina.nl en zijn zo ook met Yahoo! te vinden (voor fora over puzzels bijvoorbeeld: puzzels.site.pagina.nl linkdomain.prikkpagina.nl; voor het hoe van de zoekopdracht, zie blz. 153 e.v.). Een andere mogelijkheid om forums te zoeken is via websites. Als die een forum hebben, is de kans niet gering dat dit deel van de site een adres heeft met het woord forum in de URL, meer speciaal in het deel van de URL dat als *path* wordt aangeduid. Zo brengt de zoekvraag "alcohol en drugs" `inurl:forum site:nl` bij Google u onder meer naar de forums van de Jellinek Kliniek (<http://www.jellinek.nl/forum/>). (Zie blz. 101 e.v. over de 'taal' waarmee u in Google zoekt.)

2.6.2. FAQs

Tot de wereld van de nieuwsgroepen, fora en de hieronder besproken mailing lists behoren ook de zogeheten FAQs, of *Frequently Asked Questions*. Dit zijn documenten die speciaal zijn opgesteld om al die groepen van steeds dezelfde beginnelingenvragen te ontlasten. Maar er zijn erbij die een diepgaande behandeling van een onderwerp geven. Voor mensen die zich net op een groep hebben geabonneerd, maar ook voor wie zich op een onderwerp wil oriënteren zonder meteen met een lawine van links te worden geconfronteerd, zijn zulke FAQs verplichte kost. Een verzameling van Usenet FAQs is te vinden in de Internet FAQ Archives (<http://www.faqs.org/faqs/>), die op dit adres ook full-text kunnen worden doorzocht. Een andere toegang tot Usenet FAQs biedt de pagina Utrecht CS News.Answers FAQ Access Methods (<http://www.cs.uu.nl/cgi-bin/faqwais>). FAQs van fora die niet tot Usenet behoren zult u bij die fora of met een algemene zoekmachine moeten zien te vinden.

Twee voorbeelden van FAQs:

FAQ: How to find people's E-mail addresses
(<http://www.cs.queensu.ca/FAQs/email/bigfinding.html>)
Internet Search FAQ (<http://www.search-faq.com/>)

2.6.3. Mailing lists

Mailing lists (rondzendlijsten) zijn georganiseerde gespreksgroepen waarop u zich moet abonneren om de berichten per e-mail aan u toegezonden te krijgen. Men rekent er zowel *discussielijsten* toe, waarop de deelnemers met elkaar communiceren, als *nieuwsbrieven* (newsletters) of *e-zines*, die niet interactief zijn maar alleen bedoeld om

de abonnees op de hoogte houden. Er zijn honderdduizenden mailing lists, en het aantal berichten dat daarop per jaar wordt gewisseld wordt op 30 miljard geschat.

Anders dan de archieven van nieuwsgroepen zijn de archieven van mailing lists niet centraal doorzoekbaar. Voor zover de hier uitgewisselde berichten gearhiveerd worden en openbaar zijn – en dat is lang niet altijd het geval --, moeten ze worden gevonden door de servers te vinden waarop de betreffende lijsten worden onderhouden. Die servers, ook bekend als 'sites' of 'hosts', heten meestal naar het programma dat voor het creëren en beheren van lijsten wordt gebruikt, zoals 'listserv', 'majordomo', 'listbot', 'listproc', 'mailbase' en 'smartlist' (voor een redelijk compleet overzicht zie The Mailing List Gurus Page: Mailing List Management Programs (List Servers):

<http://lists.gurus.com/mlms.html>). De openbare lijsten zijn gelukkig meestal wel via databases van zulke lijsten te vinden. Kent men daar eenmaal het adres van, dan is het meestal ook niet moeilijk het bijbehorende archief op te sporen. Maar de enige route naar de archieven van rondzendlijsten loopt via de lijsten zelf, niet zoals bij de nieuwsgroepen via een zoekmachine voor alle archieven.

Ook voor mailing lists bestaan een aantal overzichten, hoewel deze niet centraal te doorzoeken zijn. Van de (openbare) lijsten die met het programma 'listserv' worden beheerd bestaat een volledig overzicht bij 'CataList, the official catalog of LISTSERV lists' (<http://www.lsoft.com/catalist.html>), waar zowel naar sites (hosts) als naar lijsten kan worden gezocht en waar niet alleen sites en lijsten per land kunnen worden bekeken, maar ook een overzicht van lijsten met meer dan 10.000 dan wel 1000 deelnemers kan opgevraagd; bij elke lijst staat trouwens het aantal deelnemers vermeld, dat een aanwijzing vormt voor de exclusiviteit van de lijst en de hoeveelheid berichten die de deelnemers te verwachten hebben. Elke lijst heeft een naam en een zeer summiere beschrijving. Zoeken in dit overzicht is lastig omdat de zoektermen als strings



L-Soft

Search results

AEPORTAL@SI-LISTSERV.SI.EDU

Archaeology and Ethnology Portal (5 subscribers)

STANDPIPE-L@LISTS.VCU.EDU

Discussion of Caribbean Archaeology and Ethnology (163 subscribers)

2 lists matched your search string.

Learn more about: [Announcement Lists](#) • [Discussion Groups](#) • [Email Marketing](#)

[LISTSERV](#) is a registered trademark licensed to [L-Soft international, Inc.](#)
[CataList](#) is a service mark of [L-Soft international, Inc.](#)

[Last update:](#) 17 Mar 2006 16:00 -0500 (75,465 lists, 2,511 sites)
© L-Soft international, Inc. 2006



Resultaat van een zoekactie met de term 'ethnology' in Catalist, een database van ruim 58.000 rondzendlijsten. Het oog achter de twee lijsten biedt een link naar uitgebreidere informatie over de betreffende lijst, met onder meer een mogelijkheid het archief te doorzoeken.

(reeksen van letters die woorden kunnen, maar niet hoeven te zijn) worden opgevat en met de term opera bij voorbeeld ook een aantal lijsten worden gevonden die iets met 'cooperation' te maken hebben. Verder kun je zoektermen alleen met OF (weergegeven door ;) verbinden, niet met EN. Je kunt niet zoeken naar een lijst waarin het woord music in de titel van de lijst voorkomt en het woord review in de beschrijving.

Daartegenover staat dat Catalist goed te gebruiken is om via de sites de webtoegang tot de archieven op te sporen. Direct achter de naam van lijsten met een openbaar archief staat een plaatje van een oog (zie vorige blz.) met daaronder een link naar de archieven van de lijst in kwestie. Sites zijn op verschillende manieren te vinden, onder andere doordat bij elke lijst ook een link staat naar een overzicht van alle andere lijsten bij dezelfde host.

Tile.net (<http://tile.net>) geeft een doorzoekbaar overzicht van listserv en niet-listserv rondzendlijsten. De beschrijvingen van de opgenomen lijsten zijn uiterst beknopt, maar wel is er ook een adres van de moderator (beheerder) opgenomen en over hoe je je abonnement bij een lijst moet beëindigen (een veel voorkomend probleem bij 'newbies'

Home : Travel & Leisure

In the Spotlight



[Gardenwise](#)

Not everyone was born with a green thumb. Check out this newsletter for tips and guidance on making that garden grow!



[Voyages](#)

It's a wide, wide world. Get invaluable tips and great articles on where to travel in it from this weekly 'zine. Don't leave home without it!

Travel Newsletters

[Disney World Trip Planning](#)

Prepare the family for Mickey & friends.

[Traveler Tips by Andy Graham](#)

Thorough insight by an expert traveler.

[Alaska Travelgram](#)

Cheap flights, cool sights in Alaska.

[Exotic Vacations in India](#)

Try something different on your next trip.

[Puerto Rico](#)

For Spanish-speaking visitors to the beautiful isle.

Hobbies Newsletters

[Rare Books](#)

Dealers list their rare books for sale.

[Journals & You](#)

Making the most of your diary.

[StreetFairAmerica.com](#)

Newsletter on selling your arts & crafts.

[Antiques & Collecting Newsletter](#)

Good info for antique collectors.

[Amateur Photography](#)

Share tips with fellow photogs.

Discuss Destinations & Pastimes

[Travel Guidebook Writers](#)

Learn tricks of the travel-writing trade.

[Portugal Travel Information](#)

Info on Portugal from those who've been there.

[Travel & Leisure Magazine's HotDeals](#)

Deals and discounts for the savvy traveler.

[The Kreative Komer](#)

Meet and greet experienced arts & crafters.

[National Online Quilters](#)

This discussion is a virtual quilting bee!

[Reef List](#)

Chat with your fellow reef aquarists.

[Manifest Astrology](#)

Students of astrology study the stars.

[Internet Amateur Radio Club](#)

Ham & other communications hobbyists.

Want More? Search or Browse These Leisure Topics

[Travel](#)

[Adventure Travel](#)

[Air Travel](#)

[Budget Travel](#)

[Cruises](#)

[Destinations](#)

[General Resources for Travelers](#)

[Leisure Travel](#)

[Lodging & Resorts](#)

[Motorcycle Travel](#)

[News](#)

[Animals & Pets](#)

[Regional & Travel](#)

[Countries](#)

[Regions](#)

[Travel Newsletters](#)

[Wilderness & Survival Training](#)

[Video Games](#)

[Help](#)

Mailing lists over reizen en vrije tijd bij Topica (zie volgende blz.). Let op het onderscheid tussen nieuwsbrieven en discussielijsten.

oftewel groentjes). Tevens heeft Tile.net onder meer een overzicht van nieuwsgroepen. Twee uitgebreide overzichten van rondzendlijsten bieden verder Topica (<http://www.topica.com>) en Yahoo! Groups (<http://groups.yahoo.com>). Beide hebben een directorystructuur à la Yahoo!, in combinatie met een zoekmogelijkheid, wat het vinden van lijsten heel eenvoudig maakt. Beide 'hosten' ook eigen rondzendlijsten, maar hebben daarnaast selecties van lijsten elders. Yahoo! Groups is de grootste van de twee en houdt ook eigen archieven van de opgenomen lijsten (voor zover gearchiveerd) bij, maar compleet en helemaal bij zijn die niet; Topica biedt waar dat kan een link naar de oorspronkelijke archieven.

Het bestaan van deze overzichten zou de indruk kunnen wekken dat het hele terrein van mailing lists behoorlijk ontsloten is. Dat is een illusie: er zijn ook nationale overzichten, die weer geheel nieuwe perspectieven openen. Zo bevat alleen al de Franse Annuaire liste de diffusion, mailing liste, newsletter (<http://www.francopholistes.com/>) meer dan 6000 lijsten, waar Catalist er maar ongeveer een tiende deel van kent (en Catalist kent er ook die de Annuaire niet kent!).

Meer over rondzendlijsten: Introduction to Electronic Mailing Lists (Hopkins) (<http://www.uta.fi/FAST/PK5/LISTS/e-lists.html>); Web Search Guide (<http://www.websearchguide.ca/>) over nieuwsgroepen, rondzendlijsten, webfora.

2.6.4. Nieuwsbrieven en e-zines

Een aparte plaats onder de rondzendlijsten nemen de newsletters en e-zines in, tijdschriften en nieuwsbrieven die elektronisch worden gepubliceerd en via rondzendlijsten gedistribueerd. Magazines, bladen en tijdschriften die ook in gedrukte vorm verschijnen, horen hier strikt genomen niet bij, maar omdat de overzichten van dit materiaal het onderscheid tussen 'alleen elektronisch' en 'ook elektronisch' ook niet altijd maken, behandel ik ze hier ook. Een overzicht is te vinden in Go-Ezines: Ezine directory (<http://www.go-ezines.com/>), dat onder meer een categorie E-zine Resources, Directories, met links naar andere overzichten bevat. The Book of Zines (<http://www.zinebook.com/>) vertelt u alles wat u maar over zines zou willen weten, en bevat bovendien een verzameling links naar zinesarchieven. Dé bron voor e-zines is echter John Labovitz's E-Zine-List (<http://www.e-zine-list.com/>), die al sinds 1993 bestaat. Dit is geen directory, met een nette hiërarchische indeling op algemene en bijzondere onderwerpen, maar elke zine is van een aantal trefwoorden voorzien en er is een zoekmogelijkheid; daarnaast is er overzicht van de 200 belangrijkste trefwoorden, met de bijbehorende zines.

In dit hoofdstuk behandel ik nieuwsbrieven en e-zines als bron om informatie in te vinden, met het accent op archieven. Het feit dat bij dit materiaal, net als op het Web als geheel, veel kaf tussen het koren zit, hoeft ons hier dus niet bezig te houden, dat moet het alleen doen als we ons op een nieuwsbrief of e-zine abonneren. Dan zult u misschien tot uw ongemak merken dat de nieuwsbrief waarop u zich zo enthousiast hebt geabonneerd omdat er een aardig artikel in stond, voor het merendeel uit 'spam' bestaat, reclame waarop u met uw toch al overvolle postbus niet zit te wachten. Al te pessimistisch over de kwaliteit van dit soort materiaal hoeft u overigens ook weer niet te zijn: zo vond ik in Go-Ezines bijvoorbeeld onder Reference – Knowledge Management de newsletter van de ACM Special Interest Group on Knowledge Discovery and Data Mining, die meer van een wetenschappelijk tijdschrift dan van een nieuwsbrief wegheeft.

Wie speciaal in dit soort serieuze materiaal geïnteresseerd is, dus meer in wetenschappelijke tijdschriften dan in zines, kan overigens beter terecht in overzichten die speciaal daarop betrekking hebben. De twee beste zijn de Directory of open access journals (<http://www.doaj.org/>), met meer dan 1700 vrij toegankelijke tijdschriften,

waarvan er enige honderden op artikelniveau zijn ontsloten, en de Elektronische Zeitschriftenbibliotheek van de Universitätsbibliothek Regensburg (<http://rzblx1.uni-regensburg.de/ezeit/>), een op onderwerp en alfabetisch ingerichte en doorzoekbare verzameling van ruim 26.000 elektronische tijdschriften, waarvan een niet onaanzienlijk deel vrij toegankelijk.

Directories en lijsten van nieuwsbrieven en zines zijn voor ons in dit hoofdstuk een middel om via de nieuwsbrieven en zines bij de archieven van dit soort mailing lists uit te komen, waar we artikelen over een onderwerp dat ons interesseert hopen te vinden. Die artikelen kunnen we natuurlijk ook op een andere manier proberen te benaderen, direct, en niet via de nieuwsbrieven, zines en tijdschriften waarin ze gepubliceerd zijn. Als we zo te werk gaan is MagPortal (<http://www.magportal.com>) een goed startpunt. Deze doorzoekbare directory bevat artikelen uit meer dan 200 tijdschriften, in sommige gevallen teruggaande tot 1998. Overigens zijn de meeste van de tijdschriften enkel gedurende een gedeelte van de jaren die sindsdien verstreken zijn, vertegenwoordigd. Hetzelfde geldt voor FindArticles (<http://www.findarticles.com>), dat vijf miljoen artikelen uit een nog groter aantal tijdschriften bevat, die echter maar voor een deel gratis te raadplegen zijn.

Om wetenschappelijke artikelen op het Web te vinden kunt u Google Scholar (<http://scholar.google.com>, zie blz. 132 e.v.) of Elseviers Scirus (<http://www.scirus.com>) gebruiken, artikelen uit door commerciële uitgevers gepubliceerde tijdschriften, voor het lezen en downloaden waarvan u de creditcard moet trekken, vindt u in Ingenta (<http://www.ingenta.com>). Onderzoek voordat u hier uw goede geld aan uitgeeft de mogelijkheid om lid te worden van een wetenschappelijke bibliotheek (zie blz. 132). Overigens: net zoals we via zines en tijdschriften interessante artikelen kunnen trachten op te sporen, kunnen we via artikelen interessante tijdschriften opsporen.

2.7. 'Bloggers en 'experts'

Discussiefora met FAQs en archieven, en met hun mogelijkheden tot directe communicatie vormen een enorm reservoir aan kennis waaruit u kunt putten als u dat terrein verkend hebt. Ze bieden ook goede mogelijkheden om op de hoogte te blijven van de gewoonten en ontdekkingen van gelijkgezinden en om u door het stellen van vragen (en natuurlijk eigen ervaring) te ontwikkelen van een 'newbie' (groentje) tot een 'guru' (expert). Maar voordat het zover is biedt het Web nog andere mogelijkheden om u te informeren.

In de eerste plaats zijn er duizenden mensen op het Web die ook buiten het verband van nieuwsgroepen en dergelijke bereid zijn antwoord op uw vragen te geven. Daaronder zijn natuurlijk bureaus die zich daarvoor laten betalen, en er zijn ook wel diensten waar informatie wordt verhandeld – dat wil zeggen dat er websites zijn waar zelfbenoemde 'experts' zich aanbieden om voor een onderhandelbaar bedrag het antwoord op vragen van bezoekers te zoeken --, maar vooral in de educatieve hoek van het web zijn ook tal van vrijwilligers te vinden die u gratis willen helpen. Ook hierbij is in de taal van het Web sprake van 'experts', maar het is goed te bedenken dat iedereen die op het Web expert heet, in de eerste plaats een zelfbenoemde expert is.

Experts zijn mensen aan wie we vragen kunnen stellen. *Bloggers* zijn van een ander soort. Het zijn mensen die – individueel of samen met anderen – een openbaar dagboek op het Web bijhouden. Het Web zelf als onderwerp hoeft daarin geen rol te spelen, al kunnen de actualiteit en soms omzwervingen op het Web wel de katalysator voor de ontboezemingen van de blogger vormen.

Weblogs zijn in veel opzichten te vergelijken met nieuwsgroepen, rondzendlijsten en nieuwsbrieven. Ze verschijnen met een zekere regelmaat en oude berichten verdwijnen meestal naar een archief. Verder bieden ze lezers veelal de mogelijkheid te reageren,

iets wat ze met rondzendlijsten gemeen hebben. Ten slotte bieden ze ook vaak de mogelijkheid je erop te abonneren, hetzij per e-mail, hetzij door middel van RSS-feeds (zie blz. 284 e.v.).

2.7.1. Bloggers

Doordat blogs in essentie initiatieven van één persoon zijn (al bestaan er ook community blogs waar meer personen of zelfs tienduizenden mensen aan bijdragen) is de wereld van de blogs veel minder goed georganiseerd dan die van de nieuwsgroepen. Qua organisatie is ze het best te vergelijken met de wereld van de rondzendlijsten: net als daar zijn blogs georganiseerd rond de gebruikte software, in dit geval die van de diensten die ruimte voor een Weblog bieden.

Blogs zijn er over alle mogelijke onderwerpen, van zeer algemeen tot buitengewoon gespecialiseerd, en een directory bestaat er niet van. Er zijn wel overzichten die zich zo noemen, maar die verdelen ofwel de hele blogwereld in tien of twaalf zeer algemene categorieën, zonder verdere onderverdeling, ofwel ze dekken maar een heel klein gedeelte van de meer dan de meer dan 60 miljoen blogs die er momenteel zijn. Dat betekent niet dat ze niet te vinden zijn: er zijn diverse speciale zoekmachines die het mogelijk maken blogs op hun titel of inhoud (de 'threads') te doorzoeken, hetzij om een post of thread over een bepaald feit of een bepaalde gebeurtenis te vinden, hetzij om blogs te vinden over een onderwerp waarover men op de hoogte gehouden wil worden.


Phil Bradley's weblog
Internet searching, web design, search engine developments and anything that will interest librarians!



About

[Add me to your TypePad People list](#)

Categories

- [Books](#)
- [Current Affairs](#)
- [Google specific](#)
- [Image/people search engines](#)
- [Internet searching](#)
- [Meta/multi search engines](#)
- [MSN specific](#)
- [Personal stuff](#)
- [Search engines](#)
- [Utilities](#)
- [Weblogs](#)
- [Webpage design/utilities](#)
- [Web/Tech](#)
- [Yahoo specific](#)

March 14, 2006

Hot Daddy

Link: [Hot Daddy](#). Another Google 'all in one' search interface. This one is slightly different because it allows you to also search Wikipedia, Digg, Creative Commons, IMDb, AltaVista, Wisenut, MSN, Teoma(?) or Yahoo! I suppose that it's helpful, but you do need to know what all the little icons mean in order to use it quickly (though a mouseover tells you). What's odd with it though is that once you've done a search and seen what you've got - you can't use your back button to return to the interface. At least in Firefox, it is though the page is invisible.

[March 14, 2006 in Internet searching](#) | [Permalink](#) | [Comments \(0\)](#) | [TrackBack \(0\)](#)

Recent Posts

- [Amazon Web Services Store](#)
- [Newswine](#)
- [WebFetch.com](#)
- [Hot Daddy](#)
- [Apology from the Spammers!](#)
- [Windows Live Beta](#)
- [Find Similar Sites](#)
- [Google Calendar inches over the horizon](#)
- [Silicon Cloud 12 Ways to Irritate Your Visitors](#)
- [Internet resources Newsletter: Issue 138 March 2006](#)

Subscribe!

SUB NEWSISFREE

+ newsgator

SUB BLOGLINES

Subscribe to this blog's feed

March 09, 2006

Find Similar Sites

Link: [Find Similar Sites](#) with similicio.us. I like this - simply type in a URL and it gives you a list of similar sites, based on del.icio.us and easyutil.com. I tried both my main site and this weblog and it gave me some really good results. This is going to be a great tool if you're having problems finding other sites, and it's far superior to the Google options.

[March 09, 2006 in Internet searching](#) | [Permalink](#) | [Comments \(1\)](#) | [TrackBack \(0\)](#)

March 2006

Sun	Mon	Tue	Wed	Thu	Fri	Sat
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Voorbeeld van een weblog.

Aan het karakter van blogs is verder inherent dat je ze veelal kunt vinden door naar de onderlinge relaties ('wie citeert wie') te kijken. Dat kan via zoekmachines die links analyseren, maar veel bloggers hebben op hun site ook gewoon een lijst staan van andere blogs (een zogeheten *blogroll*) die ze volgen.

Vanuit een oogpunt van zoeken zijn blogs in de eerste plaats een middel om bij te blijven. Dat feit, gecombineerd met het feit dat er speciale zoekmachines voor zijn om ze op te sporen, brengt me ertoe ze niet hier verder te behandelen, maar in het hoofdstuk Bijblijven, nadat we dieper op het onderwerp zoekmachines zijn ingegaan.

2.7.2. 'Experts'

Te vaak wordt vergeten dat het Web niet alleen uit computers en bestanden bestaat, maar ook uit mensen en contacten. In die wereld is plaats voor alle mogelijke vormen van menselijk contact, waaronder ook voor vragen en antwoorden. En dat hoeft ook niet per se in het verband van discussiegroepen.

Experts (in de hierboven gedefinieerde betekenis) zijn op het Web niet moeilijk te vinden. Zoals bij vrijwel elk onderwerp op het Web zijn er verzamelaars actief die alle AskAnExpert diensten (of in elk geval de Engelstalige) in één mooie collectie bij elkaar hebben gezet. Twee voorbeelden zijn Marcus P. Zillman's Internet Experts Subject Tracer Information Blog (<http://www.internetexperts.info/>) en AskA+ Locator (<http://www.vrd.org/locator/subject.shtml>). Dit zijn uitgebreide overzichten van mensen en instanties die vragen willen beantwoorden over recepten, haarproblemen, liefdesperikelen, make-up, computerhardware en -software, varen, aardbevingen, lichamelijke en psychische gezondheid, huiswerk, loodgieten, mythologie, voeding en wat niet al. Zillman's verzameling omvat zowel betaalde als gratis diensten, de Ask-a-Locator uitsluitend gratis experts.

Ask-an-Expert-diensten zijn vooral in de VS, met zijn wat andere cultuur, een wijdverbreid verschijnsel, vandaar dat men daar ook de moeite neemt verzamelgidsen van zulke diensten te maken. In Nederland bestaat bij mijn weten geen werkende website of -pagina waarop Nederlandse vraagbaken bij elkaar gezet zijn, maar ook hier zijn voorbeelden van dit soort diensten te vinden.

Voor feitenvragen, die u anders bij de informatiebalie van een openbare bibliotheek zou stellen, kunt u terecht op de site van de openbare bibliotheken, Al@din, de vraagbaak van bibliotheek.nl (<http://aladin.bibliotheek.nl/>). Deze vragen worden dan door bibliothecarissen beantwoord, soms met een verwijzing naar literatuur waarin het antwoord te vinden is, maar Al@din kan ook een beroep doen op expertise van buitenaf. Een andere site waar feitenvragen worden beantwoord is die van het Knowledge Explosion Network (KEN), <http://www.steljevraag.nl>. Hoewel de site zijn oorsprong heeft bij de TU Delft en het merendeel van de vragen in de bètahoek ligt, worden vragen op andere terreinen niet uitgesloten, en in de vragendatabase zitten er ook een paar over taal. KEN kan ook een beroep doen op experts in andere landen van Europa, als men daarvan gebruik wil maken dient men zijn vraag in het Engels te stellen.

Veel vragen die bij expertdiensten binnenkomen, zijn afkomstig van scholieren die een werkstuk moeten maken of met vragen zitten die op school onbeantwoord blijven, en het is dan ook niet verwonderlijk dat veel expertdiensten zich speciaal op hen richten.

In het biologielokaal van De Digitale School (<http://www.digischool.nl/bi/Vragen/>) kunnen vragen over allerlei biologische onderwerpen worden gesteld, die door docenten worden beantwoord. De biologie is trouwens goed bediend, wat ook op de site van Natuurinformatie (<http://www.natuurinformatie.nl>) wordt de gelegenheid geboden om vragen te stellen, zoals een die mij persoonlijk altijd heeft geïntrigeerd: hoe slaagt een spin erin de horizontale afstand tussen twee punten te overbruggen die soms meer dan een meter uit elkaar liggen en beide op anderhalve meter hoogte. (Het antwoord is: met behulp van de wind.)

Ook de universiteiten, met een schuin oog naar potentieel talent, spelen in dit opzicht een rol. In een Landelijk B-steunpunt (profiel)werkstukken

(<http://www.betasteunpunt.nl/>) bieden de universiteiten van Amsterdam, Delft, Groningen, Leiden en Nijmegen hulp in de vorm van aanvullende informatie, literatuurverwijzingen, aanwijzingen, tips, etc. aan, evenals experimenteerfaciliteiten. De TU Twente doet dat op eigen houtje ook op de Schoolsite Profielwerkstukken (<http://pws.schoolsite.utwente.nl/index.html>).

Alle sites die dit soort vragen beantwoorden stellen wel beperkingen aan zowel de vragen die gesteld mogen worden als de antwoorden die ze daarop zullen geven. Medische en juridische vragen zijn hier niet welkom, en men wordt verondersteld zelf al enige moeite te hebben gedaan om een antwoord te vinden (bijvoorbeeld het raadplegen van een encyclopedie of het zoeken op internet). Heel uitgebreide en diepgravende antwoorden moet u ook niet verwachten, als daaraan behoefte zou bestaan, kan het zijn dat men u op verder voerende literatuur wijst.

Naast deze diensten die overwegend kennisvragen beantwoorden en educatie als hoofdtak hebben, zijn er tal van professionele organisaties die vragen via e-mail (of direct in een chatsessie) beantwoorden als deel van hun voorlichtingstaak. Ik noem hier een paar terreinen waarop professionals of beroepsorganisaties een expertdienst aanbieden.

- Medischestartpagina.nl (http://www.medischestartpagina.nl/Algemene_informatie/Medische_vragen_stellen_per_e-mail/) bevat links naar een aantal sites waar men via e-mail medische vragen kan stellen.
- Food-Info (Wageningen Universiteit) beantwoordt vragen over voeding (<http://www.food-info.net/nl/question.htm>)
- De Jellinek Kliniek (<http://www.jellinek.nl/advies/chatten.php>) biedt vraagstellers de mogelijkheid vragen over alcohol en drugs te mailen, ofwel direct met een hulpverlener te chatten.
- Ouders Online (<http://www.oudersonline.nl/vraagbaken.htm>) heeft een aantal vraagbaken over kinderziekten, opvoeding en gedrag, borstvoeding, seksuele opvoeding etc. waar vragen per e-mail door experts worden beantwoord.
- In veel gevallen bieden bedrijven, instellingen of personen hulp of antwoord op vragen aan in de vorm van een forum, waarop leden van een beroepsgroep of mensen met een bepaalde belangstelling met elkaar communiceren en vragen proberen te beantwoorden. Een voorbeeld is Politie: vraag en antwoord (<http://www.infopolitie.nl/>), waar vragen kunnen worden gesteld over alles wat met het werk van de politie te maken heeft.

Zoals gezegd bestaat er voor Nederland geen handzaam overzicht van Vraag-en-antwoorddiensten, en dus zult u die zelf moeten vinden. Een paar tips hoe je dat doet:

- startpagina's hebben nog wel eens een rubriek Helpdesk, Vraagbaak, Advies enz. waarin dit soort diensten te vinden is. Gebruik StartNederland.nl als grootste collectie van startpagina's, zoek het onderwerp waarover u advies wilt en kijk op de diverse pagina's die daaraan gewijd zijn of u een vraag-en-antwoorddienst vindt. Als die er niet is, kijk dan uit naar een forum (zie hierboven, blz. 31).
- pagina's die toegang tot een expertdienst bieden, bevatten vaak teksten als "hebt u (heb je) een vraag", "vraag per e-mail" of "stel uw (je) vraag". Met die termen kunt u dus in een zoekmachine op zoek gaan, waarbij u de aard van de vraag in een andere zoekterm kunt toevoegen. Voorbeeld: "stel uw vraag" notaris als u iets over nalatenschappen wilt weten. Andere termen die je vaak op dit soort pagina's tegenkomt zijn raadpleeg, FAQ, database en "beantwoorde vragen".

- veel pagina's waarop een vraagformulier wordt aangeboden hebben in de de URL termen die daarop wijzen: vragen, loket, vraagbaak of helpdesk. Ook deze kunt u in combinatie met een andere zoekterm voor het onderwerp gebruiken.

Hierboven noemde ik Marcus P. Zillman's Internet Experts Subject Tracer Information Blog, een overzicht waarin experts aan wie men per e-mail vragen kan stellen bijeen zijn gebracht. Dat waren Amerikaanse sites, en de mogelijkheden om in de VS vragen gratis beantwoord te krijgen zijn inderdaad zo uitgebreid dat een eerste vraag zou kunnen zijn tot welke expert uit de velen die zich beschikbaar stellen men zich in eerste instantie zou moeten wenden. Vanzelfsprekend hangt het antwoord op die vraag af van wie u bent (een volslagen leek of iemand met enige kennis van zaken) en van het antwoord dat u hoopt te krijgen. Een bibliothecaris zal u op dezelfde vraag een ander soort antwoord geven dan een wetenschapper, een leraar aan een middelbare school zal meestal met een ander antwoord komen dan een postdoc aan een universiteit. En weer andere antwoorden zult u krijgen van professionele internetzoekers zoals die bij Google Answers (<http://answers.google.com>) te vinden zijn. Als voorbeelden van diensten met een goede reputatie noem ik een paar sites om u een idee te geven wat u daar te verwachten hebt.

Voor bibliografische vragen en vragen die vermoedelijk door het gebruik van een – al dan niet gespecialiseerd – naslagwerk kunnen worden beantwoord, en voor vragen over controversiële zaken kunt u het best bij bibliothecarissen zijn. Bedenk daarbij dat bibliothecarissen meestal generalisten zijn, met een brede, maar niet noodzakelijk heel diepgaande kennis in het algemeen of op hun vakgebied. Een Ask-an-Expertdienst uit de Amerikaanse bibliotheekwereld is Internet Public Library: Ask a Question (<http://www.ipl.org/div/askus/>). Laat u niet afschrikken door het tamelijk uitgebreide vragenformulier dat men u hier laat invullen, daaruit blijkt alleen maar dat uw vraag serieus wordt genomen. Deze dienst, die wordt bemand door studenten van een aantal Amerikaanse en Canadese bibliotheek- en documentatieopleidingen en professionele vrijwilligers, heeft inmiddels in de loop van tien jaar meer dan 50.000 vragen beantwoord. De tijd die men nodig heeft is drie dagen tot een week, voor vragen die een sneller antwoord behoeven verwijst men u door naar onder meer de Britse dienst Ask a Librarian (<http://www.ask-a-librarian.org.uk/>), die net als het Nederlandse Al@din wordt bemand door medewerkers van openbare bibliotheken en die continue in de lucht is. Hier krijgt u binnen twee dagen antwoord. De beantwoording van vragen bij beide diensten strekt zich ook uit tot gedrukte naslagwerken. Net als Ask a Question bedient Ask a Librarian vraagstellers van over de hele wereld. Tegen de tijd dat u dit leest zult u hier overigens worden doorgeleid naar een nieuwe, nog fraaiere dienst, People's Network Enquire (<http://www.peoplesnetwork.gov.uk/>), waar uw vragen in een directe onlineverbinding met een bibliotheekmedewerker worden beantwoord.

Naast bibliothecarissen zijn ook professionals, met een meer specialistische kennis op de vragenmarkt actief. Nogal wat van deze 'experts' in de VS hebben zich aangesloten bij AskAnExpert-diensten (ook wel 'Q&A services' genoemd), die zorg dragen voor de administratieve afhandeling van vragen, een database (knowledgebase) van vragen en antwoorden onderhouden en de experts soms via een puntensysteem belonen.

AllExperts.com (<http://www.allexperts.com/>) is een voorbeeld van zo'n dienst.

AllExperts is een filiaal van de 'directory plus' About.com ('The Human Internet'), een hoog aangeschreven dienst die het internet via menselijke 'guides' tracht te ontsluiten (zie blz. 85 e.v.); eigenaar van About.com is tegenwoordig The New York Times. Ook AllExperts is opgezet als een doorzoekbare directory, met tal van meer of minder gespecialiseerde onderwerpen, die elk een of meer experts hebben. Prettig is dat van elke expert een profiel beschikbaar is en dat bij dat profiel ook een staatje wordt bijgehouden van tevredenheidsbetuigingen van eerdere vraagstellers.

Van de professionals die bij diensten als AllExperts zijn aangesloten mag je andere antwoorden verwachten dan van bibliothecarissen, en u zelf als vraagsteller zult moeten uitmaken van welk type dienst u gebruik maakt. De keuze is zelfs nog iets groter omdat u zich ook nog tot Google Answers (<http://answers.google.com>) kunt wenden. De experts die u hier aantreft zijn dat vooral in het gebruik van zoekinstrumenten op het Web, in de eerste plaats natuurlijk Google, maar ook andere zoekmachines en zoekinstrumenten, en zelfs ook het onzichtbare Web. Uit de antwoorden blijkt dat een aantal van hen ook de papieren bronnen niet uit de weg gaan. In meer dan een opzicht biedt Google Answers de aantrekkelijkste formule van de diensten die hier vermeld worden. Weliswaar is de dienst niet geheel gratis, maar Google Answers is een echte markt waarop vraagstellers en experts elkaar kunnen vinden en waar beide partijen een maximale vrijheid genieten. Op die markt zijn drie partijen actief:

Question

Subject: **Poem quoted in The 6th lamentation by William Broderick**
Category: [Arts and Entertainment](#)
Asked by: **euterpe7-ga**
List Price: \$4.00

Posted: 18 Apr 2005 06:56 PDT
Expires: 18 May 2005 06:56 PDT
Question ID: 510785

What is the poem quoted in The 6th Lamentation? It is quoted near the end of the book and is about burning autumn leaves.

Request for Question Clarification by [pinkfreud-ga](#) on 18 Apr 2005 10:09 PDT

It would be very helpful if you could give us a few lines of the poem.
Not everyone has access to the book.

Sponsored Links:

Poetry Nu legaal Poetry luisteren, downloaden of branden www.planet.nl	War Poetry Contest \$3,000 in prizes. \$1,500 top prize. \$12 entry fee. May 31 deadline. www.WinningWriters.com	Poetry: we need writers Publish, be read, and get paid. Start writing poetry instantly! www.blogit.com/	Poetry Writing Contests Share your poetry. Have fun. Writing contests, reviews and fun http://www.fanstory.com
---	--	--	--

Answer

Subject: **Re: Poem quoted in The 6th lamentation by William Broderick**
Answered By: [jab-ga](#) on 18 Apr 2005 16:01 PDT

Here's is the poem:

http://www-biol.paisley.ac.uk/bulletin/Poem_leaves.htm

The Burning of the Leaves, by Laurence Binyon

Comments

[Log in](#) to add a comment

There are no comments at this time.

Important Disclaimer: Answers and comments provided on Google Answers are general information, and are not intended to substitute for informed professional medical, psychiatric, psychological, tax, legal, investment, accounting, or other professional advice. Google does not endorse, and expressly disclaims liability for any product, manufacturer, distributor, service or service provider mentioned or any opinion expressed in answers or comments. Please read carefully the [Google Answers Terms of Service](#).

If you feel that you have found inappropriate content, please let us know by emailing us at answers-editors@google.com with the question ID listed above. Thank you.

Search Google Answers for all questions

Een beantwoorde vraag bij Google Answers. Van boven naar beneden: de vraag, met daaronder een verzoek om een nadere toelichting van een van de researchers, en daaronder het antwoord van een andere expert. Er zijn bij deze vraag geen comments (suggesties voor de beantwoording) binnengekomen.

- in de eerste plaats vraagstellers, waarvoor als enige eis is dat ze zich (gratis) bij Google registreren
- experts, door Google aangeduid als 'researchers', die na een test op de markt worden toegelaten en die ook nadien door de dienst op de kwaliteit van hun werk worden gecontroleerd
- geregistreeerde gebruikers, dat zijn personen die bij Google geregistreeerd zijn en die – behalve vragen stellen – ook suggesties voor de beantwoording van de vragen van anderen mogen leveren. Ze mogen geen antwoorden geven, dat mogen alleen researchers.

Google Answers is niet gratis, voor iedere vraag betaalt u Google als aanbieder van de dienst 50 dollarcent, Daarbovenop komt een bedrag van minimaal \$2 en maximaal \$200 voor de researcher die uw vraag beantwoordt, maar voor het overige bepaalt u zelf hoeveel u voor de beantwoording van uw vraag over hebt. U bent ook degene die bepaalt of een vraag naar uw tevredenheid beantwoord is. Ook bepaalt u bij het stellen van de vraag binnen welke termijn het antwoord gegeven dient te worden. De standaard termijn is een maand.

Een aantrekkelijk aspect van de formule van Google Answers is dat in principe iedereen, researchers en geregistreeerde gebruikers in de vorm van suggesties ('comments') aan de beantwoording van uw vraag kan meewerken. Daaraan komt pas een eind als een researcher de beantwoording van uw vraag op zich heeft genomen en uw vraag daardoor de status 'locked' heeft gekregen. Vanaf dat moment heeft de researcher 4-8 uur om met een antwoord te komen en kunnen andere researchers en geregistreeerde gebruikers geen bijdragen aan de beantwoording ervan leveren. Alle deelnemers kunnen om een nadere toelichting vragen, researchers en geregistreeerde gebruikers op de vraag, en de vraagsteller op het antwoord.

Alvorens een vraag te stellen kunt u de database van vragen raadplegen om te zien of het antwoord al eerder gegeven is. Het aantal vragen dat al gesteld is bedraagt meer dan 90.000, maar niet alle vragen zijn beantwoord.

Vragen beantwoorden is een vak, vragen stellen een kunst. Als u een expert inschakelt om gratis of voor een geringe vergoeding het antwoord op een vraag te vinden, motiveert u hem of haar om een optimaal resultaat te bereiken door:

- eerst zelf moeite te doen om een antwoord te vinden bij uw vraag te vermelden welke stappen u zelf al ondernomen heeft
- de vragendatabase te raadplegen en de Help/FAQ te lezen voordat u de vraag stelt
- de vraag zo concreet mogelijk te stellen, met vermelding van wat u wel en wat u niet wilt weten, en verzoeken om een toelichting prompt te beantwoorden.
- de expert beleefd te behandelen, ook als hij/zij geen antwoord op uw vraag kan verschaffen.
- een oordeel over het antwoord en/of de expert te geven als u daarom gevraagd wordt.

Als u antwoord krijgt, kan het zijn dat dat gebeurt in de vorm van een berichtje van de dienst (niet van de 'expert' zelf) dat het antwoord op uw vraag op de server staat en daar een paar dagen zal blijven. Wees er dan wel snel bij om het te lezen, anders is het weg. Overigens: als u na lezing van dit boek nog vragen hebt over zoeken op het Web, bij AllExperts.com zijn ook experts beschikbaar die u daarbij willen helpen: Allexperts Searching the Internet Q&A (<http://www.allexperts.com/getExpert.asp?Category=1557>).

3. Het hoe van het zoeken

In dit hoofdstuk kijken we naar de keuze tussen de diverse zoekinstrumenten – wanneer gebruiken we een directory, wanneer een zoekmachine, wanneer een metazoekmachine? – en naar de syntax – velden, operatoren enz. – van het gebruik van deze zoekinstrumenten. Doel van dit hoofdstuk is u een idee te geven van de verschillende doeleinden waarvoor de verschillende zoekinstrumenten kunnen worden gebruikt, en van de zoekmogelijkheden die ze in het algemeen gesproken bieden.

In het vorige hoofdstuk zijn de belangrijkste middelen opgesomd die het Web (of, als het om nieuwsgroepen gaat internet) ons biedt om antwoorden op vragen te krijgen. In dit hoofdstuk gaan we kijken hoe we die middelen kunnen gebruiken. Daarbij beperken we ons tot het gebruik van directories of webgidsen, startpagina's en zoekmachines. De zoekmogelijkheden van andere instrumenten die in het vorige hoofdstuk zijn besproken, zijn doorgaans veel eenvoudiger dan die van beide genoemde soorten zoekinstrumenten. Als u daarmee kunt zoeken, kunt u het in (de archieven van) nieuwsgroepen en rondzendlijsten, FAQs en databases of directories van experts en dergelijke zeker.

Zoals zoveel zaken die met internet, WWW, en meer in het algemeen informatie- en communicatietechnologie (ICT), te maken hebben, hebben zoekinstrumenten hun eigen jargon. Als u dat eens wilt bekijken of er iets over wilt nakijken, maak dan gebruik van de volgende verklarende woordenlijsten:

Search Terms Glossary (<http://www.searchtools.com/info/glossary.html>)

Lexicon van *Search Engine* vaktermen (NL-versie) (<http://www.cetest.nl/nlgloss.htm>)

Een aantal termen zijn achterin dit boek opgenomen en verklaard.

3.1. De keuze van het juiste zoekinstrument

Wel is het belang er nog eens op te wijzen dat een goede zoekactie begint met de keuze van het juiste zoekinstrument. De diverse zoekinstrumenten zijn in het vorige hoofdstuk al opgesomd, hier noem ik ze nog eens met als uitgangspunt verschillende typen vragen.

Als u

- op zoek bent naar een of enkele documenten over een niet al te specifiek en al te specialistisch onderwerp
- zich wilt oriënteren op een terrein waarop u niet thuis bent, maar dat u meer in detail wilt leren kennen
- een specialistische vraag hebt, maar u weet niet in welke termen die moet worden gegoten

gebruik dan een mondiale, algemene webdirectory zoals Yahoo! of, als uw onderwerp daarin behandeld wordt, een verzameling van webgidsen zoals about.com.

Als u

- op zoek bent naar een of enkele documenten op een niet al te specifiek en al te specialistisch terrein dat u niet kent, in uw eigen taal of een andere taal dan het Engels
- informatie zoekt binnen een beperkt geografisch gebied (musea in de provincie Groningen, wandeltochten in de Ardennen)

begin dan te zoeken met een nationale of regionale directory of startpagina zoals het Open Directory Project Nederlands of Startnederland.nl.

Als u

- zeer gerichte en gedetailleerde informatie zoekt, waarover vermoedelijk niet zo heel veel relevante documenten te vinden zullen zijn, en u weet welke zoektermen u moet gebruiken om die informatie met een zoekmachine te vinden

gebruik dan een (meta)zoekmachine zoals Google, Yahoo! of Jux2. Gebruik die ook om snel iets op te zoeken.

Hier een aantal voorbeelden ter verduidelijking van de keuze tussen een directory en een zoekmachine

Directory	Zoekmachine
Een directory brengt u naar de homepage van een Website. Daar gebruikt u de eigen zoek- en navigatiemogelijkheden van die site om precies te vinden wat u zoekt. U vindt die informatie dan in een context.	Een zoekmachine brengt u direct naar de pagina('s) waarop de door u opgegeven zoektermen gevonden worden. Het kan zijn dat u direct raak schiet, maar als dat niet zo is mist u de context.
Gebruik een directory om u te oriënteren , zowel wat het precieze onderwerp betreft als wat de termen betreft waarin de vraag moet worden geformuleerd.	Gebruik een zoekmachine als u precies weet wat u wilt vinden , hoe u de vraag moet formuleren en hoe u de zoekmachine moet gebruiken . Gebruik een zoekmachine ook voor vragen waarop een directory geen antwoord heeft.
Als u een site zoekt over de Boerenoorlog in Zuid-Afrika en u weet weinig over de Zuidafrikaanse geschiedenis, gebruik dan een directory.	Als u informatie zoekt over een incident dat in 1880 de aanleiding vormde tot die oorlog en dat de ' Bezuidenhout affair ' wordt genoemd, zoek dan met een zoekmachine
Gebruik een directory om een overzicht te vinden van Nederlandse kranten op het Web.	Gebruik een zoekmachine om het gedicht ' Herinnering aan Holland ' van H. Marsman te vinden.
Gebruik een directory om een biografie van Beethoven te vinden.	Gebruik een zoekmachine om uit te vinden of er een opname van Beethovens vioolconcert met de violiste Mila Georgieva bestaat
Gebruik een directory als u in de stemming bent voor serendipiteit .	Gebruik een zoekmachine om snel een exact gegeven te vinden.
Gebruik een directory om de tekst van het bijbelboek Esther te vinden.	Gebruik een zoekmachine om uit te vinden waar het citaat ' For now we see through a glass, darkly ' vandaan komt.
Gebruik een directory als u wilt weten hoe het auteursrecht op het Web geregeld is .	Gebruik een zoekmachine om uit te vinden of uw webpagina's geplagieerd worden .
Gebruik een directory om een site met biografieën van beroemde filmsterren te vinden.	Gebruik een zoekmachine om uit te vinden in welk toneelstuk Han Bentz van den Berg en Ank van der Moer in 1962 samen speelden .

Ten dele gebaseerd op: Bob Paton, Doing Research on the Search Engines (<http://studentnurses.com/search/searchtips.html>).

Zelfs met deze voorbeelden blijft de keuze tussen een directory en een zoekmachine er toch vaak nog een die meer te maken heeft met de vraagstelling dan met de vraag. Specifieke vragen kun je immers vaak ook op een wat algemener niveau bij een directory stellen, net zoals algemenere vragen soms met (een combinatie van) zeer specifieke zoektermen in een zoekmachine kunnen worden beantwoord. De specifieke vraag naar het lot van Balthasar Gerards kan ook worden gesteld als een algemene vraag naar de biografie van Willem van Oranje of zelfs de geschiedenis van Nederland.

Met de meeste mondiale (meta)zoekmachines kunt u in alle talen in Latijns schrift

zoeken, en uw zoekactie desgewenst ook tot een bepaald land of een bepaalde taal beperken. Dat laatste hoeft u natuurlijk alleen te doen als de zoekterm(en) met een andere dan de bedoelde betekenis in andere talen (speciaal het Engels) voorkomen. Als u op zoek bent naar een Nederlands gedicht waarin de woorden kind en lover (gebladerte) voorkomen, gooien de Engelse woorden kind (zachtaardig) en lover (minnaar) roet in het eten als u niet aangeeft dat u op zoek bent naar Nederlandstalige documenten.

Bij veel vragen die (in eerste instantie) eenvoudig en heel goed door directories zouden kunnen worden beantwoord, gebruiken onervaren zoekers ten onrechte een (meta)zoekmachine. Als u moeite hebt uw vraag te formuleren en/of u kunt die maar in één zoekterm vatten, begin dan een directory te gebruiken. Directories helpen u namelijk niet alleen doordat u daarin een route van algemeen naar bijzonder kunt volgen, maar ook doordat hun zoekmachine bij de door u opgegeven zoektermen eerst de rubrieken zoekt waarin die termen voorkomen. Zo'n rubriek helpt u meteen uw vraag beter te formuleren en suggereert u ook meer of andere zoektermen. Een paar voorbeelden van 'vragen' die begin mei 2005 bij de Nederlandse zoekmachine Kobala werden gesteld, zijn:

```
korte verhalen
+kinderen
+Libië
morning after pil
voetbalshirts
+hotel +onderweg
file
Georgië
+all +inclusive +turkije
5 mei
+acacia
```

Het is duidelijk dat de vragen die achter deze zoektermen schuilgaan – welke dat ook zijn – ofwel veel preciezer aan de zoekmachine hadden moeten worden gesteld, ofwel op een algemener niveau aan een directory of een startpagina.

Directories geven een betekenisloze vraag betekenis en context. Zo komt de Nederlandse afdeling van het Open Directory Project bij de zoekterm *kinderen* met de volgende rubrieken:

1. World: Nederlands: Gezondheid: Kinderen
2. World: Nederlands: Maatschappij: Organisaties: Ontwikkelingssamenwerking
3. World: Nederlands: Webwinkelen: Kinderen
4. World: Nederlands: Huis en Tuin: Familie: Kinderen
5. World: Nederlands: Onderwijs

Wie meer over korte verhalen, voetbalshirts, files, Libië of haarproblemen wil weten, kan daarover via Startnederland.nl (<http://www.startnederland.nl>) ettelijke startpagina's vinden die het onderwerp verder onderverdelen.

Zoekmachines daarentegen zijn eerder geschikt voor de volgende vragen:

```
Monitoren Test
catherine de neuve
dutch english dictionary
Vlaamse Squashfederatie
```

```
"jack nicholson"  
Anwb Online Routeplanner Europa  
antwerps woonburo  
Erasmusbrug geschiedenis  
middeleeuwse steden
```

Het gaat hier om vragen met een specifiek karakter of waarbij de vraagsteller het bestaan van een bepaalde website al kent.

Terwijl het voor vragen met een nationale of regionale achtergrond beslist aan te raden is een Nederlandse directory of startpagina te gebruiken in plaats van een internationale, is dat advies niet zonder meer van toepassing als het gaat om de keuze tussen mondiale en nationale zoekmachines. De bekendste Nederlandse zoekmachine is Ilse, de grootste Vindex. Elk van deze beide zoekmachines heeft zijn eigen sterke punten (Ilse is selectief, Vindex is omvangrijk), maar zowel wat de dekking van het Nederlandse Web als wat de zoekfunctionaliteit betreft blijven ze achter bij een mondiale metazoekmachine als Jux2 of een zoekmachine als Google . De keuze van het verkeerde zoekinstrument is een van de meest gemaakte fouten bij het zoeken op het Web.

3.2. Zoeken in directories

Directories (in Nederland ook wel webindexen of webgidsen genoemd) zijn redelijk eenvoudig te gebruiken. Zoals al eerder gezegd voeren ze hun gebruikers van zeer algemene naar steeds specifiekere onderwerpen. Men noemt zo'n systeem wel een zoekboom: vanuit de stam waaiert de boom uit in een zich steeds verder vertakkend netwerk van takken en twijgen. Door steeds keuzes te maken op pagina's van rubrieken en subrubrieken kom je uiteindelijk bij een aantal links uit die betrekking hebben op het (tamelijk brede) onderwerp dat je interesseert.

Op het eerste niveau vind je tussen tien en twintig hoofdrubrieken ('main categories'), die vaak al onderverdeeld zijn in subrubrieken (in een kleiner lettertype). Deze geven al een idee van de onderwerpen. Klikken op een hoofdrubriek brengt je naar een volgend niveau, waar je alle subrubrieken binnen die hoofdrubriek aantreft, bij veel directories gepaard gaande met al een aantal links op dat nog tamelijk algemeen niveau.

Een van de aantrekkelijke eigenschappen van directories is dat ze de zoekroute weergeven die men bij het maken van keuzes volgt. Ze doen dat zowel in de URL als op de gepresenteerde webpagina's. Daardoor blijf je niet alleen zien hoe je bij een bepaalde subrubriek bent aangeland, maar kun je desgewenst ook gemakkelijk weer naar een hoger niveau teruggaan om vandaar andere keuzes te maken ('een andere tak van de boom in te gaan'). Ook is het mogelijk de URL die bij een bepaalde subrubriek behoort in je bookmarks of als link in een eigen webpagina op te nemen. Bij het aanklikken daarvan kom je dan meteen op de juiste plaats in de directory terecht.

Bij gebruik van de zoekboom laat het Open Directory Project de daarbij gevolgde weg zowel in de URL als bovenin het resultaatvenster zien. Het eerste kan worden gebruikt om de zoekactie als een link op te slaan, het tweede om door het aanklikken van een deel van de route op een algemener niveau in de directory uit te komen.

De presentatie van de route bij het zoekresultaat in directories heeft nog een voordeel: als u de zoekfunctie van de directory gebruikt (dus met woorden gaat zoeken in de zoekregel), krijgt u bij de beste directories bij het zoekresultaat ook de rubrieken en subrubrieken (als een route) gepresenteerd; u kunt dan kijken welke andere sites de betreffende rubriek bevat naast de site die u met uw zoektermen hebt gevonden. En dit opent dan weer nieuwe perspectieven voor verdere zoekacties. Meestal biedt de zoekmachine van een directory u de mogelijkheid de hele directory te doorzoeken, dan

[over dmoz](#) | [URL aanmelden](#) | [bijwerken vermelding](#) | [woord redacteur](#) | [meld misbruik/spam](#) | [help](#)

in de gehele index

Top: [World](#): [Nederlands](#): [Kunst](#): [Schilderkunst](#): [Kunstschilders](#) (729) [Omschrijving](#)

- ♦ [Abstract](#) (27)
- ♦ [Portretten](#) (62)
- ♦ [Realisme](#) (65)

Zie ook:

- ♦ [World: Nederlands: Kunst: Beeldende Kunst: Kunstenaars](#) (157)
- ♦ [World: Nederlands: Kunst: Geschiedenis: Kunstenaars](#) (49)

Deze categorie met sites in andere talen:

[Catalaans](#) (119) [Duits](#) (525) [Engels](#) (4.336)
[Fins](#) (44) [Frans](#) (762) [Italiaans](#) (1.316)
[Japans](#) (133) [Pools](#) (82) [Roemeens](#) (33)
[Spaans](#) (736)

- ♦ [Aanstoot, Jan](#) - Leven en werk van de Wierdense kunstenaar.
- ♦ [AbsArt - Moderne kunst](#) - Permanente expositie van het werk van Eddy Sonnenberg en Gerrit de Niet
- ♦ [Acampora, Elisabetta](#) - Werk en info van deze Nederlands/Italiaans beeldend kunstenaar.

De rubriek Kunstschilders in het Open Directory Project. Binnen deze categorie zijn de schilders alfabetisch geordend.

wel de rubriek waar u zich op dat moment bevindt (met alle onderliggende rubrieken). Sommige directories gaan nog verder: ze bieden meteen ook de mogelijkheid de zoekactie op het hele Web uit te voeren, dus gebruik te maken van een mondiale zoekmachine. U zoekt dan dus in twee linkdatabases: die van de directory, met de links die door de redacteurs zijn geannoteerd en in rubrieken ingedeeld, en die van de zoekmachine, met de door de spiders van de zoekmachine verzamelde en door de indexer geïndexeerde links.

Zoals gezegd werkt een directory met een 'zoekboom'. In de takken daarvan moeten alle onderwerpen worden ondergebracht die de directory te bieden heeft. Daarbij komt het natuurlijk regelmatig voor dat onderwerpen op twee of meer verschillende plaatsen in de boom verschijnen. Een onderwerp als 'computer science' bij voorbeeld kan als subrubriek zowel in de rubriek 'Computers & Internet' als in de rubriek 'Science' voorkomen. In zo'n geval zetten de directories zo'n onderwerp op één plaats, met een verwijzing daarnaar van uit de nadere 'tak' van de boom. Zo'n verwijzing is doorgaans te herkennen aan het @-teken bij de naam van het onderwerp. Als u op zo'n onderwerp klikt, zult u zien dat de route naar het onderwerp verandert: zo kan in Yahoo! de door u gekozen route naar de Openbare Bibliotheek Den Haag:

Directory > Reference > Libraries > Public Libraries > By Region > Countries > Netherlands > Provinces > Zuid-Holland@>

op het moment dat u in deze rubriek 'Zuid-Holland@' kiest veranderen in:

Directory > Regional > Countries > Netherlands > Provinces > Zuid-Holland > Community > Libraries

Verwijzingen in een directory voegen veel toe aan de kwaliteit daarvan, omdat de kans dat je als gebruiker verdwaald raakt, erdoor vermindert. Voor de kwaliteit van een directory is verder ook van belang of de diverse typen resources die het Web te bieden heeft op diverse plaatsen in de boom consequent benoemd worden. Bij voorbeeld: wat *Journals* heet in de rubriek News and Media, moet niet opeens *Periodicals* heten in de rubriek Science en *Serials* in de rubriek Health.

Bij de websites die een directory binnen een rubriek presenteert kan een verdere indeling gemaakt worden tussen diverse typen sites. Directories zijn commerciële bedrijven en zien er dus geen been in om bepaalde sites tegen betaling extra in het zonnetje te zetten ('Sponsor listings') of om bedrijven in een rubriek te laten adverteren ('Sponsor results'). Bij goede directories zijn die duidelijk onderscheiden van de niet gesponsorde sites. Daarnaast kan het zijn dat de directory de sites die binnen een rubriek gevonden worden op twee manieren kan presenteren: alfabetisch en in volgorde van 'popularity' (voor wat daaronder kan worden verstaan, zie blz. 55).

Rubrieken hoeven niet altijd gevuld te zijn. Speciaal op de algemenere niveaus treft men nog wel eens rubrieken aan die geen of alleen 'sponsored' resultaten bevatten.

3.3. Zoeken met zoekmachines en metazoekmachines

Als u een directory gebruikt om een document te vinden over een onderwerp dat u interesseert, wordt u als het ware bij de hand genomen en naar dat onderwerp toe geleid. Gebruikt u een zoekmachine, dan krijgt u die hulp niet. U zult de zoekmachine opdrachten moeten geven om duidelijk te maken wat u wilt. Daarbij zult u zich (doorgaans) moeten aanpassen aan de taal die de zoekmachine verstaat. En dat is niet de taal waarin u zich normaal gesproken uitdrukt.

Als wij een ander mens een vraag stellen, doen we dat in wat bekend staat als 'natuurlijke taal'. Daarbij gebruiken we woorden en zinnen die uitdrukken wat we willen en andere woorden en zinnen die bedoeld zijn om onze gesprekspartner ertoe te brengen aan die wens te voldoen. En omdat wij met intelligente gesprekspartners te maken hebben, zijn er tal van manieren waarop je dezelfde vraag kunt stellen.

'Wie is de schrijver van "Oorlog en vrede"?"

"Kunt u mij zeggen wie de schrijver is van de roman 'Oorlog en vrede'?"

'Wie heeft het boek "Oorlog en vrede" geschreven?"

'Van welke grote Russische schrijver is ook weer het boek dat over Napoleons Russische veldtocht in 1812 gaat?'

Een mens kan in al deze vraagstellingen onderscheiden wat de essentiële vraag is, en wat de franje. Een computer kan dat veel minder goed. Als u een zoekmachine een van deze vragen zou stellen, zou de zoekmachine de woorden 'oorlog' en 'vrede' in principe als net zo belangrijk of onbelangrijk voor de vraagstelling behandelen als de woorden 'kunt', 'zeggen', 'wie', 'geschreven' enz. De enige remedie die zoekmachines voor dat probleem hebben is er een van statistische aard: zoekmachines kunnen wel berekenen of het woord 'vrede' zeldzamer (en dus waarschijnlijk betekenisvoller) is dan 'wie'.

Natuurlijk is het ook mogelijk een computer zo te programmeren dat bepaalde aan het programma opgegeven woorden worden genegeerd.

Alles wat een computer kan is rekenen en vergelijken. Als u met een zoekmachine een vraag stelt in de hoop een document te vinden dat het antwoord op die vraag geeft, vergelijkt de computer de bewoordingen van uw vraag met de bewoordingen in de documenten waarin hij zoekt – en dat niet op basis van de betekenis van woorden, maar op basis van de bits en bytes van de letters waaruit die woorden bestaan. Wat de

computer van u verwacht is dus eigenlijk dat u hem de termen opgeeft die in (het document met) het antwoord moeten voorkomen. In zekere zin moet de vraag die u aan de zoekmachine stelt dus zoveel mogelijk gelijk zijn aan het antwoord op de vraag.

Het kan nog wel eens voorkomen dat een zoekmachine documenten vindt die de door u opgegeven zoekterm(en) helemaal niet (lijken) te bevatten. Daarvan kunnen diverse oorzaken zijn, waarvan ik er hier twee noem:

- de pagina die u gevonden hebt is gewijzigd sinds ze door de zoekmachine geïndexeerd is.
- de zoekterm(en) is (zijn) er wel, maar in een niet zichtbaar deel (vooral de METAtags) van de pagina. Controleer dit met View (Page) Source in het View-menu van uw browser.

Zoeken met zoekmachines is dus het opgeven van zoektermen met als doel relevante documenten te vinden en niet relevante documenten uit te sluiten. Hoe meer termen een zoekvraag bevat die overeenkomen met de termen die relevante documenten bevatten, des te groter de waarschijnlijkheid dat u het antwoord op uw vraag zult vinden. En hoe meer termen de zoekvraag bevat die niet in relevante documenten voorkomen maar wel in niet-relevante, des te groter de waarschijnlijkheid dat u het antwoord op uw vraag niet vindt.

In het geval van bovenstaande vraag 'Wie is de schrijver van "Oorlog en vrede"?' zou de zoekvraag die aan de zoekmachine wordt gesteld in elk geval de termen oorlog en vrede moeten bevatten, want die verwachten we in relevante documenten met het antwoord op onze vraag 'wie is de schrijver van...' te zullen vinden. Maar die termen zijn niet voldoende, want er zijn natuurlijk legio documenten die de woorden 'oorlog' en 'vrede' kunnen bevatten zonder dat het daarin over de roman van Tolstoj gaat. De zoekvraag zal dus moeten worden uitgebreid met een term die aanduidt dat de vraag over een literair werk gaat, dus bijvoorbeeld de term roman. Schrijver, auteur en boek zijn ook mogelijk, maar als we ons afvragen welk woord het meest waarschijnlijk voorkomt in documenten die het antwoord op onze vraag bevatten, dan komt roman het meest in aanmerking. Nog beter zou natuurlijk de term Tolstoj zijn, verondersteld dat u die weet.

Nu u weet wat zoekmachines met uw vraag doen, zult u het hopelijk voortaan uit uw hoofd laten uw vraag in één zoekterm te vatten, zoals veel onervaren zoekers helaas nog altijd doen. Stel u als regel:

Een zoekvraag dient die zoektermen te bevatten die we in relevante documenten met het antwoord op onze vraag verwachten aan te treffen – niet meer en niet minder.

3.3.1. *Booleaanse operatoren*

Als we een zoekvraag stellen door de zoektermen oorlog, vrede en roman in de zoekregel van de zoekmachine in te typen, gaan we er stilzwijgend van uit dat de zoekmachine ons documenten zal leveren waarin al die woorden voorkomen. Maar dat is niet vanzelfsprekend. Sterker nog: het is ook niet altijd wat we willen. Als u op het Web op zoek gaat naar documenten over de reparatie van fietsen, kunt u de zoekmachine de zoektermen reparatie en fietsen opgeven. Maar in plaats van de term fietsen zouden we ook rijwielen kunnen gebruiken. En als de zoekvraag dan zou luiden `reparatie fietsen rijwielen`, zou u daarmee documenten verwachten te vinden waarin ofwel de combinatie 'reparatie EN fietsen' voorkwam, ofwel de combinatie 'reparatie EN rijwielen'. Dat alle drie de termen in een document voorkomen is weliswaar mogelijk, maar niet zo waarschijnlijk.

De regel die hierboven is geformuleerd, houdt dus niet alleen in de zoektermen die we in combinatie met elkaar verwachten tegen te komen, maar ook de zoektermen die we als alternatief of als synoniem voor elkaar hopen te vinden. Dat is net zo belangrijk als het gebruik van zoektermen in combinatie met elkaar, want u vindt er andere documenten over hetzelfde onderwerp mee.

Hier rijst echter een probleem: aangezien een zoekmachine niet over menselijke intelligentie beschikt, kan hij niet uitmaken of u met de zoekvraag *reparatie fietsen rijwielen* bedoelt:

reparatie EN rijwielen EN fietsen

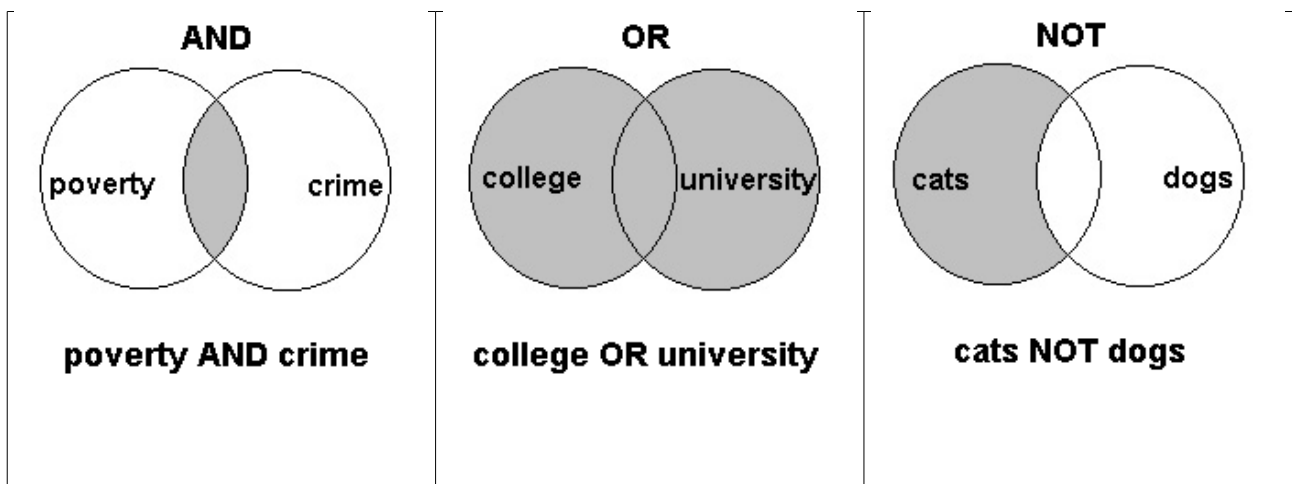
reparatie OF rijwielen OF fietsen

reparatie EN rijwielen OF fietsen

reparatie OF rijwielen EN fietsen

met andere woorden: moeten de termen in combinatie voorkomen, of als alternatief, of misschien beide?

De simpelste manier om dat probleem op te lossen is dat een zoekmachine u als gebruiker de gelegenheid geeft aan te geven wat u bedoelt, anders gezegd, de relatie tussen termen te expliciteren. Dat gebeurt door de woordjes AND (EN) en OR (OF), die in dit verband bekend staan als 'logische' of 'Booleaanse operatoren'. Naast AND en OR is er nog een derde Booleaanse operator, namelijk NOT (NIET, ook wel AND NOT, soms ANDNOT). Booleaanse operatoren drukken de relatie tussen zoektermen uit.



De Booleaanse operatoren AND, OR en NOT. Het grijs gekleurde gedeelte is de verzameling ('set') die bij elk van de operatoren gevonden wordt. Daar waar de twee cirkels elkaar in de middelste tekening overlappen gaat het om documenten die zowel de term 'college' als de term 'university' bevatten. Als u op zoek bent naar documenten die een van beide termen bevatten, vindt de zoekmachine er daarbij ook waarin beide voorkomen.

Tip: als u met Booleaanse operatoren zoekt, wen u dan aan die altijd met hoofdletters te schrijven AND, OR, NOT. Sommige zoekmachines vereisen dat, en bovendien, het brengt visueel meer structuur in uw vragen.

De meeste zoeksystemen gaan er stilschijgend van uit dat als u meerdere zoektermen opgeeft, de door u bedoelde operator wel EN zal zijn. Hierbij hoeft u AND dus niet op te geven. Er zijn er echter ook die standaard OR gebruiken. Als een zoekmachine u uitnodigt uw zoektermen zonder operator in te typen, ga dan altijd even na wat die standaard is.

Een andere mogelijkheid, die u bij veel zoekmachines vindt, is dat u de keuze wordt geboden tussen zoeken met *All of the words* en *Any of the words*. Zoals u al begrijpt is dit de keuze tussen de operator AND en de operator OR.

Meer over Booleaanse operatoren: Boolean Searching on the Internet
(<http://library.albany.edu/internet/boolean.html>)

3.3.1.1. Nesting

Kunt u zich nog de zin 'Mijnheer Van Dale wacht op antwoord' herinneren? Het was een ezelsbruggetje om de volgorde van rekenkundige bewerkingen te onthouden: machtsverheffen, vermenigvuldigen, delen, worteltrekken, optellen, aftrekken. Op een heel simpel niveau betekende dat dat je in een som die zowel een optelling als een vermenigvuldiging bevatte de op te tellen getallen tussen haakjes moest zetten. Als je dat niet deed ging vermenigvuldigen voor:

$$8 \times 6 + 5 = 53$$

$$8 \times (6 + 5) = 88$$

Zo is het ook met Booleaanse operatoren: er is een volgorde van prioriteit. In de regel komt (AND) NOT voor AND, en AND voor OR. Bij gebruik van twee verschillende operatoren in een zoekvraag moeten de met OR verbonden termen dan dus tussen haakjes worden geplaatst: reparatie AND (fietsen OR rijwielen). Het gebruik van haakjes om in een zoekvraag met twee verschillende Booleaanse operatoren de combinatie van zoektermen in een andere dan de standaardvolgorde te doen verlopen heet *nesting*. Als zoekmachines nesting mogelijk maken, worden daarvoor vrijwel altijd ronde haken gebruikt.

Bij zoekmachines die zoektermen standaard met AND verbinden is nesting niet nodig. Als ze naast dat verzwegen AND ook het gebruik van de operator OR toelaten, betekent het voorkomen van OR in de zoekvraag automatisch dat de OR-bewerking prioriteit heeft.

3.3.1.2. Nabijheidsoperatoren ('proximity operators' of 'adjacency operators')

Bij zoekmachines die met Booleaanse vragen werken vinden we nog wel eens de extra mogelijkheid om nabijheidsoperatoren te gebruiken, waarmee we de zoekmachine niet alleen opdragen een AND-relatie tussen zoektermen te leggen maar daarbij ook nog de voorwaarde kunnen stellen dat die zoektermen a. in een bepaalde, opgegeven volgorde in de te vinden documenten voorkomen en b. een maximaal aantal woorden of letters uit elkaar staan. Als varianten vinden we ook wel dat de zoektermen in dezelfde zin dan wel in dezelfde alinea moeten voorkomen.

Van de grote mondiale zoekmachines is er nog maar één, Exalead, die nabijheidsoperatoren (NEAR) ondersteunt; bij AltaVista, dat vroeger die operator ook had, werkt hij niet meer. Niettemin doet u er goed aan bij een onbekende zoekmachine uit te zoeken of u er met nabijheidsoperatoren kunt zoeken, want ze maken precieze zoekacties mogelijk. Kijk dan ook of ze de mogelijkheid bieden de nabijheid verder te specificeren: gevolgd door een getal (al dan niet met een slash tussen de operator en het getal) geeft de nabijheidsoperator dan het maximale aantal woorden of letters aan dat tussen de zoektermen mag voorkomen. NEAR/10 betekent dan dus: met een maximale tussenruimte van tien woorden.

Dat nabijheidsoperatoren nog zo zelden gebruikt worden heeft overigens wel een reden: bij de meeste zoekmachines speelt de afstand tussen de opgegeven zoektermen in de gevonden documenten een grotere (Google) of minder grote (MSN Search) rol in de ranking (zie blz. 54 e.v.) van het zoekresultaat.

Een bijzondere vorm van nabijheidszoeken vinden we bij Google (en tot voor kort via een trucje ook bij Yahoo!). Deze bestaat erin dat in een phrase (zie de volgende paragraaf) een woord door een wildcard (bij Google een *) kan worden vervangen.

“Berend Botje ging * varen” levert dus zowel “Berend Botje ging uit varen” als “Berend Botje ging eens varen” op. De truc bij Yahoo! was dat Yahoo! met stopwoorden (zie onder) werkte, die niet helemaal verwaarloosd werden, maar in phrases als een willekeurig woord werden opgevat.

Meer over nabijheidsoperatoren: Boolean Searching on the Internet (<http://library.albany.edu/internet/boolean.html>)

3.3.1.3. Phrases

Nu we het toch over nabijheid hebben, een heel speciaal geval van nabijheid (dat overigens niets met Booleaans zoeken te maken heeft) vormen phrases. "Derde Wereld", "Tweede Kamer", "vitamine C" zijn daar voorbeelden van: termen die uit meer dan een woord bestaan en een eigen betekenis hebben. Vooral het Engels is er rijk aan, doordat deze taal zelfstandige naamwoorden die samen één term vormen, meestal uit elkaar houdt: real estate developer (projectontwikkelaar), vacuum cleaner (stofzuiger), enz.

Doordat het Engels zo rijk is aan phrases en doordat zoekmachines een van oorsprong Amerikaans verschijnsel zijn behoort het zoeken met phrases gelukkig tot de standaardmogelijkheden van zoekmachines. Gelukkig, want nu kunnen we ze niet alleen gebruiken voor de samengestelde termen waarvoor ze eigenlijk bedoeld zijn, maar ook om langere zinsnedes en zelfs zinnen als zoekvraag op te geven. Bij documenten die full-text geïndexeerd zijn, kan dat immers. En naarmate het aantal documenten op het Web toeneemt, gaan phrases, tot hele zinnen toe, een steeds beter middel vormen om relevante documenten te vinden dan woorden.

Bij alle zoekmachines die phrases ondersteunen gaat dat door de phrase tussen dubbele aanhalingstekens te zetten, of door uit een menu *the (exact) phrase* te kiezen.

To be or not to be

Hoewel vrijwel alle grote zoekmachines phrase-searching mogelijk maken, kunnen ze in sommige gevallen met heel verschillende resultaten komen, die laten zien dat een phrase niet altijd op dezelfde manier wordt opgevat. Dat heeft te maken met *stopwoorden*.

Stopwoorden zijn woorden (meestal woordjes) die zoveel voorkomen dat sommige zoekmachines ze om redenen van economie niet indexerend. Voorbeelden zijn 'and', 'or' en 'not', die een aparte betekenis hebben als Booleaanse operatoren, en andere Engelse woordjes als 'a', 'the', 'is', 'of', 'for', 'be', 'in', 'about' en dergelijke.

Meestal is dat geen probleem, omdat u met zulke woordjes toch niet zult willen zoeken. Maar als ze in een phrase voorkomen, zult u er wel mee moeten zoeken. En dan kan blijken dat een zoekmachine met zo'n phrase niets kan.

De phrase “to be or not to be”, met twee Booleaanse operatoren en twee woordjes die vaak als stopwoord gelden is een goed voorbeeld. Zoekmachines zoals [Google](#), [Yahoo!](#) en [MSN Search](#) kunnen daar goed mee overweg; maar WiseNut, toch ook een grote zoekmachine, kan weinig met zo'n vraag. Het enige wat er dan opziet is: een andere zoekmachine gebruiken (of de vraag anders stellen).

Bij sommige zoekmachines (zoals [Wisenut](#)) vind je woordjes zoals 'of' binnen een phrase alleen als je er een + voor zet: “nothing +of +the sort”.

Dubbele aanhalingstekens duiden een phrase aan. Gebruik ze in alle gevallen waarin u met samengestelde termen werkt, want je kunt er veel preciezer mee zoeken dan met AND of met de unitaire operator + (zie hieronder, paragraaf 3.3.3.). Sommige zoekmachines, zoals Google en Yahoo!, zullen treffers waarin meerdere zoektermen als phrase voorkomen ook zonder het gebruik van de aanhalingstekens hoger ranken (zie blz. 54-5) dan die waarin ze dat niet doen. Toch is het verstandig daar niet op te

vertrouwen, want niet alle zoekmachines geven voorrang aan phrases, en voor die welke dat wel doen is het voorkomen van zoektermen als een phrase maar één van de vele factoren die de ranking beïnvloeden. Dus: zet een phrase tussen aanhalingstekens. Phrases zijn doorgaans dé manier om personen (met voornaam of voornamen en achternaam) te zoeken. Denk er alleen om dat u 'George W. Bush' niet vindt met de phrase "George Bush".

3.3.2. 'Best match' en andere niet-Booleaanse zoeksystemen

Het gebruik van Booleaanse operatoren lijkt voor alle vraagstellingen een oplossing te bieden. De vraagsteller geeft een aantal zoektermen op, verbonden met AND en/of OR, en de zoekmachine presenteert een aantal documenten die aan die zoekcriteria voldoen. En omdat de gevonden documenten aan de zoekcriteria voldoen, zijn ze relevant en heeft de vraagsteller gekregen waar hij om vroeg.

Helaas, zo simpel is het niet. Nog afgezien van de enorme afstand die er kan bestaan tussen de zoektermen die u gebruikt en de termen die het document met het antwoord op uw vraag bevat, zijn er een heleboel redenen waarom het Booleaanse model niet voldoet.

Ik noem er een paar:

1. in de eerste plaats vinden veel mensen het bij ingewikkelder vragen lastig al die ENs en OFs uit elkaar te houden en (bij nesting) de haakjes op de juiste plaats te zetten
2. ten tweede is lang niet altijd duidelijk of een zoekterm in een EN-relatie eigenlijk wel moet worden toegevoegd. Bij voorbeeld: het aspect verslaving is soms impliciet aan de term drugs. Als we geïnteresseerd zijn in documenten over drugsverslaving kunnen we de term 'verslaving' aan de term 'drugs' toevoegen om het verslavingsaspect tot uiting te brengen, maar het is wel zeker dat we daardoor documenten zullen missen die alleen de term drugs bevatten maar toch ook over verslaving gaan. Booleaans zoeken verdeelt documenten heel strikt in die welke aan de Booleaanse expressie (zoekvraag) voldoen, en die welke daar niet aan voldoen. EN-relaties hebben daardoor iets heel dwingend dat niet altijd strookt met de werkelijkheid van de taal
3. in (strikt) Booleaanse zoeksystemen is het alleen van belang of zoektermen in een document voorkomen, niet hoe vaak ze dat doen, met andere woorden: de *frequentie* van de zoektermen. Daardoor maken deze systemen geen onderscheid tussen documenten waarin (bij twee termen in een EN-relatie) ieder van de zoektermen tien maal voorkomt en die waarin ze elk maar één keer voorkomen. En toch is dat wel van belang: hoe vaker termen in een document voorkomen, des te groter de waarschijnlijkheid dat dat document gaat over het onderwerp dat die termen aanduiden. Als regel geldt: hoe frequenter de zoektermen in een document voorkomen, des te *relevanter* dat document is voor de vraag
4. een ander probleem is dat een Booleaans OR tussen twee zoektermen u niet de gelegenheid geeft de ene zoekterm meer gewicht mee te geven dan de andere. Een voorbeeld is het geval waarin u op het Web op zoek gaat naar een tweedehands vouwfiets, liefst van het merk Brompton. De Booleaanse zoekvraag is dan "te koop" AND tweedehands AND (Brompton OR vouwfiets), en de zoekmachine komt met een resultaat van advertenties waarin allerlei vouwfietsen worden aangeboden, met her en der een Brompton ertussen; maar het liefst zou u de Bromptons bovenin de lijst zien
5. Booleaanse systemen houden geen rekening met de plaats in het document waar de gezochte termen voorkomen. En dan heb ik het niet over de titel of de URL van het document – want alle grote mondiale zoekmachines geven u de mogelijkheid aan te geven dat zoektermen daarin moeten voorkomen --, maar over bijvoorbeeld de eerste honderd woorden of een grote kop.

In het algemeen gesproken lossen zoekmachines deze en soortgelijke problemen op door de resultaten van een Booleaanse zoekactie op een intelligente manier te sorteren, maar bij de meeste vind je ook nog sporen van een zoekstelsel dat jaren lang met het Booleaanse model geconcurrereerd heeft: het *best match* (of *partial match*) and *ranking* model. Dit model maakt het de gebruiker mogelijk al zijn zoektermen zonder operatoren in te voeren. Het zoekstelsel zoekt de documenten dan alsof tussen de zoektermen een OF-relatie bestaat, maar presenteert ze in de volgorde: hoe meer zoektermen een document bevat, hoe eerder het gepresenteerd wordt. Eerst komen de documenten die alle zoektermen (x) bevatten, dan die welke x-1 bevatten, enz. Dat lijkt het ei van Columbus, maar een ander kenmerk van 'best match' systemen is dat ze ook rekening houden met de frequentie van de zoektermen ('best match and ranking'), en dan ontstaat er een nieuw probleem: welk document is nu relevanter voor een vraag met drie zoektermen:

- een document dat alle drie de zoektermen één keer bevat, of
- een document dat twee van de drie zoektermen tien keer bevat.

Misschien zegt u: natuurlijk het eerste, maar dat is in lang niet alle gevallen zo. Bij de vraag `reparatie fiets rijwiel` bijvoorbeeld. Daar is een document met de termen 'fiets' en 'reparatie' tien keer, voor u, die iets over het repareren van fietsen wilt weten, hoogst waarschijnlijk relevanter dan een document met alle drie de termen één keer. Aan de andere kant: bij de vraag `vakantie auto Griekenland` zullen documenten waarin de termen 'auto' en 'vakantie' tientallen malen voorkomen u waarschijnlijk minder interesseren dan die welke alle drie de termen maar één keer bevatten.

Het verschil tussen Booleaans EN en OF blijft bij 'best match'-systemen dus nog wel degelijk een rol spelen.

Meer over retrievalssystemen: Ray Larson, The Technology of search engines (http://books.nap.edu/html/protecting_children/ch4.html)

Hoe zoekmachines 'ranken'

Het voorkomen van alle zoektermen in een document, dan wel het meermalen voorkomen van sommige zoektermen zijn maar twee van de vele factoren die verschil kunnen maken voor de volgorde waarin zoekmachines zoekresultaten 'ranken'. Bij de overweldigende hoeveelheid documenten op het Web is het vinden van webpagina's die aan de zoekcriteria voldoen voor zoekmachines geen probleem. Maar hoe plaats je de 15.000 documenten die je als zoekmachine bij een vraag gevonden hebt, in zo'n volgorde dat de vraagsteller al bij de eerste twintig het antwoord op zijn vraag vindt? Daarvoor passen zoekmachines op de 'ranking' allerlei, voor een deel geheim gehouden, regels toe. Deze zijn in twee groepen te verdelen: document-interne en document-externe.

Document-interne regels zijn regels die uit de structuur van de documenten zelf voortkomen. Daartoe behoren:

- de plaats van de zoektermen in het document: het voorkomen van uw zoektermen in de URL, de titel (wat u in de titelregel van uw browser ziet), de trefwoorden en de beschrijving van een document in de zogeheten METAtags (die uw browser u niet laat zien, maar zelf wel 'ziet'), de koppen, de eerste x-honderd woorden in het document kan een rol spelen bij de ranking
- de volgorde waarin u uw zoektermen hebt ingevoerd. `reparatie rijwielen fietsen` levert dan een ander resultaat op dan `fietsen rijwielen reparatie`.
- de nabijheid van zoektermen kan een rol spelen (dus ook zonder dat u, waar dat kan,

de NEAR-operator gebruikt); speciaal treffers waarin de zoektermen als phrase voorkomen staan bij sommige zoekmachines hoog aangeschreven.

- de frequentie waarmee de zoektermen in het document voorkomen, absoluut en/of in verhouding tot alle termen in het document, en soms ook in verhouding tot de frequentie waarmee ze in de hele database van de zoekmachine voorkomen: hoe zeldzamer een term in de hele database voorkomt en tegelijk hoe frequenter dat in een individueel document gebeurt, des te relevanter dat document wordt geacht te zijn

Document-externe regels zijn regels die met de plaats en de omgeving van het document te maken hebben. Daaronder zijn:

- het belang van een document te midden van concurrerende documenten. Dit belang kan worden gemeten aan het aantal en de duur van de bezoeken (*click popularity*), of aan het aantal verwijzingen die het document 'trekt' (in de vorm van links die ernaar verwijzen) van andere documenten die zelf ook veel verwijzingen trekken (*link popularity*). Zoekmachines die deze rankingmethodes gebruiken zijn resp. (vermoedelijk) Yahoo! (<http://dir.yahoo.com>), in zijn directory, en Google (<http://www.google.com>).
- sommige zoekmachines die ook een directory hebben plaatsen de documenten die ze in de directory vinden bovenin de lijst
- het niveau waarop een document in de directorystructuur van de server verstopt zit: een document op het niveau www.host.org/voorbeeld.htm scoort hoger dan een op het niveau www.host.org/sub/subsub/subsubsub/voorbeeld.htm
- last but not least: een hoge ranking tegen betaling. Is bij vlagen in de mode en lokt dan veel protest uit. Een bedrijf koopt dan een hoge ranking bij een aantal zoektermen die het product van dat bedrijf beschrijven. De meeste zoekmachines die deze 'sponsored' of 'featured' e.d. links aanbieden, winden daar geen doekjes om, maar als je een metazoekmachine gebruikt, verdwijnt die informatie nog wel eens.

Meer over ranking: Fagan Finder > Searching > Popularity Ranking
(<http://www.faganfinder.com/search/popularity.shtml>)

3.3.3. Unitaire operatoren

In paragraaf 3.3.2. hebben we gezien dat 'best match'-systemen, om een zoekresultaat op een zinvolle manier te kunnen ranken, toch over enige informatie moeten beschikken welke zoektermen in een vraag onmisbaar zijn, en welke eventueel kunnen worden gemist. In het hierboven gegeven voorbeeld kan de term 'rijwiel' worden gemist als in het gevonden document maar 'fiets' en 'reparatie' voorkomen, en kan de term 'fiets' worden gemist als het document maar de termen 'rijwiel' en 'reparatie' bevat.

Ik geef nog een iets ander voorbeeld. Vergelijk de volgende vragen met elkaar:

```
reparatie rijwielen fietsen  
verslaving drugs alcohol
```

Schijnbaar hebben we hier twee vragen met zowel een EN- als een OF-relatie. Maar in de eerste vraag neemt de zoekterm 'reparatie' een veel belangrijker plaats in dan de zoekterm 'verslaving' in de tweede: als je het over fietsen en rijwielen hebt is het aspect reparatie in die twee termen zonder meer relatief minder prominent dan het (bijna impliciete) aspect verslaving als je het over drugs en alcohol (en zeker de combinatie) hebt. Het gevolg is dat de term reparatie in de eerste vraag een groter gewicht heeft dan de term verslaving in de tweede. Zoeksystemen die alleen met 'best match' werken kunnen dit onderscheid in gewicht niet maken. Om dat wel mogelijk te maken zijn in de 'best match' wereld unitaire operatoren geïntroduceerd, die u bij alle grote mondiale zoekmachines en bij tal van andere zoekinstrumenten en databases kunt tegenkomen.

Unitaire operatoren verbinden geen zoektermen met elkaar, maar kwalificeren afzonderlijke zoektermen in een 'best match' zoekstelsel. Er zijn er twee: + en -, en die staan *altijd* onmiddellijk, zonder spatie, voor de zoekterm waarop ze betrekking hebben. + wil zeggen dat de ermee verbonden zoekterm in het te vinden document aanwezig *moet* zijn, - wil zeggen dat de ermee verbonden zoekterm *niet* aanwezig *mag* zijn. Hierboven heb ik al even genoemd dat 'best match' systemen ook rekening houden met de frequentie waarmee zoektermen in een document voorkomen: documenten waarin de opgegeven zoektermen vaker voorkomen, komen hoger in de resultatenlijst (de 'ranking') dan documenten die de zoektermen maar één of enkele malen bevatten. Dat blijft zo, ook als de zoektermen beide 'geplust' zijn en er dus feitelijk een Booleaanse EN-relatie tussen die zoektermen wordt gelegd. Maar omdat het Booleaanse model de frequentie van zoektermen in een document niet laat meewegen in de 'ranking', en het 'best match'-model wel, is er – in theorie – toch een verschil: het resultaat is even groot, maar bij +term1 +term2 is de volgorde van de treffers anders dan bij term1 AND term2. Tot een jaar of vijf geleden had je inderdaad zoekmachines die het onderscheid tussen +...+ en AND maakten, maar inmiddels is dat niet meer zo. Echte Booleaanse en echte 'best match'-systemen kom je als webzoekmachines nauwelijks meer tegen (wel als databases). De toonaangevende grote zoekmachines, Google, Yahoo! en MSN Search bieden allemaal de mogelijkheid van Booleaans zoeken aan (al doet Google dat zonder 'nesting' en met een impliciet EN tussen zoektermen), maar ranken de resultaten vervolgens op basis van 'best match' en andere niet-Booleaanse retrievalmodellen (met name zogeheten 'vector-space'-modellen). De reden waarom ik dit onderwerp hier toch behandel, is dat de sporen van het 'best match'-model nog altijd bij zoekmachines terug te vinden zijn. Zo bieden ze allemaal de mogelijkheid om zoektermen te 'plussen' of te 'minnen', ook al leidt dat niet tot andere resultaten dan wanneer je Booleaans AND en NOT zou gebruiken; daarbij is de betekenis van + nog eens extra gereduceerd doordat alle grote mondiale zoekmachines zoektermen in een vraag tegenwoordig met een impliciet EN verbinden. In de regel geeft +term1 +term2 dus dezelfde resultaten als term1 term2. Maar helemaal waar is dit ook weer niet, want een zoekmachine als Google gebruikt de +-operator nog wel degelijk, alleen niet meer om zoektermen verplicht te stellen, maar om er een andere speciale functie aan te geven, bijvoorbeeld om u naar de precieze spelling 'électricité' te laten zoeken in plaats van naar treffers waarin het woord ook zonder of met een deel van de accenten gespeld is. Bij Google geeft +électricité +elba dus een ander resultaat dan électricité elba. Een ander verschil tussen Booleaans en unitair zoeken is natuurlijk dat je in een vraag met unitaire operatoren niet alle zoektermen hoeft te 'plussen'. In theorie zou de 'best match'-vraag:

+verslaving drugs

een verzameling documenten op moeten leveren waarin de term 'verslaving' verplicht aanwezig is en de term 'drugs' facultatief. In die set kunnen dus ook documenten zitten waarin de facultatieve term 'drugs' niet voorkomt, maar in plaats daarvan andere termen, zoals 'heroïne', 'cocaïne', 'XTC' wel. Door niet een Booleaanse vraag te stellen maar een 'best match'-vraag vergroten we onze kans relevante documenten te vinden met zoektermen waaraan we zelf niet gedacht hadden. Anders gezegd: 'best match'-zoeken vergroot de *serendipiteit* – de kans iets te vinden waar je (oorspronkelijk) niet naar op zoek was – van het zoeken.

Ook deze mogelijkheid, om een zoekterm facultatief te maken, wordt door vrijwel geen enkele grote zoekmachine meer aangeboden. De bij mijn weten enige grote zoekmachines die die mogelijkheid nog hebben zijn MSN Search, dat de mogelijkheid biedt een zoekterm facultatief te maken door hem te laten voorafgaan door `prefer:`, Exalead, dat in de Advanced Search een optie *Preferably containing* (of de operator OPT) heeft, en Gigablast, dat de vraag `cat OR dog AND pig` opvat als `cat +dog`

+pig (dus met dog en pig als verplicht aanwezige en cat als facultatieve zoekterm).

Veel zoekmachines bieden zoeken met de unitaire operatoren aan via menu's waarin de + en de – worden weergegeven door de opties *Containing* en *Not containing* (ook wel in de vorm *Must include/have/contain* en *Must not include/have/contain*). Exalead, een Franse zoekmachine, biedt daarnaast nog de optie *Preferably containing* waarmee kan worden aangegeven dat het voorkomen van de ingevulde term(en) in de te vinden documenten wenselijk maar niet verplicht is.

Unitaire (en ook Booleaanse) operatoren kunnen niet alleen bij woorden worden gebruikt, maar ook bij phrases (3.3.1). Net als enkelvoudige zoektermen kan een phrase worden voorafgegaan door een unitaire operator:

+"Arabische landen", -"Verenigde Arabische Emiraten"

Hetzelfde geldt voor de combinatie van Booleaanse operatoren met phrases:

ginseng AND "geneeskrachtige eigenschappen"

“Derde Wereld” NOT Afrika

3.3.3.1. Term Weighting

In paragraaf 3.3.2. noemde ik twee voorbeelden van vragen waarbij het voor een zoekmachine uiterst lastig is om te bepalen welke treffers hij hoger in de ranking moet plaatsen: die met alle zoektermen of die waarin slechts een deel van de zoektermen prominent voorkomt. De oplossing voor dit probleem zoeken de beste onder de zoekmachines langs twee wegen:

ADVANCED SEARCH OPTIONS [back to home page](#)

FIND PAGES	containing	the words		
	containing	the words		
	containing	the words		
	containing not containing preferably containing	the words		
LANGUAGE	Only	search pages written in:	any language	
COUNTRY	Only	search for websites in:	any country	
FILE FORMAT	Only	search for files of type:	any type	
DOMAIN OR SITE	Only	search for pages from site or domain:		
LINKS	Search for pages that link to domain or page:			
SEARCH METHOD	Exact search			
DOCUMENT SORTING	By relevance			
MODIFICATION DATE	Only keep documents modified after:			

(e.g., "exalead.com" or just "com")

(e.g., "exalead.com" or "www.exalead.com/help.html")

(e.g., "20/04/1994")

OK

Het Advanced Search-scherm van Exalead, een Franse zoekmachine met een database van ruim vier miljard documenten. Dit is een van de weinige grote mondiale zoekmachines die, in de vorm van de optie 'preferably containing', een 'best match'-resultaat geeft. Daarbij sorteert de zoekmachine het resultaat eerst op de treffers die de facultatieve term bevatten, en toont vervolgens treffers die dat niet doen.

- in de eerste plaats proberen ze te bepalen of u als gebruiker misschien synonieme zoektermen (*fiets, rijwiel*) hebt opgegeven; zo ja dan laten ze een van die termen uit de vraagstelling weg als die heel weinig resultaten oplevert. Dit is wat Google bij sommige vragen doet.
- daarnaast bieden ze de gebruiker de mogelijkheid zelf relatieve gewichten aan zijn zoektermen toe te kennen. Google doet dat bijvoorbeeld door u een zoekterm te laten herhalen: bij de vraag: vakantie auto auto auto auto Griekenland Griekenland Griekenland Griekenland Griekenland Griekenland krijgen treffers waarin de woorden 'auto' en 'Griekenland' meermalen voorkomen een hogere plaats in de ranking dan die waarin het woord 'vakantie' frequent figureert.

3.3.4. *Goochelen met zoektermen: trunceren, maskeren (jokers, wildcards), word stemming, fonetisch en fuzzy zoeken; regular expressions en zoeken in ranges*

Wie wel eens in de index van een catalogus of een database heeft gekeken, zal hebben opgemerkt dat daarin tal van woorden voorkomen die dezelfde stam hebben:

therapeut
therapeute
therapeutisch
therapeutische
therapie

Als je met al die termen zou willen zoeken – bij voorbeeld in combinatie met een andere zoekterm – zou je al die termen in een OR-relatie moeten verbinden: (therapie OR therapieën OR therapeutisch OR therapeutische OR therapie enz.) AND depressie.

In zo'n geval is het veel eenvoudiger in één keer met alle termen te zoeken die met de string (letterreeks) 'therap' beginnen. Dat kan als een zoekmachine *trunceren* ondersteunt. Dat wil in dit geval zeggen dat u bij therap afbreekt en met een speciaal teken (een wildcard of joker) aangeeft dat er nog een willekeurig aantal willekeurige letters kunnen volgen. Het meest gebruikte teken daarvoor is *, maar ? of \$ kunt u ook tegenkomen.

In het hierboven genoemde voorbeeld ging het om een rechtse truncatie, en die komt inderdaad het meest voor. Linkse truncatie (*kanker voor longkanker, darmkanker enz.) komt als mogelijkheid in de praktijk veel minder vaak voor. Een derde mogelijkheid is een wildcard in een woord, bij voorbeeld publi*atie om zowel publicatie als publikatie te vinden. Als zo'n wild card één letter vervangt spreekt men van maskeren, zijn het er meer dan heet het intern trunceren. Als trunceren en maskeren beide mogelijk zijn, staat het truncatieteken (bij voorbeeld *) doorgaans voor 0-x letters, en het maskeringsteken (bij voorbeeld %) voor één letter.

Zoals we hierboven (in paragraaf 3.3.1.2) hebben gezien kan een wildcard ook een teken zijn dat niet als truncatieteken dient, maar een willekeurig woord vervangt.

Sommige zoekmachines trunceren automatisch. Dat kan lastig zijn als je bij voorbeeld op zoek bent naar iemand met de naam Filo, en dan allerlei sites vindt met het woord 'filosofie'. Probeer het dan eens met de phrase "Filo " (dus met een spatie achter de term binnen de aanhalingstekens).

Als u een truncatieteken gebruikt, let dan goed op of het teken voor 1-x letters staat dan wel voor 0-x letters. In het eerste geval vindt u met product* wel productie maar niet product, in het laatste geval vindt u allebei de termen.

Denk er ook om dat trunceren ertoe kan leiden dat het rankingmechanisme van een zoekmachine niet meer werkt!

Een variant op trunceren is *word stemming*. Dit werkt iets anders, en praktisch gesproken bij de mondiale zoekmachines alleen in het Engels. Met word stemming 'aan' geeft u een zoekterm op, waar de zoekmachine zelf dan vervoegde (change – changing) en verbogen vormen (child – children) van dat woord aan toevoegt.

Bij sommige zoekmachines is word stemming automatisch. Goede zoekmachines maken het dan ook mogelijk dat uit te schakelen.

Dan zijn er zoekmachines die *fonetisch* en *fuzzy* zoeken. In het eerste geval zoekt de zoekmachine de zoekterm plus woorden die net zo klinken als deze. Een voorbeeld is het Engelse woord 'comely' dat bij de zoekmachine Exalead als zoekterm ook de naam 'Comley' oplevert. Onder de grote, mondiale zoekmachines is Exalead een uitzondering, maar bij speciale zoekmachines gebeurt fonetisch zoeken soms standaard. Bekijk maar eens het resultaat van een zoekactie naar de naam Petrus Eisinga bij [FamilySearch](#).

The screenshot shows the FamilySearch website interface. At the top, there is a navigation bar with links for HOME, SEARCH, SHARE, LIBRARY, Sign On, Register, Site Map, and Help. Below this is a secondary navigation bar with categories: Search for Ancestors, Research Guidance, Research Helps, Web Sites, and Family History Library Catalog. The main content area displays search results for 'Petrus Eisinga' with a 'refine search' link. A 'Sources Searched' box on the right indicates that matches were found only in the sources listed below, with a link to 'IGI/Continental Europe - 8'. The search results list 8 entries, each with a link to the individual's record and brief biographical information. A 'Matches: International Genealogical Index/Continental Europe - 8' link is provided at the bottom of the results. A 'Return to top of page' link is also visible. A yellow callout box on the left side of the page features a woman's profile picture and text: 'If you don't find what you are looking for in these records, I can provide [Research Guidance on additional sources to search.](#)'

Resultaat van een zoekactie met de naam Petrus Eisinga in Nederland in de database van FamilySearch. Exact spelling is 'off', waardoor zowel fonetisch als fuzzy gezocht is.

Bij zoekmachines die *fuzzy* zoeken aanbieden, worden termen die één of enkele letters verschillen van de opgegeven zoekterm(en) (licensie – license – licensed – licentie) in de zoekactie meegenomen. Met Exalead kunt u zowel fonetisch als fuzzy zoeken, maar

doorgaans zijn die mogelijkheden anders geïmplementeerd: diverse zoekmachines doen aan spellingcontrole en komen dan met suggesties voor anders gespelde zoektermen dan u hebt ingetypt.

Een zeer grote flexibiliteit bij het zoeken naar woordpatronen bieden zoekmachines die u met *regular expressions* laten zoeken. Regular expressions zijn woord- of getalpatronen waarin bepaalde zaken vastliggen en andere niet of maar ten dele. Letters en cijfers in die woord- of getalpatronen kunnen worden vervangen door een (een geheel vrije) joker of door een keuze uit bepaalde letters of cijfers (bijvoorbeeld de letter a of e of o, of het cijfer 4 of 7 of 9); jokers en die keuzes kunnen dan ook nog optioneel of herhaalbaar worden gemaakt. Vroeger had je editors en tekstverwerkers waarmee je zo fantastische zoek- en vervangacties kon ondernemen. Als u aan kruiswoordpuzzels denkt, hebt u één doel voor ogen waarvoor een zoekmogelijkheid met regular expressions een oplossing kan bieden.

Een apart geval vormt ten slotte de mogelijkheid van sommige zoekmachines en databaseprogramma's om met *ranges* te zoeken, d.w.z. met alle getallen tussen twee opgegeven waarden. Jaartallen, huisnummers, telefoonnummers en data zijn maar enkele van de getallen waarop met ranges gezocht kan worden.

3.3.4.1. Google's ~ (tilde)-operator

Alle hierboven genoemde middelen zijn bedoeld om een zoekresultaat minder afhankelijk te maken van de precieze spelling van een zoekterm. Maar in veel vragen is het natuurlijk niet de spelling, maar de zoekterm zelf die het zoekresultaat beperkt. Denk maar aan gevallen waarin u met de term 'ouderdomsdementie' treffers had kunnen vinden die u niet met 'alzheimer' gevonden hebt, of waarin u naar 'eetstoornissen' hebt gezocht en ook met 'boulimie OR anorexie' had kunnen zoeken. In literatuurdatabases en bibliotheekcatalogi wordt de gebruiker vaak door de software geholpen door hem of haar gebruik te laten maken van een 'onderwerpsthesaurus', waarin hij of zij bij een zoekterm synoniemen, algemenere en meer specifieke termen kan vinden (zie blz. 253 e.v.). Binnen bepaalde, nauw omschreven vakgebieden bestaan ook wel speciale zoekmachines die die faciliteit kunnen bieden (zie bijvoorbeeld MedHunt (<http://www.hon.ch/MedHunt/>)), en op het Web in zijn geheel vervullen directories natuurlijk voor een kleine fractie van alle sites een soortgelijke functie (ze laten u termonafhankelijk naar sites over onderwerpen zoeken), maar een thesaurus van het hele Web blijft voorlopig een mooie droom.

Toch laat Google u bij bepaalde (Engelse) termen zien wat een thesaurusfunctie voor een grote mondiale zoekmachine zou kunnen doen. Als u bij Google '~anorexia' als zoekterm opgeeft, vindt u ook treffers die de term 'anorexia' niet bevatten, maar wel 'eating disorders', 'bulimia' en/of 'anorexic'. Mooie resultaten krijgt u ook bij termen als '~tutorial', dat door Google als een vraag naar 'help', 'guide', 'support' wordt opgevat.

3.3.5. Hoofdletters/kleine letters

Plaats- en persoonsnamen worden doorgaans met hoofdletters geschreven en zijn daardoor gemakkelijk te onderscheiden van gewone zelfstandige naamwoorden. Sommige zoekmachines bieden de mogelijkheid dat onderscheid ook tussen zoektermen te maken: 'Leiden' is iets anders dan 'leiden'.

Het verschil tussen kleine letters en hoofdletters beperkt zich niet tot initialen. 'HOOP' (een acroniem voor Hoger Onderwijs en Onderzoeks Plan) is iets (heel) anders dan 'hoop' en ook iets anders dan 'Hoop' (bij voorbeeld in De Hoop Scheffer).

Van de grote mondiale zoekmachines is er inmiddels niet één meer die dit zogeheten 'case-sensitive' zoeken mogelijk maakt. Het verschil zult u dus zelf moeten duidelijk maken door context te verschaffen (voeg aan 'Leiden' bijvoorbeeld toe 'gemeente', of het bijvoeglijk naamwoord 'Leids(e)').

3.3.6. *Bijzondere tekens (diacrieten) en 'character encodings'*

Andere talen dan het Engels werken veelal met allerlei bijzondere tekens, zoals geaccentueerde letters (accent grave, accent aigu, accent circonflexe, tilde, Umlaut, cedille). In gevallen waarin zulke tekens in een zoekterm voorkomen, is het van groot belang om u ervan te vergewissen hoe de door u gebruikte zoekmachine daarmee omgaat. Het kan zijn dat u met bij voorbeeld de naam 'Bronte' ook de documenten met de correcte vorm 'Brontë' vindt, maar zeker is dat niet. Doorgaans geldt de ongeaccentueerde vorm als de standaardvorm, waarmee ook de geaccentueerde vormen worden gevonden, maar is het andersom niet zo. Maar er zijn ook zoekmachines die 'Brontë' en 'Bronte' op exact dezelfde manier behandelen. Ik ken geen zoekmachines meer die bij 'Brontë' meer treffers vinden dan op 'Bronte', maar het blijft oppassen. Hoe een zoekmachine met diacrieten omgaat vindt u in de Help, of, waar dat niet het geval is, door goed naar de zoekresultaten te kijken. De meeste zoekmachines beelden daarin de zoektermen vet af, dus als u met 'Bronte' gezocht hebt en u vindt de vorm 'Brontë' in het zoekresultaat vet afgebeeld, betekent dat dat in dit geval de vorm 'Brontë' in de zoekactie is meegenomen.

Desondanks blijft het, tenzij u uitdrukkelijk de geaccentueerde dan wel de ongeaccentueerde vorm van een zoekterm wilt uitsluiten, het veiligst om bij iedere zoekmachine zowel met de geaccentueerde als met de ongeaccentueerde vorm te zoeken. De reden is dat geaccentueerde tekens in HTML op diverse manieren kunnen worden aangemaakt. Daardoor is de ene ã de andere niet en zegt het weinig als u met de zoekterm 'bronte' ook documenten vindt waarin alleen de vorm 'brontë' voorkomt. Een testje kan uitwijzen of het zinvol is zowel met de geaccentueerde als met de ongeaccentueerde vorm van een woord te zoeken. Als zowel de vraag:

+Bronte –Brontë

als de vraag

+Brontë –Bronte

treffers opleveren, zoek dan met beide vormen. En probeer dan ook eens Bront¨

Let er wel op dat zoekmachines in de behandeling van diacrieten niet consequent zijn. Uit het feit dat een zoekmachine op de zoekterm 'Bronte' ook treffers met 'Brontë' vindt, mag u niet opmaken dat met 'frobelen' ook 'fröbelen' wel zal worden gevonden.

Sommige zoekmachines laten u ook met vreemde lettertekens ('character encodings') (Chinees, Tsjechisch, Grieks etc.) zoeken. Zie hierover: Multilingual Fonts and Typography - A Fountain of Fonts / Meertalige fonts en Typografie - Fontavond/turen: Standaarden voor lettertekens / Standards for character encoding (<http://www.xs4all.nl/~dziewon/fontein/standard.htm>)

3.3.7. *Leestekens en andere niet-lettertekens*

Voor de Pietjes-precies onder ons is een interessante vraag ook hoe zoekmachines met lees- en andere niet lettertekens (voor zover ze niet als unitaire operatoren of jokers worden gebruikt) omgaan. Daarbij zijn vijf mogelijkheden te onderscheiden:

- het teken wordt net zo behandeld als andere tekens, dus bijvoorbeeld $\$ = \$$
- het teken wordt geheel verwaarloosd, bijvoorbeeld $\$5,000 = 5,000$
- het teken wordt behandeld als een spatie en brengt zo tussen twee termen aan weerszijden een impliciete EN-relatie tot stand; voorbeeld I/O
- het teken wordt behandeld als een spatie in een phrase, bijvoorbeeld care-taker (vgl. “care taker” en caretaker)
- het teken wordt verwaarloosd én als spatie in een phrase opgevat, bijv. care-taker (vgl. “care taker” en caretaker)

Voor een aantal voorbeelden, zie de overzichten bij de grote mondiale zoekmachines.



Παγκόσμιος Ιστός [Εικόνες](#) [Ομάδες \(usenet\)](#) [Κατάλογος](#)

Εθνική Βιβλιοθήκη της Ελλάδος

Αναζήτηση

[Σύνθετη Αναζήτηση](#)
[Προτιμήσεις](#)

Αναζήτηση: στον παγκόσμιο ιστό σελίδες γραμμένες στα Ελληνικά σελίδες από Ελλάδα

Παγκόσμιος Ιστός Αποτελέσματα **1 - 10** από περίπου **34.000** για **Εθνική Βιβλιοθήκη της Ελλάδος**. (0,19 δευτερόλεπτα)

[Εθνική Βιβλιοθήκη της Ελλάδος](#)

Ιστορία και αποστολή του ιδρύματος, τα τμήματα και οι προσφερόμενες υπηρεσίες προς τους Πολίτες, οι συλλογές και οι εκδόσεις. Πρόσβαση στους ηλεκτρονικούς ...

[www.nlg.gr/](#) - 12k - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

[Η Εθνική Βιβλιοθήκη της Ελλάδος - Δημόσιος κατάλογος \(OPAC\)](#)

Βρίσκεστε στην αρχική σελίδα του Δημόσιου Καταλόγου της Εθνικής Βιβλιοθήκης της Ελλάδος. Ο κατάλογός μας ... Copyright © 2004, **Εθνική Βιβλιοθήκη της Ελλάδος** ...

[www.nlg.gr/opac.htm](#) - 42k - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)
[[Περισσότερα αποτελέσματα από το www.nlg.gr](#)]

[Υπουργείο Εθνικής Παιδείας και Θρησκευμάτων - Εθνική Βιβλιοθήκη ...](#)

Εθνική Βιβλιοθήκη της Ελλάδος. Νομοθεσία Εθνικής Βιβλιοθήκης της Ελλάδος - Ν. 814/1943 Περί οργανισμού της Εθνικής Βιβλιοθήκης της Ελλάδος ...

[www.yperth.gr/el_ec_page687.htm](#) - 51k - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

[Horizon Information Portal](#)

Εκτελείτε αναζήτηση για: **Εθνική Βιβλιοθήκη της Ελλάδος**. Καλώς ήρθατε στο Δημόσιο Κατάλογο της Εθνικής Βιβλιοθήκης της Ελλάδος ...

195.134.102.45/irac20/irac.jsp?profile=ebe0-2&reloadxsl=true - 25k - 16 Μαρτ 2006 - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

[PICTURES DESCRIPTION](#)

Αθήνα, **Εθνική Βιβλιοθήκη της Ελλάδος**. Αποβάθρα στο λιμάνι της Χίου (λεπτομέρεια).

Dapper O., Description exacte des isles de l'Archipel, Άμστερνταμ, 1703, ...

[www.fhw.gr/projects/notaras/gr/about/lezantes.html](#) - 25k - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

[Εθνική Τράπεζα της Ελλάδος](#)

Εθνική Τράπεζα και Γ.Κ.Π.Σ. Ευκαιρίες για ανάπτυξη των Μικρομεσαίων ... Εθνικής Τράπεζας της **Ελλάδος** Α.Ε. • Μαρφωτικό Ίδρυμα, Εθνικής Τράπεζας ...

[www.nbg.gr/](#) - 69k - 16 Μαρτ 2006 - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

[Μεγάλη Μουσική Βιβλιοθήκη της Ελλάδος - ΛΙΑΝΑ ΒΟΥΔΟΥΡΗ](#)

ΑΠΘ: Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, ΒΒ: **Βιβλιοθήκη της Βουλής**, ΒΕΒ:

Βιβλιοθήκη Εργαστηρίου Βιβλιογραφίας, ΕΒΕ: **Εθνική Βιβλιοθήκη της Ελλάδος**, ...

[www.mmb.org.gr/page/default.asp?id=445&la=1](#) - 34k - [Αποθηκευμένη Σελίδα](#) - [Παρόμοιες σελίδες](#)

De Griekse Google, doorzocht met Griekse zoektermen.

3.3.8. *Extra's*

De hierboven genoemde mogelijkheden van zoekmachines: zoeken met Booleaanse en unitaire operatoren, nabijheidsoperatoren, phrases, wildcards, nesting en case-sensitive searching zijn bij de meeste zoekmachines op het Web wel in de een of andere vorm aanwezig. Soms zitten ze een beetje verstopt, vooral bij die zoekmachines die hun gebruikers een menugestuurde interface bieden, die alle 'techniek' een beetje verhuult. Bij veel zoekmachines worden bij *All of the words* alle zoektermen met AND verbonden en bij *Any of the words* met OR. Maar met enige oefening leert u wel door deze verschillen in benamingen heen kijken.

Sommige zoekmachines hebben nog aanzienlijk meer mogelijkheden. Aangezien die doorgaans een speciaal doel hebben, behandel ik ze hier naar hun functie (blz. 67 e.v.).



Advanced Search

Search for...

all of these words

Search

this **exact phrase**

and this **exact phrase**

any of these words

none of these words

Restrict to this URL

Pages that link to this URL

Site Clustering

yes no

Number of summary excerpts

0 1 2 3 4 5

Results per Page

10 20 30 40 50 100

Restrict to these Sites

Het Advanced Search-scherm van een middelgrote mondiale zoekmachine, Gigablast. De logische operatoren AND, OR en NOT zijn hier vertaald in All of these words, Any of these words en None of these words.

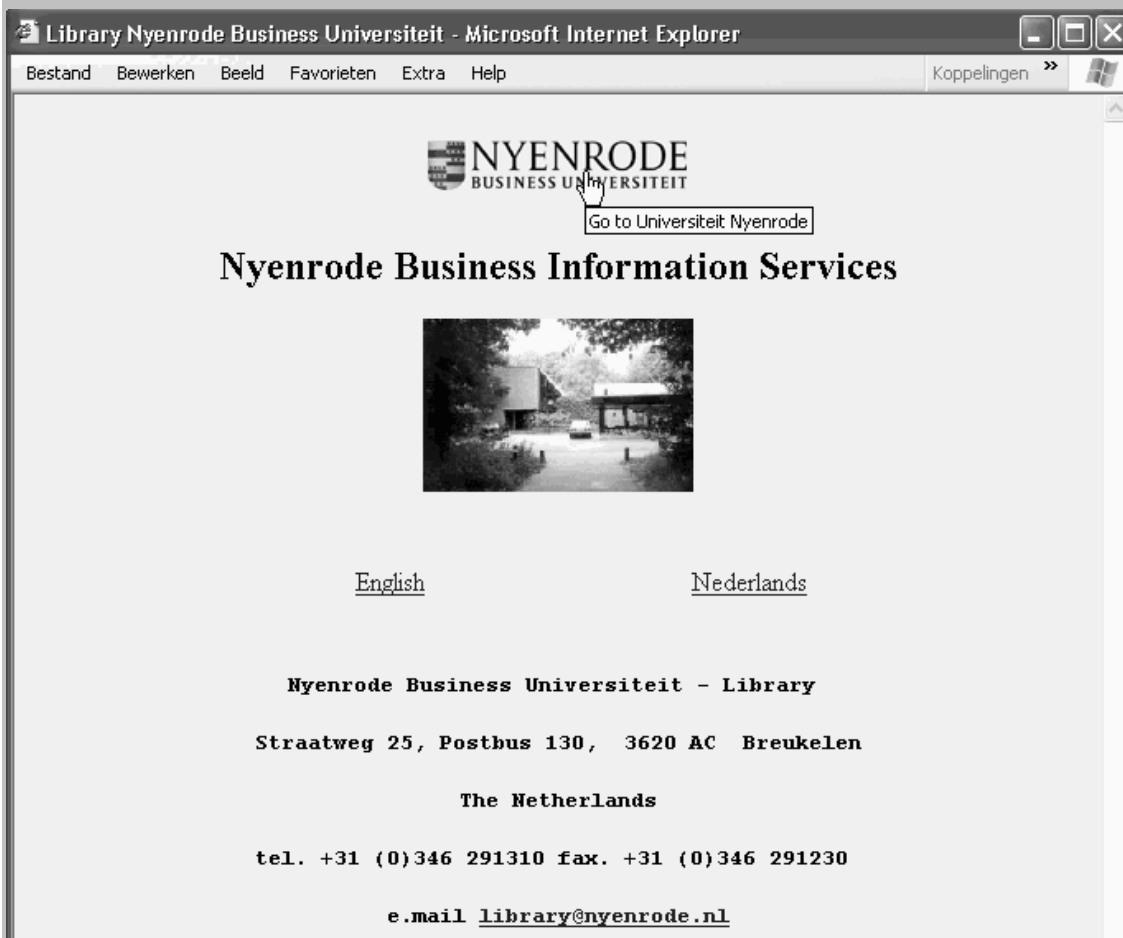
De structuur van een webpagina

Om een optimaal gebruik te kunnen maken van zoekmachines is het nodig iets van HTML te weten, de opmaaktaal van webdocumenten.

Als u een webpagina op uw computerscherm bekijkt, laat de browser, het programma waarmee u vanaf uw eigen PC het Web opgaat, u een keurig opgemaakt document zien waarin teksten en afbeeldingen netjes over de pagina verdeeld zijn, sommige stukken tekst groter, of in een ander lettertype, zijn dan andere, er een achtergrondkleur is etc. De browser kan u die pagina zo laten zien doordat in de tekst van de pagina zelf, voor u onzichtbaar, allerlei codes verscholen zitten die de browser vertellen hoe de eigenlijke tekst en afbeeldingen, die u wel ziet, moeten worden getoond. De uitgebreide tekst, inclusief de codes, kunt u bij een webpagina bekijken door in het topmenu van de browser de optie View en vervolgens (Page) Source aan te klikken.

De codes in deze tekst, die u dus normaal gesproken niet ziet, heten 'tags', d.w.z. etiketten of labels. Er zijn er per commando meestal twee van, een begin- en een eindtag. Samen 'etiketteren' ze de tussenliggende, meestal wel in de browser zichtbare tekst, die daarmee tot een soort veld (zoals in een databaserecord) wordt. Alle tekst die tussen twee tags staat is onderworpen aan het commando (om iets dat in de tags besloten ligt op deze of gene manier en op deze of gene plaats af te beelden).


In het algemeen wordt bij een webpagina onderscheid gemaakt tussen tekst, d.w.z. in de browser zichtbare tekst, en 'HTML-code', de opdrachten in de Markup Language. Hieronder is een (door mij enigszins bewerkte) heel simpele webpagina afgebeeld, met daaronder de tekst die daaraan ten grondslag ligt.




The screenshot shows a Microsoft Internet Explorer browser window. The title bar reads "Library Nyenrode Business Universiteit - Microsoft Internet Explorer". The menu bar includes "Bestand", "Bewerken", "Beeld", "Favorieten", "Extra", "Help", and "Koppelingen". The main content area features the Nyenrode Business Universiteit logo at the top, followed by a link "Go to Universiteit Nyenrode". Below this is the heading "Nyenrode Business Information Services" and a photograph of a building. At the bottom, there are links for "English" and "Nederlands", and contact information for the library.

Library Nyenrode Business Universiteit - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help Koppelingen »

 [Go to Universiteit Nyenrode](#)

Nyenrode Business Information Services



[English](#) [Nederlands](#)

Nyenrode Business Universiteit - Library

Straatweg 25, Postbus 130, 3620 AC Breukelen

The Netherlands

tel. +31 (0) 346 291310 fax. +31 (0) 346 291230

e.mail library@nyenrode.nl

```

<!DOCTYPE HTML PUBLIC "-//W3C//DTD 4.0//EN">
<html>
<head>
<title>Library Nyenrode Business Universiteit</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<meta name="keywords" content="business school library,management,executive,incompany,
education, program, research, MBA, doctoraal,nivra,bedrijfskunde,accountancy,controlling">
<meta name="description" content="Site Nyenrode Business Information Services @ Michiel
Cleijne 2001">
</head>
<body link="#0000ee" text="#000000" vlink="#551a8b" alink="#ff0000" bgcolor="#E6F2FF">
<h2 align="center"><A HREF="index.html">
</A></h2>
<h2 align="center"><b><font face="Times New Roman, Times, serif">Nyenrode Business
Information Services</font></b></h2>
<div align="center">
<p></p>
<table height="32" width="463">
<tr>
<td valign=TOP height="10" width="148">
<p align="right">&nbsp;</p>
</td>
<td valign=TOP height="10" width="153">&nbsp;</td>
<td height="10" width="146"><font face="weidemann-Book"></font></td>
</tr>
<tr>
<td valign=TOP height="7" width="148">
<div align="right"><a href="http://www.library.nijenrode.nl/INT/index.html">
English</a></div>
</td>
<td valign=TOP height="7" width="153">&nbsp;</td>
<td height="7" width="146"><a href="http://www.library.nijenrode.nl/NL/index.html">
Nederlands</a></td>
</tr>
<tr>
<td valign=TOP height="2" width="148">&nbsp;</td>
<td valign=TOP height="2" width="153"><font face="weidemann-Book"></font></td>
<td height="2" width="146">
<div align="right"></div>
</td>
</tr>
</table>
<pre><b>Nyenrode Business Universiteit - Library</pre>
<pre><b>Straatweg 25, Postbus 130, 3620 AC Breukelen<br></pre>
<pre>The Netherlands</b></font></pre>
<pre><font face="Times New Roman, Times, serif"></pre>
<pre><b>tel. +31 (0)346 291310 fax. +31 (0)346 291230</pre>
<pre>e.mail <a href="mailto:library@nyenrode.nl">library@nyenrode.nl</a></b></font>
</pre>
<p> <br>
</p>
</div>
</font></p>
</body>
</html>

```

Zoals u ziet begint de webpagina zoals de browser u die laat zien, al met de titel van de pagina in de titelbalk van uw browser. Het stukje tekst + code in het HTML-document dat zegt: `<title>Library Nyenrode Business Universiteit</title>` wordt door uw browser gelezen als de opdracht: zet in de titelbalk de tekst: Library Nijenrode Business Universiteit.

De pagina zelf opent met een logo van de Universiteit Nijenrode. Dit logo omvat zowel het wapen als de bijbehorende tekst en is een plaatje dat door de code in de pagina van elders wordt opgeroepen ('geladen'). Het plaatje heet `kleurlogo_library.gif` en wordt opgeroepen door de code `'img src'`. In dit geval staat het bestandje met de naam `kleurlogo_library.gif` op dezelfde server en in dezelfde map als de pagina die u hier ziet, maar dat hoeft niet: de kracht van het Web is dat plaatjes ook uit andere mappen, ja zelfs van andere servers kunnen worden opgeroepen. Zoals u ziet staan de afbeeldingen waarmee het plaatje in de webpagina moet worden getoond, ook opgegeven. Overigens is dit niet zomaar een plaatje: het dient tegelijkertijd als een hyperlink, een verwijzing naar een andere pagina, met het adres: <http://www.nijenrode.nl>. Als je op het plaatje klikt opent de browser de homepage van de Universiteit Nijenrode. Die verwijzing is neergelegd in de code ` ... `. Als die code in een document voorkomt zie je ofwel een plaatje, zoals in dit geval, ofwel een tekst. In het eerste geval gaat dat plaatje vaak ook gepaard met een tekst, hier 'Go to Universiteit

Nyenrode'. Zoals u ziet wordt deze tekst (gedurende enkele seconden) door de browser getoond als de muisaanwijzer op het plaatje wordt gezet; de 'tag' die dit bewerkstelligt in de HTML-pagina is de alt-tag.

Verder naar beneden in de pagina ziet u de teksten 'English' en 'Nederlands', die als hyperlinks herkenbaar zijn door hun blauwe kleur en onderstreping (andere kleuren en niet-onderstreept komt overigens ook voor). Ook hier ziet u in de achterliggende tekst de code ` ... ` staan. Net als bij het plaatje van het logo staat in de begintag `` de URL waarheen de pagina linkt, maar in plaats van een 'image source' (`img src`) vinden we hier tussen begin- en eindtag (``) de teksten English, resp. Nederlands. Doorgaans bestaan hyperlinks dus uit twee elementen: het adres (de URL) waarheen gelinkt wordt (de link), en een naam voor de pagina die achter die link te vinden is (het zogeheten 'anchor'). Dat kan de naam zijn die die pagina van haar auteur heeft gekregen, maar het kan ook een naam zijn die de auteur van de verwijzende pagina eraan geeft.

Een derde type hyperlink zien we onderaan de pagina, in het mailadres `library@nyenrode.nl` (ingeleid met `mailto:`). Deze link is vergelijkbaar met de andere twee, alleen voert klikken hierop naar een formulier met een mailadres waar men tekst kan intypen en dat vervolgens aan de geadresseerde verzonden kan worden. Het anchor (`library@nyenrode.nl`) is in dit geval identiek aan het onderliggende mailadres, maar het had natuurlijk ook een naam van een persoon kunnen zijn.

Direct onder het logo is de tekst Nijenrode Business Information Services groter afgebeeld dan de andere tekst op deze pagina. In het onderliggende HTML-bestand wordt dat effect bereikt door de tag `<h2>...</h2>`, die de browser vertelt de tussenliggende tekst groot af te beelden; met `align="center"` wordt de browser opgedragen de tekst midden op de pagina te plaatsen.

Ten slotte ziet u in de HTML-tekst direct onder de titel binnen de tags `<meta name="keywords" content=...>` en `<meta name="description" content=...>` teksten staan die u niet in de webpagina ziet verschijnen. Deze teksten zijn ook niet bedoeld om door u en mij gelezen te worden, maar wel door de spiders van zoekmachines. De informatie die ze bevatten is zogeheten meta-informatie, dus informatie die *over* de webpagina wordt gegeven.

Globaal genomen valt een HTML-bestand uiteen in twee gedeelten, de tekst binnen de tags `<head>` en `</head>` en die binnen de tags `<body>` en `</body>`. Het eerste gedeelte bevat de meta-informatie, inclusief de titel, die u immers niet in het document zelf, maar in de titelbalk van uw browser te zien krijgt; het tweede gedeelte bevat de eigenlijke tekst, zij het dat daar natuurlijk ook codes, scripts en dergelijke in kunnen voorkomen, die u niet als tekst te zien krijgt. Als een zoekmachine u met het veld (of prefix) `intext:` laat zoeken, is daarmee bedoeld: in de 'body' van de pagina.

In de HTML-code van een webpagina ligt dus heel wat informatie besloten die door zoekmachines kan worden gebruikt. Achtereenvolgens:

- de informatie dat bepaalde woorden de titel van de webpagina vormen. Dat maakt het mogelijk dat ze als zodanig worden geïndexeerd en zoekbaar (`title:` of `intitle:`) worden gemaakt en geeft ze extra gewicht in de ranking
- de informatie dat een pagina naar andere pagina's linkt. Dat maakt het mogelijk dat zoekmachines linkrelaties tussen pagina's analyseren en geeft ons de gelegenheid via die relaties (met `link:` en `anchor:`) geassocieerde webpagina's, plaatjes (`img src`) en personen (`mailto:`) op te sporen. Overall waar in de HTML-tekst `...` voorkomt heeft de pagina die we bekijken raakpunten met andere pagina's, plaatjes of personen.
- De informatie over een plaatje in de alt-tag. Deze is van grote waarde voor zoekmachines die afbeeldingen indexeren en zoekbaar maken. Aangezien de afbeeldingen zelf nog niet (op grote schaal) door zoekmachines op onderwerp

kunnen worden geïndexeerd, moeten die het hebben van de begeleidende tekst, en de tekst in de alt-tag speelt daarbij een belangrijke rol.

- de informatie dat woorden tot een kop behoren. Dat maakt ze niet als zodanig zoekbaar, maar de zoekmachine kan het wel mee laten wegen in de ranking van een zoekresultaat.
- de informatie in de metatags. In principe zouden 'keywords' en de tekst van een 'description' voor zoekmachines uiterst waardevol kunnen zijn om te bepalen waar een document over gaat, maar helaas wordt dat feest door zogeheten spammers verstoord. Spammers zijn mensen die, onder meer in de metatags, desinformatie over een pagina of site verstrekken, meestal met de bedoeling daardoor (meer) bezoekers te trekken. Toch zijn er zoekmachines die woorden in de metatags tot de titel rekenen en ze als titelwoorden zoekbaar maken. Waarschijnlijk betekent dat ook dat die zoekmachines ze een rol laten spelen in de ranking.

Voor nog meer informatie die zoekmachines over webdocumenten kunnen krijgen, zie blz. 10.

3.3.8.1. Heel precies zoeken: Title en URL

Webdocumenten hebben doorgaans een titel, die u bovenin uw browservenster ziet verschijnen. Meestal geeft die titel de essentie aan van waar een webpagina over gaat, en het zou dus prettig zijn als je niet alleen maar met zoektermen in een heel document zou kunnen zoeken, maar die zoekactie zou kunnen beperken tot die titel. Bij de meeste grote zoekmachines kan dat, waarbij de mogelijkheden van zoeken met operatoren of phrase- zoeken ook nog bewaard kunnen blijven.

Een voorbeeld biedt Google, waar je bijvoorbeeld kunt vragen naar:

- [intitle:"Harry Mulisch"](#) – levert 626 documenten met de phrase 'Harry Mulisch' in de titel
- [intitle:"Harry Mulisch" bruidsbed](#) – levert 529 documenten met de phrase 'Harry Mulisch' in de titel en het woord aanslag ergens in het document
- [allintitle:"Harry Mulisch" bruidsbed](#) – levert 73 documenten met in de titel zowel de phrase "Harry Mulisch" als het woord aanslag

(Bovenstaande voorbeelden bij Google werden begin maart 2006 gevonden met de instelling pagina's in het Nederlands.)

Een precieze zoekactie is ook die waarbij u opgeeft dat een woord in de URL (Google: `inurl:...`) moet voorkomen. Termen als 'rapporten', 'verslagen', 'images', 'notulen', 'archieff', 'brieven' enzovoort in (het pad van) de URL kunnen in combinatie met andere zoektermen elders in het document helpen een bepaald type document te vinden. De URL omvat niet alleen het domain en de site, maar ook het 'pad' (foldernamen) en de bestandsnaam van een document en biedt dus op diverse niveaus mogelijkheden om zoekacties te preciseren. Denk alleen maar aan het gebruik van jaartallen in URL's. De combinatie `inurl:archieff inurl:2004` in combinatie met `site:...` en een zoekterm voor een onderwerp kan u direct bij een gezocht document in een archief van een krant of een organisatie brengen.

Zoals al eerder vermeld is zoeken binnen de URL ook een manier om personen op te sporen die een website onderhouden en wier naam in het pad van de URL van die site voorkomt.

`intitle:` en `inurl:` zijn voorbeelden van zogeheten 'prefixen' of 'limiters' (begrenzers). Prefixen heten ze omdat ze direct vóór een zoekterm worden geplaatst, en limiters omdat ze de ruimte waarbinnen gezocht wordt, in het document of in zijn omgeving, beperken.

3.3.8.2. Meer van hetzelfde: link en anchor, related e.d.

Er is nog een andere manier om documenten met zoektermen in de titel op te sporen.

Webdocumenten verwijzen via hyperlinks naar andere documenten, en benoemen die documenten daarbij. Die naam, d.w.z. de tekst die u als hyperlink in een afwijkende kleur (doorgaans blauw) op uw scherm ziet, heet een *anchor*. Meestal is die tekst de titel die de auteur van het document waarnaar verwezen wordt, daaraan gegeven heeft (en die door de auteur van het verwijzende document is overgenomen), maar het kan ook zijn dat de auteur van het verwijzende document een andere tekst gekozen heeft.

Door naar zoektermen in een anchor te zoeken spoort u documenten op waarnaar andere pagina's met de gezochte termen in het anchor verwijzen. De documenten die u zo vindt hoeven zelf de zoektermen niet in de titel of zelfs in de tekst te hebben, bepalend is enkel dat andere documenten er met die termen naar verwijzen. Eén toepassing zou kunnen zijn om sites over de Nederlandse geschiedenis te vinden waarnaar vanuit pagina's in de Engelse taal verwezen wordt. Google is de enige grote mondiale zoekmachine die dit prefix ondersteunt, en bij deze zoekmachine zou de opdracht om het bovenstaande te doen als volgt luiden: `inanchor:"Dutch history"`.

Documenten waarnaar in andere pagina's met de phrase "Dutch history" wordt verwezen en die die phrase zelf niet in de titel hebben, zoekt u met `inanchor:"Dutch history"-intitle:"Dutch history"`. Tot de sites die op deze manier gevonden worden behoren het Nationaal Archief en het Instituut voor Nederlandse Geschiedenis. `inanchor:` vormt daarmee een interessante aanvulling op `intitle:`, en het bestaan van een zoekterm in het anchor van een verwijzend document speelt dan ook terecht een belangrijke rol in de ranking die Google aan een document waarnaar verwezen wordt toekent.

Bij een anchor hoort een *link*, d.w.z. het adres waarnaar het anchor verwijst. Hebt u eenmaal het adres van een relevant document gevonden, dan kunnen sommige zoekmachines (bij voorbeeld Google, Yahoo!, MSN Search) voor u uitvinden welke andere documenten naar dit uitgangsdokument linken (dus het adres van die pagina bevatten). Dit is een andere benadering van hetzelfde principe dat bij anchor-zoeken wordt toegepast. Het verschil zit 'm in het feit dat een adres – een URL – een uniek document aanduidt en dat met zoektermen of phrases in een anchor natuurlijk verwijzingen naar meerdere documenten kunnen worden gevonden. Zoeken met een link is dus preciezer dan zoeken met een anchor. Maar beide methodes bieden interessante mogelijkheden om gerelateerde documenten op te sporen. Denk erom dat `link:` niet bij alle zoekmachines die het ondersteunen dezelfde resultaten geeft; Google toont alleen sites met een PageRank die een bepaalde minimale waarde te boven gaat. Een interessante uitbreiding van de mogelijkheden van `link:` biedt Yahoo!'s `linkdomain:`, dat u naar pagina's laat zoeken die niet alleen naar de opgegeven pagina, maar naar alle pagina's van een opgegeven site linken. Zo kunnen we met Yahoo! `linkdomain:dbnl.nl -site:nl` achterhalen hoeveel en welke buitenlandse sites naar een pagina van de Digitale Bibliotheek voor de Nederlandse Letteren linken.

Sommige zoekmachines bieden daarnaast ook nog een mogelijkheid aan om naar soortgelijke pagina's te zoeken als een al bekende of net gevonden pagina. Dat gebeurt met opdrachten als `related:www.etc.` in de zoekregel of met een link naar Related sites of iets dergelijks bij de beschrijving van een link. Dit kan nuttig zijn als de zoekactie zelf maar een enkele relevant document heeft opgeleverd.

3.3.8.3. Een zoekresultaat beperken door toevoeging van speciale zoekcriteria

Enkele zoekmachines bieden allerlei extra mogelijkheden om het resultaat van een zoekactie te beperken. Dat kan op diverse manieren:

- door een optie te bieden om binnen een al eerder verkregen zoekresultaat verder te zoeken. Sommige zoekmachines bieden deze mogelijkheid expliciet aan. Maar in feite kunnen alle zoekmachines dit doen doordat de gebruiker een of meer

zoektermen aan de al eerder gestelde vraag toevoegt (zo nodig met Booleaans AND, OR of NOT of de unitaire operatoren + en -)

- door een taal, een domein (.edu, .gov, .com, .nl, .de enz.), een continent of een host op te geven beperkt u uw zoekresultaat tot webdocumenten die in een bepaald land, continent, type organisatie of op een bepaalde computer zijn gepubliceerd

Een grote mondiale zoekmachine gebruiken voor site search

Diverse zoekmachines maken het mogelijk een zoekactie te beperken tot de eigen site. Dat kunt u doen door aan de zoekopdracht host:... of site:... met de URL van uw site (www.etc.) toe te voegen. Alleen de resultaten van uw site worden dan gepresenteerd. Vanzelfsprekend kunt u deze zoekmogelijkheid ook gebruiken om uw zoekactie tot een andere site dan uw eigen te beperken, of, nog mooier, diverse sites tegelijk, inclusief of exclusief uw eigen, te doorzoeken. De zoekvraag wordt dan: zoekterm1 zoekterm2 site: ... OR site: ... (Googeliaans) of zoekterm1 AND zoekterm2 AND (site:... OR site:...) (Booleaans). Zie verder blz. 231 e.v. over een eigen zoekmachine.

- door op te geven dat de te vinden documenten voor het laatst na een bepaalde datum moeten zijn bewerkt kunt u uw zoekresultaat beperken tot die documenten die waarschijnlijk nog worden onderhouden. Talloze documenten op het Web zijn door hun ouders in de steek gelaten of zijn verwaarloosde kinderen. Links zijn al meer dan een jaar niet meer gecontroleerd, de informatie die ze bevatten is aan het verouderen, er zijn betere pagina's voor in de plaats gekomen, enz. Door een tijds criterium toe te voegen kunnen we althans een deel van dat soort documenten uit ons resultaat uitsluiten.

Kan een zoekmachine zien wanneer een document voor het laatst bewerkt is? Ja, dat kan. Denk maar aan de documenten op uw eigen computer, daarvan kunt u ook zien wanneer ze voor het laatst zijn opgeslagen. Maar het kan niet altijd: soms wordt de datum niet op de server bijgehouden, en dan heeft een zoekmachine daarover ook niets te melden. Zie verder blz. 128 e.v., What's in a date

- door een *Page Depth* op te geven – een mogelijkheid die Yahoo! biedt – beperkt u uw zoekresultaat tot pagina's die tot op een bepaald niveau in een site verborgen zitten. Sites kunnen tientallen, honderden, duizenden pagina's of meer omvatten, die over diverse niveaus (van sub- en sub-sub-directories) verdeeld zijn. Hoe hoger een document in die hiërarchie van subdirectories voorkomt, des te groter de kans dat de zoektermen relevant zijn voor de site, of een gedeelte daarvan, en niet alleen maar voor één enkele pagina. Een Page Depth van 3 bij Yahoo! betekent dat een bestand drie subdirectories diep zit (anders gezegd: de URL bevat na de domeinnaam (bij voorbeeld amazon.com), die wordt afgeloten met een slash, nog drie extra slashes). Als u een zoekmachine een 'Page Depth' opgeeft, is dat een maximale Page Depth. Page Depth 3 is dus Page Depth 0-3. Een variant is de optie om één pagina per website te tonen. Dit maakt een snel overzicht van sites mogelijk waar een zoekterm te vinden is.
- door een *feature* op te geven kunt u uw zoekresultaat beperken tot documenten die niet alleen de opgegeven zoektermen, maar tevens bepaalde extra's bevatten. Die extra's kunnen zijn:
 - video, audio, image, MP3, Shockwave, Java applets, objects, Javascript, ActiveX, VMRL, Acrobat, VB Script, Win Media, Real Audio/Video en bestandsformaten met een bepaalde extensie (bij voorbeeld .GIF, .JPG, .PPT). Het gaat hierbij merendeels om grafische en geluidsformaten, stukjes programmacode en scripts die het Web tot een multimediaal kanaal maken. Een simpel voorbeeld van een

zoekactie die je hiermee kunt doen is dat je plaatjes van Aziatische olifanten zoekt door met de zoektermen Asian en elephants te zoeken en daarbij op te geven dat de te vinden documenten naar een of meer bestanden met de extensie .GIF verwijzen. Beethoven als zoekterm in combinatie met het criterium MP3 wijst hopelijk de weg naar een klassieke geluidsfile. Een zoekmachine die in dit opzicht veel verschillende mogelijkheden biedt is Yahoo!.

- daarnaast kan – in combinatie met zoektermen – ook naar die pagina's worden gezocht waarin forms, frames of tables in de HTML voorkomen, waarin plug-ins, of meer speciaal flash plug-ins zijn opgenomen (embedded) of waarin (in het algemeen) scripts zijn opgenomen. Handig voor gespecialiseerde vragen, misschien ook wel eens om modern vormgegeven pagina's op te sporen, maar voor de meeste zoekacties geen belangrijk extra. Bovendien bestaan voor veel speciale bestandsformaten speciale zoekmachines.
- een apart zoekfilter bij veel zoekmachines is ten slotte een 'family filter', dat de bedoeling heeft obscene documenten uit het zoekresultaat te weren. Nuttig, maar het filter verwijdert niet alle obscene materiaal en verwijdert ook niet-obscene documenten. (Het woord 'hardcore' bij voorbeeld slaat niet alleen op sex, maar ook op rock.)
- veel zoekmachines hebben tegenwoordig een automatische spellingcontrole, die u waarschuwt als u een typfout maakt (en ook wel eens met gekke voorstellen komt als u dat – in een andere taal dan het Engels – niet doet). Zulke spelling checkers zijn slim genoeg met woorden, maar van phrases bakken ze meestal niet veel. Ook hier zijn weer uitzonderingen op: probeer Google maar eens een vorm van de naam “Britney Spears” te voeren waarbij deze zoekmachine niet met een suggestie voor de juiste spelling komt. Een goede spellingcontrole kan [wonderen doen](#).

3.3.8.4. Sortering

De meeste zoekmachines werken met één standaard sortering van het zoekresultaat, maar er zijn er die vanaf het zoekscherm (MSN Search) of in de Preferences (Exalead) de keuze tussen verschillende sorteringen bieden. Het verst daarin gaat MSN Search, dat u vanaf het zoekscherm de te verwachten treffers op actualiteit, populariteit en relevantie laat sorteren. Dat gebeurt door middel van een mengpaneel met schuifregelaars, zodat men de diverse sorteringen ook nog kan 'mengen'.

3.3.8.5. Shortcuts en rekenfuncties

Grote mondiale zoekmachines zoals Google, Yahoo! en MSN Search hebben nog een aantal extra's die vermelding verdienen: zogeheten shortcuts en rekenfuncties. *Shortcuts* zijn prefixen waarmee snel allerlei praktische vragen kunnen worden opgelost, zoals het omrekenen van maten en gewichten, het opvragen van weersinformatie voor alle grotere steden van de wereld en het volgen van postpakketten.

Rekenen kunnen zoekmachines ook, en sommige hebben zeer geavanceerde rekenfuncties. Belangstellenden hiervoor verwijs ik naar de helpfiles van de betrokken zoekmachines, de shortcuts behandel ik wel in het kort.

3.3.9. Standaard search en advanced search

Vrijwel alle zoekmachines hebben twee verschillende zoekschermen, een met één simpele zoekregel (met links naar een 'advanced search', 'preferences' (voorkeursinstellingen die via een cookie op uw PC worden geparkeerd), helpteksten en eventueel ook speciale of andere zoekmachines) en het andere met een aantal 'geavanceerde' opties. In het algemeen zijn de extra opties van het geavanceerde zoekscherm bedoeld om de gebruiker

- een zoekvraag te helpen bouwen (door de keuze tussen EN, OF en NIET, het

invoeren van een phrase en het gebruik van prefixen te vereenvoudigen of uit handen te nemen), om preciezer zoeken mogelijk te maken en om de gebruiker menukeuzes te bieden.

- om hem zijn vraag tot een land, taal, documenttype en/of site of domain te helpen beperken
- om extra mogelijkheden, zoals `link:` en `related:` of fonetisch en fuzzy zoeken, aan te bieden

Ze bieden niet noodzakelijk meer functionaliteit dan de standaard search, maar bieden een aantal mogelijkheden van de standaard search waar vaak bij die zoekmethode geen gebruik wordt gemaakt, op een meer expliciete manier aan.

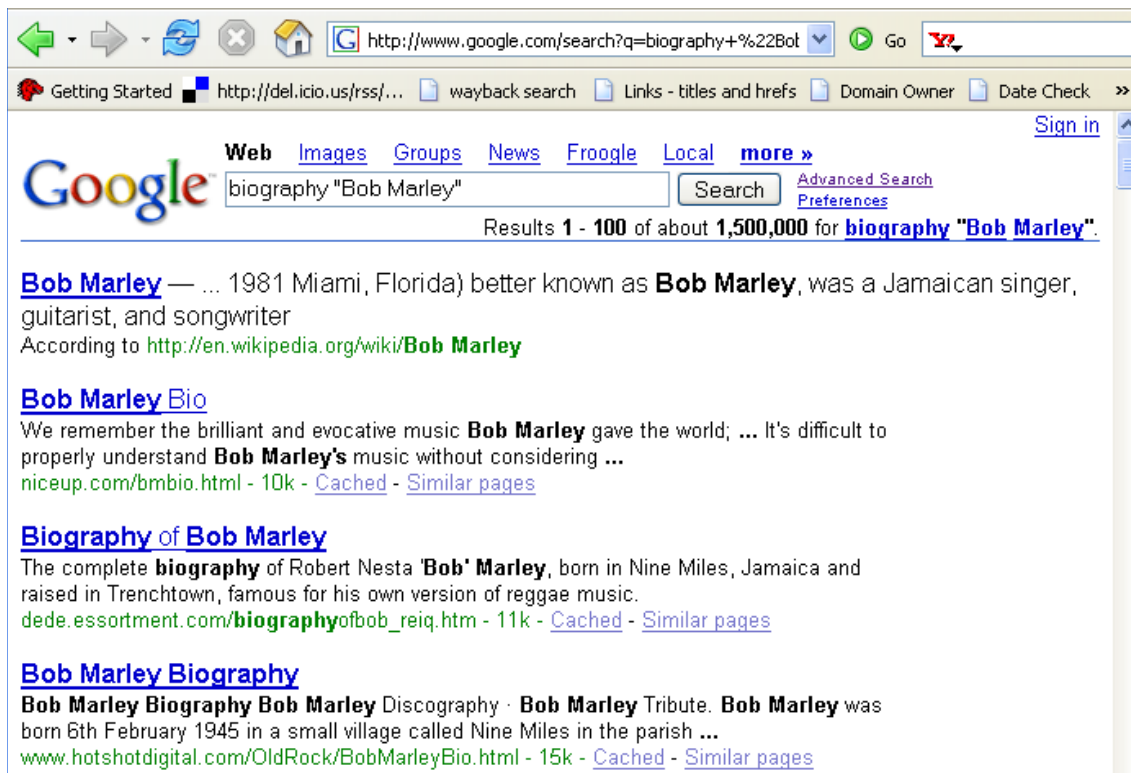
Daarnaast biedt de 'advanced search' de gebruiker soms een grotere vrijheid om instellingen aan zijn wensen aan te passen. Die kunnen betreffen de sortering van de gevonden documenten, een optie om (in eerste instantie) maar één of enkele documenten per site te zien ('clusteren'), het al dan niet vertonen van een beschrijving en/of de lengte daarvan, de aantallen documenten per pagina die getoond moeten worden enz. Meestal kunnen deze instellingen ook in de Preferences worden aangepast. Een probleem van Advanced-search schermen is dat ze gebruikers weinig flexibiliteit bieden en bepaalde opties (bijvoorbeeld het gebruik van twee phrases, zoeken naar sommige woorden in de titel en andere in de tekst) ontnemen. Een oplossing biedt dan de mogelijkheid om een zoekvraag in het resultatenscherm verder aan te passen; de grote mondiale zoekmachines tonen in het resultatenscherm van een geavanceerde zoekactie de zoekregel van de Basic Search met daarin de vraag zoals die in de Basic Search geluid zou hebben. Maar aangezien de zoektaal van zoekmachines nu ook weer niet zo moeilijk of uitgebreid is, kunt u zich ook aanwennen het Advanced Search-scherm van een zoekmachine links te laten liggen en al uw vragen – met operatoren, prefixen, jokers etc. – direct in de standaard zoekregel in te typen.

3.3.9.1. Manipulatie van de URL

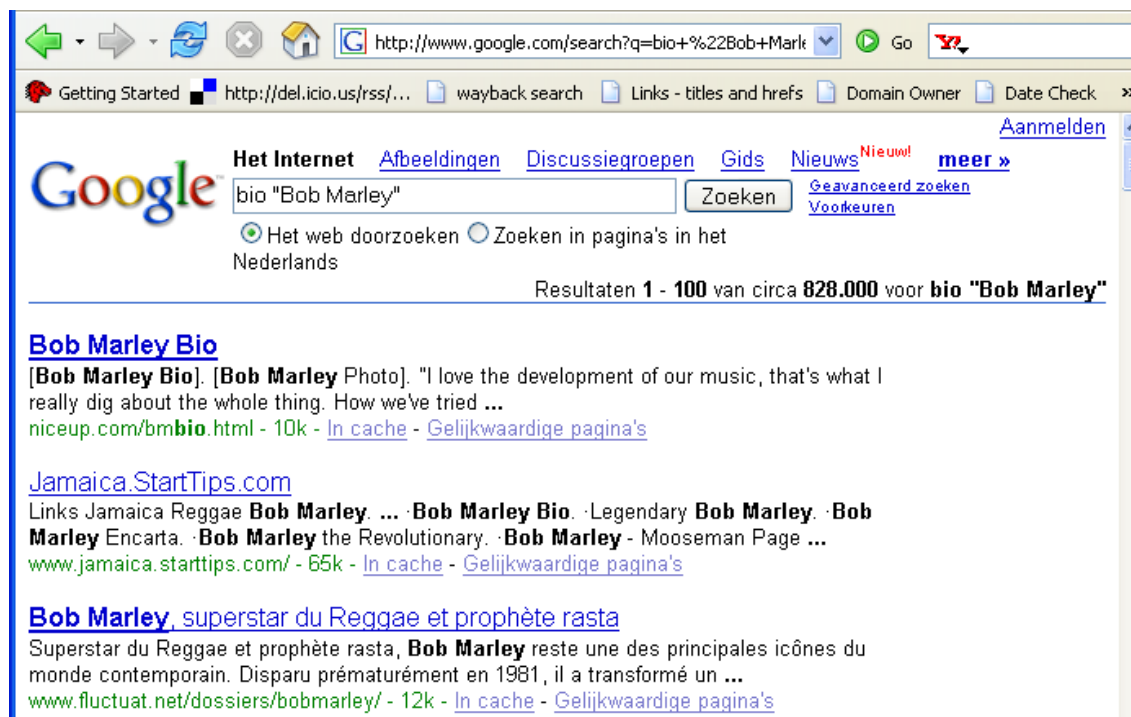
Naast de drie hierboven genoemde manieren van zoeken – Basic Search, Advanced Search en commandogestuurd zoeken via Basic Search – biedt ook de URL van een zoekresultatenpagina nog enkele mogelijkheden die hier moeten worden genoemd. In de eerste plaats is de URL die de meeste zoekmachines als het resultaat van een zoekactie opleveren, als een zelfstandig adres te gebruiken, d.w.z. als u zo'n URL als bookmark opslaat of per e-mail aan uzelf of iemand anders verstuurt, voert de zoekmachine de zoekactie bij het aanklikken van de link op precies dezelfde manier en met dezelfde instellingen uit als de oorspronkelijke zoekvraag. Dat is handig voor didactische doeleinden. Als iemand u meldt dat hij niet kan vinden hoe de reggaemuzikant Bob Marley aan zijn eind is gekomen kunt u hem natuurlijk melden dat de man aan een hersentumor is overleden, maar u kunt hem – bij een geslaagde zoekactie – ook de resultatenpagina van een zoekmachine toesturen.

<http://www.google.com/search?hl=en&lr=&q=biography+%22Bob+Marley+%22&btnG=Search>

Als de ontvanger deze link aanklikt krijgt hij het volgende te zien (zie volgende bladzijde):



Een andere mogelijkheid is de URL te wijzigen. Door het element hl=en te veranderen in hl=nl dwingt u Google naar de interfacetaal Nederlands over te gaan, iets wat u normaal gesproken via de Preferences zou doen. Ook kunt u de term biography veranderen in bio. Als de link hieronder wordt aangeklikt <http://www.google.com/search?hl=nl&lr=&q=bio+%22Bob+Marley+%22&btnG=Search> ziet het resultaat er zo uit:



Als u bij het bekijken van resultaten naar een volgende pagina doorbladert, verandert ook de URL:

http://www.google.com/search?q=bio++%22Bob+Marley%22&num=100&hl=nl&lr=&newwindow=1&c2coff=1&safe=off&as_qdr=all&start=100&sa=N

Door het element `start=100` te veranderen in `start=900` kunt u vervolgens direct naar de laatste treffers in Google's zoekresultaat bladeren. Google toont maximaal 1000 treffers.

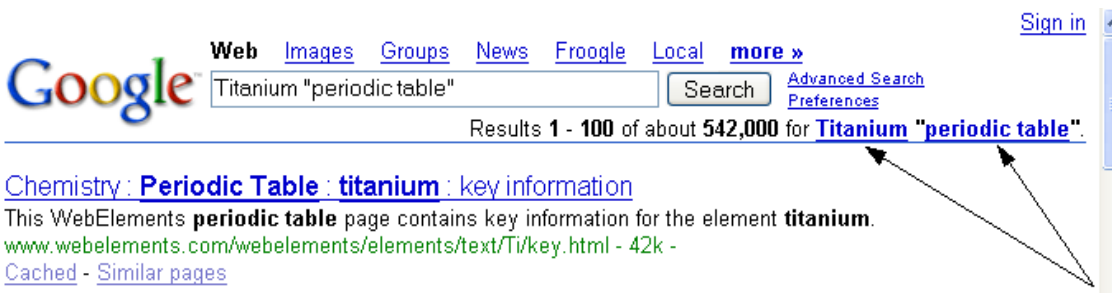
Elke zoekmachine heeft zijn eigen manier om zoekresultaten en instellingen in zijn URLs zichtbaar te maken, maar met enig geëxperimenteer komt u er wel achter met welke elementen in die URLs u kunt manipuleren en met welke niet. Manipulatie van de URL kan bij sommige zoekmachines mogelijkheden bieden die via de andere interfaces niet beschikbaar zijn. Zo kon tot voor kort bij MSN Search het maximale aantal treffers dat wordt getoond van 250 naar 350 worden verhoogd.

3.3.10. *Het zoekresultaat*

Een zoekactie resulteert in een zoekresultaat, van 0 (zero-hit) of meer treffers. Zulke treffers heten in het Engels 'hits' of 'matches'. Met treffers worden in de wereld van de zoekmachines de gevonden documenten bedoeld, niet de gevonden termen. Dat onderscheid is wel van belang, want een document kan een zoekterm vanzelfsprekend meermalen bevatten.

De kwaliteit van een zoekmachine wat de presentatie van de zoekresultaten betreft is onder meer afhankelijk van:

- de volgorde waarin hij gevonden documenten presenteert (op basis van relevancy ranking, aantallen bezoekers en duur van de bezoeken ['click popularity'], aantallen links die naar een document worden gelegd ['link popularity'], link popularity van de documenten die naar het document linken, enz.)
- de informatie die hij over de gevonden documenten verschaft (datum last update, lengte samenvatting/beschrijving, melding welke zoektermen gevonden zijn of markering van de zoektermen in het document)
- de vraag of hij ook rubrieken uit een bijbehorende directory toont, waardoor de gebruiker zich kan oriënteren en zijn vraag kan herformuleren
- of en hoe hij na het eerste zoekresultaat nog een verdere bewerking van dat resultaat mogelijk maakt (andere sortering, suggestie verdere zoektermen, mogelijkheid soortgelijke documenten te vinden [More like this, Find similar pages], zoeken binnen het al verkregen resultaat)
- de vraag of een zoekmachine alle met de zoektermen gevonden pagina's van een site in het zoekresultaat presenteert, met de mogelijkheid dat tot één pagina per site te beperken ('clustering'), dan wel één of enkele pagina's per site presenteert ('automatic clustering'), met de mogelijkheid er meer te zien als men dat wenst
- de aanwezigheid van extra's, zoals de mogelijkheid om van documenten in vreemde talen online een ruwe vertaling te krijgen ('translate'), de mogelijkheid om het aantal gevonden documenten per pagina in te stellen, een andere dan de standaard sortering, het presenteren van 'related searches' (zoekacties van anderen met ongeveer dezelfde term(en), die u op een idee kunnen brengen uw vraag nog eens anders te formuleren), en de mogelijkheid om te kiezen uit een lange, een korte of geen beschrijving van die documenten. Populair is het *categoriseren* van zoekresultaten, waarbij de zoekmachine de gebruiker een aantal aspecten of rubrieken presenteert waarin een zoekresultaat verder kan worden ingedeeld. Dit werkt vooral goed bij vragen met een of twee zoektermen zonder operatoren, die een heel groot resultaat opleveren. Een extra waar u niet moet nalaten gebruik van te maken is de link die Google en Yahoo! leggen naar naslaginformatie ('fast facts') over de term(en) die u hebt ingevoerd. Ze doen dat in de gemarkeerde regel waarin ze de omvang van het zoekresultaat melden.



Bovenin de resultatenlijst heeft Google een regel waarin het het aantal treffers bij de zoekactie meldt. Daarin zijn links naar naslaginformatie (encyclopedie, woordenboek etc.) over de opgegeven zoekterm(en) opgenomen.

- last but not least bieden enkele zoekmachines de mogelijkheid van afzonderlijke treffers de cachekopie te bekijken, dat wil zeggen de kopie van een pagina die de spider van de zoekmachine heeft opgevraagd en opgeslagen. Dat is vooral van belang omdat de oorspronkelijke pagina inmiddels verwijderd of veranderd kan zijn, of niet bereikbaar. Sommige zoekmachines die een cachekopie laten zien, markeren daarin de zoektermen.

Ten slotte blijken uit het zoekresultaat nog een paar dingen die niets met de presentatie van de documenten te maken hebben, ook niet met de zoekmogelijkheden, maar met processen achter de schermen van de zoekmachine:

- hoe goed een zoekmachine *spam* buiten de deur weet te houden
- wat het percentage 'dead links' (verwijzingen naar verdwenen adressen) is dat een typische zoekactie oplevert; anders gezegd: hoe frequent 'ververst' de zoekmachine zijn database

Wat is 'spam'?

Het Engelse woord spam betekent eigenlijk 'gekookte, ingeblikte ham', met de bijbetekenis 'slappe troep'. In de internetwereld wordt het woord gebruikt voor ongewenst reclamemateriaal op mailing lists en voor pogingen om met oneigenlijke middelen een hoge ranking te verkrijgen bij zoekmachines. Zoekmachines voeren een voortdurende strijd om spam buiten de deur te houden. Ze doen dat door pagina's die ze indexeren ook te analyseren op mogelijke kenmerken die op spam wijzen.

Zie: What, Exactly, is Search Engine Spam

(<http://searchenginewatch.com/searchday/article.php/3483601>)

3.3.11. Customizing

Onder *customizing* wordt verstaan het aanpassen van diverse zoek- en presentatiemogelijkheden van een zoekmachine aan de persoonlijke wensen van de gebruiker, incidenteel of permanent. Voor het laatste is meestal nodig dat uw browser een 'cookie' accepteert, d.w.z. een bestandje met informatie over uw keuzes dat in het geheugen van uw PC of op de harde schijf wordt geparkeerd en dat door een server wordt herkend en gelezen als u weer verbinding met die server maakt. In zo'n cookie legt u vast in welke taal en welk alfabet u zoekt, hoeveel treffers een resultatenpagina moet bevatten, hoe de treffers moeten worden gesorteerd (relevantie, populariteit, actualiteit), welke informatie een treffer moet bevatten, of zoektermen daarin gemarkeerd moeten worden, of 'Family Filtering' aan of uit moet staan enzovoort. U vindt deze functionaliteit onder 'Customizing', 'Preferences', 'Personalizing' en dergelijke.

3.3.12. *Samenvatting*

Zoekmachines bieden hun gebruikers een breed spectrum van middelen aan om een vraag te stellen en het zoekresultaat gepresenteerd te krijgen. Aan het eind van dit hoofdstuk is het goed die vanuit een functioneel oogpunt op een rijtje te zetten.

- De meeste zoekmachines doorzoeken standaard het hele document op zoektermen en werken met relevancy ranking. Dat betekent dat het altijd mogelijk (en in sommige gevallen zelfs wenselijk) is een vraag zonder meer in de vorm van een aantal steekwoorden in de zoekregel van een zoekmachine in te typen en die dan voor u te laten beslissen welke documenten daar als het meest relevant bij passen.
- Veel van de zoekmogelijkheden die zoekmachines hun gebruikers bieden zijn bedoeld om hen in staat te stellen het resultaat van een vraag te beperken en te preciseren: beperken in de zin van een kleiner zoekresultaat, en preciseren in de zin van een relevanter (en ook beter geordend) zoekresultaat. Tot deze mogelijkheden behoren, in volgorde van effectiviteit:
 - zoeken met een beperking tot bepaalde velden (URL, title, anchor, link enz.)
 - zoeken met phrases
 - zoeken met nabijheidsoperatoren (NEAR etc.)
 - zoeken met de Booleaanse operatoren AND en (AND) NOT
 - zoeken met de unitaire operatoren + en -
 - zoeken met unitaire operatoren in combinatie met 'niet gepluste' of 'niet geminde' zoektermen (waar dat nog mogelijk is)
 - zoeken binnen (een) bepaald(e) domein(en): .edu, .gov, .nl, .be
 - zoeken met hoofdletters en kleine letters ('case-sensitive')
 - zoeken naar pagina's die (verwijzingen naar) bepaalde extra's (plaatjes, geluidsfiles, scripts, niet door de zoekmachines geïndexeerde tekstdocumenten met een bepaald bestandsformaat) bevatten
- Een aantal zoekmogelijkheden is bedoeld om gebruikers heel uitgebreid te laten zoeken en zoveel mogelijk documenten in het zoekresultaat te presenteren. Dit zijn:
 - de Booleaanse operator OR (in combinatie met AND)
 - trunceren en maskeren
 - Google's tilde (~)
- Zoekmogelijkheden die het zoekresultaat beperken en zoekmogelijkheden die het resultaat uitbreiden, lijken bedoeld voor verschillende doeleinden. In werkelijkheid zijn ze dat niet: beide, heel vaak in combinatie met elkaar, zijn bedoeld om een zowel zo compleet als zo relevant mogelijk zoekresultaat te verkrijgen. Dus: als u in titels zoekt en met de Booleaanse operator AND, vergeet dan niet om te trunceren waar dat zinvol is, enz.

Een compleet en relevant resultaat is wat u wenst, of het daarbij nu om één document gaat, of om tientallen. In de informatiekunde worden de begrippen compleetheid en relevantie van een zoekresultaat uitgedrukt in de termen *recall* en *precision*. Daarbij is recall de verhouding tussen het totale aantal relevante documenten dat u gevonden hebt, en het totale aantal relevante documenten dat er is: hoeveel van wat er is hebt u ook gevonden? Precision is de verhouding tussen het aantal relevante documenten dat u gevonden hebt en het totale aantal documenten dat u gevonden hebt: hoeveel van wat u gevonden hebt is ook relevant? Met andere woorden: recall meet (in een percentage) de effectiviteit van een zoekactie, precision de efficiency. Wat niet relevant is, heet 'ruis'.
100% recall: u hebt alle relevante documenten gevonden die er zijn
100% precision: alle documenten die u gevonden hebt zijn relevant
Efficiency en effectiviteit staan niet los van elkaar: een goede zoekactie scoort op beide criteria zo hoog mogelijk.

4. De afzonderlijke zoekinstrumenten

In dit hoofdstuk gaan we nader in op de specifieke mogelijkheden en bijzondere eigenschappen van de voornaamste vertegenwoordigers van de diverse zoekinstrumenten: directories en verzamelingen van Webgidsen, zoekmachines en metazoekmachines. We gaan er daarbij van uit dat de vorige hoofdstukken u een algemene indruk hebben gegeven van wat je met de verschillende zoekinstrumenten kunt.

Maar voordat we de afzonderlijke zoekinstrumenten behandelen, wijs ik u op enkele webpagina's die u een idee kunnen geven van de diversiteit aan zoekinstrumenten en de factoren die een rol kunnen spelen bij de beslissing welke van die zoekinstrumenten u bij een bepaalde vraag wilt gebruiken. Het gaat hierbij om

- Choose the Best Search for Your Information Needs
(<http://www.noodletools.com/debbie/literacies/information/5locate/adviceengine.html>)
- Phil Bradley: Finding what you need with the best search engines
(<http://www.philb.com/whichengine.htm>)
- NoodleQuest - The Search Strategy Wizard
(<http://www.noodletools.com/noodlequest/>)

De eerste twee geven bij een aantal verschillende typen van vragen – 'I need images and sounds...', 'I need statistical data depth', 'I need a map', 'I want to get advice and opinions on current issues' – de geschiktste zoekinstrumenten. NoodleQuest laat u een aantal keuzes aankruisen, zoals 'I want a search engine based outside the United States', 'I want to be able to search by a phrase' enz. en komt dan op basis van het door u gegeven profiel met enkele adviezen voor zoekinstrumenten die u kunt gebruiken. Ook zonder dat u deze hulpmiddelen ook werkelijk gebruikt geven ze u al een idee van de mogelijkheden.

Precieze en actuele informatie over het gebruik en de mogelijkheden van de afzonderlijke zoekinstrumenten vindt u in de Help bij ieder zoekinstrument. Ook in dit boek wordt daar natuurlijk het nodige over gezegd, maar die informatie verouderd snel. Daarnaast zijn er een aantal plaatsen op het Web waar de informatie over dit onderwerp zelfstandig wordt bijgehouden, en waar ook handige overzichten worden gegeven van waar u bepaalde zoekmogelijkheden kunt aantreffen ('Zoeken met Booleaanse operatoren kunt u met ..., ..., ... en'). De belangrijkste zijn:

Geordend per zoekinstrument/zoekmachine:

Search Engine Features (<http://www.notess.com/search/features/>)

Best Search Tools Chart (<http://www.infopeople.org/search/chart.html>)

Comparison of Search Engine User Interface Capabilities by Gillian Westera
(<http://library.curtin.edu.au/staff/personal/gwpersonal/compare.html>)

Geordend per zoekmogelijkheid:

How to Choose a Search Engine or Directory

(<http://library.albany.edu/internet/choose.html>)

Search Engine Features Chart (<http://www.searchengineshowdown.com/features/>)

Recommended Search Engines: Tables of Features

(<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html>)

De hierboven genoemde overzichten waren begin 2006 goed bijgehouden, maar veranderingen in de wereld van de zoekinstrumenten gaan snel. Greg Notess' Search Engine Showdown (<http://searchengineshowdown.com/>) is een betrouwbare bron voor nieuws over zoekinstrumenten. Wilt u meer, dan vindt u dat bij de Search Engine Detective, search engine news intelligence tools and resources (<http://www.pandia.com/searchworld/detective.html>). Zie ook het hoofdstuk Bijblijven.

4.1. Directories, (verzamelingen van) webgidsen en startpagina's

Directories en (verzamelingen van) Webgidsen kunnen doorgaans op twee manieren worden gebruikt, door erin te browsen, d.w.z. de onderwerpsindeling van algemeen naar bijzonder te volgen, dan wel door de interne zoekmachine te gebruiken. Als die er is, zijn de zoekmogelijkheden van zo'n zoekmachine meestal minder groot dan die van de pure zoekmachines.

Diverse directories bieden naast de interne zoekmachine ook een externe zoekfaciliteit aan. Meestal gebruiken ze daarvoor de eigen zoekmachine, met een externe database van webpagina's. Denk erom dat de zo gevonden pagina's niet door de redacteurs van de directory zijn geselecteerd en beschreven.

4.1.1. Algemene, mondiale directories

De bekende algemene mondiale directories zijn:

- Yahoo! (<http://dir.yahoo.com/>)
- Open Directory Project (ODP; <http://dmoz.org>)

Hiervan is de tweede een vrijwilligersproject dat bovendien 'Open Content' is, dat wil zeggen: door iedereen die dat wenst gedownload en (met bronvermelding) opnieuw gepubliceerd mag worden. U kunt deze directory, geheel of gedeeltelijk, dan ook onder tal van namen op tal van plaatsen en in tal van gedaanten terugvinden, onder meer bij de grote mondiale zoekmachines Google en Exalead en de Nederlandse zoekmachine Track. Ik bespreek hier alleen het oorspronkelijke ODP.

4.1.1.1. Yahoo!

In kort bestek:

URL: <http://dir.yahoo.com>

Omvang: ca. 3 miljoen websites

Aantal rubrieken: onbekend.

Aantal rubrieken op de startpagina: 14, met 48 voorbeelden van subrubrieken eronder

Maximaal aantal niveaus: 8.

Onderhouden door: ca. 100 redacteurs in dienst van Yahoo!

Zoekmachine intern: ja, met keuze tussen doorzoeken van het Web, de hele directory of de rubriek ('category') waarin men vertoeft; sites worden getoond met de route in de directory die erheen leidt

Zoekmachine extern: ja (doorzoekt Yahoo!'s eigen database van ca. 8 miljard documenten)

Advanced Search: Yahoo! Advanced Directory Search

(<http://search.yahoo.com/dir/advanced>)

Beschrijvingen: niet altijd aanwezig; indien aanwezig een of twee regels lang. Tussen de 1000 en 2000 sites hebben een review, van Yahoo! Picks. Deze hebben ook een pictogrammetje met de tekst PiCK! en zijn door de Yahoo!-redactie als apart en origineel beoordeeld. Voor een overzicht van alle PiCK-sites zie:

<http://picks.yahoo.com/picks/>. Daarnaast heeft een beperkt aantal sites in niet-commerciële rubrieken een zonnebrilletje als pictogram, als teken van een uitzonderlijk goede redactionele beoordeling.

Typen resources: ja: o.a. conferences, education, employment, events, institutes, journals, mailing lists, news and media, organizations, Web directories, Usenet. Deze zijn niet overal beschikbaar, maar als ze er zijn, heten ze (meestal) zo.

Sponsored/featured links (commercie): ja (in commercial categories), in (sub)rubrieken onder de kop Sponsor Results onderaan de pagina tegen achtergrondkleur.

Nationale versies: Argentinië, Australië en Nieuw-Zeeland, Azië, Brazilië, Canada

(Engels en Frans), China, Duitsland, Frankrijk, Groot-Brittannië en Ierland, Hongkong, India, Italië, Japan, Korea, Mexico, Noorwegen, Singapore, Spanje, Taiwan, Zweden; daarnaast ook in diverse talen anders dan het Engels (zie: <http://world.yahoo.com/>).

Help: Yahoo! Help Central (<http://help.yahoo.com/help/us/dir/>)

Meer over Yahoo!: Yahoo! FAQ (<http://docs.yahoo.com/info/pr/faq.html>)

Yahoo! Review on Search Engine Showdown

(<http://www.searchengineshowdown.com/dir/yahoo/review.html>)

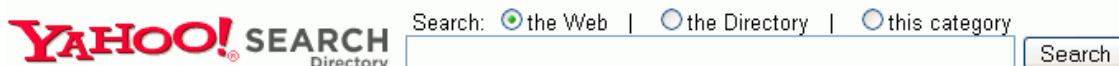
Abondance: Fiche descriptive de l'annuaire Yahoo! (<http://outils.abondance.com/yahoo-annuaire.html>)

Yahoo! is de oudste en de meest gebruikte van de directories, en zelfs het meest gebruikte van alle zoekinstrumenten: zo'n honderd miljoen surfers op het Web bezoeken Yahoo! (of een van zijn nationale versies) een of meer malen per maand, en een aanzienlijk deel van hen doet dat om via de directory te zoeken.

Yahoo! is dan ook een van de meest gebruikersvriendelijke zoekinstrumenten. Het biedt een compacte, duidelijke interface, waarin zowel de classificatie als zoekresultaten op een consequente manier worden gepresenteerd. Bovendien heeft Yahoo! als portal ook veel te bieden, onder meer e-mail, een bookmark manager waarmee u bij Yahoo! een eigen bookmarkbestand kunt opbouwen dat u dan op elke computer kunt raadplegen, Yahoo! Groups als list server, My Yahoo!, enzovoort. Voor een compleet overzicht van Yahoo!-diensten zie: Everything Yahoo! (<http://docs.yahoo.com/docs/family/more/>). Yahoo!'s directorystructuur is praktisch en bruikbaar. Handig is dat op de beginpagina onder de 14 hoofdruubrieken al bij wijze van voorbeeld links naar subrubrieken staan. Opklimmend in de boom zie je de route waarlangs je dit doet zowel in de URL als bovenin het Yahoo!-venster verschijnen.

Veel, maar niet alle, 'categories' of subrubrieken (maar niet de hoofdruubrieken) in Yahoo! zijn met links gevuld, op veel niveaus tussen algemeen en zeer specifiek vindt u dus links, de 'Site listings'. Dat heeft ook gevolgen voor het voorkomen van zie-verwijzingen: Yahoo! is een van de directories met zie-verwijzingen (aangegeven door een @ teken).

Nieuwe subrubrieken worden bij Yahoo! kennelijk toegevoegd als een rubriek voor praktisch gebruik te groot dreigt te worden.



Wie in Yahoo! de boomstructuur volgt, houdt steeds de zoekregel tot zijn beschikking. Daarbij heeft men de keuze tussen zoeken op het Web, in de hele directory of in de rubriek waarin men op dat moment is. Bij het laatste worden overigens ook de onderliggende rubrieken doorzocht.

Bij het browsen van de directory blijft men binnen Yahoo!; er worden dus geen websites gepresenteerd die niet in Yahoo! zijn opgenomen. Alle sites binnen een rubriek worden op één pagina gepresenteerd, eerst de in de directory opgenomen sites, in volgorde van populariteit (met als alternatief een sortering op alfabet), daarna, als ze er zijn, de Sponsor results tegen een gekleurde achtergrond.

Wie in Yahoo! niet wil browsen, maar de zoekmachine gebruikt om de directory te doorzoeken, heeft steeds de keuze tussen zoeken in heel Yahoo! of in de rubriek (plus onderliggende rubrieken) waarin hij of zij op dat moment vertoeft. En een zoekactie biedt vrijwel altijd een gefaseerd resultaat. Achtereenvolgens presenteert Yahoo! (op de eerste resultatenpagina):

- (maximaal vijf) eigen rubrieksnamen waarin uw zoekterm(en) voorkomt(en)

- (Categories) (met een link naar 'More')
- enkele Sponsor results, zoekresultaten waarvoor Yahoo! zich door de geadverteerde bedrijven laat betalen
- (10 van) de websites in de directory. Via de Preferences kan dit aantal worden verhoogd tot maximaal 100.
- nog meer Sponsor results, ook in een kolom aan de rechterkant van het browservenster

Wie niet in de directory maar op het Web zoekt, maakt gebruik van de eigen zoekmachine van Yahoo! (zie blz. 152 e.v.) en verlaat de directory. Deze blijft echter als menukeuze beschikbaar in de resultatenpagina van de zoekmachine. Yahoo!'s zoekmachine dekt naar schatting tegen de acht miljard webdocumenten en is daarmee een van de grootste zoekmachines op het Web. Via de directory krijgt u op deze manier dus toegang tot een wereld die de grenzen van de directory verre te buiten gaat. Zoals al eerder vermeld onderscheidt Yahoo! onder meer Web directories als een aparte categorie (of type resources). Dit zijn sites op een vakgebied of over een onderwerp die als gidsen voor dat vakgebied of onderwerp fungeren en meestal grote aantallen links naar meer gespecialiseerde onderwerpen bevatten. Ook deze Web directories, die doorgaans een veel vollediger overzicht van resources over een onderwerp of op een vakgebied bieden dan Yahoo! zelf, breiden de capaciteit van Yahoo! uit tot ver over zijn eigen grenzen heen.

The screenshot shows the Yahoo! search interface. At the top, there are navigation links for Web, Images, Video, Directory, Local, News, and Products. The search bar contains the word 'drugs' and a search button. Below the search bar, there are links for Shortcuts, Advanced Search, and Preferences. The main content area is titled 'Directory' and shows 'Results 1 - 100 of about 2,716 for drugs. (About this page...)'. On the left, there are 'RELATED DIRECTORY CATEGORIES' including 'Drugs and Medications', 'Specific Drugs and Medications', and 'More...'. In the center, there are 'SPONSOR RESULTS' for 'Clinical Drug Trials Treatments' and 'Canada Pharmacy - Official Site'. On the right, there are more 'SPONSOR RESULTS' for 'Canada Pharmacy - Official Site', 'CanadaMedicineShop.com - Canada Pharmacy', and 'Your Health First - Canada Rx Drugstore'. At the bottom, there are two numbered search results: '1. Drugs.com' and '2. Food and Drug Administration (FDA)'. Each result includes a brief description, a category, and a URL.

Als u de zoekmachine van Yahoo! gebruikt om de directory te doorzoeken, vindt die de zoektermen in achtereenvolgens de rubriksaanduidingen (Categories) en de websites in de directory; de laatste worden boven, onder en rechts omgeven door Sponsor results.

De interne zoekmachine van Yahoo! staat op elke pagina. Op de homepage doorzoekt

hij standaard de hele directory van Yahoo! zelf (d.w.z. de categorie-aanduidingen, de titels van de websites en hun beschrijvingen), en bij het aanklikken van het keuzerondje 'the Web' de database van 4 miljard webdocumenten. Eenmaal in een categorie hebt u naast de optie 'zoeken op het Web' de keus tussen zoeken in de directory (d.w.z. de hele directory), en zoeken in de actuele rubriek en de daaronder vallende subrubrieken. De zoekmachine ondersteunt:

- de unitaire operatoren + en -
- `title: (t:)` en `url: (u:)` (de prefixen title en url)
- phrases: "..."
- rechts trunceren met * (Yahoo! truncateert woorden van zes of meer letters zelfs automatisch, een functie die kan worden uitgeschakeld door een term als phrase op te geven: serendip vindt naast serendip ook serendipity, "serendip" vindt alleen serendip.).

Standaard zoekt de zoekmachine bij meer zoektermen met een AND-relatie.

N.B. Als u gewend bent met prefixen in de 'Basic search' te zoeken, is het wel zaak de syntax van de directory en die van de webzoekmachine uit elkaar te houden. Zoeken met de velden `title:` en `url:` werkt wel in de Yahoo! Directory zelf maar niet in de webzoekmachine, waar u met prefixen als `intitle:` en `inurl:` (zie bij de zoekmachine van Yahoo!, hoofdstuk 4.2.3) moet werken.

Daarnaast is er een Advanced Search voor alleen de directory, die meer opties biedt. Dit zijn de keuzes tussen:

- *Show results with all of these words, the exact phrase, any of these words* en *none of these words* wat de relatie tussen zoektermen betreft
- *Updated anytime, within the last 3 months, within the last 6 months* en *within a year* wat de actualiteit van de opname van sites in de directory betreft
- *All Categories* of een van de 14 hoofdcategorieën
- het 'family filter' aan of uit
- 10-100 hits per pagina wat de presentatie betreft

Wat in de Advanced Search van de directory ontbreekt is een optie om de zoekactie tot woorden in de titel of in de URL te beperken. Maar in plaats daarvan kunt u de prefixen `title:` en `url:` ook in de Advanced Search gebruiken.

Yahoo! zoekt niet 'case-sensitive', maakt dus geen onderscheid tussen hoofdletters en kleine letters, maar maakt wel een onderscheid tussen geaccentueerde en niet geaccentueerde letters: Bronte vindt Bronte en Brontë, maar Brontë vindt alleen Brontë.

Moet een goede directory ook groot zijn?

Directories dekken maar een zeer klein gedeelte van het Web. Daartegenover staat dat ze hun gebruikers de weg wijzen naar informatie over een onderwerp. Met een zoekmachine kunt u bij de eerste vraag feilloos in de roos schieten of hopeloos verdwaald raken. Een directory zet u bij wijze van spreken niet voor de deur af, maar bij een bushalte vanwaar u in staat wordt geacht uw weg zelf verder te kunnen vinden. Om dat te kunnen doen dient een directory over elk onderwerp een representatief overzicht te bieden, dat wil zeggen dat elk van de rubrieken een aantal links bevat dat het onderwerp van die rubriek zo economisch en zo volledig mogelijk dekt. (Denk aan recall en precision, blz. 75.) Dat kan net zo goed één link zijn als twintig. (Meer dan twintig wordt onoverzichtelijk, dan wordt het tijd een rubriek verder te gaan onderverdelen.)

Hoe kan één link in een rubriek voldoende zijn, en zelfs beter dan twintig? Wel, als u op zoek bent naar een museum in Zuid-Holland waar poppenhuizen te zien zijn, hebt u in

de rubriek 'Musea' meer aan één link naar een site waar je alle Nederlandse musea kunt vinden, dan aan twintig links naar sites die naar diverse musea verwijzen, maar niet naar het museum dat u zoekt. En die ene link is zelfs beter dan het twintigtal waarin die ene misschien op de twaalfde plaats verscholen zit.

Conclusie: een grote directory is niet per se beter dan een kleine. Representativiteit en kwaliteit van de links zijn belangrijker dan kwantiteit, mits de gebruiker van zo'n directory de vraagstelling maar daarop afstemt. Wat Yahoo! betreft, het is een relatief omvangrijke directory, maar zijn kwaliteit ontleent het aan zijn onderscheid tussen de Web directories die het apart vermeldt en de overige websites die het opneemt.

4.1.1.2. The Open Directory Project (ODP)

In kort bestek:

URL: <http://www.dmoz.org>

Omvang: ca. meer dan 5 miljoen.

Aantal rubrieken: meer dan 590.000.

Aantal rubrieken op de startpagina: 15 rubrieken, met 47 subrubrieken als voorbeelden; plus verwijzingen naar versies van de directory in andere talen (World), die overigens ook in de internationale directory zijn opgenomen.

Maximaal aantal niveaus: 8

Onderhouden door: ruim 71.000 vrijwilligers, van wie ca. 9000 actief

Zoekmachine intern: ja, met keuze tussen doorzoeken van de hele directory of de rubriek ('category'; inclusief onderliggende subcategorieën) waarin men vertoeft; sites worden in hun subcategorie (met de route daarheen) getoond

Zoekmachine extern: n/a

Advanced search: http://search.dmoz.org/cgi-bin/advanced_search.html

Beschrijvingen: meestal aanwezig; 1 à 4 regels lang.

Sponsored/featured links (commercie): nee

Typen resources: ja: databases, directories, conferences enz.; minder consequent dan Yahoo!

Nationale versies: Argentinië, België, Duitsland/Oostenrijk/Zwitserland, Frankrijk, Nederland, Polen, Zweden, Zwitserland

Help: Open Directory Project Help Central (<http://dmoz.org/help/helpmain.html>); Open Directory Search Guide (<http://dmoz.org/searchguide.html>)

Meer over ODP: FAQ General Info (<http://dmoz.org/help/geninfo.html>)

Open Directory Project – Wikipedia, the free encyclopedia

(http://en.wikipedia.org/wiki/Open_Directory_Project)

Open Directory Review on Search Engine Showdown

(<http://searchengineshowdown.com/dir/dmoz/review.html>)

Abondance: Open Directory (<http://www.abondance.com/outils/opendirectory.html>)

Webverzeichnis ODP - Open Directory Project (<http://www.at-web.de/odp/index.htm>)

Tussen de grote algemene directories neemt The Open Directory Project een heel aparte plaats in. Het zou wel eens de enige algemene directory van het Web kunnen blijken te zijn die levensvatbaar is.

The Open Directory Project ontstond begin 1998 onder de naam Gnuhoo, een combinatie van GNU en (Ya)hoo. GNU (Gnu's Not Unix) is een in 1984 door Richard Stallman gestichte beweging voor de vrije verspreiding van software; ook het besturingssysteem Linux, dat een steeds geduchtere concurrent wordt voor Microsofts Windows, staat in de traditie van die beweging.

Bedrijven als Yahoo! hebben een probleem: bij een groei van 3 miljoen webpagina's per dag, zoals die de afgelopen jaren heeft plaatsgevonden, raken 100 of 200 betaalde redacteuren allengs bedolven onder de lawine. 20.000 redacteuren zouden het werk

misschien nog wel aankunnen, maar dan zou Yahoo! onbetaalbaar worden. Commerciële directories als Yahoo! raken dus steeds verder achterop. In verband daarmee staan nog twee andere problemen: het onderhoud van de links en de verwerkingstijd van aangemelde pagina's. Weliswaar heeft Yahoo! daarvoor een oplossing door websites tegen betaling van de auteurs met voorrang te behandelen, maar die oplossing staat niet iedereen aan. Vandaar Gnuhoo.

Why I Must Write GNU

I consider that the golden rule requires that if I like a program I must share it with other people who like it. I cannot in good conscience sign a nondisclosure agreement or a software license agreement.

So that I can continue to use computers without violating my principles, I have decided to put together a sufficient body of free software so that I will be able to get along without any software that is not free.

Uit Richard Stallmans aankondiging van het GNU-project in de nieuwsgroepen net.unix-wizards en net.usoft, 1983

In de bijna acht jaar van zijn bestaan is The Open Directory Project een doorslaand succes gebleken. Op het moment van ODP's ontstaan telde Yahoo!, dat toen al bijna vier jaar bestond, ongeveer 750.000 sites; thans is ODP Yahoo! al ruim gepasseerd. De organisatie van de directory (heel praktisch gebaseerd op de onderwerpen die door Usenet newsgroups worden bestreken), de kwaliteit van de selectie en het gebruiksgemak van ODP doen nauwelijks onder voor die van Yahoo!, en er zijn punten waarop ODP duidelijk hoger scoort.

Op twee punten blijft ODP enigszins achter bij zijn commerciële concurrent. Het eerste is dat ODP alleen de eigen directory doorzoekt; er is hier dus geen combinatie met een zoekmachine, zoals bij Google en Yahoo! Wel biedt het resultaat scherm van ODP de mogelijkheid een vraag ook nog eens door verschillende zoekmachines te laten uitvoeren, maar die vraag wordt niet altijd correct vertaald. Wie het ODP mét een webzoekmachine wil hebben kan dus het beste Google's versie van het ODP gebruiken. Het tweede punt waarop ODP onderdoet voor Yahoo! is de consequentie van de naamgeving aan de categorieën. Net als Yahoo! kent ook ODP 'Web directories', maar deze staan ook bekend als 'Internet directories', 'Web directories and indexes', 'Directories and link collections', 'Internet directories and tools' enz. Evenzo komt naast 'Mailing lists' 'Discussion lists' voor.

Ten slotte is het bij een product van tienduizenden vrijwilligers onvermijdelijk dat er kwaliteitsverschillen tussen de verschillende categorieën zijn.

Daar staan voordelen van ODP tegenover. Zo groeit deze directory niet alleen door aanmelding, maar ook doordat de redacteurs zelf op zoek gaan naar goede sites.

Voorts bevat ODP niet alleen zie-verwijzingen (net als bij Yahoo! aangeduid met een @) maar ook zie-ookverwijzingen. Handig bij het zoeken is ook dat veel rubrieken een 'Description' en een FAQ hebben waarin beschreven staat wanneer een site in de betrokken categorie thuishoort en wanneer niet; hiervan kan ook bij het zoeken gebruik worden gemaakt. De redacteurs bij ODP zijn weliswaar vrijwilligers, maar bij iedere rubriek staan de aliases vermeld waaronder ze bekend zijn, met een link naar hun mail-en/of chatadres en een kort profiel.

Sommige sites in ODP hebben een ster. Deze zijn door de betreffende redacteur als opvallend goed aangemerkt.

Editor: kh7

Name: **Katinka Hesselink**
 Email: [Send to kh7](#)
 Home Page: www.katinkahesselink.net
[kh7's bookmarks](#) (239)

Categories

[Kids and Teens: International: Nederlands: Mens en Maatschappij: Religie en Spiritualiteit](#) (8)
[Reference: Quotations: Religion and Spirituality](#) (30)
[Regional: Europe: Netherlands](#) (1,801)
[Regional: Europe: Netherlands: Groningen](#) (26)
[Society: Religion and Spirituality](#) (102,748)
[Society: Religion and Spirituality: Enlightenment: Teachers: Krishnamurti, Jiddu](#) (31)
[Society: Religion and Spirituality: Esoteric and Occult: Personalities](#) (339)
[Society: Religion and Spirituality: Esoteric and Occult: Theosophy](#) (178)
[World: Netherlands](#) (61,813)
[World: Nederlands: Maatschappij: Religie en Spiritualiteit: Theosofie](#) (35)
[World: Nederlands: Regionaal: Nederland: Groningen](#) (894)

Profile

I am a Dutch math and chemistry teacher, very interested in Krishnamurti, theosophy and spirituality in general.

I edit in Dutch categories, like my hometown, and in English categories about spirituality.

You can see the result of my spiritual studies at least partly on my homepage (above).

(I am a volunteer editor on the Open Directory Project (ODP). If you see this text on any other search engine or directory, it simply means that they use ODP data, not that I am affiliated with that search engine or directory.)

Het Open Directory Project werkt met tienduizenden vrijwilligers, die anders dan die van Yahoo! niet anoniem zijn. Onderaan veel rubriekspagina's staat de naam of een alias van degene die de rubriek onderhoudt. Klikken op dat alias levert een pagina informatie op, met onder meer een mailadres, de door de redacteur onderhouden rubrieken en een korte persoonsbeschrijving.

Net als Yahoo! biedt ODP op elk niveau (behalve het hoogste) een aantal sites. Een verademing is dat de ODP-schermen niet door advertenties worden ontsierd. De zoekmachine van ODP ondersteunt Booleaans zoeken met de operatoren AND, OR en ANDNOT, unitair zoeken met + en – en rechts trunceren met *. Nesting is niet mogelijk, en EN- en OF-relaties in dezelfde vraag zijn taboe: een vraag als `elections Bush OR Clinton`, zoals die bij Google wel kan worden gesteld, leidt in het ODP niet tot het gewenste resultaat. Zoeken met phrases is mogelijk, maar er zijn geen nabijheidsoperatoren beschikbaar.

Net als bij de directory van Yahoo! kan ook in het ODP met prefixen worden gezocht.

t: en d:, direct voor de zoekterm, geven aan dat deze in de titel of in de beschrijving van een site moet staan, met c: als prefix kunt u aangeven dat de zoekterm of phrase in een categorienaam moet voorkomen. Verder kunnen vragen door het gebruik van u: tot een domain ([bush u:whitehouse.gov](http://bush.whitehouse.gov)) worden beperkt. (Andersom kunnen met -: natuurlijk ook bepaalde domeinen van de vraag worden uitgesloten.)

De zoekmachine maakt geen onderscheid tussen hoofdletters en kleine letters, maar wel tussen wel en niet geaccentueerde letters: egalite vindt niet égalité, en andersom.

Behalve de Basic Search is er ook een Advanced Search, die u de zoekactie tot enkel categoriën dan wel enkel sites laat beperken, een hoofdcategorie laat opgeven en een leeftijdscategorie (kids, teens, mature teens) laat instellen.

De Franse zoekmachine Exalead toont bij zijn resultatenlijst de categorieën van de ODP-classificatie ("Related categories") voor de treffers die hij op de opgegeven zoektermen heeft gevonden.

Omdat ODP Open Content is, zijn er tal van bedrijven en personen (meer dan 200) die een (eventueel gewijzigde) versie van ODP op hun eigen site publiceren. Daarbij zitten bekende zoekmachines zoals Google, die zo hun zoekfaciliteiten aanvullen met een directory, maar ook minder bekende zoekmachines die de omvangrijke en vrij beschikbare ODP-database gebruiken om er hun speciale zoekalgoritmen mee te testen of te demonstreren. Een daarvan is de Franse zoekmachine Exalead, dat de ODP-classificatie (of met een deftiger woord: 'ontologie') gebruikt om bij zijn resultatenlijsten 'Related Categories' te tonen.

4.1.2. Directories Plus

Directories hebben op zoekmachines voor dat je zonder al te veel zorgen over hoe je een vraag precies moet inkleden naar een overzienbare selectie van websites wordt geleid. Maar zowel bij Yahoo! als bij het Open Directory Project is van een geregeld contact tussen gebruikers en redacteurs geen sprake.

In dat manco voorzien enkele diensten die ik bij gebrek aan een betere term maar als Directories Plus aanduidt: als directories – hier 'channels' of 'communities' geheten – opgebouwde verzamelingen resources die per onderwerp worden samengesteld, besproken en onderhouden door een redacteur met affiniteit voor het onderwerp, over wiens kwalificaties niet geheimzinnig wordt gedaan en die een directer contact met de gebruikers van de dienst onderhoudt dan de redacteurs van de hierboven besproken directories. Twee van deze diensten zijn About.com, dat zich The Human Internet noemt, en Suite101.com, met als ondertitel 'Real People Helping Real People'.

4.1.2.1. About.com - The Human Internet

In kort bestek:

URL: <http://about.com>

Omvang: naar schatting ca. 1 miljoen artikelen van de gidsen, waarvan de meeste met een aantal externe links

Aantal rubrieken: 475 guides; de lijst is beschikbaar onder de link About A-Z; >50.000 onderwerpen

Aantal rubrieken op de startpagina: 23 rubrieken (channels)

Maximaal aantal niveaus: 5

Onderhouden door: 475 gidsen uit meer dan 20 landen

Zoekmachine intern: ja; daarnaast is er een alfabetische site index

Zoekmachine extern: nee

Sponsored/featured links (commercie): ja

Beschrijvingen: altijd aanwezig; 1-2 regels lang

Typen resources: nee

Nationale versies: nee

Hoe langer je met een dienst als About.com werkt, hoe meer je het idee achter dit soort diensten gaat waarderen. Zeker, met de ruim een miljoen sites die About.com dekt, wordt bij lange na niet het hele Web bestreken. Maar de informatie die je over een onderwerp vindt is praktisch, veelzijdig en stelt je in staat je in vrij korte tijd over diverse aspecten van een onderwerp te informeren.

De guides (zo heten zowel de redacteurs als de onderwerpsguides, die hier bedoeld zijn) hebben doorgaans een algemeen en praktisch karakter, zoals Attention Deficit Disorder, Special Education, Mental Health Resources, Parenting of K-6 Children. Dit zijn vier van de guides die je vindt als je binnen About.com naar een term als ADHD (aandachtstekortstoornis en hyperactiviteit) zoekt. Het verschijnsel behoort tot ieder van de terreinen die door de verschillende guides bestreken worden en wordt zo vanuit diverse gezichtspunten, door verschillende redacteurs, belicht. Dit is een aantrekkelijke wijze van informatievoorziening, omdat het de gebruiker enerzijds in staat stelt de gebieden (en de bijbehorende guides) die hem niet interesseren links te laten liggen en anderzijds ook kennis te nemen van terreinen die aan zijn belangstellingssfeer raken: in het geval van ADHD bij voorbeeld Special Education. Er is echter ook een keerzijde: in sommige gevallen maakt het voorkomen van één onderwerp in verschillende gidsen een ongecoördineerde indruk. Zo komt Shakespeare vier keer voor: onder Shakespeare, Quotations, Literature: Classic en German Language (Duitse vertalingen van Shakespeare), ieder met hun eigen redacteur en links naar resources.

De gidsen zijn volgens een vast patroon ingericht, in vier kolommen met elk een aantal rubrieken, bestaande uit folders. Daarvan bevat de eerste kolom de rubrieken Essentials, Offers, Topics en een Buyer's Guide, plus links naar forums, latest articles, recent updates en een algemene helptekst. Helemaal onderin vindt u nog de mogelijkheid om u op een gratis nieuwsbrief te abonneren en een link naar gratis online cursussen. Het

interessantst in deze kolom zijn de Essentials, die bijvoorbeeld een FAQ of tips van de dag kunnen omvatten, de Topics, bestaande uit folders met artikelen en verwijzingen, en de forums, de rest is voor een groot deel commercie.

In de tweede, centrale kolom, verschijnt de inhoud van de folders, bestaande uit de titel van bijvoorbeeld een artikel, met een korte beschrijving. Hier vindt u dus echte inhoud van de gidsen, naast Sponsored Links. De homepage van de gids bevat naast links naar de meest recente artikelen ook een link naar het profiel van de auteur van de gids, plus een link naar een nieuwsbrief, waarop men zich kan abonneren.

De derde en vierde kolom zijn gereserveerd voor meer reclame, en voor een aantal rubrieken waarin op verschillende manieren (Most Popular, What's Hot, Headlines, Related Topics) nog eens op de inhoud van de gids wordt geattendeerd. In de derde kolom kunnen ook More Categories verschijnen bij de inhoud van een folder die men in de linker kolom onder Articles & Resources heeft aangeklikt.

De About-gidsen kunnen je snel een beeld geven van wat er met betrekking tot een bepaald onderwerp speelt, want de informatie wordt wekelijks ververs en aangevuld.

De guides hebben geen wetenschappelijke pretentie en sommige ervan behandelen maar een beperkt aantal thema's, maar de geïnteresseerde leek zal er heel veel van zijn gading in vinden. De redacteurs zijn enthousiaste amateurs of mensen die op het vakgebied goed thuis zijn, bij voorbeeld leraren; sommigen van hen publiceren ook elders over hun onderwerp(en) en hebben door hun werk voor About.com de status van expert gekregen. Het idee van de 'community' wordt sterk benadrukt.

Bij gebruik van de zoekmachine doorzoekt About.com eerst het Topic waar u zich bevindt en als het daar niets vindt, ook de hele site. Gepresenteerd worden de artikelen in de Guide, plus, alweer, een aantal Sponsored Links.

Aanzienlijk kleiner en in een aantal opzichten beperkter is Suite101.com (<http://www.suite101.com>), dat wel ongeveer dezelfde opzet heeft als About.com. Ook deze 'directory plus' wordt samengesteld en onderhouden door vrijwilligers-redacteurs, ongeveer van hetzelfde niveau als die van About.com. Suite101 heeft 22 'community channels', ingedeeld in 562 'topics', waaraan de redacteurs bijna 90.000 artikelen hebben gewijd. Een deel van de topics is afgesloten ('retired'). Een bijzonderheid van Suite101 is dat bij een groot aantal topics een betaalde cursus wordt aangeboden. Deze worden verzorgd door de redacteurs van de topics en zijn tezamen ondergebracht op een website die zich als 'Suite University' presenteert.

Een belangrijk gedeelte van de inhoud van Suite101 wordt uitgemaakt door de door de redacteurs over hun onderwerp geschreven artikelen; deze verschijnen meestal eens in de 2 à 3 weken. Net als bij About.com leveren die een redelijk volledig beeld op van de verschillende facetten die een onderwerp heeft, en stellen ze gebruikers in staat de actualiteit te volgen. 'Gewone' directories doen dat niet, of niet zo efficiënt.

Elke gids bevat een welkomspagina, met daarop een overzicht van de artikelen over het onderwerp en een menu van opties dat op elke volgende pagina terugkeert: Welcome, Recent Articles, Recent Discussions. De discussies zijn meestal naar aanleiding van de artikelen, maar er is ook gelegenheid om vragen te stellen. Ook wat de deelname aan de Discussions aangaat maakt Suite101 een 'stillere' indruk dan About.com.

Zoeken in Suite101 levert in eerste instantie topics en articles op, maar desgewenst kan vervolgens ook (of uitsluitend) op 'discussion threads' en persoonlijke websites van de redacteurs worden gezocht. De treffers worden op 'relevantie' gesorteerd, maar sorteren op datum is ook mogelijk. Suite101 heeft één ding dat About.com niet heeft: er is een rating systeem waarmee lezers artikelen kunnen beoordelen.

Rechter bladzijde: Deel van de eerste drie kolommen van de homepage van een Guide over hoofdpijn in About.com.



Search

- Home
- Essentials
 - Forum, Chat, & Site Index Links
 - Headache & Migraine 101
 - Ask the Clinician
 - Is it Migraine? Tension? What?
 - Essential Information Pages
- Headaches Offers
- Articles & Resources
- Buyer's Guide
 - Before You Buy Top Picks
 - Headache and Migraine Disease Books, Part II
 - Medication Organizers and Carrying Cases
 - Super Calendars for 2005
 - Product Reviews
- Articles
- Forums
- Help

EarthLink
Dial-Up, High Speed & Wireless Internet access for all your needs!
[Learn More](#)
High Speed

VONAGE
THE BROADBAND PHONE COMPANY
PLANS STARTING AT JUST
\$14.99 / MO.
[START SAVING >>](#)

FREE Newsletter
Sign Up Now for the Headaches / Migraine newsletter!
Enter email address

[See Online Courses](#)

Headaches / Migraine



From [Teri Robert](#),
Your Guide to [Headaches / Migraine](#).
FREE Newsletter. [Sign Up Now!](#)

Migraineurs Bill of Rights



Many interesting ideas for what should be on our site come from email from our readers. Yesterday, one of our forum members wrote to me about a "Bill of Rights" for bipolar disorder patients that she had found on another web site. She wondered if I'd be interested in pursuing including such a document here. Such a document already exists for Migraine and headache sufferers... [Read more...](#)
[Tuesday May 17, 2005](#) #

Ask the Clinician: Expert Answers



This week, Dr. Krusz and I have answered another batch of new questions to add to the hundreds already in our [Ask the Clinician Archives](#). As I prepare the column, I can never help noticing that the questions could apply to many of us. This week's questions include:
[Read more...](#)
[Monday May 16, 2005](#) #

More from your Guide below

Sponsored Links

Migraines

Updates on migraine information
www.joshgreene.com/migraine

Beating Headaches

Overcome & treat headaches now Fast, easy, effective, 100% natural
www.beating-headaches-migraines.com

Online Headache Diary

For those with chronic headaches research your headaches on here.
www.myheadachediary.com

More from your Guide

Choosing Our Battles

Choosing our battles is important to everyone. For someone with a chronic illness, including chronic headaches or Migraine disease, it can be especially important because we often have less strength for any battles. Mental Health Resources Guide Leonard Holmes, Ph.D. offers an insightful article on this issue... [Read more...](#)
[Sunday May 15, 2005](#) #

A Blooming Onion to Larry King Live!



This week, I have a problem. I've thought long and hard about the Larry King Live show from April 26 that dealt with Migraine Disease, read and reread the transcript of the show, and read all of your feedback. While I'd like to give them a Rose for doing a show on the subject, I'd also like to give them an Onion for some of the content of and fallout from the show. Thus, I've combined the two and present Larry King Live with a Blooming Onion! Some of you have asked what I thought of the show, and I've asked you to be patient. There's been a [discussion going on in our forum](#), and I didn't want my opinion to influence it in any way... [Read more...](#)
[Friday May 13, 2005](#) #

Around About!



If you're like me, it's easy to get into a habit of regularly visiting some web sites, but not exploring many new ones. Did you know that About has **more than 70** topic sites just on health topics? Surprised? Once a week, I look around and try to pick out a few articles I think might interest you.... [Read more...](#)
[Thursday May 12, 2005](#) #

Advertisement

Learn more about . . .

Migraine prevention

CLICK HERE

FREE from our Partners PROGRAM

- » [Manage diabetes more easily.](#)
- » [Learn about a treatment for nail fungus infection](#)
- » [Permanent solution for hair loss. Get your FREE DVD!](#)
- » [Earn your Healthcare Degree 100% Online](#)

Most Popular

- [Noni Juice](#)
- [Anatomy of a Migraine](#)
- [Migraine, Tension, What?](#)
- [Head Pain 101: Diagnosis, Education, Support](#)
- [Sneaky Exercise](#)

Partner Content

- [Chronic Pain Management](#)
- [Back and Neck Pain](#)
- [Diabetes](#)
- [Alzheimer's Disease](#)
- [Erectile Dysfunction](#)

4.1.2.2. Communities

About.com en Suite101.com zijn twee vroege voorbeelden van een concept van informatievoorziening op het Web waarin allerlei elementen bijeen zijn gebracht die elders in dit boek apart behandeld worden. In dat concept staan mensen centraal, niet alleen de auteurs van de gidsen als informatieverstrekkers en de lezers als informatieafnemers, maar ook de andere lezers als vraagstellers en vraagbeantwoorders. Wat het zoeken naar informatie betreft komen zoekmachine, directory, webgids, expert, nieuwsbrief, online cursus en forum hier bij elkaar, en dank zij de activiteiten van de auteurs van de Guides wordt ook in de behoefte om bij te blijven (zie blz. 273 e.v.) voorzien. In deze constellatie zijn vraag en antwoord ingebed in een 'community' die zich gezamenlijk en wederzijds informeert en geïnformeerd probeert te blijven. Artikelen van speciaal About.com-gidsen duiken regelmatig in de zoekresultaten van Google en andere zoekmachines op, wat erop wijst dat ze links trekken en dus waarschijnlijk door velen hoog worden aangeslagen. Maar wie met een zoekmachine een geïsoleerd artikel heeft gevonden en ermee volstaat dat te bekijken, doet zichzelf beslist tekort. Vandaar dat ik hierboven tamelijk wat aandacht aan deze 'directories plus' heb besteed.

Het idee van communities is ook elders verbreid en nu ik toch over dit onderwerp te spreken ben gekomen noem ik in de marge nog een paar voor de zoeker interessante voorbeelden van webresources waarbij de 'community'/'collaboration'-gedachte een rol speelt.

Een her en der op het Web verspreid verschijnsel is dat van de zogeheten wiki's, websites die door de samenwerking van gebruikers tot stand worden gebracht. Het bekendste voorbeeld is Wikipedia, een online encyclopedie die door de gezamenlijke inspanningen van de – in principe – hele webgemeenschap wordt bijgehouden. Voor een overzicht van soortgelijke projecten, zie Complete list of Wikimedia Projects, http://meta.wikimedia.org/wiki/Complete_list_of_Wikimedia_projects.

Samenwerking in communities wordt natuurlijk sterk bevorderd door het gezamenlijk gebruik van online software die het mogelijk maakt samen aan documenten te schrijven, al dan niet in wiki's. Een voorbeeld daarvan is Writeboard.

Een voorbeeld onder de zoekmachines is Eurekster (<http://eurekster.com/>), dat de mogelijkheid biedt gespecialiseerde, zelflerende zoekmachines op te zetten (op basis van de URL-database van Yahoo!). Zoekresultaten zijn daarbij het resultaat van drie factoren: een door de vervaardiger van de zoekmachine opgegeven combinatie van termen en sites die voor het terrein dat de zoekmachine moet bestrijken relevant zijn; het aanklikgedrag van de gebruikers van de zoekmachine; de swicki-techniek die dit gedrag vertaalt in de ranking van de zoekresultaten. Voor een voorbeeld zie de Bibliotheca 2.0 swicki, voor meer uitleg zie de swicki FAQs.

Bijdragen en profiteren tegelijk kan ook op een site waar gezamenlijk bookmarks worden bewaard. del.icio.us (<http://del.icio.us/>) is een zogeheten 'social bookmark manager', die een stap verder gaat dan 'gewone' bookmark managers zoals Backflip. Die laten je je bookmarks (via een bookmarklet bij je browser, zie blz. 264 e.v.) op hun site opslaan (zodat je ze op elke computer tot je beschikking hebt) en ze meestal ook allemaal of voor een deel met anderen delen (dat wil zeggen dat je bookmarkbestand op die site voor iedereen toegankelijk is). In dat concept is het bookmarkbestand wel openbaar, maar heeft maar één auteur. Bij del.icio.us voorziet u, als u een pagina daar via de bookmarklet bookmarkt, deze van een of meer 'tags', trefwoorden waarmee u de bookmark later hoopt terug te vinden. Ideaal is als diverse gebruikers van del.icio.us aan dezelfde bookmark dezelfde tag of tags uitdelen, want via de tags kunt u vervolgens niet alleen de pagina's opsporen die u zelf gebookmarkt hebt, maar ook anderen die dat ook hebben gedaan. U vindt dan andere webpagina's met dezelfde tag, en de bookmarks van andere personen die kennelijk hetzelfde soort pagina's bookmarken als u doet. Een

middel om bij te blijven is del.icio.us ook, en dat zelfs zonder dat je er regelmatig op bezoek hoeft, want op alle pagina's van del.icio.us is een RSS-feed (zie blz. 284 e.v.) beschikbaar.

In paragraaf 2.7.2 zijn de 'experts', die vragen beantwoorden, al even ter sprake gekomen. Een type diensten dat ik daar terloops genoemd heb, zijn die waar de rollen van vraagstellers en vraagbeantwoorders niet zo duidelijk vastleggen als bij 'Ask-an-Expert'-diensten, maar waar alle bezoekers, geregistreerd en soms ook ongeregistreerd, zowel vragen kunnen stellen als beantwoorden. Ook deze reken ik tot de 'communities'. Voorbeelden hiervan zijn Wondir (<http://www.wondir.com>) en het Duitse wer-weiss-was – Experten-Netzwerk zum Austausch von Wissen auf Gegenseitigkeit (<http://www.wer-weiss-was.de/content/start.shtml>). Verwacht er niet te veel van.

Een laatste voorbeeld van samenwerking tussen personen en in communities zijn Group en Community Blogs. Voor de zoeker zijn die niet interessanter dan blogs sec, en ik ga er hier dan ook niet verder op in. Voor blogs in het algemeen, zie het hoofdstuk Bijblijven, blz. 273 e.v.

4.1.3. *Als directories georganiseerde verzamelingen van webgidsen*

Wie op onderwerp in een directory zoekt, hetzij door te zoeken hetzij door te browsen, komt uit op een aantal webpagina's of websites die door een of meer redacteurs geselecteerd zijn. Iets anders dan het onderwerp en het feit dat ze deel uitmaken van de selectie hebben deze sites niet gemeen. Voor een deel bestrijken ze hetzelfde terrein, voor een deel betreffen ze heel uiteenlopende typen resources of hebben ze een heel verschillende 'scope'. Eenmaal in de rubriek aangeland waarin hij de resources vindt, moet de gebruiker zelf maar uitzoeken wat wat is en waar hij het antwoord op zijn vraag het best kan vinden.

Directories plus doen al iets meer. Doordat de onderwerpen bij een deskundige redacteur zijn ondergebracht die niet alleen aangemelde sites beoordeelt, maar ook actief resources verzamelt (en daarbij zowel selectief als uitputtend zal proberen te zijn) organiseren ze de informatie beter en plaatsen die door de artikelen die de redacteurs ook over het onderwerp schrijven in een verband.

Nog een stap verder gaan verzamelingen van webgidsen (en op een meer specifiek niveau natuurlijk ook webgidsen zelf). Deze verzamelingen bieden een min of meer compleet overzicht van de informatie die op het Web te vinden is, georganiseerd in gespecialiseerde gidsen over vakgebieden of onderwerpen. De gidsen zelf beschrijven de resources over die vakgebieden of onderwerpen. Ze plaatsen de resources in een context en bieden samenhang en onderscheid. Zowel voor wie zich breed wil oriënteren, als voor degene die snel wil weten in welke omgeving hij een goede kans heeft een antwoord op zijn vraag te vinden, vormen verzamelingen van webgidsen een goed startpunt.

4.1.3.1. Usenet FAQ's by Category

URL: <http://www.faqs.org/faqs/by-category.html>

In hoofdstuk 2.6 is in verband met nieuwsgroepen en rondzendlijsten al even sprake geweest van FAQs. FAQs (Frequently Asked Questions) zijn een verschijnsel dat onlosmakelijk met het communicatie-aspect van internet verbonden is. Het zijn documenten die worden gemaakt om gebruikers, speciaal nieuwe gebruikers, van een nieuwsgroep of rondzendlijst een eerste introductie tot het onderwerp en de groep of lijst te geven. Daartoe behoren praktische aanwijzingen over hoe berichten te 'posten' en netiquette ('Hoe hoort het eigenlijk op internet?'), maar FAQs kunnen ook uitgebreide overzichten van resources (en niet alleen webresources) over een onderwerp geven. Ze kunnen verouderd zijn, maar niettemin heeft het bij veel vragen zin eerst eens te kijken of er een FAQ bestaat en die te raadplegen. De meeste geven een eenvoudige, maar toch

gedegen inleiding op een onderwerp. Aan het eind van uw zoekactie komt u er trouwens misschien weer bij terug: om een vraag te stellen in de groep of lijst over het onderwerp waarover u iets wilt weten.

'Usenet FAQ's by category' wordt gepubliceerd door het Internet FAQ Consortium, dat ook andere mogelijkheden dan 'by category' biedt om FAQs te vinden: op auteur, nieuwsgroep of archiefnaam. Zie <http://www.faqs.org/>. Het gaat hierbij alleen om FAQs van nieuwsgroepen. FAQs van rondzendlijsten zal men via de overzichten van rondzendlijsten moeten opsporen. Zie hiervoor hoofdstuk 2.6.

Meer over FAQs: Infinite Ink's

Finding Periodic Postings ('FAQs' & 'PIPs') (<http://www.ii.com/internet/faqs/>)

4.1.3.2. The Argus Clearinghouse en de Internet Public Library

URLs: <http://www.clearinghouse.net/> en <http://www.ipl.org/>

Voor wie in goede onderwerpsgidsen op het Web geïnteresseerd is, vormt de verzameling die de redactie van het Argus Clearinghouse tot begin 2002 heeft aangelegd, nog altijd een goed startpunt. Hoewel de verzameling is afgesloten en een aantal van de gidsen niet meer wordt bijgehouden of verdwenen is, biedt het Clearinghouse nog altijd wel een handige toegang tot onderwerpsgidsen van goede kwaliteit die de tand des tijds hebben doorstaan en nog altijd regelmatig geactualiseerd worden.

The Argus Clearinghouse is een verzameling van zelfstandige webgidsen die door de auteurs bij Argus Associates zijn aangemeld, voor opname zijn beoordeeld en in de verzameling zijn opgenomen. De onderneming heeft als achtergrond de School for Library and Information Science van de University of Michigan, waaruit een groot deel van de redactie en veel van de auteurs van de webgidsen voortkomen. Niettemin kon iedere auteur van zo'n gids zijn werk aanmelden; slechts 5% à 10% werd echter geaccepteerd.

Het Clearinghouse bevat ca. 1500 Webgidsen, te vinden in een directory van 13 rubrieken, bijna 100 subrubrieken en ca. 1000 keywords. Elke webgids heeft 1 à 3 keywords, per keyword kunnen er meerdere webgidsen zijn.

Zoals gezegd werd maar een klein gedeelte van de aangemelde webgidsen geaccepteerd. Voor die gidsen hanteerde Argus een zeer gedegen (en gelukkig ook beschreven) beoordelingssysteem, dat naar de beschrijving en evaluatie van de in de gids opgenomen resources keek en ook het ontwerp, de organisatie en de herkomst van de gids in de beoordeling betrok. Bovendien zag Argus er tot 2002 door middel van een jaarlijkse controle van de gidsen op toe dat deze ook onderhouden werden. Dit alles resulteerde in een waardering met 1 tot 5 sterren voor de verschillende criteria en als algemeen eindoordeel. (Zie: <http://www.clearinghouse.net/ratings.html>)

Voor zover ze nog bestaan en worden bijgehouden, zijn de tot het Clearinghouse toegelaten gidsen vaak van uitstekende kwaliteit en bieden een gedetailleerde toegang tot grote aantallen resources binnen het onderwerp dat de webgids bestrijkt.

Voor wie zich grondig en breed wil informeren, daarbij gebruik makend van wat anderen al op het Web hebben verzameld en ingedeeld, vormt The Argus Clearinghouse een goede portal. Hetzelfde geldt voor de Internet Public Library (IPL;

<http://www.ipl.org>), een eveneens vanuit de University of Michigan opgezette openbare bibliotheek van het Web. In de Subject Collections van deze site vindt u, net als in het Clearinghouse, geannoteerde links naar een per vak ingedeelde verzameling omvangrijke resources, in veel gevallen webgidsen, maar daarnaast heeft deze bibliotheek nog meer te bieden. Evenals in een 'echte' openbare bibliotheek vindt u er (geannoteerde links naar) een goede verzameling naslagwerken, zoals adresboeken, woordenboeken, encyclopedieën, citatenwoordenboeken enz., en in de 'Reading Room' treft u (links naar verzamelingen van) boeken, tijdschriften en kranten aan. De 'Special

Collections' bevatten varia zoals een uitgebreide verzameling links naar literaire kritieken op de werken van Amerikaanse maar ook andere auteurs, een geheel aan blogs gewijde site, biografieën en bibliografieën van indiaanse auteurs, en POTUS, een geheel aan de Amerikaanse presidenten gewijde site in de beste traditie van de openbare bibliotheek. En natuurlijk heeft de Internet Public Library ook een sectie 'Searching Tools', met onder meer een vijftigtal Frequently Asked Reference Questions (met veel verwijzingen naar resources op het onzichtbare Web) en meer dan honderd 'Pathfinder Research Guides', korte handleidingen over het verrichten van, bijvoorbeeld, genealogisch en statistisch onderzoek. Zoals het een goede bibliotheek betaamt is er ook een inlichtingenbalie die vragen beantwoordt. Meer daarover in paragraaf 2.7.2.

4.1.3.3. The WWW Virtual Library en het Resource Discovery Network

URLs: <http://vlib.org/> en <http://www.rdn.ac.uk>

De WWW Virtual Library (VL) is de oudste verzameling van webgidsen; het initiatief ertoe is afkomstig van de 'vader' van HTML en het Web zelf, Tim Berners-Lee, destijds verbonden aan het Europese laboratorium voor Deeltjesfysica CERN. De VL is een doorvrijwilligers samengestelde collectie van meer dan 300 Webgidsen, verdeeld over 16 hoofdrubrieken. Doel van de Virtual Library is de ontsluiting van het serieuze, academische, niet-commerciële Web.

De meeste gidsen zijn zeer uitgebreid en worden goed bijgehouden. (Ettelijke ervan zitten ook in de collectie van het Argus Clearinghouse.) Niettemin zijn er ook gidsen die enigszins verouderd zijn.

Als het (niet al te gedetailleerde) onderwerp waarover u informatie wenst binnen het bestek van een gids valt, is de VL een goed, misschien wel het beste, startpunt voor uw zoekactie. Webgidsen zoals die van de WWW Virtual Library zijn bovendien uitstekende hulpmiddelen om het onzichtbare Web op te sporen.

Een andere verzameling gidsen van het wetenschappelijke Web is het Britse Resource Discovery Network. Hierin zijn acht grote 'gateways' (in de hier gebruikte zin is een gateway 'een via WWW-technologie opgebouwde database van geselecteerde Internetbronnen (bijvoorbeeld voor onderwijsdoeleinden)') opgenomen voor de volgende vakgebieden:

- Arts & Creative Industries
- Engineering, Mathematics & Computing
- Geography & Environment
- Health, Medicine & Life Sciences
- Hospitality, Leisure, Sport & Tourism
- Humanities
- Physical Sciences
- Social Science, Business & Law.

Bij elk van deze 'gateways' gaat het om een doorzoekbare vakspecifieke directory van door vakmensen geselecteerde internetbronnen.

4.1.4. *De kleinere subject directories*

Als we iets over een onderwerp willen weten zonder er nu direct een hele studie van te willen maken, kijken we in een encyclopedie of raadplegen een naslagwerk in de Openbare Bibliotheek.

Op het Web zijn voor zulke situaties web-encyclopedieën en door bibliotheken of bibliothecarissen onderhouden subject directories handig. Deze zijn kleiner dan de grote



The WWW Virtual Library

[Agriculture](#)

Irrigation, Livestock, Poultry Science, ...

[The Arts](#)

Art History, Classical Music, Theatre and Drama, ...

[Business and Economics](#)

Finance, Marketing, [Transportation](#), ...

[Communications and Media](#)

Broadcasters, Publishers, Telecommunications, ...

[Computing and Computer Science](#)

Artificial Intelligence, Cryptography, Logic Programming, ...

[Education](#)

Primary, Secondary, Tertiary, ...

[Engineering](#)

[Architecture](#), Electrical, Mechanical, ...

[Humanities and Humanistic Studies](#)

[History](#), [Languages and Linguistics](#), [Museums](#), ...

[Information and Libraries](#)

Information Quality, Knowledge Management, Libraries, ...

[International Affairs](#)

International Relations and Security, Sustainable Development, ...

[Law](#)

Arbitration, Forensic Toxicology, Legal History, ...

[Natural Sciences and Mathematics](#)

[Biosciences](#), [Earth Science](#), [Medicine and Health](#), [Physics](#), ...

[Recreation](#)

Gardening, Recreation and Games, Sport, ...

[Regional Studies](#)

African, [Asian](#), Latin American, European, ...

[Social and Behavioral Sciences](#)

Anthropology, [Archaeology](#), [Population and Development Studies](#), ...

[Society](#)

[Peoples](#), [Religion](#), Gender Studies, ...

Mirrors: [vlib.org](#) (CH), [East Anglia](#) (UK).

[About](#) | [Contact](#) | [Donors](#)

De WWW Virtual Library, de oudste en meest serieuze wetenschappelijke directory van het Web.

subject directories als Yahoo!, maar hebben het voordeel dat ze over elk onderwerp wel een of twee artikelen (in een encyclopedie) of verwijzingen naar websites bieden waarinde belangrijkste informatie over dat onderwerp te vinden is. Over een onderwerp als 'celebrities' (Julia Roberts, Leonardo DiCaprio) zult u er niets in vinden, maar tot de meeste wat serieuze onderwerpen bieden ze een goede introductie.

4.1.4.1. Librarians' Internet Index

In kort bestek

URL: <http://lii.org>

Omvang: links naar >16.000 websites, voorzien van een annotatie en een of meer trefwoorden

Aantal rubrieken: ca. 2000; daarnaast meer dan 18.000 trefwoorden

Aantal rubrieken op de startpagina: 15, met 52 voorbeelden van subrubrieken eronder

Maximaal aantal niveaus: 5

Onderhouden door: ca. 100 bibliothecarissen in de Amerikaanse staten Californië en Washington

Zoekmachine intern: ja

Zoekmachine extern: nee

Advanced Search: ja

Beschrijvingen: altijd aanwezig; 3-6 regels lang.

Typen resources: ja, bijvoorbeeld directories, database, statistics, dictionary

Nationale versies: nee

Help: <http://lii.org/search/file/help>; <http://lii.org/advanced/file/help>

Meer over Librarians' Internet Index: Péter's Digital Reference Shelf (February 2001):
Librarians' Index to the Internet

<http://www.galegroup.com/servlet/HTMLFileServlet?imprint=9999®ion=7&fileName=reference/archive/200103/lii.html>

Op de Librarians' Internet Index is volop van toepassing wat ik hierboven schreef over de betrekkelijke waarde van de omvang van een directory (zie blz. 80-1). De Index dekt maar 1/250e deel van het aantal sites in het Open Directory Project, maar de kwaliteit van de geselecteerde sites wordt gegarandeerd door een zorgvuldige selectieprocedure met openbare selectiecriteria (zie: <http://lii.org/search/file/pubcriteria>). Doordat een van de toegepaste criteria de inhoudelijke 'coverage' (dekking) van sites is, is het werkelijke bereik van de directory veel ruimer dan het getal van 16.000 sites zou doen vermoeden. Daarbij komt dat de sites voortreffelijk beschreven zijn.

Een bezwaar tegen deze directory is wel dat ze voor een aantal onderwerpen erg Amerikaans georiënteerd is: in de rubriek "statistics: labor" vind je 14 treffers die allemaal betrekking hebben op de VS; de statistieken van de ILO en de OESO zul je in deze directory vergeefs zoeken, al is de ILO als zodanig er wel in te vinden. Anderzijds is in de rubriek 'statistics: religion' wel weer het internationale Adherents opgenomen. Sites in andere talen dan het Engels worden niet opgenomen, en hoewel sites van bedrijven en personen niet uitgesloten worden, heeft de Index een sterke voorkeur voor institutionele en overheidssites.

De zoekmachine

- verbindt meer zoektermen impliciet met AND
- accepteert de Booleaanse operatoren AND, OR, NOT
- werkt met nesting (haakjes) – reparatie AND (fiets OR rijwiel)
- gebruikt dubbele aanhalingstekens voor phrases ("...")
- maakt geen onderscheid tussen hoofd- en kleine letters
- past automatic word stemming (record ↔ records ↔ recordings) toe; in de Advanced Search kan dit worden uitgeschakeld
- staat rechts trunceren toe, met de asterisk (*) als truncatieteken
- zoekt in de Basic Search in een combinatie van de velden Description/Title/Subject (met het laatste zijn de toegekende trefwoorden bedoeld)
- biedt in de Advanced Search de mogelijkheid de zoekactie tot Description of Title of Subject te beperken; daarnaast de mogelijkheid extra velden te doorzoeken, namelijk: naam van de auteur van de opgenomen site, de voor de site verantwoordelijke instelling of organisatie, de URL van de site (domain, pad en bestandsnaam) en de initialen van de bibliothecaris die de site heeft opgenomen

Als u de zoekmachine gebruikt om sites te vinden, treft u bij ieder record de trefwoorden aan die aan een site zijn toegekend. Door daarop te klikken vindt u andere sites met hetzelfde trefwoord.

De Index heeft een gratis nieuwsbrief waarin nieuw opgenomen sites worden gemeld, en ondersteunt RSS-feeds (zie blz. 284 e.v.).

4.1.5. Nederlandse directories en startpagina's

Wie op zoek gaat naar informatie die beperkt blijft of moet blijven tot een bepaald geografisch of taalgebied, kan in veel gevallen beter gebruik maken van een nationaal zoekinstrument dan van een van de mondiale zoekinstrumenten. Dat geldt in elk geval voor directories en startpagina's: de hierboven besproken mondiale directories dekken sowieso een klein deel van het Web, en het aandeel van Nederlandse websites daarin is minimaal (met als enige uitzondering het door vrijwilligers gerunde Open Directory Project, dat een Nederlandse afdeling heeft). Wat zoekmachines betreft ligt de zaak anders: hier bieden Nederlandse zoekmachines maar in enkele opzichten voordelen boven zoekmachines als Google en Yahoo!.

De interessantste startpagina's zijn die met een heleboel 'dochters'. Daarbij dient de algemene startpagina als toegangspoort tot pagina's met dezelfde interface over een groot aantal verschillende onderwerpen en op allerlei verschillende terreinen. Vanuit dit oogpunt zijn Startpagina.nl, Startkabel.nl en Boogo.nl de grootste en interessantste. We zullen deze echter niet apart bespreken, omdat hun 'dochters' geschaakt zijn door Startnederland.nl, dat zich tot taak heeft gesteld alle dochters van diverse startpagina's te verzamelen en via een menu toegankelijk te maken. Een beetje sneu voor de vaders en moeders van die dochters misschien, maar het Web bestaat nu eenmaal bij de gratie van links, en het is beslist zo dat de compilatoren op het Web nuttig werk verrichten.

Directories en startpagina's hebben beide hun sterke kanten: directories beschikken nogal eens over een interne zoekmachine, die het traceren van een bepaalde link met beschrijving vergemakkelijkt. Bovendien is de navigatie simpeler dan bij startpagina's, die de meest actuele en meest gevraagde onderwerpen meestal naar voren halen (ten koste van de consistentie van de onderwerpsstructuur) en op de eerste pagina vaak een nogal overladen menu van onderwerpen bieden. Daartegenover staat dat de dochters van startpagina's, samengesteld als ze zijn door verzamelaars, vaak een veel completer behandeling van een onderwerp bieden dan het betreffende deel van een directory, dat het in veel gevallen van aanmeldingen moet hebben. Een startpagina met veel dochters is eigenlijk te vergelijken met een verzameling van webgidsen.

4.1.5.1. De dekking van Nederlandse directories en startpagina's

Het Nederlandse deel van het Web omvat naar schatting 120 à 160 miljoen webpagina's, waarvan het merendeel met het topdomein .nl (de rest heeft een generieke topdomeinnaam zoals .com). Daarvan dekt het Nederlandse ODP er 100.000, minder dan een tiende procent dus.

Startnederland.nl biedt een alfabetisch overzicht van meer dan 100.000 startpaginadochters en andere onderwerpsoverzichten. Veel daarvan gaan – afkomstig als ze zijn uit verschillende startpagina's – in twee-, drie-, vier- of meervoud over dezelfde onderwerpen. Maar zelfs zo mogen we aannemen dat ze te zamen een veel groter deel van het Nederlandse Web dekken dan welke directory ook.

4.1.5.2. Open Directory Project Nederlands

URL: <http://www.dmoz.nl/>

Voor deze directory verwijst ik korthedshalve naar de bespreking van het internationale ODP (blz. 81 e.v.), waarvan de Nederlandse versie onder World deel uitmaakt.

4.1.5.3. Startnederland.nl

In kort bestek:

URL: <http://www.startnederland.nl/>

Omvang: ca. 100.000 gidsen ('portals' genaamd), alfanumeriek ingedeeld naar onderwerp, plaatsen in Nederland, landen, personen met een website (Nederlanders en buitenlanders), eigen dochterpagina's en kaarten (van Nederlandse plaatsen); daarnaast

zijn er overzichten van gidsen per regio (op kengetal), de hoofdrubricken van diverse Nederlandse webgidsen (DirNL), overkoepelende verenigingen en bonden (Ver), sites of pagina's van en over Google (Google), allerlei zaken met een internationaal karakter, zoals geld, kranten, luchthavens, regeringen (Wwijd), de hoofdrubricken van een aantal Engelstalige directories, zoals Google en Yahoo! (DirEng), varia, onder meer aandoeeningen, kunstenaars, medicijnen, alles alfabetisch (Meer indexen), plus nog reeksen van snellinks naar veelgevraagde informatie en veelbezochte sites.

Aantal onderwerpen: ettelijke duizenden (veel gidsen komen in x-voud voor)

Aantal gidsen op de homepage: >800 (dit is een selectie uit het totale aantal van ca. 100.000).

Maximaal aantal niveaus: n.v.t.

Onderhouden door: RECALC Computersystemen, Lichtenvoorde

Zoekmachine intern: nee; wel alfabetische index

Zoekmachine extern: nee

Advanced Search: n.v.t.

Beschrijvingen: nee

Sponsored/featured links (commercie): ja

Typen resources: n.v.t.

Nationale versies: n.v.t.

Help: niet beschikbaar

Portals en zoekmachines worden tegenwoordig geassocieerd met beursgangen en het grote geld, en dus is het een genoegen zo nu en dan een project tegen te komen dat de kenmerken van klantvriendelijkheid en uit de hand gelopen hobby met elkaar verenigt. Zo'n project is Startnederland.nl, het product van een tweemansbedrijf in het Gelderse Lichtenvoorde, dat computers assembleert en bij zijn afnemers installeert.

Startnederland.nl is oorspronkelijk bedoeld als homepage voor die klanten.

Startnederland.nl noemt zichzelf een 'metaportal', en dat is een juiste omschrijving van wat bij mijn weten de enige startpagina is die geheel is samengesteld uit links naar de dochters van andere startpagina's en rubrieken van andere directories. Andere startpagina's en directories gaan zo nu en dan ook wel eens met elkaars dochters aan de haal, maar bij Startnederland.nl is dit tot een missie verheven. Zelfs de ongeveer 570 eigen dochters van Startnederland zijn alleen maar verzamel pagina's van links naar de dochters van andere startpagina's. Het resultaat is een plezierige overdaad waarin – op een tamelijk algemeen niveau – naar informatie over een zeer groot aantal onderwerpen kan worden gezocht.

Startpagina's hebben doorgaans geen uitgewerkte directorystructuur, en ook geen interne zoekmachine. Bij Startnederland.nl zult u het dan ook met een alfabetisch overzicht van onderwerpen moeten doen. Erger nog, sommige onderwerpen, zoals Games, komen tientallen malen voor. Maar voor de echte liefhebber zal dat geen bezwaar zijn.

Enige eigen inspanning – buiten het opsporen en het uitstellen van al die mooie dochters natuurlijk – hebben de makers van Startnederland.nl gelukkig nog wel verricht: sommige pagina's zijn omgedoopt om beter in het onderwerpsoverzicht van Startnederland.nl te passen. Niettemin blijft enige creativiteit van de kant van de gebruiker geboden. Als u in het overzicht niet vindt wat u zoekt, zult u misschien met een verwante, algemenere of specifiekere term op een andere plaats in het alfabet moeten zoeken. Zo bestaan er naast startpagina's over kamers ook ettelijke over studentenkamers, huren en verhuur.

De homepage van Startnederland.nl bevat maar een keuze uit de dochters. Het alfabetische menu geeft toegang tot alle dochters (of portals) die de startpagina heeft opgenomen, op trefwoord.

Als extra's bevat Startnederland.nl links naar gemeentelijke sites, zowel alfabetisch als

Welkom in Nederland... **www. Start Nederland .nl** Welcome to the Netherlands...

Countrymap maandag 30 mei Gastenboek
[Home](#) [Info](#) [Aanmelden](#) [Nieuws](#) [Portalen](#) [Uitgaan](#) [Zoeken](#)

oel Auto TXT Radio TV Populair Nieuws Sport Gidsen Zoeken PC Vrijetijd Weer Winkelen Handig Muziek Overheid Verkeer Banken
 Wereld Landelijk Regionaal Lokaal Diversen Medisch Nederland Portalen Spelen Eten Uitgaan Relatie Vacatures Internet Vrouw Actueel

2Link A1 B9 Boogo Dmoz Via Eigen Jouw Jump Leuk GoogleDir Startkabel Startpagina MSN Ikwil Lookup StartBe Startbew Start2000 Verz.gids Zappsite
 SCourant AD Boerderij Emea FD dFT Gelderlander NedD Nos NRC Nu Parool Planet Spits Telegraaf Trouw Tubantia Voetbal Volkskr Web Cbs CNN BBC
 GeenStijl Agd Ameist GIC DVHN Eindh Elsevier Friesch Leeuwarder Haagsche Stentor R'dam .pdf NHD PZC Stand Utrecht Brabants Limb LD Nws Nws TisWan Omlaag

Plaatsen in NL:	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	Regios DirNL
Onderwerpen:	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	0-9 Ver Google
Buitenland:	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	Wwijd DirENG
Personen:	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	Meer indexen
Dochters:	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	Alle dochters
Kaarten:	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	Route TXT Tel

Fast : Google: Afb:
 Track: Naam: Afb:
 Vivisimo: Plaats: Wb:
 Yahoo: 101 Bel TV Tel Pers Journ R Hotm ? Ixq:

- PSV bekroont seizoen met dubbel meesterwerk
- Verliezen luchtvaart lopen op
- Nieuwe worm schakelt Windows Firewall uit
- Betere meting verdubbelt files
- TNO: kinderziekten met UMTS
- 4233 Webcams in Europa ...
- Media: politieke tsunami en mas.
- Franse volk en reg. begrijpen elkaar niet meer
- 'Belangen Frankrijk zijn moeilijk te verdedigen'
- 'Balkenende teleurgesteld over Franse 'nee'
- 'Nederlanders moeten stemmen'
- Referendumwijzer
- Verdr. tot vaststelling van een Grondwet voor Eur.
- ADSL naar 100 Mb/s
- Kennisnet steunt 'onmisbare' Wikipedia
- China schrapt exportheffingen op textiel
- Vernieuwde uitgave StartNederland.nl op komst
- Meer nieuws / nieuwsarchief ...

A B C D E F G H I J K L M N O P R S T U V W X Y Z Omlaag

Windows Update Windows XP Service Pack 2 NL Win XP SP2 Support Center Office Update NL Office 2003 Service Pack 1 Office 2000 Service Pack 3	Vorige kolom, omlaag ... Bedrijven zoeken in NL : Internet Registratie Dienst ABC-D Goudengids Wie Levert ? Bedrijven FBG	Vorige kolom, omlaag ... Kleding.StartNederland Kleurplaten.nl Koninklijkhuus.nl Korrelatie.nl K.StartNederland L Laboratorium.nl	Kalender 2005-2006 Nieuwsoverzichten: Nieuwstracker.nl Laatstenieuws.net Nieuwskaart Nederland NOS Mediatheek Actuele verkeerssituatie Weerradar Bliksem
--	---	--	---

Deel van de homepage van Startnederland.nl.

op kengetal (Regios).Onder Wwijd zijn allerlei zaken met een internationaal karakter bij elkaar gebracht, en als ware metaportal biedt Startnederland.nl onder de kopjes DirNL en DirEng ook nog eens een directe toegang tot de hoofdruibrieken van enkele Nederlandse en mondiale directories. Landelijk overkoepelende verenigingen en bonden zijn handig onder Ver gerangschikt, en onder Meer indexen vindt u albetische lijsten van links naar zaken die het simpelst alfabetisch kunnen worden geordend, zoals aandoeningen, cabaretiers, kunstenaars, lyrics, computertermen en ziektebeelden. Bij de afsluiting van dit boek was men bezig met een verbouwing, die mogelijk ook consequenties heeft voor de juistheid van de hier gegeven beschrijving. Startnederland.nl is typisch een site voor mensen die nooit kunnen onthouden waar je de telefoongids, de NS-dienstregeling en routekaarten ook alweer moet vinden, want links daarnaar staan op de homepage en alle andere pagina's klein afgedrukt bij elkaar. Daarnaast biedt de site via het startpagina-concept een overweldigende hoeveelheid toegangen tot alle mogelijke onderwerpen, die zich, door het karakter van de onderliggende startpagina's ook zeker niet tot Nederland beperken.

4.1.6. Web directories en startpagina's van andere landen

Nationale directories zijn veelal ware schatkamers van webpagina's binnen een beperkt geografisch domein. Helaas te vaak ten onrechte achtergesteld bij zoekmachines. U vindt ze ook tussen de portals bij MAIN PORTALS (TM): Guide to the Main Internet Portals by subject and by countries (<http://www.mainportals.com/>). Overzichten van

nationale zoekmachines maken vaak geen onderscheid tussen zoekmachines en directories, dus zie ook hoofdstuk 4.4.2.

Startpagina's van andere landen spoort u op via Startpaginas.StartTips.com

(<http://startpaginas.starttips.com/>) en [www.Start4all.com](http://www.start4all.com/) (<http://www.start4all.com/>).

Startpagina's in andere landen hebben echter bij lange na niet de omvang en het bereik van die in Nederland.

4.2. Zoekmachines

Hoewel directories en startpagina's nog altijd zeer populair zijn, zijn zoekmachines 'in'. Er wordt veel over geschreven en er worden hele congressen aan gewijd. Begrijpelijk is dat ook, om verschillende redenen:

- de groei van het Web, en de steeds dringender wordende vraag hoelang en in hoeverre zoekmachines in staat zullen zijn die groei bij te houden
- het feit dat zoekmachines in korte tijd honderden miljoenen documenten fulltext kunnen indexeren, een tot voor kort onvoorstelbare gedachte
- de concurrentie tussen de diverse zoekmachines, zowel wat hun dekking als wat de 'intelligente' presentatie van hun resultaten betreft
- last but not least: het grote commerciële belang van zoekmachines voor bedrijven, speciaal wat hun 'position' in de resultatenlijst betreft als de zoeker met bepaalde woorden gezocht heeft.

Goed zoeken met een zoekmachine kan geweldige resultaten opleveren. Zowel de keuze van de juiste zoekmachine(s) als het juiste gebruik ervan zijn dus van groot belang.

Als u regelmatig meer dan een zoekmachine gebruikt, waarom zou u dan steeds opnieuw een afweging maken welke u zult gaan gebruiken, en hoe? Er bestaan handige sites waar je ofwel kunt vinden welke zoekmachine(s) je voor een bepaalde vraag moet gebruiken, ofwel hoe je ze moet gebruiken. Enkele daarvan zijn hierboven al genoemd, op blz. 76. Hier volgen er nog enkele. Kijk eens op deze adressen en overweeg of u een ervan niet als startpagina wilt gebruiken of in elk geval in uw bookmarks wilt opnemen:

- Infopeople's Search Tool Chart (<http://infopeople.org/search/chart.html>)
- Search Engines: Table of Features (<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html>)
- Tool Kit for the Expert Web Searcher (<http://www.ala.org/ala/lita/litaresources/toolkitforexpert/toolkitexpert.htm>)

Bij de hieronder besproken grote mondiale zoekmachines heb ik de volgende opzet gevolgd. Ik begin met essentiële informatie over de zoekmachine, met wat literatuurverwijzingen, onder meer naar sites waar die informatie wordt bijgehouden. Daarna volgen tabellen met de zoekmogelijkheden, van de grote mondiale zoekmachine en – indien aanwezig – enkele geselecteerde speciale zoekmachines. In de eerste kolom vindt u de syntax, in de tweede voorbeelden van het gebruik en in de derde de resultaten. Samen met de algemene uitleg van de diverse mogelijkheden van zoekmachines moeten deze voorbeelden duidelijk maken in welke gevallen en hoe deze mogelijkheden gebruikt kunnen worden. Daarbij komen ook de shortcuts ter sprake, maar niet de mogelijkheden van de zoekmachines als rekenmachine. Voor dat laatste verwijs ik naar de helpfile van de zoekmachines zelf.

Na de tabellen behandel ik de diverse zoekinterfaces en de resultatschermen en bespreek vervolgens de diverse sterke en zwakke punten van de zoekmachine. Ten slotte besteed ik aandacht aan enkele speciale zoekmachines.

4.2.1. *Wat maakt een zoekmachine tot een goede zoekmachine?*

Vier factoren bepalen hoofdzakelijk de kwaliteit van een zoekmachine:

- de snelheid waarmee de spider de database ververst, nieuwe sites toevoegt en dode links (verwijzingen naar inmiddels verdwenen webpagina's) verwijdert
- omvang en kwaliteit van de database
- de functionaliteit en het gebruiksgemak van de zoekinterface
- de ranking en de presentatie van de gevonden sites

De verschillen op deze punten tussen de diverse zoekmachines zijn aanzienlijk:

- de snelheid waarmee spiders de databases verversen, varieert van minder dan twee weken tot maanden;
- de omvang van de databases varieert bij de grote mondiale zoekmachines van drie miljard tot bijna 9 miljard (maart 2006); en hun kwaliteit varieert afhankelijk van de door de spiders gevolgde routes;
- de zoekmogelijkheden variëren van uiterst sober tot zeer geavanceerd; en het gebruiksgemak varieert van eenvoudig tot ingewikkeld;
- en er zijn zoekmachines die irrelevante zoekresultaten in een onbegrijpelijke volgorde presenteren, met summaries die nauwelijks informatie verschaffen over het onderwerp van de gevonden sites, en zoekmachines die meteen in de roos lijken te hebben geschoten en je laten zien waarom de gevonden documenten gevonden zijn.

Een van de belangrijkste punten waarop zoekmachines met elkaar concurreren is de omvang van hun database. In dat opzicht gaan de veranderingen ook het snelst: met een zekere regelmaat worden de uitkomsten van onderzoeken gepubliceerd waarin een schatting van de omvang van het Web wordt gemaakt; en korte tijd later treden er dan weer een of meer zoekmachines in de publiciteit met een bericht dat ze hun database hebben opgewaardeerd en nu de grootste zijn. De getallen die u bij de hieronder behandelde grote mondiale zoekmachines ziet dateren van mei 2005. Recentere schattingen worden – met nogal wat vertraging – bijgehouden door Danny Sullivans Search Engine Watch (<http://searchenginewatch.com/reports/sizes.html>) en Greg Notess' Search Engine Showdown

(<http://searchengineshowdown.com/stats/sizeest.shtml>).

Meer over de omvang van zoekmachines: The Indexable Web is more than 11.5 billion pages (<http://www.cs.uiowa.edu/~asignori/web-size/>). De nieuwste schattingen vindt u in een onderzoek, getiteld A Comparison of the Size of the Yahoo and Google Indexes (<http://vburton.ncsa.uiuc.edu/indexsize.html>). Wordt vervolgd...

Nieuws over zoekmachines is onder meer te vinden bij Webmaster World World

(<http://www.webmasterworld.com/>), Pandia's Search Engine News, on Web searching and search engine optimization (<http://www.pandia.com/searchworld/index.html>) en in Greg Notess' hierboven genoemde Search Engine Showdown (<http://www.searchengineshowdown.com/>). Zeer de moeite waard is ook Tara Calishains Researchbuzz (<http://www.researchbuzz.com/>).

Daarnaast bieden alle grote zoekmachines extra's. Zo bieden drie van de vier grootste mondiale zoekmachines bij de link naar een gevonden webpagina ook een 'cached' kopie van die pagina in de eigen database, een geweldige service in die gevallen waarin een site voor het moment onbereikbaar is; de op vier na grootste mondiale zoekmachine, Gigablast, biedt naast een 'archived copy' ook nog een handige 'stripped copy' zonder plaatjes en een link naar oudere kopieën van pagina's in het Internet Archive (zie blz. 5). Daarnaast biedt Google bij elke link ook 'Similar pages'. Yahoo!, Ask en Exalead analyseren het zoekresultaat automatisch en doen suggesties voor andere of meer zoektermen, waarbij Ask ook nog eens gerelateerde persoonsnamen opspoort, en Exalead de bijbehorende rubrieken van het Open Directory Project vermeldt als treffers uit het zoekresultaat in die directory voorkomen. MSN Search laat u van tevoren een zoekresultaat op verschillende manieren sorteren, Google beschikt

over een ongeëvenaarde spellingcontrole. Een veel voorkomende optie bij zoekmachines geldt ook de presentatie van de gevonden webpagina's: zowel het aantal per pagina gepresenteerde treffers van de zoekmachine, als de optie om de zoekresultaten in een nieuw venster of onder een nieuwe tab te tonen.

Hoe snel verversen zoekmachines hun database?

Aanbieders en afnemers van informatie op het Web hebben er belang bij dat de informatie die gegeven en gevonden wordt up-to-date is. Op dat punt zijn er tussen zoekmachines grote verschillen. Dat blijkt als je het resultaat van diverse grote mondiale zoekmachines voor pagina's die de afgelopen drie maanden geupdated zijn vergelijkt met het resultaat dat ze zonder die beperking vinden. De zoekterm is hier tantaluskwelling, en de aantallen zijn die van de werkelijk getoonde treffers.

	<i>Google</i>	<i>Yahoo!</i>	<i>Ask</i>	<i>Exalead</i>
alle	<u>663</u>	<u>558</u>	21	<u>55</u>
laatste 3 maanden	<u>656</u>	<u>322</u>	1	<u>2</u>

Al te veel conclusies mag je uit zo'n testje echter ook weer niet trekken. In de eerste plaats updaten zoekmachines hun databases op uiteenlopende momenten, maar misschien nog belangrijker is dat ze waarschijnlijk ook de datering van webpagina's (speciaal van niet gedateerde) verschillend interpreteren.

Zoals al eerder gezegd hebben vrijwel alle grote mondiale zoekmachines ook een directory, hetzij een eigen, hetzij een die ze van het Open Directory Project hebben overgenomen en waarmee ze de zoekmachine aanvullen.

Van deze en soortgelijke factoren hangt de kwaliteit van zoekmachines voor een groot deel af. Een beste zoekmachine is er niet. Dat is niet alleen zo omdat geen enkele zoekmachine aan alle wensen voldoet en omdat hun zoekmogelijkheden aanzienlijk verschillen, maar ook omdat zoekresultaten mede afhankelijk zijn van de actuele vraag: een zoekmachine die bij de ene vraag kwantitatief en kwalitatief relatief goed scoort, kan dat bij een andere vraag minder doen. En per slot van rekening dekken ze geen van alle meer dan een deel van het Web en indexeren ze lange documenten niet allemaal even volledig. Het gebruik van meer dan een zoekmachine, mits met inachtneming van de verschillen tussen hun zoektaken, wordt dan ook door de deskundigen sterk aanbevolen.

4.2.1.1. Overlap

In de afgelopen jaren is er nogal wat onderzoek gedaan naar de overlap van de databases van zoekmachines, en een gangbare opvatting die daaruit naar voren is gekomen dat die overlap betrekkelijk gering is. Na de laatste herindexeringen van Google en Yahoo! tegen het eind van 2004 en in augustus 2005 en de introductie van de eigen zoekmachine van MSN Search is het de vraag of dat nog waar is. Maar voor de zoeker is dat in elk geval niet relevant. De meeste vragen leveren immers veel meer treffers op dan de enkele tientallen die we bereid zijn te bekijken, en voor die treffers is niet de overlap tussen de zoekmachines maar zijn de verschillen in hun ranking essentieel. Wie met één zoekmachine zoekt, loopt het risico het antwoord op zijn vraag te missen, niet omdat die zoekmachine misschien toevallig het ene document dat dat antwoord bevat niet geïndexeerd heeft, of niet helemaal geïndexeerd heeft, maar vooral omdat het document, dat wel in de database van zijn zoekmachine voorkomt, het in de ranking niet tot de eerste dertig heeft gebracht.

4.2.2. Google

4.2.2.1. In kort bestek

URL (Basic Search): <http://www.google.com/ncr> of <http://www.google.com/webhp>; als u het ncr (webhp) weglaat opent uw browser automatisch – of als u het in de Preferences zo hebt ingesteld – de lokale versie van Google, die naast mondiaal ook nationaal zoeken (op taal en binnen een land) mogelijk maakt.

NB: Denk erom dat sommige speciale zoekmogelijkheden van shortcuts, waarvoor Google externe (Amerikaanse) databases gebruikt, niet werken in de lokale versie van Google!

Omvang : volgens eigen opgave (3-11-2004) ruim 8 miljard (waarvan naar schatting 7 miljard geïndexeerd; van een onbekend aantal pagina's met alleen 'outgoing' links is alleen de URL geïndexeerd; verder behoort een deel tot een aparte database van zogeheten 'Supplemental results'; volgens een studie van Gulli en Signorini (<http://www.cs.uiowa.edu/~asignori/web-size/>) indexeert Google iets meer dan 76% van het zichtbare Web.

Indexeert tot: 101 K (HTML), ca. 2 Mb (PDF) (voorbeeld); maar sommige pagina's hebben een limiet tot ca. 500K (voorbeeld). Bij TXT-bestanden ligt de indexeerlimiet nog veel hoger, tot meerdere megabytes (voorbeeld).

Maximale omvang zoekargument in het zoekvenster: 32 zoektermen, wildcards (*) niet meegerekend; **NB:** deze beperking kan niet worden omzeild door de URL te manipuleren

Indexeert andere formaten dan HTML: ja, onder meer .txt, Adobe PDF, Adobe Postscript, Microsoft Word, Works, Write, PowerPoint en Excel (zie ook:

<http://www.faganfinder.com/filetype/>)

Toont: nooit meer dan 1000 treffers

Updating: enkele minuten (News) – enkele weken

Google Sitemap: <http://www.google.com/sitemap.html>

Help en info: <http://www.google.com/help/index.html>

Advanced Search: http://www.google.com/advanced_search

Directory: <http://www.google.com/dirhp> (Open Directory Project)

Speciale zoekmachines: <http://www.google.com/imghp?hl=en&tab=wi&q=> (Google Images; afbeeldingen), <http://news.google.com/nwshp?hl=en&gl=us> (Google News; nieuws),

<http://groups-beta.google.com/?sourceid=ghpp> (Google Groups; discussiegroepen, met de mogelijkheid een eigen groep te starten),

<http://answers.google.com/answers/> (Google Answers; vraagbaak),

<http://scholar.google.com/> (Google Scholar; wetenschappelijke informatie),

<http://books.google.com/> (Google Book Search; gedigitaliseerde boeken),

<http://www.google.com/psearch> (Google Personalized Search), Google Blog Search (<http://blogsearch.google.com/blogsearch>) en diverse andere; daarnaast Google API-zoekmachines (zie onder)

Testomgeving: <http://labs.google.com>

Blog: <http://www.google.com/googleblog/>

Meer over Google:

Google Guide: Making Searching Even Easier: <http://www.googleguide.com/>

google.htm: How to search the web, by fravia+ google's own

(<http://www.searchlores.org/google.htm#schola>)

Google Indicateur: <http://google.indicateur.com/>

Google: Tips, Tricks and More:

<http://www.library.auckland.ac.nz/docs/handouts/google.pdf>

Moteurs de recherche; le comparatif Google-Yahoo
<http://www.journaldunet.com/0411/041116googleyahoo.shtml>

Overig:

<http://www.dsi-info.ca/moteurs-de-recherche/Google/fiche-technique.html>

<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Google.html>

<http://www.abondance.com/outils/google.html>

http://c.asselin.free.fr/french/google_tips.htm

<http://www.searchengineshowdown.com/features/google/review.html>

<http://www.at-web.de/google/google.htm>

En voor wie niet genoeg van Google kan krijgen: de echte Google-fanaten vindt u hier:

<http://www.startnederland.nl/nl/g/google/index.html>

Overzicht

Soort zoekactie:	Voorbeelden:	Vindt:
Natuurlijke taal (alleen Engels!)	can I use natural language in Google searching?	Probeer!
<i>Prefixes</i>		
allintitle: (komt overeen met de optie 'Occurrences: in the title of the page' in de Advanced Search)	allintitle:geschiedenis brabant allintitle:geschiedenis Brabant -"Noord Brabant" allintitle:geschiedenis Brabant -"Noord Brabant" filetype:pdf	documenten met de opgegeven zoektermen in de titel van de pagina NB: allintitle: kan niet met zoektermen elders worden gecombineerd: het heeft betrekking op alle zoektermen in de zoekexpressie en dient daaraan vooraf te gaan; wel kunnen termen met de – operator in de expressie worden opgenomen en kan allintitle: worden gecombineerd met filetype:, numrange:, site: Google neemt allintitle letterlijk: u vindt hiermee geen pagina's op zoektermen in de META-tags
intitle:	intitle:philately intitle:"stamp collecting" intitle:"stamp collecting" faq faq intitle:"stamp collecting"	documenten met de opgegeven zoekterm of phrase in de titel van de pagina (en eventuele andere zoektermen op een willekeurige plaats in de pagina) Google neemt intitle: letterlijk: u vindt hiermee geen documenten op zoektermen die in de META keywords en description voorkomen
allinurl: (komt overeen met de optie 'Occurrences: in the URL of the page' in de Advanced Search)	allinurl:ub uva allinurl:ub uva -aanwinsten allinurl:ub uva -aanwinsten filetype:ppt	documenten met de opgegeven zoektermen in de URL NB: allinurl: kan niet met zoektermen elders worden gecombineerd: het heeft betrekking op alle zoektermen in de zoekexpressie en dient daaraan vooraf te gaan; wel kunnen termen met de – operator in de expressie worden opgenomen en kan allinurl: worden gecombineerd met filetype:, numrange:, site:

<p>inurl:</p>	<p>anthrax inurl:minvws</p> <p>notulen inurl:minbuza</p> <p>intitle:notulen inurl:pdf</p> <p>site:hva.nl inurl:handleiding</p> <p>intitle:"digital library" inurl:search</p> <p>inurl:umass integrity OR misconduct "record keeping" OR notebook OR notebooks</p>	<p>documenten die de opgegeven zoekterm in de URL bevatten (en eventuele andere zoektermen op een willekeurige plaats in de pagina); voor grote precisie te combineren met site: of intitle:</p> <p>het antwoord op de vraag hoe lang minimaal bij de University of Massachusetts onderzoeksverslagen moeten worden bewaard</p>
<p>intext: (zie paragraaf 3.3.8, blz. 66)</p>	<p>intext:oneiromancy</p> <p>oneiromancy -intext:oneiromancy</p> <p>oneiromancy -intext:oneiromancy -inurl:oneiromancy -intitle:oneiromancy</p>	<p>webpagina's met de opgegeven zoekterm in de 'body' (in tegenstelling tot de URL en/of de titel) van de pagina</p> <p>webpagina's met de opgegeven term ergens anders dan in de body van de pagina; dat kan zijn in de URL, in de titel, of in een anchor van een verwijzende pagina (zie bij inurl:, intitle: en inanchor:)</p> <p>webpagina's die de opgegeven zoekterm zelf niet bevatten maar waarbij deze alleen in een anchor van naar de betreffende pagina verwijzende andere pagina's voorkomt (d.w.z. die gevallen waarin Google bij de gecacheerde versie van een treffer de melding geeft: 'These terms only appear in links pointing to this page:')</p> <p>NB: Deze melding heeft alleen betrekking op de gecacheerde versie van de pagina waarnaar wordt verwezen; de huidige versie kan de zoektermen inmiddels wel bevatten!</p>
<p>allintext: (komt overeen met de optie 'Occurrences: in the text of the page' in de Advanced Search)</p>	<p>allintext:geschiedenis brabant</p> <p>allintext:geschiedenis Brabant -"Noord Brabant"</p> <p>allintext:geschiedenis Brabant -"Noord Brabant" site:be filetype:pdf</p>	<p>webpagina's met de opgegeven zoektermen in de 'body' (in tegenstelling tot de URL en/of de titel) van de pagina</p> <p>NB: allintext: kan niet met zoektermen elders worden gecombineerd: het heeft betrekking op alle zoektermen in de zoekexpressie en dient daaraan vooraf te gaan; wel kunnen termen met de – operator in de expressie worden opgenomen en kan allintext: worden gecombineerd met filetype:, numrange:, site:</p>
<p>inanchor: (zie paragraaf 3.3.8.2, blz. 68)</p>	<p>inanchor:mennonieten</p> <p>inanchor:"Dutch history" -intitle:"Dutch history"</p>	<p>documenten die worden gevonden doordat er vanuit een andere pagina of pagina's met de opgegeven zoekterm naar verwezen wordt</p> <p>documenten waarnaar verwezen wordt met het anchor 'Dutch history', maar die zelf een andere titel hebben</p>

allinanchor:	allinanchor:geschiedenis brabant allinanchor:geschiedenis Brabant -"Noord Brabant" allinanchor:geschiedenis Brabant -"Noord Brabant" site:be filetype:pdf	documenten die worden gevonden doordat er vanuit een andere pagina of pagina's met de opgegeven zoektermen naar verwezen wordt NB: allinanchor: kan niet met zoektermen elders worden gecombineerd: het heeft betrekking op alle zoektermen in de zoekexpressie en dient daaraan vooraf te gaan; wel kunnen termen met de – operator in de expressie worden opgenomen en kan allinanchor: worden gecombineerd met filetype:, numrange:, site:
link: (link: komt overeen met de optie 'Link:' onder Page-Specific Search op de Advanced Search pagina van Google)	link:www.searchability.com variant: link:http://www.searchability.com	documenten met de opgegeven URL als verwijzing; NB: bij Google kan link: niet worden gecombineerd met een andere zoekterm of zoekopdracht; zoeken naar een document dat naar twee andere, al bekende documenten linkt, kan dus bij voorbeeld niet link: verzamelt de verwijzingen naar een pagina, niet naar een hele site; gebruik voor dat laatste linkdomain: van Yahoo!
related: (related: komt overeen met de optie 'Similar:' onder Page-Specific Search op de Advanced Search pagina van Google)	related:www.searchability.com	webpagina's die overeenkomen met de pagina met de opgegeven URL (similar, gelijk(w)aardig)
info: id:	info:www.academicinfo.com www.academicinfo.com	de informatie die Google over de opgegeven URL kan bieden (cache, similar to, link to, from the site, contain the term) het prefix kan ook worden weggelaten
site: (komt overeen met de optie 'Domain:' in de Advanced Search)	handleidingen site:hva.nl handleidingen site:ie.hva.nl handleidingen site:htsa.ie.hva.nl faq site:advies.overheid.nl site:advies.overheid.nl faq	documenten met de opgegeven term(en) bij de opgegeven site (domain of subdomain), m.a.w. beperkt de zoekactie tot het opgegeven domain (inclusief subdomains) of subdomain kan worden gecombineerd met zoektermen elders, volgorde (meestal) niet van belang
filetype: ext: (komt overeen met de optie 'File Format' in de Advanced Search)	classified filetype:pdf filetype:txt vertrouwelijk inurl:leidenuniv ext:ps	(tekst)bestanden of webpagina's met een bepaalde extensie, onder meer de extensie PDF (Acrobat), maar ook b.v. TXT, TEX, XML, ASP; kan alleen in combinatie met zoekterm; in het voorbeeld ('classified'): PDF-files met die zoekterm (zie ook: http://www.faganfinder.com/filetype/)
safesearch: (alleen Engels) (komt overeen met de optie 'SafeSearch' in de Advanced Search)	safesearch:sex education	volgens Google's documentatie zou dit prefix Google's 'adult filter' inschakelen; maar het werkt niet (altijd); gebruik in plaats hiervan Advanced search

<p>numrange:[#-#] [#]..[#]</p>	<p>23..30 concerten amsterdam "23..30 december 2006"</p> <p>16.3..21.7</p> <p>10000..11000</p> <p>10.000..11.000</p> <p>"10.000..11.000 inhabitants" "10.001..10.999 inhabitants"</p> <p>"waarden tussen 2999..3999 en 2499..4999"</p> <p>"temperaturen tussen numrange:21-25 C en numrange:21-40 C"</p>	<p>documenten met daarin vermeld de getallen tussen 23 en 30 (inclusief 23 en 30, en breuken volgens de Angelsaksische notatie); kan in phrases worden gebruikt en gecombineerd met andere zoektermen; werkt alleen numeriek</p> <p>documenten met getallen tussen 16,3 21,7, inclusief de hele getallen NB: de breuken werken alleen in de Angelsaksische notatie, met de punt; breuken in de continentale notatie, met een komma, worden door Google niet als zodanig herkend</p> <p>vindt getallen tussen 10000 en 11000, inclusief 10,000-11,000 (maar niet 10.000-11.000) en inclusief breuken zoals 10,481.60</p> <p>vindt 10 of 11 en alle tussenliggende breuken; daarnaast ook de getallen tussen 10.000 en 11.000 in onze notatie, met de punt</p> <p>aantallen inwoners tussen 10.000 en 11.000, maar ook de aantallen 10 of 11 aantallen inwoners tussen 10.001 en 10.999, maar ook bijv. "10.5 inhabitants per square kilometer"</p> <p>NB: gebruik numrange bij getallen in onze notatie altijd in combinatie met een zoekterm (lieft in een phrase) of zoektermen waaruit blijkt dat u op zoek bent naar een duizendtal en niet naar een breuk; zie ook onder getallen</p> <p>vindt documenten met bijv. de phrase "waarden tussen 3500 en 3100" (NB: in de zoekexpressies #..# en numrange:#-# dient het laagste getal altijd voorop te staan!)</p> <p>documenten met bijv. de phrase: "temperaturen tussen -25°C en +25°C"</p>
<p>cache:</p>	<p>cache:www.leidenuniv.nl cache:www.leidenuniv.nl nieuws</p>	<p>de door Google gecacheerde versie van een webpagina</p> <p>de opgegeven zoekterm(en) (in het voorbeeld 'nieuws') gemarkeerd in de cachekopie</p> <p>NB: de zoekterm komt na cache:-opdracht</p>
<p><i>Operatoren</i></p>		
<p>term1 term2 (komt overeen met de optie 'with all the words' in Advanced Search)</p>	<p>Verdi Requiem</p>	<p>de ene EN de andere zoekterm (standaard)</p> <p>NB: zoektermen hoeven niet altijd in het document zelf voor te komen maar kunnen dat ook doen in het anker van een verwijzend document</p>

<p>+term NB: het +-teken dient bij Google niet meer als unitaire operator, maar kan in een aantal gevallen worden gebruikt om Google te dwingen een standaardbewerking niet uit te voeren</p>	<p>movies stars Malawians +movies +stars +Malawians</p> <p>child xenophobic counterfactual family +child xenophobic counterfactual family</p> <p>in the solicitation of +in +the solicitation +of "in the solicitation of"</p> <p>électricité +électricité</p> <p>e-merge +e-merge</p>	<p>NB: om de resultaten met de gepluste termen in deze voorbeelden te zien moet u de plussen zelf in de zoekregel aanbrengen! Door een zoekterm te 'plussen' elimineert u treffers die die zoekterm alleen in het anchor van een verwijzend document hebben. in dit voorbeeld wordt de automatic stemming (zie hieronder) van child (children) voorkomen door de term te 'plussen' + vindt stopwoorden (zoals 'of', 'the', 'in') die anders niet in de zoekactie meegenomen zouden worden NB: stopwoorden in phrases hoeven niet meer te worden geplust + kan ten slotte worden gebruikt om Google uitsluitend naar de geaccentueerde of exacte vorm van een woord te laten zoeken (électricité)</p> <p>e-merge, niet emerge (zonder de plus levert e-merge ook emerge op)</p>
<p>-term (komt overeen met de optie 'without the words' in de Advanced Search)</p>	<p>Bach -Johann</p> <p>Bach -"Johann Sebastian"</p> <p>intitle:"Roman empire" -site:com</p>	<p>de ene maar niet de andere term (of phrase)</p> <p>documenten met de opgegeven zoekterm of -phrase in de titel, maar afkomstig van sites die niet het TLD com hebben</p>
<p>OR (komt overeen met de optie 'with at least one of the words' in de Advanced Search)</p>	<p>Beethoven OR Mozart</p> <p>trio Beethoven OR Mozart</p> <p>Beethoven OR Mozart trio</p> <p>"Oost-Indische Compagnie" OR "West-Indische Compagnie" bewindvoerders</p> <p>Beethoven OR Mozart trio OR kwartet</p>	<p>de ene OF de andere zoekterm, OF beide; in combinatie met een derde zoekterm wordt met deze derde zoekterm een EN-relatie gelegd; OR verbindt twee woorden of twee phrases, en wel alleen die welke direct links of rechts van de OR-operator staan</p> <p>documenten met de naam Beethoven of de naam Mozart in combinatie met de term trio of de term kwartet</p>
<p>~term (alleen Engels)</p>	<p>~anorexia</p> <p>~disability</p>	<p>documenten met de opgegeven zoekterm en/of verwante termen Gebruik -[term] om te achterhalen wat de verwante termen zijn: (Uw zoekbewerking - ~anorexia -anorexia -"eating disorders" -bulimia -anorexic - heeft geen overeenkomstige documenten opgeleverd) (Uw zoekbewerking - ~disability -disability -ADA -accessibility -disabled -SSI -wheelchair -"cerebral palsy" - heeft geen overeenkomstige documenten opgeleverd)</p>

<p><i>Nabijheid</i></p> <p>Phrase ("...") (komt overeen met 'with the exact phrase' in de Advanced Search)</p>	<p>"Willem van Hanegem"</p> <p>"part time"</p> <p>"I O"</p> <p>"I/O"</p>	<p>de opgegeven naam of zinsnede als reeks van opeenvolgende woorden, met tussen elke twee woorden al dan niet een spatie of een ander leesteken part-time, maar ook parttime</p> <p>io, maar niet I/O (zie ook bij Lees- en andere niet-lettertekens, hieronder) I/O</p>
<p>wildcard (*)</p>	<p>"Peter * de Vries"</p> <p>"Peter * de Vries" -"Peter R de Vries"</p> <p>"in the * of the"</p> <p>"leerde * Lindelaan"</p> <p>"leerde * * Lindelaan"</p> <p>"van * * Westerflier"</p>	<p>* staat voor één willekeurig woord documenten met onder meer de naam Peter R. de Vries</p> <p>documenten met onder meer de naam van Peter Paul de Vries of Peter Rudolph de Vries (= Peter R. de Vries), maar niet met de naamsvorm Peter R. de Vries</p> <p>documenten met naast phrases als "in the spirit of the", "in the wake of the" etc. ook phrases "in the Inaugural Session of the" en "in the largest American campaign of the", waarin de joker niet één maar meer (1-3?) woorden representeert</p> <p>geen treffers</p> <p>Documenten met de phrase "leerde Lotje lopen langs de lange Lindelaan" de naam van Van Hövell tot Westerflier en variante vormen, maar ook de naam (van Hövell) van Wezeveld en Westerflier; daarnaast ook: van Nijenhuis, Peckedam en Westerflier en de phrase "van hypotheekverstrekker Hövell tot Westerflier" (3 woorden)</p>
<p><i>Overige</i></p>		
<p>.. results²⁾</p>	<p>10/20/30/50/100</p>	<p>kies het aantal treffers per pagina van de resultatenlijst</p>
<p>language²⁾</p>	<p>Icelandic</p>	<p>kies een van de 35 talen in het menu onder Any language in de Advanced search; of gebruik de opties Preferences of Language Tools vanaf het Basic Search-scherm</p>

country	Singapore bij de keuze Taal: Nederlands en Land: Singapore levert de zoekactie Nederland -site:sg onder meer deze treffer op:	deze optie zit verstopt onder de Language tools ' van Google; kiest een van de 69 landen in het menu onder Search pages located in: any country NB: deze wijze van zoeken verschilt van die met de beperking site:sg (voorbeeld: Singapore) en levert een aanzienlijk groter resultaat (doordat country ook generieke TLD's in een land omvat). XpatXperience - Informatie voor Nederlandstaligen in Singapore (http://www.xpatxperience.com/communities/nl.shtml)
date (Return web pages updated in the) ²⁾	anytime / past 3 months / past 6 months / past year	documenten met als laatste wijzigingsdatum een datum binnen een van de opgegeven periodes
Usage Rights ²⁾	not filtered by license /free to use or share / free to use or share, even commercially / free to use, share or modify / free to use, share or modify, even commercially "How computers work" course	documenten waarvoor het volledige dan wel een gedeeltelijk (" Creative Commons ") copyright geldt materiaal voor een cursus over het opgegeven onderwerp onder de gekozen voorwaarden
SafeSearch ²⁾	No Filtering / Filtering using SafeSearch	schakelt het Family Filter van Google uit / aan Standaard is Moderate filtering; gebruik de mogelijkheden van Advanced Search of Preferences om die instelling te veranderen
accenten	é, ï, ö, û etc. électricité oneiromancy electricite oneiromancy +électricité oneiromancy oologie site:nl oölogie site:nl ode site:de öde site:de "Munchausen by proxy syndrome" site:org "Münchhausen by proxy syndrome" site:org "Munchausen by proxy syndrome" OR "Münchhausen by proxy syndrome" site:org	Google maakt soms wel en soms geen onderscheid tussen wel en niet geaccentueerde letters. In de regel geldt dat geaccentueerde zoektermen ook ongeaccentueerde treffers vinden, en andersom. Als dat het geval is, kan de geaccentueerde vorm zonder de ongeaccentueerde worden gevonden door de term te 'plussen' in dit geval geven de eerste twee zoekexpressies hetzelfde resultaat, de derde een lager in dit voorbeeld vindt oologie niet oölogie, maar oölogie wel oologie ode vindt niet öde, maar öde wel ode 396 treffers 406 treffers 408 treffers (een van eigenaardigheden die soms bij zoekmachines opduiken)
hoofdletters/ kleine letters (case-sensitivity)	next NeXT	Google maakt geen onderscheid tussen hoofdletters en kleine letters

<p>lees- en andere niet-lettertekens</p>	<p>-,/_&+#\$</p> <p>part-time part-time -"part time" +part-time c+</p> <p>c++ c+ -c++ &c single_user a# c# f# g# a&b&c a_b_c \$5000</p> <p>€5000 £5000 \$5000..\$6000</p> <p>I/O #2 >2</p>	<p>Google verwaarloost de meeste lees- en niet-lettertekens, of vat ze op als spaties binnen een phrase; maar in een aantal gevallen worden ze mee geïndexeerd en zijn als losse tekens of als onderdeel van een string van tekens vindbaar; probeer en let op de presentatie van de zoekresultaten (vet en evt. onderstreept wil zeggen: gevonden) parttime, part-time, "part time" alleen parttime alleen part-time c+; Google geeft hierbij, heel handig, ook een verwijzing naar treffers bij c++ c+ maar niet c++ &c single_user a# c# f# g# a&b&c a_b_c \$5,000 (of \$5000); NB: in de resultatenlijst worden de komma's veelal niet getoond! 5,000 (5000) (dus geen € en £ tekens)</p> <p>NB: het dollarteken werkt niet (meer) met numrange: (zie boven) I/O 2 2</p>
<p>getallen (zie ook: numrange: en Codes onder Shortcuts)</p>	<p>10000</p> <p>10.000</p> <p>25.32 25,32</p>	<p>10000, inclusief 10,000 (dat lijkt anders te zijn als je naar de resultatenlijst kijkt, maar dat komt doordat Google in de snippet de komma verwijdert) vindt 10.000, daarnaast ook 10.000.000, 10.000m, 10.000ste 25.32 25 32 (dus 25 AND 32) (Google gebruikt de Angelsaksische notatie)</p>
<p>stopwoorden</p>	<p>from Amsterdam to Warsaw +from Amsterdam +to Warsaw</p> <p>"from Amsterdam to Warsaw" "from Amsterdam" "to Warsaw" "from Amsterdam * * to Warsaw"</p>	<p>Google heeft wel stopwoorden, maar die kunnen toch worden gevonden door ze te 'plussen'; ook binnen een phrase zijn ze, al dan niet 'geplust', vindbaar Amsterdam [EN] Warsaw from [EN] Amsterdam [EN] to [EN] Warsaw "from Amsterdam to Warsaw" "from Amsterdam" [EN] "to Warsaw" "from Amsterdam Schiphol Airport to Warsaw", maar ook: "from Amsterdam Schiphol airport on 20 November 2002 to Warsaw"; zie boven bij wildcard</p>
<p>(vertaling) [Web Page Translation]</p>	<p>"histoire de la France"</p>	<p>treffers in het Frans, Duits, Italiaans, Spaans en Portugees hebben als extra in de snippets een keuze 'Translate this page' (automatische vertaling in het Engels)</p>

definition	women "Alexander the Great"	<p>over veel zoektermen kan Google nalaginformatieverschaffen via de balk waarin het de omvang van het zoekresultaat meldt; dat gebeurt op twee manieren:</p> <ul style="list-style-type: none"> als u met één zoekterm hebt gezocht, staat die zoekterm in de balk met daarachter een link onder (definition) als u meer zoektermen hebt gebruikt zijn die allemaal of voor een deel zelf in de balk van een link voorzien <p>klikken op de link brengt ons naar de dictionary van Answers.com, waar naast een definitie nog aantal andere mogelijkheden, waaronder synoniemen en vertalingen, worden aangeboden. Google heeft deze informatie niet voor alle woorden</p> <p>NB Dit is een andere dienst dan die welke Google met het prefix define: aanbiedt!</p>
(spellingcontrole) [Spell Checker]	<p>catagory (spelfout)</p> <p>wife-swlling (typfout)</p> <p>koningin (typfout)</p> <p>"Berend Botje ging uit varen"</p> <p>"Berend Botje ging eens varen"</p>	<p>verschijnt bovenaan de resultatenlijst in de vorm van de vraag: Did you mean: ...?</p> <p>Did you mean category? (gevolgd door de vele treffers met de foutieve spelling)</p> <p>Did you mean wife-swapping (nee, ik bedoelde: wife-selling)</p> <p>Did you mean koningin?, maar ook twee treffers met de foute spelling</p> <p>318 treffers, geen spellingsuggestie (correct)</p> <p>1 treffer, spellingsuggestie: Did you mean: "Berend Botje ging eens waren"</p> <p>Google controleert niet de phrase, maar de afzonderlijke termen; de suggestie 'waren' is statistisch bepaald</p>
suggesties voor verder zoeken/categorisatie	n/a	n/a
[term weighting]: volgorde zoektermen	<p>kanker voeding voeding kanker</p> <p>nieuwe lichtung rekruten lichting nieuwe rekruten nieuwe rekruten lichtung</p>	<p>als eerste treffers in de resultatenlijst, afhankelijk van de volgorde van de zoektermen, documenten waarin het ene dan wel het andere aspect van de vraag overweegt (op basis van de door Google gebezigde statistische criteria)</p> <p>De volgorde van de zoektermen is ook van belang omdat Google automatisch naar phrases zoekt.</p> <p>NB: het effect is, te midden van alle andere rankingfactoren, gering</p>
[term weighting]: herhaling van zoektermen	<p>voeding voeding voeding voeding kanker kanker kanker kanker kanker voeding</p>	<p>als eerste treffers in de resultatenlijst, afhankelijk van de relatieve ranking en frequentie van de zoektermen, documenten waarin het ene dan wel het andere aspect van de vraag overweegt</p> <p>Frequentie zoektermen op eerste resultatenpagina van 10, resp.:</p> <p>37 maal voeding / 31 maal kanker</p> <p>30 maal voeding / 37 maal kanker</p>

truncatie	n/a	trunceren van termen is bij Google niet mogelijk; zie voor andere mogelijkheden (automatic) stemming en nabijheid
[(automatic) stemming]; (automatische) toevoeging van zoektermen met dezelfde stam als de opgegeven zoektermen; voor de voorwaarden waaronder dit gebeurt, zie de volgende kolommen	<p>child (nee)</p> <p>child bicycle (ja: ook children, bicycles, bicycling, bicyclist, bicyclists) +child +bicycle (nee) child +bicycle (ja/nee: wel children, niet bicycles)</p> <p>child bicycle helmet (ja: ook children, bicycles etc., helmets) child "bicycle helmet" (ja/nee: wel children, niet "bicycle helmets") child bicycle OR bike children bicycles</p> <p>rank company fortune Belarussian ranked company fortune Belarussian ranking company fortune Belarussian pet lemur dietary needs</p>	<p>geen automatic stemming bij één zoekterm</p> <p>wel automatic stemming bij meer zoektermen, behalve:</p> <ul style="list-style-type: none"> als ze worden voorafgegaan door een + bij phrases bij zoektermen die door OR met elkaar gecombineerd zijn <p>NB</p> <ul style="list-style-type: none"> werkt alleen in het Engels werkt bij zelfstandige naamwoorden bij gebruik van enkel- en meervouden; maar het enkelvoud bicycle vindt ook bicycling, het meervoud bicycles doet dat niet werkt bij werkwoorden bij diverse vervoegde vormen, maar met uiteenlopende resultaten <p>vindt documenten met deze combinatie van zoektermen, maar ook met pet lemur diet needs</p>
<i>Shortcuts</i>		
Woordenboek		
define: (alleen Engels!)	define:serendipity define:borderline personality disorder	omschrijvingen van de opgegeven term uit woordenboeken en dergelijke
Filmbesprekingen		
movie:	movie:weeping camel	recensies van de film "The Story of the Weeping Camel"
Nieuws		
(automatisch) Typ: [term]; bij een (combinatie van) zoekterm(en) die Google op dat moment in zijn Engelstalige nieuwsbronnen heeft zitten, verschijnen enkele van de daar gevonden documenten als eerste in de resultatenlijst	war terrorism Iran nuclear weapons	als eerste treffer enkele berichten uit de nieuwsbronnen van Google News
(Beurs) [Stock Quotes] Typ: stocks: [ticker]	stocks:GOOG 1)	als eerste treffer financiële informatie over de opgegeven ticker van resp. Yahoo! Finance, MSN MoneyCentral, Fool.com en ClearStation
Lokale informatie		
(Gouden Gids) [Local Search] Typ: [bedrijfstype] [plaats of postcode]	Bookstore 94501 hotel Newark beauty salon Berkeley	als eerste treffer informatie over winkels en bedrijven in de opgegeven plaats of met de opgegeven postcode; gebruik als alternatief: Google Local, http://local.google.com/

(Telefoongids) [Phone Book]; alleen VS, alleen white pages Typ: [voornaam of initiaal][achternaam][stad of staat of postcode of kengetal] [achternaam] [stad + staat, of postcode] [telefoonnummer met kengetal]	E. Tamm Alameda E. Tamm CA E. Tamm 94501 E. Tamm 510 Tamm Alameda CA Tamm 94501 5107481252 510 748 1252 (510) 748-1252	het telefoonnummer bij de opgegeven abonnee, of de abonnee bij het opgegeven telefoonnummer
(Kaarten) [Map]; alleen VS Typ: [nummer] [straat] [plaatsnaam] [staat] [kengetal]	165 University Ave Palo Alto CA 415 418	als eerste treffer een kaart van de omgeving van dit adres; gebruik als alternatief: Google Maps, http://maps.google.com/ (ook Canadees)
Reisinformatie		
(Flight Tracker); Amerikaanse binnenlandse vluchten Typ: [maatschappij] [vluchtnummer]	American Airlines 24	als (een van de)eerste treffer(s) informatie over de opgegeven vlucht
(Luchthaveninformatie) [Airport Information]; alleen VS Typ: [plaatsnaam of luchthavencode] airport	New Orleans airport JFK airport	als eerste treffer diverse info (kaartjes, weer, vluchtinfo etc.) over de betreffende luchthaven Voor luchthavencodes, zie: http://www.world-airport-codes.com/
Overige		
(Postpakkettenvolgers) [Tracking Numbers]; UPS, Fedex, USPS Typ: [volgcode]	1z9999999999999999 (UPS) fedex 7917 2567 0102 USPS 9101128882300016879545 (USPS)	als eerste treffer de status van het pakket met de opgegeven volgcode
(Voertuigidentificatie- nummer) [Vehicle ID Number (VIN)] Typ: [VIN]	Zffyt53a510122724	als eerste treffer informatie over de auto met de opgegeven VIN-code
(UPC code) [UPC Codes]; dit is een twaalfcijferige streepjescode op producten Typ: [UPC-code]	636920005506	als eerste treffer info over het artikel met de opgegeven code in de UPC database
(Amerikaanse octrooi-code) [Patent Numbers] Typ: [nummer]	patent 5123123	als eerste treffer informatie over het octrooi met het opgegeven nummer in de USPTO patent database

(Vliegtuigregistratienummer) [Aircraft Registration]; alleen in de VS geregistreerde nummers, beginnend met N Typ: [registratienummer toestel]	n158ua	als eerste treffer het record van het vliegtuig met het betreffende staartnummer in de FAA registry
[FCC Equipment ID's] Typ: [equipment ID]	fcc EJM386S303	(in de VS goedgekeurde communicatieapparatuur)
(Omrekenen van maten en gewichten) [Weights and Measures Converter] Typ: [aantal] [eenheden] in [eenheden]	6 pounds in kilograms 160 pounds * 4000 feet in Calories	de bekende maathoeveelheid omgerekend naar de gewenste maathoeveelheid; kan worden gecombineerd met de rekenmachine van Google
<i>Speciale zoekmachines</i>		
Google Groups		
<i>Prefixes</i>		
allintitle: allinsubject:	allintitle:tiramisu recept allintitle:tiramisu recipe group:rec.food.baking allinsubject:tiramisu recipe author:monahan allinsubject:bakkers zondag author:jole allinsubject:bakkers zondag author:"francisco van jole" allinsubject:bakkers zondag group:nl.net.politiek allinsubject:bakkers zondag group:nl.net*	berichten met de opgegeven zoektermen in de titel (het subject) van de pagina NB: allintitle: (allinsubject:) kan niet met zoektermen elders worden gecombineerd: het heeft betrekking op alle zoektermen in de zoekexpressie en dient daaraan vooraf te gaan; wel kunnen termen met de – operator in de expressie worden opgenomen en kan allintitle: onder bepaalde voorwaarden worden gecombineerd met group: en author: Die voorwaarden zijn: author: mag alleen worden gevolgd door één woord of phrase, dus wel author:jole of author:"francisco van jole", maar niet author:francisco van jole; group: mag alleen worden gevolgd door de volledige naam van de group of een rechts getrunceerde naam.
intitle: insubject:	intitle:tiramisu recept insubject:bakkers insubject:zondag insubject:bakkers group:*politiek	berichten met de zoekterm direct na het prefix in de titel en andere zoektermen elders; intitle: of insubject: kan met andere zoektermen en prefixen worden gecombineerd

allintext:	allintext:cemetery "Colleville sur Mer" veteran	berichten met de opgegeven zoektermen in de tekst van het bericht, en niet ergens anders (bijvoorbeeld Quoted text) NB: allintext: kan niet met zoektermen elders worden gecombineerd: het heeft betrekking op alle zoektermen in de zoekexpressie en dient daaraan vooraf te gaan; wel kunnen termen met de – operator in de expressie worden opgenomen en kan allintext: onder bepaalde voorwaarden worden gecombineerd met group: en author: Die voorwaarden zijn: author: mag worden gevolgd door voor- en/of achternaam of phrase, dus zowel author:jole (of author:francisco), als author:francisco van jole (author:”francisco van jole”); group: mag alleen worden gevolgd door de volledige naam van de group of een rechts getrunceerde naam.
intext:	intext:Volkskrant hetze Zalm	berichten met de zoekterm direct achter intext: in de tekst van het bericht (en niet bijvoorbeeld in de Quoted text) en de andere zoektermen waar dan ook
author: (komt overeen met de optie: 'Author: Return only messages where the author is' in de Advanced Search	author:jole author:francisco van jole	posts van de opgegeven auteur
group: (komt overeen met de optie: 'Group: Return only messages from the group' in de Advanced Search	group:rec.music group:rec.music* group:*music group:*music -group:rec.music* "Bob Dylan" group:*music -group:rec.music* group:nl* group:*linux Microsoft	berichten in de opgegeven groep of groepen (bij truncatie) NB: bij sommige treffers zult u zien dat daarbij een groep vermeld staat die niet aan het zoekcriterium voldoet; dat is schijn, het bericht is naar diverse nieuwsgroepen verstuurd, waarvan er een of meer wel aan het zoekcriterium voldoen, maar alleen de eerste nieuwsgroep in de adressering staat vermeld. Toon het bericht in de complete presentatie en klik op 'Show options' om alle nieuwsgroepen te zien waarnaar het bericht verstuurd is.
inlinktext:	inlinktext:beethoven inlinktext:nl inlinktext:spam inlinktext:nl inlinktext:linux "Windows XP"	berichten in een groep met de opgegeven term in de naam van de groep, eventueel in combinatie met andere termen ook in de naam en/of elders; komt overeen met group:*...*; kan niet worden gecombineerd met allinsubject: of allintext:
[msgid:] (komt overeen met de optie: 'Message ID: Lookup the message with message ID' in de Advanced Search	3B6AAE00.45A61B38@rug.ac.be	berichten met de opgegeven identificatiecode (u vindt die code bij de meeste berichten door in de complete presentatie van een bericht eerst op Show options te klikken, en daarna op Show original).
<i>Operatoren</i>		

term1 term2 (komt overeen met de optie 'with all the words' in Advanced Search)	tiramisu recept	berichten met de opgegeven zoektermen in een impliciete EN-relatie
+term (het +-teken dient bij Google Groups buiten phrases nog als unitaire operator; daarnaast ook om Google te dwingen standaardbewerkingen niet uit te voeren)	the way of the world +the way +of +the world “ the way of the world ” (= "+the way +of +the world") +électricité +e-merge	berichten met de opgegeven zoektermen, maar niet noodzakelijk de stopwoorden berichten met de opgegeven zoektermen, inclusief de 'stopwoorden' the en of berichten met de opgegeven phrase, inclusief de stopwoorden alleen berichten met de geaccentueerde term, geen berichten met alleen de ongeaccentueerde alleen berichten met de term met het koppelteken, geen berichten met alleen de term zonder koppelteken
-term (komt overeen met de optie 'without the words' in de Advanced Search)	Bach -"Johann Sebastian" recordings Beethoven -group:rec.music*	berichten waarin de uitgesloten zoekterm of phrase niet voorkomt
OR () (komt overeen met de optie 'with at least one of the words' in de Advanced Search)	preventie alzheimer OR dementie musical group:nl.kunst.theater group:nl.kunst.film	berichten met de ene of de andere zoekterm, of beide
<i>Nabijheid</i>		
Phrase ("...") (komt overeen met 'with the exact phrase' in de Advanced Search)	"op hete kolen zitten"	berichten met de opgegeven woordgroep
* (wildcard)	"de grootste stad * Europa"	berichten met de opgegeven phrase, waarin de wildcard * een willekeurig woord vertegenwoordigt
<i>Overige</i>		
.. messages ²⁾	10/20/30/50/100	kies via Advanced Search of Preferences het aantal treffers per pagina van de resultatenlijst
sorteren (kan in de Advanced Search , maar ook nog in het resultatscherm)	Sort by relevance/Sort by date	kies de sorteervolgorde
language ²⁾	Slovak	kies een van de 35 talen in het menu onder Any language in de Advanced search
message dates	anytime / past 24 hours / past week / past month /past 3 months / past 6 months / past year; of een periode naar keuze	documenten met als laatste wijzigingsdatum een datum binnen een van de opgegeven periodes; via Advanced Search
SafeSearch ²⁾	No Filtering / Filtering using SafeSearch	schakelt het Family Filter van Google uit / aan Standaard is Moderate filtering; gebruik de mogelijkheden van Preferences om die instelling te veranderen

1) Ticker Symbol Lookup: <http://cbs.marketwatch.com/tools/quotes/lookup.asp?siteid=mktw> ; Yahoo! Finance – Ticker Symbol Lookup: <http://finance.yahoo.com/lookup>

2) Alleen Advanced Search en/of via Preferences

4.2.2.2. Basic Search

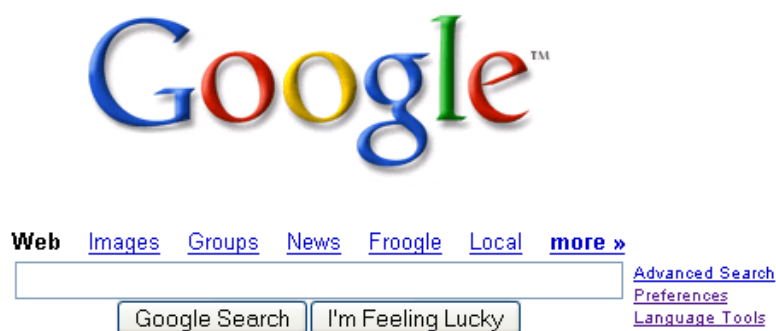
Google's Basic Search is in drie versies beschikbaar: Google Nederland (in een Nederlands- en een Engelstalige versie) die uw browser waarschijnlijk zal kiezen als u <http://www.google.com/> intypt, en de Amerikaanse (of internationale) versie die u krijgt als u daar <http://www.google.com/ncr> van maakt. Het voornaamste verschil tussen beide is dat de Nederlandse versie (<http://www.google.nl>) u de keus geeft tussen zoeken naar documenten in alle talen en zoeken naar documenten in het Nederlands, dan wel pagina's uit Nederland. (**NB:** dat is niet hetzelfde als zoeken in sites met de domeinnaam .nl)

Verder zijn de links naar andere Google-databases beperkt tot de voor ons land meest relevante: Afbeeldingen, Discussiegroepen en Gids. Aangezien de zoekvraag bewaard blijft als je die links aanklikt, is dat wel prettig: je kunt heel snel achter elkaar tussen de zoekresultaten voor de vier databases heen en weer gaan; de instelling 'in het Nederlands' blijft daarbij voor drie van de vier bewaard, alleen niet voor Afbeeldingen. Helaas is de keuze tussen 'Het Web doorzoeken' en 'Zoeken in pagina's in het Nederlands' in Discussiegroepen en Gids dan niet meer beschikbaar. Ten slotte verwijzen links naar Geavanceerd zoeken, Voorkeuren en Taalhulpmiddelen naar in het Nederlands vertaalde pagina's en zijn er nog wat kleinere verschillen. Ik bespreek hier verder de internationale versie, die ook bij de shortcuts de volledige functionaliteit biedt.

De opvallendste keuze in de Basic Search is die tussen de knoppen Google Search en I'm Feeling Lucky. Het laatste betekent zoveel als 'Ik waag het erop': Google geeft u één treffer, de beste die het de bieden heeft.

Uitleg over onder meer hoe u kunt zoeken vindt u onder About Google, en Preferences (Voorkeuren), rechts van de zoekregel, biedt u de mogelijkheid voorkeuren in te stellen die dan in een cookie bewaard blijven totdat u (of iemand anders die op uw computer werkt) ze weer verandert. Language Tools (Taalhulpmiddelen), eveneens rechts van de zoekregel, biedt land- en taalgerelateerde mogelijkheden.

In de zoekregel kunt u maximaal 32 strings typen, d.w.z. zoektermen, eventueel met een prefix ervoor, en limits zoals `site:hva.nl`. Ieder daarvan telt als één term. De joker *, die één woord in een phrase kan vervangen, telt daarbij niet mee; "Wilhelmus * * * * * bloed" geldt dus als twee termen. Als u meer dan 32 strings gebruikt, voert Google de zoekactie niet uit en meldt dat.



[Advertising Programs](#) - [Business Solutions](#) - [About Google](#) - [Go to Google Nederland](#)

©2006 Google

Google's Basic zoekscherm.

4.2.2.3. Language Tools (Taalhulpmiddelen)

Onder de [Language Tools](#) biedt Google vier interessante hulpmiddelen.

- In de eerste plaats vindt u hier de mogelijkheid pagina's in een taal en in een land te zoeken, desgewenst gecombineerd. Het is hier dus bij voorbeeld mogelijk naar Nederlandstalige pagina's in Singapore te zoeken, die niet noodzakelijk de domain name sg hebben (zie het voorbeeld in de tabellen). Let op: de mogelijkheid om naar pagina's in een land te zoeken wordt *alleen* onder Language Tools aangeboden (tenzij u de nationale versie van Google gebruikt). Het geeft aanzienlijk meer resultaten dan zoeken met de domain name (`site:`) nl. De mogelijkheid pagina's in een bepaalde taal te zoeken is ook in de Advanced Search beschikbaar en kan onder Preferences (Voorkeuren) als standaard worden ingesteld.
- In de tweede plaats kunt u Google stukken tekst (die u in een venster typt of plakt) in een andere taal laten vertalen, of de URL van een pagina opgeven waarmee dat moet gebeuren. Beschikbaar zijn de combinaties: Engels naar Duits, Spaans, Frans, Italiaans, Portugees, Japans, Koreaans en Chinees (de laatste drie in bèta) en omgekeerd, plus Duits-Frans en Frans-Duits.
- De derde mogelijkheid die Google hier heeft is de keuze van de interfacetaal. Deze is ook onder Preferences (Voorkeuren) beschikbaar.
- Ten slotte vindt u hier Google in 113 nationale versies. **NB:** bij veel daarvan betekent 'nationaal' dat u (naast de keuze: het hele Web doorzoeken) de keus krijgt tussen zoeken naar pagina's in het land en pagina's in de taal, net zoals in de eerste optie die hierboven werd genoemd. Aangezien de Advanced Search altijd zoeken in een taal mogelijk maakt is bij die nationale versies de combinatie land en taal ook mogelijk.

4.2.2.4. Preferences (Voorkeuren)

Onder [Preferences](#) kiest u achtereenvolgens:

- de interfacetaal, dus de taal waarin Google zich aan u presenteert
- de taal van de documenten waarin uw zoektermen moeten voorkomen; de keuze is hier alle talen (standaard) of een of meer – maximaal 10 – geselecteerde talen.
- SafeSearch Filtering strikt (standaard) / gematigd (alleen voor expliciete afbeeldingen) / geen filter; dit is Google's Family Filter
- aantal treffers per pagina; advies: stel dit in op minimaal twintig.
- of de resultaten waarop u klikt al dan niet in een nieuw venster moeten verschijnen.

4.2.2.5. Advanced Search

Zoals bij de meeste andere zoekmachines biedt Google's [Advanced Search](#) enkele mogelijkheden die via het gebruik van prefixen ook in de Basic Search kunnen worden benut, en andere mogelijkheden die alleen op dit scherm worden aangeboden. Het is goed te bedenken dat sommige mogelijkheden alleen in de Basic Search beschikbaar zijn; de Advanced Search is vooral bedoeld om het u gemakkelijker te maken.

Het Advanced Search-scherm is verdeeld in drie secties, elk voor een aparte manier van zoeken:

1. een sectie waarin u Google gebruikt om zoektermen en/of 'phrases' in te vullen en daartussen relaties aan te brengen, het deel van het document waarin de zoektermen worden gezocht te beperken, een type en/of de 'versheid' van het document te specificeren en/of het deel van het Web waarbinnen documenten worden gezocht te beperken. (Beperking van de ruimte waarbinnen u zoekt is altijd raadzaam.) Ook kan hier het standaard aantal van 10 treffers per resultatenpagina worden verhoogd.
2. een sectie waarin u op basis van een al bekend (relevant) document of het adres daarvan naar gelijksoortige documenten kunt zoeken
3. en een sectie waarin u speciale zoekmachines van Google over bepaalde onderwerpen en onderwerpsgebieden kunt gebruiken.

In de tabellen ga ik dieper in op de keuzes die Google op dit scherm biedt, maar hier gaat het mij erom duidelijk te maken dat het Advanced Search-scherm van Google een hulpmiddel is om de wat meer geavanceerde mogelijkheden van Google te ontdekken, maar dat het die mogelijkheden niet uitput. Met andere woorden, met Google is meer mogelijk dan het Advanced Search scherm suggereert.

Probeer eens een paar zoekacties vanuit het Advanced Search-scherm waarbij u diverse standaardinstellingen verandert, en u zult zien waar nog extra mogelijkheden liggen. Als u dat gedaan hebt en ten slotte op Enter hebt gedrukt, toont Google u een resultatenpagina met in het zoekvenster de vraag zoals Google die heeft uitgevoerd. Plotseling blijken daar niet alleen maar uw zoektermen meer in te staan, maar blijken die te worden voorafgegaan door woorden als `allintitle:` of `allinanchor:` (zogenoeten *prefixen* of *limits*) of tussen aanhalingstekens te staan. Het kan zijn dat een of meer van uw zoektermen worden voorafgegaan door een minteken (-), als gevolg van het feit dat u op het Advanced Search-scherm de zoekregel met de optie 'without the words' hebt ingevuld. En als u daar de zoekregel 'with at least one of the words' hebt ingevuld, zult u nu een of meer malen OR in het zoekvenster zien staan. Of in het zoekvenster staat `filetype:PDF`, omdat u in de Advanced Search 'any format' hebt veranderd in Adobe Acrobat PDF (.pdf). Hier volgt een voorbeeld van een in de Advanced Search gestelde vraag met een aantal gewijzigde instellingen:

The screenshot shows the Google Advanced Search interface. The search bar contains the text "science fiction" and the search button is labeled "Google Search". The search criteria are set to "with at least one of the words" and "without the words". The language is set to "French", file format to "any format", date to "anytime", occurrences to "in the title of the page", domain to "fr", and usage rights to "not filtered by license". SafeSearch is set to "Filter using SafeSearch".

De vraag in Advanced Search ...

The screenshot shows the Google search results page for the query "allintitle: âge OR période OR époque "science fiction" site:fr". The search bar shows the query and the search button is labeled "Search". The results are displayed as "Results 1 - 6 of 6 French pages for allintitle: âge OR période OR époque "science fiction" site:fr with Safesearch on."

... en het antwoord in Basic Search: `allintitle:âge OR période OR époque "science fiction" site:fr`

U ziet dat een groot deel van de veranderde standaardinstellingen in de syntax van de vraag terugkeert. In deze vraag is, behalve de SafeSearch-instelling, alleen de taalkeuze niet in de vraagstelling zelf vertegenwoordigd, maar in een apart keuzerondje: terwijl u

zelf ook in het Basic Search scherm de zoekopdrachten zo kunt formuleren zoals u ze in het resultatenscherm van een opdracht in de Advanced Search ziet verschijnen, kan dat niet voor de taalkeuze. Daar valt overigens wel een mouw aan te passen, want via Preferences (op het Basic Search scherm onder Advanced Search) kunt u uw taalkeuze anders instellen dan de standaardinstelling 'Search for pages written in any language'. Hetzelfde geldt voor de instelling 'Google SafeSearch is on', ook een melding die voortkomt uit de verandering van een standaardinstelling in de Advanced Search; per sessie stelt u die vanaf het Advanced Search scherm in, maar sessie-overschrijdend kan het ook via Preferences (Google plaatst dan een cookie op uw PC).

De enige instellingen in de eerste sectie van het Advanced Search scherm die u niet op een andere manier dan via dat scherm kunt veranderen zijn de optie: 'Date: Return web pages updated in the', waar u 'anytime' kunt veranderen in 'past 3 months', 'past 6 months' en 'past year'; en de optie Usage Rights, waar u 'not filtered by license' kunt veranderen in 'free to use or share', 'free to use or share, even commercially', 'free to use, share or modify' en 'free to use, share or modify, even commercially'.

De sectie 3 (Topic-Specific Searches) van het Advanced Search-scherm is bedoeld voor speciale doeleinden en blijft hier buiten beschouwing. Wel van belang is sectie 2 (Page-Specific Search), omdat die ons een van de beste zoekmogelijkheden geeft die het Web te bieden heeft: de mogelijkheid om bij een al bekend relevant document andere documenten over hetzelfde onderwerp te zoeken. Hier volgt een voorbeeld van een zoekactie die in deze sectie van het Advanced Search scherm is uitgevoerd, met daaronder het resultaat in het zoekvenster van de resultatenlijst.

Deze zoekactie begint met het zoeken naar een relevant document over zoekstrategieën. Hoewel Google dankzij zijn slimme rankingalgoritmes met een goed resultaat zou komen als we simpelweg `search strategies` in de Basic zoekregel zouden invullen, specificeren we nu zelf dat we onze zoektermen in een phrase willen hebben en voorafgegaan door `intitle:`, zodat we alleen treffers krijgen waarin de zoektermen als phrase in de titel van het document voorkomen. Verder willen we alleen documenten van educatieve instellingen, dus met `site:edu`; daarbij voegen we nog iets toe wat in de Advanced Search niet kan, namelijk dat de documenten ook afkomstig mogen zijn van academische instellingen in Groot-Brittannië:

[intitle:"search strategies" site:edu OR site:ac.uk](http://www.google.com/search?q=intitle:)

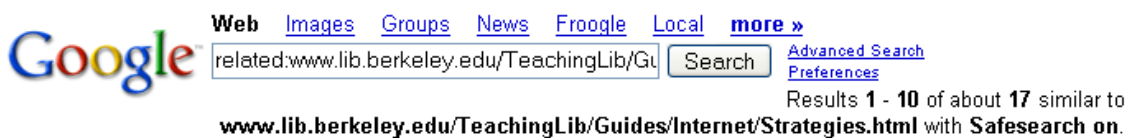


The screenshot shows the Google search interface. At the top, there are navigation links: Web, Images, Groups, News, Froogle, Local, and more. The search bar contains the query `intitle:"search strategies" site:edu OR site:ac.uk`. To the right of the search bar are links for 'Advanced Search' and 'Preferences'. Below the search bar, it says 'Results 1 - 100 of about 9,900 for intitle:"search strategies" site:edu OR site:ac.uk'.

Search Strategies: Search with Peripheral Vision - 2 visits - Apr 1
Finding Information on the Internet: A Tutorial <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html> ...
www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html - 31k -
[Cached](#) - [Similar pages](#) - [Remove result](#)

Achter de eerste treffer in de resultatenlijst ziet u een link met de tekst 'Similar pages' staan. Als u daarop klikt en vervolgens de resultatenlijst bekijkt vindt u daarin enkele tientallen documenten die ook over zoekstrategieën gaan maar niet noodzakelijk de phrase `search strategies` in de titel hebben. Maar u ziet (zie onder) ook dat in het zoekvenster de phrase `"search strategies"` wordt voorafgegaan door *related*: Hetzelfde resultaat zou u hebben gekregen als u in de URL van de hierboven genoemde eerste treffer (www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html) uit

de resultatenlijst zou hebben geknipt en in het Advanced Search scherm onder Page Specific Search in de regel **Similar** *Find pages similar to the page* zou hebben geplakt. Net zo kunnen we door de URL www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html in de regel **Links** *Find pages that link to the page* te plakken en daarmee te zoeken webpagina's vinden die naar de eerste treffer in ons oorspronkelijke zoekresultaat linken: ook dit is een manier om, uitgaande van een al bekend document, andere overeenkomstige documenten te vinden. Een nog wat verder uitgewerkt voorbeeld van dit principe, met de vraag om pagina's die naar twee al bekende pagina's linken, vindt u bij het prefix `link:` in de tabellen.



[Search Strategies: Search with Peripheral Vision](#)

Finding Information on the Internet: A Tutorial <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html> ...
www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html - 31k - [Cached](#) - [Similar pages](#)

[Types of Search Tools](#)

Tables of the BEST search engines and subject directories, as part of a comprehensive and flexible purpose for research using the Web.
www.lib.berkeley.edu/TeachingLib/Guides/Internet/ToolsTables.html - 14k - [Cached](#) - [Similar pages](#)

[Search Engine Showdown: The Users' Guide to Web Searching](#)

A user's guide to Web searching with features, comparative analysis, strategies, discussion, news and reviews.
www.searchengineshowdown.com/ - 12k - [Cached](#) - [Similar pages](#)

[How to Choose a Search Engine or Directory](#)

University librarian Laura Cohen gives advice in the format of "if you want A then choose B."
library.albany.edu/internet/choose.html - 26k - [Cached](#) - [Similar pages](#)

[The University of South Carolina Beaufort](#)

Seventeen lesson course on the various aspects of research in the library and on the web.
www.sc.edu/beaufort/library/pages/bones/bones.shtml - 21k - [Cached](#) - [Similar pages](#)

Het resultaat van een zoekactie met Similar pages (zie vorige bladzijde).

4.2.2.6. Google's resultatenlijst

Google's resultatenlijst begint met de adresregel van uw browser. Daarin staan namelijk weergegeven:

- de vraag zoals die gesteld is
- de voornaamste instellingen die daarbij gehanteerd zijn en
- het punt waarop u zich in de resultatenlijst bevindt.

Een aantal simpele wijzigingen in deze regel kunnen al deze elementen veranderen en zo een volkomen ander resultaat opleveren.

Bekijk de volgende URL eens:

<http://www.google.com/search?q=%22Janet+Jackson%22+%22breast+flash%22++fine&num=20&hl=en&lr=&newwindow=1&c2coff=1&safe=off&start=20&sa=N>

Zoals u al ziet gaat deze URL over de gebeurtenis die een paar jaar geleden heel Amerika schokte en die in een korter en krachtiger vraagstelling had kunnen worden

samengevat als 'nipplegate': hoe hoog was de boete die het Amerikaanse televisiestation CBS heeft moeten neertellen voor Janet Jacksons 'breast flash' tijdens haar pauzeoptreden met Justin Timberlake in Hollywoods Superbowl.

De vraag begint met:

```
?q=%22Janet+Jackson%22+%22breast+flash%22++fine
```

De %22 staan voor aanhalingstekens die 'Janet Jackson' en 'breast flash' tot phrases maken; haal je ze weg dan stijgt het aantal treffers bij deze vraag van 1120 naar 37.900 (en daar zitten best nog relevante bij, bijvoorbeeld met 'breast' en 'Janet Jackson flash' of 'halftime flash')

Het +-teken staat voor een spatie, zoals uit de ++ voor 'fine': hier is per abuis een spatie te veel in de zoekopdracht getypt, maar Google maakt daar geen punt van.

Dan volgt:

```
&num=20&hl=en&lr=&newwindow=1&c2coff=1&safe=off&start=20&sa=N
```

Het & staat voor 'en': het verbindt de diverse elementen in de URL met elkaar.

'num=20' geeft aan dat de instelling die we onder Preferences (Voorkeuren) hebben vastgelegd of in de Advanced Search hebben gekozen, 20 treffers per pagina is. Dit aantal kan in de URL op elk aantal tussen 1 en 100 worden ingesteld.

'hl=en' geeft de interfacetaalinstelling (Interface Language) weer die we onder Preferences hebben ingesteld. Door deze te veranderen in 'hl=nl' krijgen we het scherm dezelfde zoekresultaten met het Nederlands als interfacetaal te zien.

'newwindow=1' geeft aan dat we er onder de Preferences voor hebben gekozen Google een nieuw venster te laten openen als we op een treffer klikken. Door de waarde 1 in 0 te veranderen maken we die keuze ongedaan

Wat 'c2coff=1' betekent en voor effect heeft, weet ik niet. Verandering van 1 in 0 heeft geen waarneembaar effect.

Daarentegen heeft de verandering van 'safe=off' in 'safe=on' het effect dat het aantal treffers van 1120 tot 679 wordt gereduceerd. Hiermee wordt Safesearch, Google's Family Filter aan en uit gezet.

'start is 20' geeft aan dat de resultatenlijst met treffer no. 20 begint. Als u 'num=20' hebt veranderd in 'num=100', en 'safe=off' in 'safe=on', dan kunt u de laatste treffers van deze zoekactie bekijken door 'start=20' te veranderen in 'start=600'.

Daarbij zal dan blijken dat Google in eerste instantie lang niet alle treffers toont, want in werkelijkheid krijgt u niet de treffers 601-679 te zien, maar 201-244. Onderaan deze laatste pagina met zoekresultaten ziet u de mededeling dat Google een aantal zoekresultaten heeft weggelaten en de uitnodiging de zoekactie met de weggelaten resultaten te herhalen. Door op de onderstreepte link te klikken laat u Google op zoek gaan naar een ongeclusterd resultaat, waarbij niet, zoals in eerste instantie, maar twee pagina's per site worden getoond die aan de zoekcriteria voldoen, maar alle.

Ditzelfde effect kunt u ook bereiken door achter de URL zoals die nu is de tekst '&filter=0' te plakken. Stuurt u nu de vraag opnieuw op, dan toont Google de treffers 601-634.

De aantallen treffers dat Google meldt zijn een schatting, en hoe hoger die schatting is, des te verder zit ze ernaast.

Daarnaast doet zich nog een onverklaarbaar verschijnsel voor. Normaal gesproken geeft het aantal treffers dat Google op de laatste resultatenpagina meldt, het werkelijke aantal te bekijken treffers weer (in het hierboven genoemde voorbeeld 634). Maar er zijn ook legio gevallen waarin Google ook op de laatste resultatenpagina een veel hoger aantal

treffers blijft melden dan in werkelijkheid bekeken kan worden. Dat leidt ertoe dat Google bijvoorbeeld bij zoekacties in Open WorldCat (site:worldcatlibraries.org) in sommige gevallen een hoger aantal treffers blijft melden dan Yahoo!, maar er in werkelijkheid aanzienlijk minder kan tonen. Wat in sommige gevallen dan wel eens wil helpen, is de pagina opnieuw te laden (vernieuwen, verversen), of de zoekactie later nog eens uit te voeren.

Aan de aantallen treffers die Google meldt mag dan ook geen enkele conclusie worden verbonden. Het enige houvast voor het werkelijke aantal treffers is, praktisch gesproken, het nummer van de laatste treffer die de zoekmachine toont. Een extreem voorbeeld van het grillige verloop van de gemelde aantallen is biedt de zoekvraag: [woman liberation sunglasses earrings](#), uitgevoerd op 17 maart 2005:

Opgegeven aantal op eerste resultatenpagina	134
Opgegeven aantal op volgende resultatenpagina's	134
Aantal aanvankelijk getoond	61
Opgegeven aantal op eerste pagina na aanklikken weggelaten zoekresultaten (filter=0)	3390
Opgegeven aantal op laatste pagina na aanklikken weggelaten zoekresultaten	3390
Werkelijke aantal getoond	983

De datum waarop deze zoekactie is uitgevoerd, staat hierboven niet voor niets vermeld: als u deze zoekactie zelf uitvoert, zult u merken dat Google weer met heel andere aantallen treffers komt (18 september 2005 met filter=0: 36.900!; 23 maart 2006, idem: 43.900).

De eigenlijke resultatenlijst opent met de zoekregel met daarin de vraag zoals u die in de Basic Search of via de Advanced Search gesteld hebt. Als SafeSearch aan is en/of de interactie het Nederlands is, krijgt u direct onder de zoekregel daarvan een melding en de keuze tussen het hele Web doorzoeken of alleen pagina's in het Nederlands (dit laatste is niet hetzelfde als zoeken met site:nl!). Direct daaronder staat de resultatenbalk, waarin Google links de doorzochte database vermeldt (het Internet, Afbeeldingen etc.; als u op dit punt wilt zien waar 'nipplegate' alweer over ging, klik dan op de link Images boven de zoekregel), en rechts de resultatenscore met de nummers van de getoonde treffers, de zoektermen van de vraag en de tijd die Google aan de vraag besteed heeft. Als u de Engelstalige Google gebruikt, zijn de (meeste van die) zoektermen aanklikbaar; daaronder vindt u een definitie van de betreffende term of uitleg over een persoon (in dit geval Janet Jackson) van [Answers.com](#). In wezen gaat het hier een schat aan compacte informatie die u nuttige context kan leveren voor het vervolg van uw zoekactie en dit in sommige gevallen zelfs overbodig kan maken. (Helaas verschijnt die definitie of uitleg in Google niet in alle gevallen waarin Answers.com wel degelijk antwoorden blijkt te hebben.)

Als Google de zoekterm in zijn nieuwsdatabase of als shortcut heeft gevonden, volgen nu eerst enkele links daarnaar, anders volgen links de reguliere treffers ('natural results' of 'algorithmic results'), en rechts de 'sponsored results'; bij de laatste valt op dat Google 'ziet' waar u vandaan komt, want deze 'sponsored results', in feite advertenties die zijn aangepast aan uw vraag, komen uit het land waar u op dat moment op de computer werkt.

[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#) [Local](#) [more »](#)

New Orleans airport

Search

[Advanced Search](#)
[Preferences](#)**Web**Results **1 - 100** of about **4,780,000** for **New Orleans airport**. (0.10 seconds)News results for **New Orleans airport** - [View today's top stories](#)[Exhausted evacuees pack New Orleans airport](#) - CNN - 19 hours ago[Horrible scenes at New Orleans airport](#) - MSNBC - 22 hours ago[New Orleans airport closed to tourists](#) - CNN International - 31 Aug 2005[View conditions at Louis Armstrong New Orleans International Airport \(MSY\), New Orleans, Louisiana](#)www.fly.faa.govLocal results for **airport** near **New Orleans, LA**[Louis Armstrong New Orleans International Airport: Downtown Public Use Heliport](#) - 0.3 miles SW - 1500 Poydras St, New Orleans, 70130[Airport Shuttle Inc](#) - 1.4 miles W - 4220 Howard Ave, New Orleans, 70125 - (504) 522-3500[Limousine Livery Ltd](#) - 1.6 miles S - 624 Race St, New Orleans, 70130 - (504) 561-8777[Louis Armstrong New Orleans International Airport](#)

Information on the facilities, cargo, downtown heliport, employment, flights, general aviation, ground transportation, parking and security.

www.flymsy.com/ - 2k - 1 Sep 2005 - [Cached](#) - [Similar pages](#)[Real-Time Flight Information](#)www.flymsy.com/html/default.html - 2k - [Cached](#) - [Similar pages](#)[[More results from www.flymsy.com](#)]

Sponsored Links

[New Orleans Katrina Site](#)Resources, People Search, Donate Temp Housing / Jobs & Volunteer www.craigslist.org[Chaos in New Orleans](#)All the latest news on Katrina's Aftermath from a Unique perspective RealTruthMag.org/NewOrleans

Een resultaatenschermbalk van Google: bovenin de zoekregel met de vraag, daaronder de resultatenbalk, met de doorzochte database, het aantal treffers, de links naar Answers.com en de tijd die de vraag gekost heeft. Daaronder links drie verschillende shortcuts: uit de Google News-database, luchthaveninformatie en lokale resultaten. Rechts enkele 'sponsored results', en onder de shortcuts aan de linkerkant twee 'natural results'. Google toont maximaal twee treffers per site ('automatic clustering'), met de tweede ingesprongen.

Bij de 'natural results' geeft Google ook een plaats aan treffers die in zijn (ODP-) directory zijn opgenomen. Dat is zeker het geval als een of meer zoektermen in het directorypad van een categorie voorkomen. Bij zulke treffers toont Google die categorie met een link en is het verstandig daar ook eens in te kijken. In deze gevallen toont Google ook niet de snippet, maar de beschrijving die de ODP-redacteur van de site heeft gegeven.

De eerste regel van een treffer geeft de titel van het document weer, of de eerste kop van een pagina of de meest relevante betiteling die Google voor het document heeft kunnen vinden; minimaal toont Google de URL. Deze eerste regel van een treffer is altijd gelinkt naar het achterliggende document. Zoektermen die Google in de titel heeft gevonden zijn daarin vet gemarkeerd, evenals in de snippets onder de titel. Bij Google bestaan die snippets uit (een aantal van) de zoektermen plus wat context, maar de lengte ervan is beperkt, dus als er veel zoektermen zijn, zult u (meestal) alleen de paar eerste te zien krijgen; naast de 'term weighting' die Google op basis van de volgorde van de zoektermen toepast een reden te meer om op die volgorde te letten. Overigens toont Google ook wel eens een later opgegeven zoekterm als die toevallig in een stukje snippet voorkomt dat een andere zoekterm bevat.

Onder de snippets volgt dan nog een regel met de URL van de treffer, de grootte van het document – ook geen onbelangrijke informatie – een link naar de cacheversie en een naar 'Similar pages' (die hetzelfde resultaat geeft als de opdracht related:[URL]). Standaard toont Google maximaal twee treffers per site. Heeft een site meer dan twee treffers, dan eindigt de tweede, ingesprongen treffer met een link waarin de zoeker wordt uitgenodigd meer resultaten van de betreffende site te bekijken. Als alternatief kunt u op de laatste resultatenpagina gebruik maken van de mogelijkheid Google de zoekactie te herhalen inclusief de weggelaten zoekresultaten. Nog sneller is de hierboven geschetste methode: voeg de string '&filter=0' aan de URL van een resultatenpagina toe.

Tussen de titel van een treffer kan nog een regel staan, met daarin het File Format, als dat geen HTML is, en in veel gevallen ook een uitnodiging het document als HTML te bekijken. Voor een snelle check valt dit aan te bevelen boven het volgen van de link onder de titel, die naar het document in zijn oorspronkelijke formaat (.pdf, .doc etc.) maar voor het tonen waarvan het bijbehorende programma aanwezig moet zijn en moet worden gestart.

Onderin de resultatenpagina vindt u Googles zoekregel herhaald, met de vraag erin, zodat u van hieruit een nieuwe zoekactie kunt starten of de zoekexpressie kunt veranderen. Daaronder staan nog enkele links, waarvan een u in staat stelt binnen de resultaten van uw huidige zoekactie te zoeken. Een klik op de link brengt u naar een Basic Search-scherm waarop de vraag, zoals Google die heeft opgevat, en het aantal treffers vermeld staan. In het (lege) zoekvenster kunt u zoektermen opgeven die de bestaande zoekactie verder preciseren.

4.2.2.7. Bespreking Google

Wie het hele scala van eigenschappen en mogelijkheden van Google overziet en dat met dat van Google's concurrenten vergelijkt, kan moeilijk tot een andere conclusie komen dan dat deze zoekmachine nog altijd een voorsprong op de concurrentie heeft. Dat neemt niet weg dat ook Google zijn beperkingen heeft en dingen niet kan of doet die in potentie wel zouden kunnen of die door andere zoekmachines wel worden gedaan.

In het hoofdstuk 'Het hoe van het zoeken' heb ik uiteengezet wat zoekmachines in het algemeen allemaal kunnen. Laat ik deze bespreking daarom beginnen met wat Google allemaal niet kan en doet.

1. Google indexeert de meeste van de documenten die zijn spider van het Web ophaalt nog altijd maar tot een bepaalde limiet, nl. 101K. (HTML; 120K PDF), dat is ongeveer 32 A4-tjes. De spider haalt wel het hele document op, maar omdat maar een deel daarvan geïndexeerd wordt, zult u dat document niet meer vinden als u (een combinatie van) zoektermen gebruikt die alleen in het 33e of latere A4-tje van het document voorkomen. Met de phrase "Gejuil doorkruist de lucht; de ellenden zijn vergeten" zult u Tollens' 'De overwintering der Hollanders op Nova Zembla' dus met [Google](#) niet vinden, en met [Yahoo!](#) (HTML tot 500K) en [MSN Search](#) (tot ca. 1 Mb) (doorgaans) wel. Dit is een van de redenen om naast Google ook andere zoekmachines of metazoekmachines te blijven gebruiken.

[NB: sommige documenten die door Google's spider zijn opgehaald, vertonen een veel hogere indexeerlimiet: tot ca. 500K. Het advies naast Google ook andere zoekmachines te gebruiken blijft onverminderd van kracht.]

2. Google biedt geen mogelijkheden om zoektermen te trunceren. Hoewel de andere grote mondiale zoekmachines die mogelijkheden ook niet hebben, zijn er kleinere, zoals Exalead, die ze wel bieden. Exalead ondersteunt niet alleen * als teken om rechts te trunceren maar ook 'regular expressions' waarbij bijvoorbeeld de punt (.) voor elke letter waar dan ook in het woord kan staan.

3. Google biedt beperkte mogelijkheden tot nabijheidszoeken. Het teken * kan binnen een phrase een of meer malen als joker worden gebruikt, maar staat dan voor één woord (maar zie de tabel!); voor één woord of geen zou handiger zijn geweest: Exalead heeft een NEAR operator die de afstand tussen twee zoektermen in een document beperkt tot maximaal 16 tussenliggende woorden.
4. Google categoriseert zijn resultaten niet, wat tot gevolg heeft dat zijn gebruikers altijd enige inventiviteit aan de dag zullen moeten leggen om context aan hun centrale zoekterm te geven. Een zoekactie naar 'apple' levert een resultaat op waarin de vrucht een bescheiden plaats inneemt. Andere zoekmachines, zoals Ask en de kleinere WiseNut en Exalead, en speciaal ook metazoekmachines zoals Vivísimo, bieden de gebruiker meer hulp. Exalead biedt de gebruiker bij het zoekresultaat related terms, related categories en een geografisch filter tot op plaatsniveau en een filter voor het documenttype (PDF, TXT, DOC, PPT).
5. Google biedt niet de mogelijkheid de volledige syntax van zijn zoekmachine in een commandogestuurde zoekvraag te gebruiken. Met name de mogelijkheid tot een taalkeuze en de keuze het aantal treffers per pagina in te stellen ontbreken in de commandotaal. MSN Search kent de opdracht `language:` en biedt ook verder mogelijkheden om de sorteeropties (freshness, popularity en relevantie) in de opdracht te verwerken.
6. Google ondersteunt niet alle mogelijkheden van Booleaans zoeken. Wie met Google op zoek wil naar een benedenwoning in Amsterdam of een flat in Buitenveldert, heeft een probleem: Bij de vraag

"te koop" OR aangeboden benedenwoning Amsterdam OR flat
buitenveldert

gaat Google voor u op zoek naar documenten die benedenwoning EN (Amsterdam OF flat) EN Buitenveldert bevatten.

"te koop" OR aangeboden "benedenwoning in Amsterdam" OR "flat in
buitenveldert"

kan wel, want je kunt met Google ook phrases in een OR-relatie verbinden, maar dat levert een heel mager resultaat op, omdat daarmee alle documenten worden gemist die tussen de woning en de plaatsnaam nog een adres hebben staan. Yahoo! maakt zo'n vraag wel mogelijk:

("te koop" OR aangeboden) AND ((benedenwoning AND Amsterdam) OR
(flat AND buitenveldert))

7. Google stelt een limiet van 32 aan het aantal zoektermen waarmee het zoekt. Dat stelt onder meer beperkingen aan de mogelijkheid om Google als zoekmachine te gebruiken die een aantal zelf geselecteerde sites doorzoekt.
8. Google toont nooit meer dan 1000 resultaten; voor de meeste huis-, tuin- en keukenzoekacties is dat ruimschoots genoeg, maar voor speciaal onderzoek, bijv. naar de verspreidingsnelheid van een neologisme, is het een hinderpaal. Exalead toont er meer: maximaal 2000.
9. Google laat weliswaar zien welke pagina's naar andere pagina's linken, maar niet naar welke site ze dat doen. Het prefix `link:` werkt alleen op de opgegeven pagina, niet op een hele site. Yahoo! heeft naast `link:` ook `linkdomain:` dat de resultaten voor een hele site geeft.
10. `link:` kan niet met andere zoektermen, prefixen of met een tweede `link:` worden

gecombineerd. Daardoor is het bijvoorbeeld niet mogelijk door de combinatie <link:www.hva.nl -site:hva.nl> alle pagina's van buiten de Hogeschool van Amsterdam te zoeken die naar de homepage van die hogeschool linken. Iets anders dat niet kan is pagina's opsporen die naar twee sites linken (met <link:www.leidenuniv.nl> <link:www.uva.nl>) om een pagina te vinden die naar zowel de site van de Leidse als die van de Amsterdamse universiteit) linkt. Bij [Yahoo!](#) en [MSN Search](#) kan dit wel.

11. Google sorteert zijn zoekresultaten maar op één manier: op een combinatie van relevantie en link popularity. Andere mogelijkheden zouden zijn op omvang, actualiteit, geografische locatie of documenttype. Exalead biedt die laatste twee mogelijkheden wel, MSN laat je kiezen tussen freshness (actualiteit), popularity en relevantie. (Onder relevantie wordt hier verstaan: de mate waarin, volgens de algoritmen van de zoekmachine, een document voldoet aan de zoekvraag.).

12. Google biedt veel minder mogelijkheden dan Yahoo! (features) om webpagina's met speciale kenmerken of verwijzingen naar speciale formaten te zoeken. Zo levert de zoekvraag ["Mila Georgieva"](Mila Georgieva) <feature:image> in Yahoo!, naast veel ruis, nog ettelijke afbeeldingen meer van deze schitterende Bulgaarse violiste op dan de beeldzoekmachines van Yahoo! zelf en Google kunnen leveren.

En dan laat ik allerlei extra's waarmee experimentele zoekmachines zoals [KartOO](#) aan de weg timmeren, nog buiten beschouwing.

Daar staat echter veel tegenover dat Google in positieve zin van zijn naaste concurrenten onderscheidt.

1. Ook al is de concurrentie van Yahoo! en MSN Search de afgelopen twee jaar toegenomen, toch heeft Google vermoedelijk nog altijd de grootste database van webdocumenten en links naar webdocumenten. Dit betekent dat Google treffers vindt die geen enkele andere grote mondiale zoekmachine in zijn database heeft. Het omgekeerde is trouwens ook waar: Google meldt 3 documenten te vinden op de vraag <school site:cultuurwijs.kennisnet.nl>; MSN Search vindt er 18. En wie de eigen zoekmachine van de site gebruikt, schrikt wel even van de 2300 resultaten die deze vindt en die kennelijk door geen enkele grote mondiale zoekmachine gevonden worden.

NB: ook voor deze resultaten geldt dat ze een momentopname zijn. Als u de zoekactie naspeelt, krijgt u zeker andere resultaten!

2. Google ontsluit ook documenten in meer talen (35) en meer bestandstypen (13) dan de andere grote mondiale zoekmachines.

3. De zoektaal van Google is erg veelzijdig, veelzijdiger dan dat van zijn concurrenten. Speciaal MSN Search, met zijn (naar eigen zeggen) op een na grootste database, kan hier niet aan tippen. Speciale vermelding verdienen Google's operator ~ waarmee 'synoniemen' kunnen worden gevonden, de joker * voor een willekeurig woord in een phrase, en de prefixen inanchor: en numrange:

Welke synoniemen vindt Google met ~? Hier een paar voorbeelden, verkregen door eerst de treffers met de term zelf en daarna steeds treffers met een 'synoniem' uit te sluiten.

~**help:** support manual "how to" FAQ guide problems troubleshooting helping tips tutorial tutorials helper

~**review:** summary reviewer rating movie reviews ratings report

~**anorexia:** "eating disorder" bulimia anorexic bulimic

~**journal:** association review magazine research times news newspaper journalism

~**love:** romantic compatibility marriage lovers lovely loving

~**romantic:** love wedding luxury romance sensual romántic romàntic

Uit de laatste twee voorbeelden blijkt dat het hier niet om vaste, maar vermoedelijk om

op statistische co-occurrence gebaseerde clusters van synoniemen gaat. Aardig is verder dat ~romantic ook de Spaanse schrijfwijzen romántic en romàntic vindt, maar niet als de ~-operator wordt gebruikt in combinatie met de +, die geaccentueerde vormen uitsluit (~+romantic).

Lang niet alle 'synoniemen' zijn nuttig, maar er zitten er beslist nuttige bij.

De ~-operator werkt ook voor andere talen, zoals Nederlands en Spaans, waarbij Google bij de vraag naar synoniemen voor vrouwen, resp. mannen seksistische trekjes begint te vertonen:

~**huis**: woning makelaar

~**koop**: winkel huur prijs kopen

~**mujeres**: latinas chicas sexo hombres

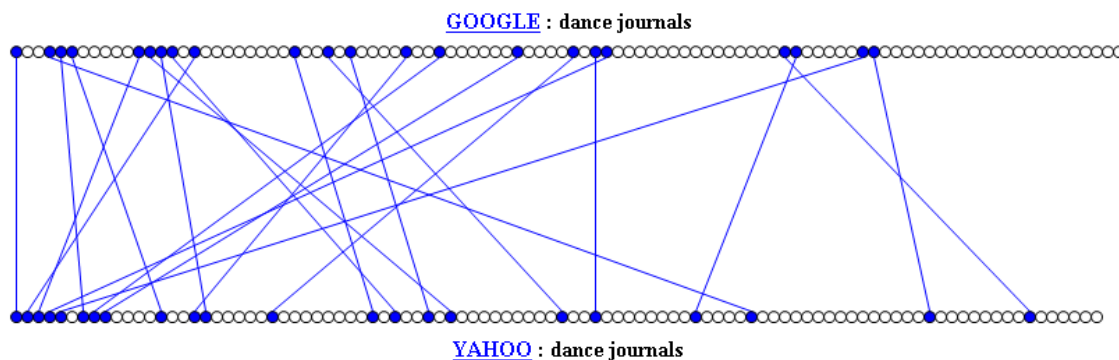
~**hombres**: chicos mujeres jovenes

4. Google gaat met een aantal zaken preciezer om dan zijn concurrenten, met name waar het om accenten, getallen en lees- en andere tekens gaat.

5. Aan het eind van een zoekactie komt een zoekmachine met een lijst van treffers, en bij de meeste zoekmachines is de lengte van de eerste pagina van die lijst standaard op 10 ingesteld. De meesten van ons zijn er inmiddels aan gewend geraakt te verwachten dat het antwoord op onze vraag al in een of meer documenten bij die eerste tien te vinden moet zijn, en dat die verwachting heeft kunnen ontstaan is in hoge mate aan het rankingmechanisme van Google te danken – en dan met name aan PageRank, Google's methode om de linkrelaties tussen documenten op het Web in de ranking een rol te laten spelen.

Veel fans van Google menen nog altijd dat Google op de meeste vragen veel relevantere resultaten levert dan andere zoekmachines. Of dat wel of niet zo is, valt niet te zeggen als er niet bij wordt gezegd wat men precies onder 'relevantie' verstaat. Uit diverse onderzoeken is gebleken dat Google weliswaar in de beleving van gebruikers constant hoog scoort wat de relevantie van zijn eerste resultaten betreft, maar dat zoekmachines zoals Yahoo! en Exalead tegenwoordig niet lager meer, en in sommige onderzoeken zelfs hoger scoren.

Maar een objectief oordeel blijft moeilijk, omdat de onderzoeksmethodes verschillen en beleving een subjectief gegeven is. Eigenlijk zou zo'n onderzoek zich op harde feiten zoals de plaats en frequentie van zoektermen (of daaraan verwante termen) in de gevonden webpagina's moeten richten, maar ook daarbij spelen subjectieve oordelen – wat is het gewicht van een zoekterm in een document als het daar in een kop staat vergeleken met het gewicht van diezelfde term als die als vet is aangegeven – een rol. Per slot van rekening zijn het precies die subjectieve oordelen die de verschillen tussen de rankingalgoritmen van de grote zoekmachines bepalen.



De eerste honderd treffers van Google en Yahoo! voor de vraag dance journals met elkaar vergeleken. Voor deze vraag hebben ze 22 treffers gemeen, waarvan drie bij de eerste tien.

Wat dit onderwerp betreft volsta ik ermee te zeggen dat Google naar mijn indruk bij zijn eerste 10 à 30 treffers (afhankelijk van de vraag) minder irrelevante treffers toont dan andere zoekmachines (speciaal MSN Search), maar ik voeg daar direct aan toe dat die indruk grotendeels gebaseerd is op de samenvatting (summary of snippets) die de zoekmachines in hun resultatenlijsten tonen, niet op inspectie van de documenten zelf. Wel is zeker dat tussen samenvatting en brondocument vaak aanzienlijke verschillen bestaan.

Voor het overige verwijs ik naar onderstaande literatuur.

Evaluation of three German search engines: Altavista.de, Google.de and Lycos.de

(<http://informationr.net/ir/9-4/paper189.html>)

Inktomi® Corp.: Web Search Relevance Test

(http://www.veritest.com/clients/reports/inktomi/inktomi_Web_search_test.pdf)

Study of the pertinence of the search engine AOL France versus Google

(http://www.veritest.com/clients/reports/aolfrance/aol_france_english.pdf)

In Search Of The Relevancy Figure (<http://searchenginewatch.com/sereport/article.php/2165151>)

Study questions Google's long-term dominance (http://news.com.com/2100-1038_3-5220295.html)

What is relevancy (<http://forums.searchenginewatch.com/showthread.php?t=3606>)

RustySearch Results – The Search Engine Relevancy Challenge at RustyBrick

(<http://www.rustybrick.com/rustysearch-results.php>)

J. Véronis, A comparative study of six search engines

(<http://www.up.univ-mrs.fr/veronis/pdf/2006-comparative-study.pdf>)

6. De informatie die de resultatenlijst verschaft. De omvang van de database en de relevantie van het zoekresultaat mogen belangrijke criteria zijn waarop gebruikers zoekmachines expliciet beoordelen, impliciet spelen ook de layout en de presentatie van de resultatenlijsten een rol, niet alleen in de perceptie van gebruikers maar ook in de snelheid waarmee ze in het zoekresultaat het antwoord op hun vraag weten te vinden. Een van de factoren die hier in het voordeel van Google spreekt is de duidelijke scheiding die de zoekmachine aanbrengt tussen de gewone treffers ('natural', 'algorithmic' of 'organic results') en de zogeheten 'sponsored results', die alleen in het zoekresultaat verschijnen omdat de zoekmachine ervoor betaald krijgt als u daarop klikt. Het is een bekend feit dat computergebruikers webpagina's volgens bepaalde vaste patronen doorlopen (eerder 'scannen' dan lezen) om de informatie te vinden waarnaar ze op zoek zijn, en de pagina's met treffers die zoekmachines tonen vormen op die regel geen uitzondering. Het is daarom van belang dat het onderscheid tussen 'natural results' en 'sponsored results' in een trefferlijst zo duidelijk mogelijk is. Van de vier grote mondiale zoekmachines is Google degene die het duidelijkste onderscheid maakt: door alle sponsored results in een aparte tabel aan de rechterkant van de pagina te plaatsen. Ask zet ze bovenaan, maar tegen een gekleurde achtergrond en onder het kopje 'Sponsored Results', en Yahoo! en MSN Search hebben boven, rechts van en onder de 'natural results' sponsored links, ten dele tegen een wat afwijkende kleur achtergrond. Daartegenover staat dat Google (net als Yahoo!) ook naar zijn winkelzoekmachine Froogle verwijst, maar die verwijzing is duidelijk onderscheiden van de eigenlijke zoekresultaten.

De summaries of snippets van de treffers bij de diverse zoekmachines verschillen niet erg in kwaliteit, maar er zijn wel detailverschillen. Zo laten Google en MSN Search de zoektermen altijd vet afgedrukt in het zoekresultaat zien, en doen Yahoo! en Ask dat niet altijd; als de zoektermen in de URL van treffers voorkomen, laten Google en Yahoo! (vaak) dat zien, MSN Search en Ask niet. Google, MSN Search en Ask vermelden bij de cacheversie van een pagina wanneer die gespiderd is, Yahoo! doet dat niet; en Google is de enige zoekmachine die je de cachepagina als 'text only' laat zien. (Als u de normale cachepagina bekijkt is de tekst weliswaar die van de cachekopie, maar komen de afbeeldingen van de server van de oorspronkelijke pagina; door de text only versie van de cachekopie te bekijken vermijdt u contact met de server die de afbeeldingen levert.)

Wat Google voorheeft op alle andere grote mondiale zoekmachines, is dat het in zijn resultatenlijst de eerste kop als titel van een webpagina toont als die pagina geen titeltag heeft.

Kortom, de indruk die de zoekfunctie van Google geeft, dat over de details bij Google net iets beter is nagedacht dan bij de concurrentie, wordt ook in de resultatenlijsten bevestigd: over de hele linie presteert Google net iets beter dan de rest. Er blijft echter ruimte voor persoonlijke voorkeuren, en het gebruik van diverse zoekmachines naast elkaar valt in elk geval aan te raden.

4.2.2.8. Wat met Google kan dat met Google niet kan

Een van de aardige dingen van Google is dat het bedrijf gewone gebruikers als u en ik de gelegenheid biedt mee te denken over wat er nog aan de zoekmachine te verbeteren valt. Onder meer daartoe dient de Google API service, die het mogelijk maakt met gebruikmaking van de zoekmachine applicaties te schrijven die functies bevatten die Google zelf (nog) niet heeft.

Om van deze dienst gebruik te kunnen maken heeft men een Google Account met een zogeheten 'license key' van Google nodig, die te verkrijgen is op het adres <http://www.google.com/apis/>. Daar kan men ook de 'developer's kit' met documentatie en voorbeeldcode downloaden, en een Help-document en een FAQ lezen.

Voor de simpele gebruiker, die niet zelf wil programmeren, is de 'license key' het belangrijkste, want om gebruik te kunnen maken van wat anderen met Google gedaan hebben, is het soms nodig daarover te beschikken.

De 'license key' krijgt u toegestuurd als u twee maal achtereen de procedure doorloopt die u op de bovengenoemde pagina onder 2. Create a Google Account vindt. Overigens zult u de 'license key' maar zelden nodig hebben, en zijn de voorbeelden die ik hieronder noem zelfs ook niet allemaal voorbeelden van sites waar een Google API is gebruikt.

Een waarschuwing: wie van een van de hieronder staande diensten gebruik maakt, krijgt niet altijd hetzelfde resultaat als hij of zij van Google zelf zou krijgen. Dat komt doordat Google beperkingen aan het gebruik van zijn API's stelt. Controleer dus altijd even aan de hand van een simpele vraag of de dienst die u gebruikt hetzelfde resultaat oplevert als Google zelf.

Er zijn diverse aardige dingen met Google gedaan, en enkele daarvan zijn terloops al even genoemd. Een handige uitbreiding van Googles mogelijkheden die al even ter sprake is gekomen, is GAPS (<http://www.staggernation.com/cgi-bin/gaps.cgi>), waarmee de joker * optimaal als nabijheidsoperator kan worden benut.

What's in a date?

Google en andere zoekmachines bieden hun gebruikers de mogelijkheid webpagina's op datum te zoeken en/of te sorteren. Maar over wat voor datum hebben we het dan?

Het is niet de datum die de auteur van de webpagina zo vriendelijk is geweest er boven- of onderaan te zetten. Als een auteur (met de hand of door middel van een script) op zijn pagina netjes bijhoudt wanneer hij of zij die pagina het laatst gewijzigd heeft, kan die datum identiek zijn aan de datum die de zoekmachine hanteert, maar dat hoeft niet. En in elk geval is de datum die u in het document ziet staan niet de datum waar de zoekmachine zijn datering op baseert.

Het is ook niet de datum die de auteur in de META-tags van een pagina opgeeft. Die datum ziet u niet (alleen als u de source-tekst bekijkt), maar de spider – die met het document ook de datum ervan ophaalt – wel. Maar ook deze datum ontbreekt meestal en wordt door de zoekmachine niet gebruikt.

Is het dan de datum die de browser opgeeft als je de pagina op je scherm hebt staan en je

vraagt (in Firefox onder de rechter muisknop) de Page Info op? Nee die is het ook niet, want die kent de zoekmachine niet; die kent alleen de datum van het document zoals dat de laatste keer door zijn spider is opgehaald. Maar we komen wel in de buurt: het is de datum die Firefox ons zou hebben opgegeven als we de pagina op ons scherm hadden gehad op het moment dat de spider van de zoekmachine er een kopie van maakte.

Het gaat zo: een auteur slaat een document op op een computer die tot het openbare internet behoort. Op dat moment krijgt dat document van die computer een datumstempel, en het is die datum die de spider met de kopie van het document aan de Google-computers doorgeeft om te indexeren.

Nadat de spider de kopie heeft gemaakt, kan het oorspronkelijke document wel of niet zijn gewijzigd. Als het wel is gewijzigd heeft het een nieuwe datum gekregen en is het – zonder verdere informatie – meestal niet mogelijk te achterhalen wat de datum van de vorige, gespiderde, versie was. Op de server staat die niet meer, en als je bij Google naar de cachekopie gaat kijken, tref je daar alleen de datum aan waarop die kopie door Google gemaakt is. Bij mijn weten is er maar één algemene mondiale zoekmachine die zijn gebruikers ook vertelt welke datum hij bij het document heeft aangetroffen toen hij het spiderde, en dat is de wat kleinere (2 miljard documenten) zoekmachine Gigablast. Deze geeft in de snippets zowel de serverdatum als de spiderdatum.

Een andere vraag is wat de datum is van de pagina die u, nadat u haar gevonden heeft, op uw scherm hebt staan. Als u met Firefox de Page Info van die pagina opvraagt en de datum en tijd zijn ouder dan de dag en het moment waarop u de link naar de pagina gevolgd hebt, dan is dat de datum die u zoekt. Maar de kans is groot dat Firefox u niet die datum en tijd opgeeft, maar die van het moment waarop de pagina werd opgevraagd; als de pagina uit een database moet komen of via een doorverwijzing (redirect) of een script op uw scherm verschijnt, krijgt ze als datum en tijd het moment waarop dat gebeurt mee. Dit geldt overigens ook voor de datum die de spider van een document te zien krijgt. De 'Last Modified'- datum die door zoekmachines wordt gebruikt om naar actuele pagina's te zoeken of pagina's op actualiteit te sorteren, is dus verre van betrouwbaar.

Waarom Firefox, en niet Internet Explorer? Omdat Internet Explorer u onder Properties (Eigenschappen; ook onder de rechter muisknop) de datum en tijd geeft waarop u de pagina hebt opgevraagd, niet die waarop het document werd opgeslagen. Wat die datum en tijd geweest zijn kunt u in Internet Explorer (en in elke andere browser die Javascript ondersteunt) achterhalen door in de adresregel van uw browser de volgende tekst te typen: `javascript:alert(document.lastModified)` en <Enter> te geven. Als u verwacht vaker in de datum van webpagina's geïnteresseerd te zijn kunt u de Javascriptregel ook als Favoriet opslaan.

De versiegeschiedenis van webpagina's zou eigenlijk veel beter moeten worden bijgehouden. Een project waarbij dat voorbeeldig gebeurt is de al eerder genoemde, gezamenlijk door vrijwilligers gecreëerde internetencyclopedie [Wikipedia](#). In de meeste gevallen waarin niet rigoureuus aan versiebeheer wordt gedaan, is maar één ding zeker: een recente datum van een webpagina, van welke aard ook, zegt niets over de actualiteit van de inhoud ervan; een oudere datum garandeert vrijwel zeker dat delen van de inhoud verouderd zijn.

Hieronder volgen nog wat voorbeelden van Google API's die ikzelf nuttig vind.

Uitbreidere overzichten zijn te vinden op:

<http://www.webhostingchat.com/forum/archive/index.php/t-3470> en via Google Indicateur (<http://google.indicateur.com/>)

4.2.2.8.1. GoogleAlert (<http://www.googlealert.com/>)

GoogleAlert is een attenderingsdienst of 'tracking service' op basis van de

zoekresultaten van Google. Het is een commerciële dienst, die wel resultaten van Google gebruikt, maar verder niets met Google heeft uit te staan. Wie zich registreert kan zich bij wijze van proef gratis op drie vragen abonneren. In de Advanced Search kunnen daarbij allerlei preciseringen van de zoekopdracht worden opgegeven, zoals beperkingen tot een bepaald land, een bepaalde taal, een bepaalde site etc. De bekende mogelijkheden van Google's Advanced Search dus. GoogleAlert stuurt dan een (beperkt) aantal resultaten van die vragen dagelijks/eenmaal per twee dagen/twee maal per week/wekelijks/eenmaal per twee weken/maandelijks per email toe, desgewenst als HTML of als RSS -feeds (daarover zie blz. 284 e.v.). Ook kan men bij GoogleAlert zelf inloggen en daar de resultaten van een aantal weken bekijken, eventueel na de vragen nog eens te hebben laten uitvoeren.

In de gratis versie monitort GoogleAlert tot 50 resultaten van Google en controleert welke daarvan nieuw of verversd zijn. Die treffers waar Google's spiders een verandering hebben gevonden, worden verstuurd of krijgt u bij GoogleAlert zelf te zien. GoogleAlert moet niet worden verward met Google Alerts, de attenderingsdienst van Google zelf. Daarover meer in het hoofdstuk Bijblijven (zie blz. 273 e.v.).

4.2.2.8.2. Soople (<http://www.soople.com/>)

Soople is een gereedschapskist met een heleboel handige spullen voor de Google-gebruiker. De veelzijdige functionaliteit van Google is hier netjes uitgesplitst in een groot aantal afzonderlijke functies, waaruit de gebruiker op een simpele manier kan kiezen. Zo kun je naar boeken, video's, nieuws en definities zoeken, op een overzichtelijke manier de uitgebreide rekenfuncties van Google bij elkaar vinden en de vertaalfunctie gebruiken.

4.2.2.8.3. FindForward (<http://www.findforward.com>)

Eveneens een verzameling van handige en minder handige uitbreidingen van Google. Een daarvan is de mogelijkheid Google-zoekresultaten als RSS- of Atomfeed geformatteerd te krijgen, zodat ze in een RSS-reader kunnen worden gelezen en bewaard (zie blz. 284 e.v.).

4.2.2.8.4. Google Total (<http://goototal.blogspot.com/>)

Het aantal zoekmachines dat Google aanbiedt is inmiddels zo groot, dat Google wel een eigen springplankzoekmachine (zie blz. 27) kan gebruiken. Dat is wat Google Total is.

4.2.2.9. Google's speciale zoekmachines

Wie de Basic Search-pagina van Google opent, ziet daar boven de zoekregel een aantal links die naar speciale zoekmachines van Google verwijzen:

Web Images Groups News Froogle Local more ».

Wie dan vervolgens op more » klikt, komt op een scherm dat Google Services en Google Tools bevat. Een paar daarvan zullen we nader bekijken.

Onder Google Services vinden we alle zoekinstrumenten die Google te bieden heeft. Naast een aantal die voor zichzelf spreken (Web, Images, Groups, News, Mobile en SMS), zijn dat:

- Catalogs, een zoekmachine die een verzameling door Google met OCR gedigitaliseerde postordercatalogi doorzoekt
- Directory, Google's versie van de algemene mondiale webgids van het Open Directory Project (ODP)
- Froogle, Google's winkelzoek- en prijsvergelijkingsmachine
- Local, Google's experiment in bèta van een (meta)zoekmachine die met behulp van kaarten lokaal in de VS naar bedrijven, producten en diensten zoekt
- Maps, een zoekmachine voor kaarten en routes

- Print, Google's zoekmachine voor gedigitaliseerde gedrukte werken
- Scholar, Google's speciale zoekmachine voor wetenschappelijke informatie
- Special Searches, speciale zoekmachines voor sites van de Amerikaanse overheid, sites die Linux en BSD (Berkeley Software Distribution, een implementatie van het Unix besturingssysteem) betreffen, sites betreffende de Apple Macintosh, en sites betreffende Microsoft en Microsoft-producten
- University Search: een grote verzameling universitaire Websites waar Google als site search engine kan worden gebruikt

Daarnaast biedt het scherm toegang tot een aantal zoekgerelateerde diensten:

- een attenderingsdienst (Alerts)
- een markt van vraagstellers en zoekers (Answers)
- Google Labs, een speeltuin waar Google met nieuwe ideeën experimenteert.

Om te beginnen behandel ik in het kort een paar van de speciale zoekmachines die niet in ander verband in dit boek ter sprake komen: Images en Scholar. (Voor Google Groups, zie blz. 28 e.v. (tabel blz. 112), voor Google Alerts, blz. 274, voor Google Answers, zie blz. 41 e.v., voor Google Personalizes Search blz. 292 e.v. Google Labs en Google Blog Search komen apart ter sprake.)

4.2.2.9.1. Google Images (<http://images.google.com/>)

Met de Google Images zoekmachine zoekt u in een database van meer dan een miljard afbeeldingen. Die database is opgebouwd uit de cachekopieën van plaatjes met hun URLs, plus eventueel aanwezige tekst in het document dat naar het plaatje linkt (en waarin het dus voorkomt), die geacht wordt de voorstellingen op die plaatjes te beschrijven. Dat zijn niet alleen de diverse elementen van de URL van het plaatje (bijvoorbeeld de bestandsnaam `cat-playing.jpg`) en eventuele tekst in een ALT-tag in het verwijzende document, maar ook andere tekst die Googlebot in de omgeving van de link naar het plaatje vindt; ook de titel van het document waarin het plaatje voorkomt speelt een rol. Voor u als zoeker betekent dat dat u een plaatje met de bestandsnaam `cat-playing.jpg` niet alleen kunt vinden met de zoektermen `cat playing`, maar ook met de termen `black cat jumping`, als die in de tekst bij de link naar het plaatje staat.

Het verband tussen die tekst en de voorstelling op het plaatje is niet altijd even nauw, met als gevolg dat plaatjes die met woorden uit de begeleidende tekst worden gevonden, niet altijd de verwachte voorstelling vertonen. Zo kan bij het plaatje met de bestandsnaam `cat-playing.jpg` de tekst staan: 'This is Nero, leisurely playing. I would have loved to get a picture with more action but getting my camera to focus on that black cat jumping is next to impossible.' Google Images laat u echter zowel het plaatje zelf als het verwijzende document zien, zodat u kunt constateren waarom het plaatje met de door u opgegeven zoektermen gevonden is.

Hoeveel tekst Google bij een plaatje indexeert hangt af van de omstandigheden; het kan variëren van enkele woorden uit een zin tot bijna een hele pagina.

In de Basic Search kunnen de meeste van de prefixen en operatoren van Google's tekstzoekmachine worden gebruikt. Met behulp daarvan kunnen ook hier zoekacties worden beperkt (gespecificeerd) of uitgebreid. Wel moet, om verwarring te voorkomen, de betekenis van enkele van de prefixen in deze database worden gespecificeerd:

- `intitle:` in de bestandsnaam van de afbeelding, bijvoorbeeld `cat.jpg`
- `inurl:` in de URL van de afbeelding, bijvoorbeeld:
http://www.iac.es/telescopes/cat/TTNN_schedule.html (deze afbeelding heeft dus niets met katten te maken, maar ergens in het pad van het adres van dit plaatje komt een folder met de naam 'cat' voor)
- `inanchor:` in de titel van het document waarin de afbeelding is opgenomen (= dat naar de afbeelding verwijst) of in de tekst van de hyperlink waarmee naar een afbeelding verwezen wordt

Het prefix `link:` en de synoniem-operator `~` kunnen in de images-database niet worden gebruikt.

Evenals bij de webzoekmachine biedt de Advanced Search de mogelijkheid de vraag met enkele specificaties te beperken. In de eerste plaats kan een bestandstype worden opgegeven, waarbij kan worden gekozen tussen any filetype (standaardinstelling), JPG, GIF en PNG ; dit kan ook via het prefix `filetype:` in de Basic Search, die als extra functionaliteit biedt dat het prefix meermalen in de zoekopdracht kan worden gebruikt, verbonden met OR of met het minteken om aan te geven dat men een bestandstype niet wenst: `cat -filetype:PNG`. Voordeel van de Basic Search is dat je ook BMP en TIF kunt opgeven, die sporadisch voorkomen maar niet in het keuzelijstje van de Advanced Search zitten.

Daarnaast biedt het Advanced Search scherm van de Images-zoekmachine enkele mogelijkheden die de Basic Search niet biedt. Er kan worden gekozen tussen afbeeldingen die any size, small, medium en large zijn, en tevens tussen afbeeldingen in kleur, grijswaarden en zwart-wit: any colors, black and white, greyscale en full color.

4.2.2.9.2. Google Scholar Bèta (<http://scholar.google.com>)

'Zelfs de beste breister laat wel eens een steekje vallen,' luidt het spreekwoord, en in het geval van Google is dat steekje Google Scholar, dat in november 2004 in bètaversie beschikbaar kwam.

Volgens de summier informatie die Google in de FAQ bij de zoekmachine geeft, omvat Google Scholar 'peer-reviewed artikelen, dissertaties, boeken, preprints, abstracts en technical reports uit alle vakgebieden', en Google voegt daaraan toe dat de literatuur afkomstig is van een groot aantal wetenschappelijke uitgevers, wetenschappelijke genootschappen, preprintarchieven en universiteiten, en daarnaast ook wetenschappelijke artikelen bevat die op het Web beschikbaar zijn.

Bij nadere inspectie blijkt Google Scholar grotendeels de bibliografische beschrijvingen en abstracts van tijdschriftartikelen te omvatten, met in een aantal gevallen een link naar de fulltext, die het wel heeft geïndexeerd maar die meestal alleen tegen betaling beschikbaar is. Die beschrijvingen en abstracts haalt Google uit databases zoals PubMed en Ingenta (en tal van andere), en direct van de sites van uitgevers die hun site voor de spiders van Google beschikbaar hebben gesteld.

Google Scholar ontsluit dus niet in de eerste plaats het wetenschappelijke gedeelte van het openbaar toegankelijke Web, maar een deel van het onzichtbare Web dat achter wachtwoorden en creditcardbetalingen verborgen ligt. (Een uitzondering vormen de preprint-archieven, die voor sommige vakgebieden de tijdschriftenliteratuur al overschaduwden en de als kool uit de grond schietende 'repositories' waarmee wetenschappelijke instellingen hun onderzoek openbaar toegankelijk maken.)

Daarmee treedt Google Scholar als fulltext zoekmachine voor een deel in concurrentie met de dienstverlening die traditioneel door wetenschappelijke bibliotheken wordt aangeboden. Die dienstverlening, die u pakweg € 30 per jaar kost, omvat het gratis gebruik van bibliografische databases (met trefwoorden en thesaurussen om efficiënt en effectief zoeken mogelijk te maken), de gratis levering van miljoenen boeken en tientallen miljoenen tijdschriftartikelen en andere documenttypen uit de eigen collectie, de toegang tot duizenden elektronische tijdschriften waarvoor de bibliotheek een licentie heeft, en (tegen betaling) de levering van praktisch alle literatuur die niet in de eigen bibliotheek aanwezig is (via het Interbibliothecair Leenverkeer, IBL). De gratis hulp die u bij het zoeken naar literatuur van ervaren professionals krijgt heb ik dan nog niet eens genoemd. Voor minimaal € 60 koopt u de full text van twee artikelen die u via Google Scholar gevonden hebt; uw bibliotheek heeft die artikelen waarschijnlijk gratis voor u in de collectie, en als dat niet zo is, betaalt u bij gebruik van het IBL per artikel ca. € 6 (de bedragen verschillen enigszins per bibliotheek).

Misschien denkt u dat dit een oratio pro domo is. Laat ik dan een voorbeeld van buiten de Nederlandse bibliotheekwereld gebruiken om u die gedachte uit het hoofd te praten. De bekendste grote database voor de literatuur over de geneeskunde en aanverwante gebieden is PubMed, een meer dan een eeuw oude database van de Amerikaanse National Library of Medicine die meer dan 15 miljoen records telt. Die records zijn de bibliografische beschrijvingen van tijdschriftartikelen met abstracts, korte samenvattingen. Daarnaast zijn echter aan elk artikel een of meer trefwoorden toegekend die de inhoud van het artikel omschrijven en het artikel beter vindbaar maken dan wanneer men het alleen van woorden in de titel, of nog veel erger, van woorden in het hele document zou moeten hebben. Tezamen vormen die trefwoorden een zogeheten thesaurus (zie ook blz. 253 e.v.), die het hele vakgebied waarvan PubMed de literatuur dekt, in een uniform vocabulaire beschrijft.

PubMed zelf heeft een zoekmachine waarmee u van deze thesaurus en trefwoorden gebruik kunt maken. Dat betekent niet alleen dat als u literatuur wilt zoeken over de oorzaken van buikpijn bij peuters, de thesaurus van PubMed en PubMed zelf u naar alle mogelijke oorzaken van buikpijn laten zoeken bij alle kinderen van 2-5 jaar (hoe ze ook benoemd zijn: peuters, kleuters, kinderen van drie jaar, kinderen van vier, kleuterschoolkinderen enz.), maar ook dat u door gebruik te maken van trefwoorden artikelen in alle talen krijgt, en niet alleen in het Engels. Dezelfde (Engelse) trefwoorden als voor Engelstalige artikelen worden gebruikt worden namelijk ook aan artikelen in andere talen toegekend.

Vergeleken met de zoekmogelijkheden van PubMed (voor die speciale database) zijn de zoekmogelijkheden van Google Scholar uiterst primitief: je kunt met auteursnaam, woord(en) in de titel of woord(en) in het hele document zoeken; als je voor peuter het Engelse toddler gebruikt, mis je de pre-schoolers en vice versa, en er is geen beginnen aan alle mogelijke oorzaken van buikpijn (constipatie, blinde-darmonsteking, nervositeit enz. enz.) in een vraag te verwerken.

PubMed is een voor iedereen gratis toegankelijke database op het gebied van de medische literatuur, en als u wetenschappelijke literatuur over een medisch onderwerp zoekt, raad ik u aan die database, en niet Google Scholar, te gebruiken. De resultaten die u krijgt zijn groter en beter. Hetzelfde geldt voor alle andere niet-vrij toegankelijke databases op andere terreinen waarvan u – zoals gezegd – in uw bibliotheek (of als u bij een universiteit werkt of studeert ook daarbuiten) gebruik kunt maken.

Ook zonder gebruik te maken van de trefwoorden waarmee bibliografische databases u het zoeken vergemakkelijken, zult u via die databases doorgaans meer literatuur vinden dan via Google. Dat komt voornamelijk doordat Google Scholar alleen kijkt naar literatuur die op het Web is gepubliceerd of waarnaar in literatuur op het Web verwezen wordt. Gespecialiseerde bibliografische databases proberen in principe alle literatuur over een onderwerp te dekken, dus zowel oudere en niet elektronisch gepubliceerde literatuur als literatuur waarnaar niet op het Web verwezen wordt.

Om dat te illustreren laat ik hier (zie volgende bladzijde) een overzichtje van 12 zoektermen op diverse vakgebieden volgen, met daarbij de resultaten voor Google Scholar en de beste bibliografische databases die mij als lid van een grote wetenschappelijke bibliotheek ter beschikking staan. Alle termen zijn gezocht met het prefix intitle: en bij alle blijft het resultaat onder de 1000, zodat het aantal controleerbaar is. In de 3e-6e kolom staan de aantallen treffers voor Google Scholar eerst als totaal en daarna uitgesplitst naar de typen treffers die Google Scholar in zijn resultatenlijst onderscheidt: artikelen, [BOOKs] en [CITATIONs]; daarover straks meer. In de 7e kolom staan de aantallen treffers bij de zoektermen in de diverse databases zoals ze in de titel van de beschrijvingen worden gevonden. In de 8e kolom ten slotte staan, ter controle van wat Google Scholar uit WorldCat vindt, de aantallen treffers van WorldCat zelf. Omdat u als Nederlandse bibliotheekgebruiker van de NCC

gebruik zou maken om een boek te lenen dat niet in uw eigen bibliotheek is, heb ik er tussen haakjes de aantallen voor de Nederlandse Centrale Catalogus (NCC) achter gezet. De cijfers dateren van begin september 2005. (Intussen – maart 2006 – zijn alle getallen met enige tientallen tot honderden toegenomen. Daarbij is de omvang van de Google Scholar-database sneller toegenomen dan die van de andere gebruikte databases.)

<i>Vakgebied</i>	<i>Zoekterm</i>	<i>Google Scholar</i>	<i>Art.</i>	<i>Boeken</i>	<i>Cit.</i>	<i>Bibliografie - titelwoorden</i>	<i>WorldCat² (NCC)</i>
geneeskunde	rofecoxib (= Vioxx)	889	573	1	315	(PubMed) 508	16 (0)
geschiedenis	historiographical	408	101	37	270	(Historical Abstracts/ America: History & Life) 904	511 (88)
taalkunde	grammaticalization	434	101	32	301	(Linguistic Bibliography) 369	148 (55)
neurowetenschappen	neuromodulators	216	123	3	90	(PubMed) 111	26 (7)
wijsbegeerte	syllogism	139	29	3	107	(Philosophers Index) 90	96 (17)
psychologie	behaviorism	571	133	41	397	(PsycINFO) 588	420 (96)
klassieke oudheid	hellenistic	780	101	165	514	(L'Année Philologique) 1304	1976 (656)
theologie	christology	845	39	181	625	(ATLA Religion Database) ¹ 3379	2190 (434)
wiskunde	"positive integers"	174	61	0	113	Web of Science 121 (Scirus 201)	51 (9)

<i>Vakgebied</i>	<i>Zoekterm</i>	<i>Google Scholar</i>	<i>Art.</i>	<i>Boeken</i>	<i>Cit.</i>	<i>Bibliografie - titelwoorden</i>	<i>WorldCat² (NCC)</i>
recht	"habeas corpus"	126	35	13	78	(Index to Legal Periodicals/ Index to Foreign Legal Periodicals) 365	1588 (47)
zoölogie	Acrocephalus	330	129	0	201	Web of Science 214	26 (14)
literatuur-wetenschap	"Harlem Renaissance"	268	37	72	159	MLA Bibliography 388	622 (72)

¹ Voor de theologie kan ook de vrij toegankelijke Theoldi-database (<http://bildi.uibk.ac.at/theoldi-search-en.html>) worden gebruikt, die met deze vraag 331 treffers oplevert.

² De aantallen voor WorldCat zijn geflatteerd doordat WorldCat meer documenttypen bevat dan alleen boeken en omdat in de PICA-versie van WorldCat niet met phrases kan worden gezocht.

Uit dit staatje blijkt (voor wat zo'n testje waard is) dat Google Scholar getalsmatig in de geneeskunde, de neurowetenschappen, de wiskunde en de zoölogie meer vindt dan de bibliografische databases die ik gebruikt heb (PubMed en Web of Science), maar ook daar is nog wel iets op af te dingen. In de eerste plaats zitten in Google's resultaat veel doublures, zoals we nog zullen zien, en in de tweede plaats vertekenen de [CITATIONS] bij Google het beeld doordat die in de resultaten van PubMed en Web of Science niet kunnen worden meegeteld.

De aantallen treffers die Google Scholar met de aanduiding [BOOK] vindt blijven ver achter bij die van WorldCat (waar het zijn treffers van dit type vandaan heeft), en komen ook niet overeen met de resultaten in de algemene Google, wat op z'n minst twijfel wekt aan de betrouwbaarheid van beide. Hoe dan ook, wat boeken over wetenschappelijke onderwerpen betreft wordt u door de Nederlandse Centrale Catalogus en het IBL beter bediend dan door Google Scholar.

Zoeken naar boeken is een hoofdstuk apart, dat in dit boek maar zijdelings ter sprake kan komen. Als het u alleen om het bestaan en de juiste beschrijving van een boek gaat, zijn er over de hele wereld duizenden bibliotheekcatalogi waarin u kunt zoeken. Een nog altijd heel waardevol overzicht daarvan vindt u onder de titel Bibliotheks-OPACs und -informationsseiten weltweit in de 'Bibliographischer Werkzeugkasten' van Hans-Dieter Hartges en Claudia Kröhnert (<http://toolbox.hbz-nrw.de/>), die ook een schat aan ander materiaal bevat waarmee gedeelten van het invisible Web kunnen worden verkend.

Twee grote catalogi die vrij toegankelijk op het Web beschikbaar zijn, zijn de Karlsruher Virtueller Katalog (<http://www.ubka.uni-karlsruhe.de/kvk.html>), een soort metzoekmachine van bibliotheekcatalogi en online boekhandels overal ter wereld, en de voortreffelijke catalogus van de Research Libraries Group, RedLightGreen (<http://www.redlightgreen.com/>), die meer dan 120 miljoen titels dekt.

Books and Book Collecting (http://www.trussel.com/f_books.htm) van de onvermoeibare Steve Trussel biedt u verder alle informatie voor wat u ook maar met boeken wilt doen.

The screenshot shows the ScienceDirect website interface. At the top left is the Elsevier logo and the ScienceDirect text. Below this is a navigation menu with buttons for Home, Journals, Books, Abstract Databases, My Profile, and Alerts. On the right, there is a 'Help' button and a 'WELCOME GUEST USER' message with an 'info' link. The main heading is 'Access Online Article'. Below this, the article title is 'Leishmania major: Promastigotes Induce Expression of a Subset of Chemokine Genes in Murine Macrophages'. The journal information is 'Experimental Parasitology, Volume 85, Issue 3, March 1997, Pages 283-295'. The authors are 'Esther L. Racoosin and Stephen M. Beverley' with a 'View Abstract' link. The interface is split into two columns. The left column is for users with a ScienceDirect User Name and Password, featuring input fields for 'User Name:' and 'Password:', 'Submit' and 'Cancel' buttons, and links for 'Login using your Athens ID', 'Login via InQueue', and 'Forgotten your User Name or Password?'. The right column is for users without a User Name and Password, showing the price 'Price: US \$ 30.00' and a 'Register to Purchase' button.

Een zoekactie naar wetenschappelijke literatuur in Google Scholar eindigt vaak op een plaats als deze. Google maakt wetenschappelijke artikelen wel fulltext vindbaar, maar als u die vervolgens als niet bij een bibliotheek geregistreerde gebruiker wilt lezen, zult u de creditkaart moeten trekken.

De dekking van Google Scholar

De belangrijkste vraag die een zoeker naar wetenschappelijke literatuur zich moet stellen als hij Google Scholar gaat gebruiken is wat de zoekmachine eigenlijk dekt. Zoals gezegd, volgens de eigen FAQ stelt deze zoekmachine 'u in staat specifiek naar wetenschappelijke literatuur te zoeken, waaronder peer-reviewed artikelen, dissertaties, boeken, preprints, abstracts en technical reports uit alle vakgebieden'. Dit soort literatuur is inderdaad via het Web te vinden, en alvorens op de vraag in te gaan in hoeverre Google Scholar die ook werkelijk vindt, volgt hier een beknopt overzicht van wat we ons daarbij moeten voorstellen.

'Peer-reviewed' artikelen zijn artikelen in (gedrukte of elektronische) tijdschriften waarvan de redactie de aan haar ter publicatie aangeboden manuscripten eerst aan vakgenoten ('peers') van de auteurs voorlegt alvorens ze te publiceren. Deze vakgenoten, de zogeheten 'referees', doen dan suggesties voor verbetering of geven in het ergste geval een negatief advies over publicatie. 'Peer-reviewed' wordt gebruikt als

een soort kwaliteitslabel, en in het beste geval is het dat ook. En aangezien de grote en veel kleinere tijdschriftenuitgevers hun tijdschriften niet meer alleen in gedrukte, maar ook in elektronische vorm uitgeven, zijn er al duizenden van die tijdschriften op het Web beschikbaar, en produceren die samen meer dan 100.000 artikelen per jaar, de meeste 'peer reviewed'. Het grootste deel daarvan blijft echter voor de gewone Googelaar onzichtbaar, doordat de toegang ertoe achter een wachtwoord verborgen ligt: het gaat hier om tijdschriften van commerciële uitgevers die voor het gebruik ervan kosten in rekening brengen. Die toegang vindt men soms bij de uitgever zelf – voorbeelden zijn Elsevier met ScienceDirect en Springer met SpringerLink –, soms bij een distributeur die de toegang voor een aantal uitgevers verzorgt – bijvoorbeeld Ingenta – en daarnaast zijn ook bibliotheekorganisaties zoals OCLC en hostorganisaties van bibliografische databases zoals InfoTrac, CSA en Ovid op dit gebied actief. Veelal bieden deze organisaties wel een gratis, openbare toegang tot de bibliografische beschrijvingen en abstracts (samenvattingen) van de artikelen in hun database, maar moet voor het downloaden (bekijken en/of opslaan) van de fulltext worden betaald. Toegang (op basis van IP-adres en in veel gevallen een wachtwoord) wordt ook verschaft aan mensen die lid zijn van een bibliotheek die voor een flink bedrag een licentie voor het gebruik van elektronische tijdschriften heeft betaald. Er zijn ook 'peer reviewed' elektronische tijdschriften (of oudere jaargangen/afleveringen daarvan) die gratis toegankelijk zijn. Zo onderhoudt het National Center for Biotechnology Information (NCBI) van de Amerikaanse National Library of Medicine (NLM) een database van zulke tijdschriften en de daarin verschenen artikelen, PubMed Central geheten (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pmc>); in die database zitten 160 tijdschriften in de biowetenschappen; het wat kleinere BioMed Central heeft er 100. Highwire Press, van Stanford University, heeft een database van meer dan 950.000 artikelen in de bèta- en biowetenschappen, en miljoenen andere op diezelfde terreinen zijn in de fulltext archieven van overwegend Amerikaanse overheidsorganisaties te vinden (zie: Earth's Largest Free Full-Text Science Archives: <http://highwire.stanford.edu/lists/largest.dtl>). Een overzicht van meer dan 2100 voor iedereen gratis toegankelijke wetenschappelijke tijdschriften op het Web (waarvan ca. bijna 600 doorzoekbaar) geeft de Directory of Open Access Journals (<http://www.doaj.org/>). Een zeer uitgebreid overzicht, waarin onder meer per vakgebied gratis en niet-gratis toegankelijke tijdschriften naast elkaar zijn ondergebracht, geeft de Elektronische Zeitschriftenbibliotheek van bijna 350 samenwerkende Duitse bibliotheken, <http://rzblx1.uni-regensburg.de/ezeit/>. Deze lijst bevat de adressen van tijdschriften op onder meer zo'n exotisch vakgebied als 'Klassische Philologie. Byzantinistik. Mittellateinische und neugriechische Philologie. Neulatein'; het aantal gratis toegankelijke tijdschriften alleen al op dit vakgebied bedraagt niet minder dan 48. Een bron van wetenschappelijke artikelen die hier beslist dient te worden genoemd, maar die over het hele Web verspreid ligt, zijn verder de cv's van wetenschappers op het Web, die niet zelden gepaard gaan met een lijst van publicaties die ze geheel of gedeeltelijk op een eigen server van hun werkgever hebben gezet. Je mag hopen dat dat gebeurd is op een plaats waar de spider van Google Scholar komt. Voorlopig blijft het in alle gevallen zinvol daarnaar met de naam van de auteur en/of de titel van de publicatie met een algemene grote zoekmachine op zoek te gaan. Om maar één voorbeeld te noemen, de bijdrage 'Theoretical neuroanatomy: Analyzing the structure and dynamics of networks in the brain.' van de auteurs Anil K. Seth en G.M. Edelman verscheen in 2004 in de de bundel *Complex Networks* van E. Ben-Naim et al. Via Google Scholar is er geen spoor van te vinden, maar via de webzoekmachine van Google vind je de fulltextversie in de folder van een van de auteurs bij het Neurosciences Institute in La Jolla.

Full-text dissertaties zijn er op het Web ook volop. Het Amerikaanse bedrijf ProQuest heeft een bibliografische database van praktisch alle Amerikaanse en een groot aantal Europese dissertaties en levert op bestelling en tegen betaling kopieën van die dissertaties aan bibliotheken en particulieren; de bibliografische beschrijvingen en de abstracts (samenvattingen) van de dissertaties van de afgelopen twee jaar zijn gratis te raadplegen en te doorzoeken via UMI ProQuest Digital Dissertations (<http://www.lib.umi.com/dissertations/>). Maar ook aan openbaar en gratis toegankelijke dissertaties ontbreekt het niet: bij heel wat universiteiten bestaan archieven van elektronisch beschikbare publicaties, en mogelijkheden om die op te sporen zijn er ook; zie bijvoorbeeld de Networked Digital Library of Theses and Dissertations, <http://www.ndltd.org/browse.en.html>. Dissertationen Online (http://www.dissonline.de/convert/index_such.htm) biedt nog ettelijke andere mogelijkheden om online dissertaties op te sporen. De dissertaties die Google Scholar vindt, zal het ongetwijfeld in de al eerder genoemde institutionele 'repositories' vinden, maar anders dan op de hierboven genoemde adressen kun je met Google Scholar niet specifiek naar dissertaties zoeken.

Wetenschappelijke boeken in elektronische vorm zijn er nog niet zoveel, en wat er is zit grotendeels achter slot en grendel van commerciële bedrijven zoals Questia.com. Een uitzondering vormt een collectie van de Amerikaanse [National Academies Press](#). Google Book Search (zie hieronder) moet in dit opzicht een aanvulling op Google Scholar gaan vormen.

Preprints zijn elektronische documenten die in een pre-publicatiestadium zijn voordat ze in een (doorgaans papieren) wetenschappelijk tijdschrift worden afgedrukt. Ze worden door hun auteurs in elektronische vorm op 'preprint servers' gepubliceerd, om uiteenlopende redenen, maar met als voornaamste motief om zo snel mogelijk aan hun inhoud bekendheid te geven. Veel van deze preprints verschijnen later, als ze na 'peer review' geaccepteerd zijn, in gedrukte en elektronische tijdschriften; overigens zijn ook veel van de preprints het stadium van de 'peer review' al voorbij. In dat geval bestaat er naast het artikel dat officieel gepubliceerd is, een preprint die niet noodzakelijk identiek is aan het artikel. Preprint archives kunnen miljoenen preprints bevatten, en zijn er tientallen van die afzonderlijk doorzoekbaar zijn. Voor een overzicht zie: Preprint Archives etc., UK Libraries (<http://www.lib.umd.edu/ETC/preprints.html>), voor een gedegen artikel over karakter en betekenis van preprints in *Searcher* van oktober 2000, zie Preprint Servers: Pushing the Envelope of Electronic Scholarly Publishing (<http://www.infoday.com/searcher/oct00/tomaiuolo&packer.htm>).

Abstracts zijn, zoals eerder gezegd, samenvattingen van artikelen. De meeste ervan worden door de auteurs zelf geschreven en bij het artikel gevoegd om de lezer in de gelegenheid te stellen snel te beoordelen of het al dan niet de moeite waard is het hele artikel te lezen; maar omdat zo'n korte samenvatting ook interessant is voor literatuurzoekers, komen ze in veel gevallen als doorzoekbare tekst ook in bibliografische databases terecht en worden soms door de bedrijven die die databases samenstellen, geproduceerd. Abstracts worden ook apart, buiten het verband van bibliografische databases gepubliceerd, bijvoorbeeld door organisaties die hun leden snel en efficiënt over de ontwikkelingen op het vakgebied willen informeren. Abstracts zijn niet hetzelfde als de artikelen die ze beschrijven. Zeker als deze laatste niet een door de auteur zelf gemaakt abstract hebben en er direct met een zoekmachine zoals Google, en niet via een bibliografische database naar moet worden gezocht, kan het zijn dat dat op een andere manier moet gebeuren (bijv. met phrases in plaats van met trefwoorden).

Een '*technical report*' is een verslag waarin de voortgang en de uitkomsten van een onderzoek of experiment worden gedocumenteerd. Technical reports zijn bedoeld voor personeel en leiding van de organisatie die het onderzoek uitvoert en voor de

opdrachtgever of financier van het onderzoek. Ze worden doorgaans niet via de gewone kanalen gepubliceerd en er komt geen peer review aan te pas. Wel zijn ze vaak te vinden op een server van een organisatie, instelling of laboratorium die bij het onderzoek betrokken is. Voor een uitgebreid overzicht, zie The Virtual Technical Reports Center (<http://www.lib.umd.edu/ENGIN/TechReports/Virtual-TechReports.html>). Een gespecialiseerde zoekmachine die de inhoud van grote aantallen Technical Reports fulltext indexeert en doorzoekbaar maakt was er tot dusver niet.

Het bovenstaande is een onvolledige opsomming van wat er op het gebied van de in Google Scholars FAQ genoemde literatuur op het Web te koop is, en een onderzoek naar de vraag of dit alles door Google Scholar wordt gedekt zou het bestek van dit boek te buiten gaan. Maar een proefje van begin september 2005 met een tiental databases waarbij de aantallen treffers bij gebruik van de eigen zoekmachine van die bestanden worden vergeleken met wat Google in die bestanden vindt, laat het volgende beeld zien:

<i>Zoekterm</i>	<i>Naam Archive/ Database</i>	<i>Aantal Archive/ Database</i>	<i>Google Scholar</i>	<i>Opmerking</i>
codeine	Highwire Press (zonder PubMed)	136	0	in de titel; niet in GS
brightness fluctuation OR fluctuations	NASA Astrophysics Data System (ADS)	241	126	in de titel
combinatorial AND optimization	MIT Dspace	451	6	in de titel
cognition	Cogprints	46	20	in de titel
"global warming"	DOE Information Bridge	1450	354	free-text
vries	PubMed Central	177	84	als auteursnaam
vries	BioMed Central	3	2	als auteursnaam; beperkt tot research articles
kovenock	IDEAS Repec	74	54	free-text
chaos	Institute of Physics / Electronic Journals	263	262	in de titel
1998-2005	PhilSci Archive	900	54	als publicatie-periode
Bayesian	The ACM Portal (The ACM Digital Library + The Guide)	1723	556	in de titel

Met alle reserves die je bij zo'n steekproefje in acht moet nemen, kun je je toch niet aan de indruk onttrekken dat het gebruik van een spider om documenten te vinden die in een database zijn opgeslagen, misschien toch niet zo'n goed idee is.

Zoekmogelijkheden

De zoekmogelijkheden van Google Scholar zijn beperkt, hoewel ze op drie punten zijn aangepast aan het afwijkende materiaal dat Google Scholar omvat. In de eerste plaats kan met `author:` (of via de Advanced Search) naar een auteursnaam worden gezocht of kan daarmee een zoekactie tot het werk van een auteur worden beperkt; net als in de algemene Google biedt de Basic Search hier weer meer mogelijkheden doordat ook een zoekactie van het type `osteoporosis author:Smith OR author:Jones` mogelijk is. In de Advanced Search wordt de mogelijkheid genoemd de auteursnamen voorletters mee te geven; als u die kent, geef de naam dan ook meteen als phrase op, want anders verbindt Google Scholar naam en voorletters in een EN-relatie. Dat betekent dat als u artikelen zoekt van een auteur J. Smith, u treffers krijgt van alle mogelijke Smithen zolang er bij de coauteurs maar één is die een J als voorletter heeft. Als een auteur twee initialen heeft – een aanduiding met meer dan twee is in de wetenschappelijke literatuur niet gebruikelijk – zult u die allebei moeten opgeven, aangezien de voorletter J in de combinatie J.A. niet afzonderlijk geïndexeerd is (anders gezegd: J.A. = JA); “J.A. Smith” zult u dus met “J Smith” niet vinden. Tenzij de betreffende auteur ook onder de naam J. Smith gepubliceerd heeft, want die publicaties zijn alleen onder “J Smith” te vinden. De juiste syntax voor een zoekactie naar literatuur over osteoporosis van de auteur P.D. Ross, die ook wel onder de naam P. Ross publiceert, zou dus zijn: `osteoporosis author:"PD Ross" OR author:"P Ross"`. Sommige tijdschriften vermelden de eerste voornaam van de auteur voluit: Philip D. Ross, maar in die gevallen vindt Google Scholar de artikelen toch als u met “PD Ross” zoekt.

Het indexeren van auteursnamen is in Google Scholar niet altijd goed gegaan. In een van mijn resultatenpagina's kwam ik enkele treffers met nogal ongewone auteursnamen tegen, die me deden denken aan de kleurrijke benamingen waarmee spammers tegenwoordig werken: D Info, F Function, C Judges, E Lobbying, SL Supply, S Design en L Bargaining. Om nog maar niet te spreken van SI Expectations, C Models en I Risk. U begrijpt het al: bij nadere inspectie bleken de complete auteursnamen Download Info, File Function, en Legislative Bargaining, Careerist Judges, Endogenous Lobbying, Strategic Liquidity Supply and Security Design, en Legislative Bargaining te zijn, die op een schijnbaar willekeurige manier door Googles spider tot auteurs waren gebombardeerd. Enfin, het verklaart in elk geval hoe Philip D. Ross toch als PD Ross kan worden gevonden.

Een tweede aanpassing van Google Scholar aan de eisen van het materiaal is dat men op tijdschrifttitel, dus op bron, kan zoeken. Ook dat is geen sinecure, want daarbij moet wel of niet met titelafkortingen (zoals 'J Cell Sci' tegenover: Journal of Clinical Densitometry) worden gewerkt. Meestal zal een tijdschrift zowel onder zijn volle als onder zijn afgekorte naam bij Google Scholar bekend zijn. Zo staat het Journal of Consulting and Clinical Psychology ook bekend als J Consult Clin Psycho. Zoekt men met de hele naam, dan is de kans groter dat men fulltext artikelen vindt (aangezien die naam doorgaans in de artikelen zelf vermeld staat, zoekt men met de afkorting, dan vindt men waarschijnlijk verwijzingen. De Clinical Psychology Review staat zelfs onder drie namen bekend: de volle naam plus twee verschillende afkortingen. Hoe de naam van een tijdschrift wordt afgekort, wordt immers niet door de uitgever van het tijdschrift of door een centrale instantie bepaald, maar door de uitgevers van bibliografische databases, en als er daarvan meer door Google Scholar worden ontsloten, kan het tijdschrift onder meer afkortingen bij de zoekmachine bekend zijn.

In PubMed zelf kun je in een index van tijdschriftnamen en -afkortingen bladeren om daaruit de gewenste te selecteren. Maar PubMed is natuurlijk niet de enige database met eigen afkortingen waaruit Google put; voor een overzicht van afkortingenlijsten op allerlei vakgebieden, zie All that JAS van metabibliothecaris Gerry McKiernan, <http://www.public.iastate.edu/~CYBERSTACKS/JAS.htm>.

Zoeken op tijdschrifttitel kan alleen in de Advanced Search, er is geen prefix voor in de Basic Search. Dat geldt ook voor de derde concessie van Google Scholar aan het bijzondere type materiaal: de mogelijkheid om een zoekactie tot tijdschriftafleveringen van een bepaald jaar of een bepaalde reeks van jaren te beperken. numrange: (zie de algemene Google), dat wel werkt, is niet voor dit doel te gebruiken. Wel verschijnen er, nadat je eenmaal in de Advanced Search Publication en Date hebt ingevuld, in de Basic Search drie extra zoekvakjes waarin de gegevens kunnen worden gewijzigd.

Helemaal goed gaat het bij het zoeken naar artikelen in bepaalde tijdschriften ook niet. De afkorting van het tijdschrift Pharmacological Reviews is Pharmacol Rev. Als je dat in de Advanced Search bij Publication invult, krijg je toch ook treffers uit een tijdschrift dat als afkorting Annu Rev Pharmacol Toxicol heeft. Google Scholar behandelt de tijdschriftafkorting dus niet als phrase, maar legt een EN-relatie tussen Rev en Pharmacol. Voer tijdschrifttitels en afkortingen daarvan dus altijd als phrase in. Zoeken op auteur en titel van een artikel en op titel en jaar van een tijdschrift zijn de minimale mogelijkheden waarvan de literatuurzoeker gebruik moet kunnen maken. Maar tijdschriften bevatten nogal divers materiaal, variërend van artikelen met onderzoeksresultaten, via hoofdredactionele artikelen en boekbesprekingen tot

Google Scholar **Advanced Scholar Search** [Advanced Search Tips](#) | [About Google Scholar](#)

Find articles with **all** of the words
with the **exact phrase**
with **at least one** of the words
without the words
where my words occur

100 results

anywhere in the article

Author Return articles written by
e.g., "PJ Hayes" or McCarthy

Publication Return articles published in
e.g., J Biol Chem or Nature

Date Return articles published between and
e.g., 1996

Subject Areas

Return articles in all subject areas.

Return only articles in the following subject areas:

- Biology, Life Sciences, and Environmental Science
- Business, Administration, Finance, and Economics
- Chemistry and Materials Science
- Engineering, Computer Science, and Mathematics
- Medicine, Pharmacology, and Veterinary Science
- Physics, Astronomy, and Planetary Science
- Social Sciences, Arts, and Humanities

Google Scholars Advanced Search-scherm.

ingezonden brieven; verder gaan medisch-wetenschappelijke artikelen altijd over mensen of dieren, mensen in het algemeen of vrouwen of mannen en mensen van alle leeftijden of van een of meer leeftijdsgroepen. Databases als PubMed bieden hun gebruikers dan ook nog diverse andere mogelijkheden om zoekacties verder te specificeren. In andere wetenschappen gelden andere specialiteiten en mogelijkheden. Google Scholar als grote vergaarbak heeft die mogelijkheden niet.

Anderzijds: hoewel dat niet gedocumenteerd is, werken een aantal zaken die de algemene Google tot zo'n veelzijdige zoekmachine maken, ook hier. Met `intitle:` en `allintitle:` kan naar publicaties worden gezocht die de zoekterm(en) in de titel hebben, en ook de andere prefixen werken, hoewel niet altijd duidelijk is hoe. `site:` en `(all)inurl:` beperken het zoekresultaat tot de opgegeven site of de sites die de opgegeven zoekterm(en) in de URL hebben. Dat is een krachtig instrument om het zoekresultaat te beperken tot die documenten die fulltext – en niet enkel als abstract in een database als PubMed, als [BOOK] of als [CITATION] – beschikbaar zijn. Als van zo'n fulltext beschikbaar document andere versies, of ook abstracts, bestaan, worden die wel vermeld, maar als u de site hebt opgegeven waar het document fulltext beschikbaar is, linkt de titel in de resultatenlijst naar die site.

Een voorbeeld:

Het American College of Physicians (ACP) geeft een tijdschrift uit, de *Annals of Internal Medicine*. Behalve onder zijn volle naam, staat het tijdschrift ook onder twee afkortingen bekend: Ann Int Med en Ann Intern Med. Artikelen die in dit (niet-gratis) tijdschrift verschijnen zijn zes maanden na verschijnen gratis voor iedereen toegankelijk. Van die artikelen bestaan in hoofdzaak vier versies: de fulltextversie bij ACP zelf, en wel onder twee adressen: dat van de *Annals* en een van de ACP, en drie abstractversies, een bij de ACP/Annals, een uit de database van PubMed en een in een van de diverse tijdschriften van de American Medical Association (AMA). De abstracts bij de ACP/Annals-site zitten in een directory met het woord 'abstract' in het pad, de fulltext artikelen in een directory met de aanduiding 'full' of 'reprint' in het pad. De zoekterm *heart* in de jaargang 1996 van dit tijdschrift (onder zijn drie titels) levert 114 op, waarvan er 99 fulltext beschikbaar blijken te zijn. Je vindt die door aan de zoekterm *heart* toe te voegen `site:annals.org -inurl:abstract` (of: `site:annals.org inurl:full OR inurl:reprint`). (Omdat het Publication-veld van het Advanced Search- en van het Basic Search-scherf niet meer dan 60 tekens kan bevatten, moet de beperking "ann intern med" OR "annals of internal medicine" OR "ann int med" in de zoekvraag via URL-manipulatie worden gerealiseerd.) Gebruik van `site:` en `inurl:` is ook mogelijk zonder zoekterm. Op de hele Annals-site toegepast levert `inurl:full` 10.700 treffers op, op heel Google Scholar (voor wat het waard is) 13,7 miljoen.

In de algemene Google kan `(all)intext:` worden gebruikt om documenten te vinden die de opgegeven zoekterm(en) in de (body van de) tekst hebben en niet in de hyperlink van een ander document dat naar de gezochte documenten verwijst. In Google Scholar kan het prefix ook worden gebruikt, en heeft een effect op de zoekresultaten; maar hoe het werkt, heb ik niet kunnen achterhalen.

`site: worldcatlibraries.org` beperkt het zoekresultaat tot de [BOOK]- en (een gedeelte van) de [CITATION]-treffers en kan dus met een minteken ervoor ook worden gebruikt om alle BOOK- en een deel van de CITATION-treffers van een zoekresultaat uit te sluiten.

Google Scholar indexeert veel publicaties die in de eerste plaats voor papier bedoeld zijn, niet als webdocumenten. Vandaar dat prefixen als `inanchor:` en `link:` in deze zoekmachine weinig betekenis hebben. Zoals we bij de algemene Google hebben gezien

vindt `inanchor:[zoekterm]` documenten waarnaar door andere documenten met de zoekterm in het anchor gelinkt wordt. Wetenschappelijke publicaties bevatten weliswaar talrijke verwijzingen naar oudere literatuur, maar dat zijn papieren verwijzingen, geen links; en als er al eens sprake is van een link, dan wordt het document waarnaar wordt verwezen vrijwel steeds aangeduid met de titel die de auteur eraan gegeven heeft. In de algemene Google kunnen we `intitle:[zoekterm]` uitbreiden met `OR inanchor:[zoekterm]` in de hoop daarmee niet alleen documenten te vinden die door hun auteur met die zoekterm benoemd zijn, maar ook die dat door anderen zijn gedaan. In Google Scholar is de kans dat we die extra documenten vinden voornamelijk niet erg groot. Dat zal anders worden naarmate de inhoud van institutionele 'repositories' een groter deel van de database van Google gaat uitmaken.

Ook het prefix `link:` heeft in Google Scholar niet veel belang, aangezien, zoals we zullen zien, de links al bij de betreffende artikelen in de resultatenlijst worden vermeld en desgewenst getoond.

`filetype:` is in Google Scholar eigenlijk alleen bruikbaar om documenten met de extensie `.PDF` of `.PS` te vinden; verreweg de meeste documenten die Google Scholar vindt hebben geen extensie, omdat ze uit databases zoals PubMed worden gegenereerd. In de algemene Google wordt bij de meeste treffers de link naar 'Similar pages' aangeboden, een mogelijkheid waarvan daar ook gebruik kan worden gemaakt door in de Basic Search direct `related:` gevolgd door de URL van een webpagina op te geven; evenzo verschaft het prefix `info:` in de algemene Google informatie over nog andere aspecten van een pagina (zie daar). Noch van `related:` noch van `info:` heb ik bij Google Scholar voorbeelden kunnen vinden.

Het prefix `numrange:` doet wat het moet doen: documenten vinden met getallen tussen de opgegeven waarden. `cache:` ten slotte werkt eveneens zoals in de algemene Google. Wat de operatoren betreft werkt Google Scholar op dezelfde manier als de algemene Google, inclusief de mogelijkheid `*` in een phrase een of meer malen als joker voor een willekeurig woord te gebruiken. De enige uitzondering vormt de synoniemoperator `~`, die door Google Scholar niet lijkt te worden ondersteund.

Zoals te verwachten biedt de Advanced Search een aantal mogelijkheden van de Advanced Search van de algemene Google niet. Het aantal treffers per pagina kan wel worden ingesteld, maar de keuzes voor taal (via de Language Tools van de algemene Google), land, datum en Safesearch ontbreken. Daartegenover staat dat gefilterd kan worden op subject area.

Wat de overige zoekeigenschappen van Google Scholar betreft verwijs ik naar het overzicht voor de algemene Google op blz. 101-114. Er zijn een paar kleine uitzonderingen:

1. bij geaccentueerde termen van het type *électricité* lijkt Google Scholar geen onderscheid te maken tussen *électricité* en *+électricité*.
2. Google Scholar heeft geen optie Translate this page bij treffers in andere talen dan het Engels.
3. Google Scholar past geen automatic stemming toe.

Tot slot: de algemene Google heeft een indexeerlimiet van 101K voor HTML-documenten, met een limiet bij 500K voor sommige documenten. Maar de omvang van een doorsnee webpagina bedraagt niet meer dan 10-20K, dus voor verreweg de meeste gevallen is dat wel genoeg. Tijdschriftartikelen zijn gemiddeld heel wat langer, dus je zou verwachten dat de indexeerlimiet voor Google Scholar hoger zou liggen. Dat doet hij ook, maar niet veel hoger: hij ligt bij ca. 250K voor HTML-documenten, en vermoedelijk even hoog voor PDF-bestanden.

Het zoekresultaat

Omdat Google Scholar diverse materiaaltypen – fulltext publicaties, database- en catalogusrecords, literatuurverwijzingen – bestrijkt en zijn treffers uit diverse bronnen haalt, en omdat het zich bovendien ten doel stelt de 'citation popularity' van publicaties te melden, is het lezen van een resultatenlijst van deze zoekmachine geen sinecure.

Op het eerste gezicht heeft een resultatenlijst veel weg van die van de grote webzoekmachine. Ook hier kan de URL worden gemanipuleerd (zie blz. 71 e.v.), komt Google met spellingsuggesties en bevat de resultatenlijst vaak een link naar naslaginformatie bij Answers.com. Maar de treffers zelf zien er anders uit.

In de eerste plaats zijn die verdeeld in drie typen: die welke worden ingeleid door de aanduiding [BOOK] of [CITATION] en die welke niet zo'n aanduiding hebben. De treffers met de aanduiding [BOOK] komen uit de WorldCat-catalogus die ook door de grote Google en Yahoo! wordt ontsloten, de [CITATIONS] zijn literatuurverwijzingen naar boeken, artikelen of andere documenttypen die Google Scholar aantreft in de fulltext publicaties die het indexeert. Deze publicaties zijn wel of niet gratis openbaar toegankelijk. Bij de [CITATIONS] zitten dus zowel literatuurverwijzingen uit artikelen als die van de *Annals of Internal Medicine*, een tijdschrift dat al zijn artikelen een half jaar na verschijning gratis toegankelijk maakt, en uit de fulltext publicaties die Google in de institutionele 'repositories' vindt, als citaties uit niet-gratis toegankelijke artikelen

[Advanced Scholar Search](#)[Scholar Preferences](#)[Scholar Help](#)**Scholar**Results 1 - 100 of about 111 for "[rhetorical argumentation](#)". (1.06 seconds)

[CITATION] Concept maps, Vee diagrams and **Rhetorical Argumentation** (RA) Analysis: Three educational theory- ...

M Ahlberg - *Third International Seminar on Misconceptions in Science and ...*, 1993

[Cited by 4](#) - [Web Search](#)

[CITATION] Concept maps, vee diagrams and **Rhetorical Argumentation** Analysis (RAA): Three educational theory- ...

M Aahlberg - *Proceedings of the Third International Seminar on ...*, 1993

[Cited by 2](#) - [Web Search](#)

[CITATION] Concept Maps, Vee Diagrams and **Rhetorical Argumentation** Analysis: Three Educational Theory-based ...

M Ahlberg - *The Proceedings of the Third International Seminar on ...*, 1993

[Cited by 2](#) - [Web Search](#)

[A mental model for a rhetorical arguer](#)

F Grasso - *Proceedings of EuroCogsci, 2003* - [csc.liv.ac.uk](#)

... Liverpool, UK Abstract This paper proposes a general mental model for an agent capable to engage in **rhetorical argumentation**. The model ...

[Cited by 3](#) - [View as HTML](#) - [Web Search](#)

[Towards a Framework for Rhetorical Argumentation](#)

F Grasso - *EDILOG, 2002* - [ltg.ed.ac.uk](#)

... Towards a Framework for **Rhetorical Argumentation** Floriana Grasso Department of Computer Science University of Liverpool [floriana@csc.liv.ac.uk](#). ...

[Cited by 1](#) - [View as HTML](#) - [Web Search](#) - [csc.liv.ac.uk](#)

[BOOK] **Rhetorical argumentation** in Philo of Alexandria

M Alexandre - 1999 - Atlanta, Ga.: Scholars Press

[Cited by 1](#) - [Web Search](#) - [Library Search](#)

Resultatenlijst van Google Scholar, met de drie typen treffers: CITATIONS, BOOKS en artikelen, abstracts etc.

bij distributeurs zoals Ingenta. Die artikelen zijn door Google Scholar wel fulltext ontsloten, maar je krijgt ze pas te zien als je bij Ingenta en andere distributeurs of uitgevers de creditkaart getrokken hebt.

Ook bij de treffers die niet als [BOOK] of [CITATION] in de resultatenlijst verschijnen kan het om verschillende documenttypen gaan. Voor zover het artikelen betreft – die het grote merendeel van de door Google Scholar gedekte publicaties uitmaken – is er sprake van ofwel abstracts/summaries uit databases als PubMed of Ingenta (dat zijn abstracts wel gratis openbaar toegankelijk maakt), ofwel fulltext artikelen.

Alle treffers in Google Scholars resultatenlijst beginnen met de titel van de gevonden publicatie, met daaronder wel of niet een link. [BOOKs] en [CITATIONS] hebben geen link.

De simpelste treffers zijn die welke als [BOOK] zijn aangeduid. Deze komen uit de WorldCat database en hebben behalve een niet doorgelinkte titel een snippet met de voornaamste bibliografische gegevens: auteur(s), plaats van uitgave en uitgever, met daaronder (meestal) drie links: 'Cited by', gevolgd door het aantal van de publicaties waarin Google Scholar het boek geciteerd heeft gevonden, 'Web Search' en 'Library Search'. Klikken op 'Web Search' leidt tot een zoekactie met een auteursnaam en een phrase van enkele titelwoorden in de zoekmachine van Google, die vaak tot resultaat heeft dat het boek bij de uitgever of een internetboekhandel wordt gevonden. Klik je op 'Library Search', dan komt Google Scholar met de titelbeschrijving van het boek uit WorldCat, met daarbij een dichtstbijzijnde (Amerikaanse) bibliotheek die het boek in de collectie heeft of de mogelijkheid daarnaar te zoeken. Daarbij kan ook op land worden gezocht, maar aangezien WorldCat maar heel weinig collecties van Nederlandse bibliotheken bevat, heeft dat voor ons weinig nut.

Auteursnaam, trefwoorden en een 'Related Title/Author Search'/Other editions of item' in de titelbeschrijving zijn gelinkt en leveren, indien aangeklikt, zoekacties met (een deel van) de termen in Google Scholar op. Het is van belang onderscheid te maken tussen de 'Web Search' die als link bij een [BOOK] wordt aangeboden, en de zoekacties die via links bij titel, auteur en 'subjects' in het WorldCat-record kunnen worden uitgevoerd. Bij het eerste wordt met een aantal zoektermen, ontleend aan auteurs- en titelveld, een zoekactie in de grote mondiale Google uitgevoerd, bij het tweede gebeurt hetzelfde in Google Scholar, met de beperking `site:worldcatlibraries.org`. Daarbij blijkt dat in Google Scholar treffers kunnen worden gevonden die in de zoekmachine ontbreken.

Net als bij de [BOOKs] vindt u in de snippet van de [CITATIONS] (meestal) de belangrijkste bibliografische gegevens (auteursnamen, plaats van uitgave en uitgever als het om een geciteerd boek gaat, auteursnamen en naam van het tijdschrift als het een geciteerd tijdschriftartikel gaat). Daaronder vindt u bij vrijwel elke treffer van dit type overeenkomstige links als bij de [BOOKs]. De eerste daarvan is altijd 'Cited by ..', de opgave van het aantal verwijzingen dat Google Scholar naar de geciteerde publicatie heeft gevonden. Onder deze link zit de lijst van de verwijzende publicaties. Naast de link naar 'Cited by ..' zijn bij de [CITATIONS] ook de twee andere links, 'Web Search' en 'Library search' aanwezig, maar alleen als de geciteerde publicatie een boek is. In dat geval heeft klikken op de beide links dezelfde resultaten als hierboven voor de [BOOK]-treffers zijn beschreven. Maar als de [CITATION]-treffer een tijdschriftartikel is, ontbreekt de link naar de 'Library Search' en geeft de 'Web Search' een ander resultaat. Nu levert de zoekactie in de grote mondiale Google vaak ofwel een abstract in een database als PubMed of Ingenta op, ofwel, als u geluk hebt het full-text artikel. Mogelijk (maar niet gegarandeerd) zit dat trouwens behalve als [CITATION] ook als 'normale' treffer in Google Scholar zelf.

Het interessantst zijn natuurlijk de 'normale' treffers. Deze hebben een link naar (meestal) ofwel een abstract in een bibliografische database of bij een uitgever (waar u

[CITATION] LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development
 ... , R Baron, BR Olsen, ML Warman, G **Osteoporosis** ... - Cell, 2001
[Cited by 149](#) - [Web Search](#)

... of osteoporotic fractures in the United States in 1995: report from the National **Osteoporosis** ...
 NF Ray, JK Chan, M Thamer, LJ Melton III... - J Bone Miner Res, 1997 - ncbi.nlm.nih.gov
 Medical expenditures for the treatment of osteoporotic fractures in the United States
 in 1995: report from the National **Osteoporosis** Foundation. ...
[Cited by 328](#) - [Web Search](#)

Glucocorticoid-induced **osteoporosis**: pathogenesis and management
 BP Lukert, LG Raisz - Ann Intern Med, 1990 - ncbi.nlm.nih.gov
 ... Glucocorticoid-induced **osteoporosis**: pathogenesis and management. ... **Osteoporosis** occurs
 in at least 50% of persons who require long-term glucocorticoid therapy. ...
[Cited by 307](#) - [Web Search](#)

... of vertebral fracture risk in postmenopausal women with **osteoporosis** treated with raloxifene: ...
 B Ettinger, DM Black, BH Mitlak, RK Knickerbocker, ... - JAMA, 1999 - ncbi.nlm.nih.gov
 ... with **osteoporosis** treated with raloxifene: results from a 3-year randomized clinical
 trial. Multiple Outcomes of Raloxifene Evaluation (MORE) Investigators. ...
[Cited by 489](#) - [Web Search](#) - [ncbi.nlm.nih.gov](#)

The diagnosis of **osteoporosis**
 JA Kanis, LJ Melton, C Christiansen, CC Johnston, ... - J Bone Miner Res, 1994 - ncbi.nlm.nih.gov
 ... 1996 Aug;11(8):1199-201. The diagnosis of **osteoporosis**. ... Humans; Male; Middle Aged;
Osteoporosis/complications; Osteoporosis/diagnosis*; **Osteoporosis/physiopathology** ...
[Cited by 511](#) - [Web Search](#) - [ncbi.nlm.nih.gov](#)

... Alendronate on Bone Mineral Density and the Incidence of Fractures in Postmenopausal
Osteoporosis
 UA Liberman, SR Weiss, J Broll, HW Minne, H Quan, ... - Obstetrical & Gynecological Survey, 1996 - obgynsurvey.com
 ... Effect of Oral Alendronate on Bone Mineral Density and the Incidence of Fractures
 in Postmenopausal **Osteoporosis**. ... Transplantation **osteoporosis**. ...
[Cited by 531](#) - [Web Search](#) - [149.142.238.229](#) - [emedicine.bc.ca](#) - [ncbi.nlm.nih.gov](#) - [all 5 versions >](#)

Find: Find Next Find Previous Highlight Match case

http://scholar.google.com/url?sa=U&q=http://149.142.238.229/k30/reading/k30_Journal_Club_Geller.pdf

Nog een resultatenpagina van Google Scholar. De eerste treffer is een [CITATION], waarin de opmerkelijke auteursnaam G Osteoporosis voorkomt. De laatste treffer linkt naar het abstract van een artikel op de site 'obgynsurvey.com', terwijl er op het adres '149.142.238.229' ook een fulltextversie van het document in PDF (zie het adres in de statusregel onderin) beschikbaar blijkt te zijn.

in veel gevallen wordt doorgeleid naar de al dan niet gratis toegankelijke fulltextversie), of direct naar de fulltext zelf.

De snippets bij de 'normale' treffers in Googles resultatenlijst zijn uitgebreider dan die bij de andere twee typen. Naast auteursnamen en tijdschrifttitel en jaar vindt u hier in de tweede regel van de treffer de site vermeld waar Google Scholar de gelinkte titel van de treffer vandaan haalt. Verder bevat de snippet nog een of meer (vet afgedrukte) zoektermen met context en ten slotte een aantal links. Twee daarvan, 'Cited by ..' en 'Web Search' doen wat hierboven bij [BOOKS] en [CITATIONS] is beschreven, maar 'Library Search' ontbreekt hier en in plaats daarvan vindt men in een aantal gevallen diverse links die naar diverse versies van het betreffende document leiden.

Bij dit laatste is het oppassen geblazen. Zoals hierboven vermeld, staat in de tweede regel van een treffer vermeld waar Google Scholar de versie waarnaar de titel van de

treffer linkt, vandaan haalt. In veel gevallen zal dat een database als PubMed (ncbi.nlm.nih.gov) zijn, die een abstract laat zien. Maar het is heel goed mogelijk dat Google Scholar ook een fulltextversie van het artikel in zijn database heeft, en dat die onder een van de links zit die onder in een treffer beschikbaar zijn.

Soms is laatste regel in de snippet te kort om alle sites te bevatten waar een artikel, in wat voor versie ook, beschikbaar is. In dat geval is de laatste link er een naar 'all ... versions', die u beslist moet volgen om u ervan te vergewissen welke en wat voor versies van het artikel bestaan.

Het naast elkaar bestaan van een abstract- en een fulltextversie in Google Scholar is niet het grootste probleem dat Google Scholars zoekresultaten opleveren. Dat is veeleer het feit dat er nog andere versies kunnen bestaan, bijvoorbeeld een preprint en/of een ongereedige/niet aan peer review onderworpen versie op de site van de auteur, die eveneens in Google Scholars zoekresultaten verschijnen. Zoek bij meer versies altijd naar een fulltextversie waarvan duidelijk is dat ze al in een tijdschrift gepubliceerd is, en vraag u als die er niet is maar een andere wel, af of het wel verantwoord is u daarop te baseren. Kunt u het erop wagen de preprint te citeren als Ingenta u het uiteindelijk gepubliceerde artikel voor \$30 te koop aanbiedt? En kunt u erop vertrouwen dat Google Scholar ook de nadere toelichting heeft opgenomen die de auteur in het volgende nummer van het tijdschrift heeft gepubliceerd nadat hij een lading kritiek van vakgenoten over zich heen heeft gekregen?

Een sterk punt van Google Scholar is dat Googles ranking ook bij deze database in bedrijf is. Net als bij de zoekmachine is die gebaseerd op een combinatie van relevantie en 'link popularity', terwijl bij de artikelen die als 'normale' zoekresultaten verschijnen ook nog eens de aantallen publicaties worden vermeld waardoor het artikel geciteerd wordt ('Cited by ..'). De gebruiker die Googles ranking niet vertrouwt kan zo tussen de lager gerankte treffers nog eens op zoek gaan naar die welke een hoog 'Cited by'-aantal hebben. Er zijn professionals die vanwege dit extra aan Google Scholar de voorkeur geven boven PubMed (zie: Google Scholar Review (<http://kidneynotes.blogspot.com/2005/05/google-scholar-review.html>)).

Als u toegang hebt tot een wetenschappelijke bibliotheek, kan het zijn dat die bibliotheek Google Scholar heeft gecompleteerd met een zogeheten 'link referrer' zoals [Openly's OpenURL Referrer](#) die een directe, gratis toegang regelt tot artikelen in elektronische tijdschriften waarop de bibliotheek geabonneerd is. Hoewel dit een zeer waardevolle uitbreiding van Google Scholars mogelijkheden om aan literatuur te komen is, neemt het de tekortkomingen van de zoekmachine als pseudo-bibliografische database niet weg.

Google Scholar is de enige wetenschappelijke zoekmachine die gratis op grote schaal databases ontsluit, iets wat van zijn Elsevier-concurrenten [Scirus](#) (gratis) en [Scopus](#) (tegen betaling) niet kan worden gezegd. Het is dus een nuttige en nobele onderneming. Maar – mede doordat de zoekmachine hier moet concurreren met databases die beter op het materiaal toegesneden mogelijkheden bieden – laat hij een flink aantal wensen nog on vervuld.

Literatuur:

Jeroen Bosman, Eric Sieverts, Wetenschappelijk Googelen eenvoudiger en beter dan ooit? *Informatie Professional* 9 (2005) nr. 1 (januari), blz. 18-26. Een auteursversie (?) van het artikel is fulltext beschikbaar op het adres: <http://www.library.uu.nl/medew/it/eric/scholar.htm>.

Péter Jacso, *Google Scholar Beta*

(<http://www.galegroup.com/servlet/HTMLFileServlet?imprint=9999®ion=7&fileName=/reference/archive/200412/googlescholar.html>); idem, Google Scholar Redux (<http://www.gale.com/reference/archive/200506/google.html>)

Greg R. Notess, Scholarly Web Searching: Google Scholar and Scirus, *Online* 29, 4 (Jul/Aug 2005) <http://www.infotoday.com/online/jul05/OnTheNet.shtml>

M. Myhill, Review of Google Scholar (<http://www.charlestonco.com/review.cfm?id=225>)

J.-P. Lardy, Google Scholar (<http://urfist.univ-lyon1.fr/GoogleScholar.pdf>)
What's New, Doc?: BMJ vs Google vs Scholar: une étude comparative
(<http://whatsnewdoc.blogspot.com/2006/01/bmj-vs-google-vs-scholar-une-tude.html>)
A look at Google Scholar, PubMed, and Scirus: comparisons and recommendations
(http://www.slais.ubc.ca/courses/libr538f/04-05-wt2/giustini_barsky.pdf)
On Google Scholar (blog) (<http://schoogle.blogspot.com/>)
UBC Google Scholar Blog (<http://careo.elearning.ubc.ca/weblogs/googlescholar/>)

4.2.2.9.3 Google Book Search Bèta (<http://books.google.com/>)

Google Book Search (voorheen Google Print) is een zoekmachine die enigszins te vergelijken is met de hieronder in 4.2.6.1 besproken zoekmachine A9 van Amazon: de zoekmachine doorzoekt geen websites maar gedrukte boeken die door Google in samenwerking met een aantal uitgevers en een aantal grote Amerikaanse wetenschappelijke bibliotheken gescand en vervolgens geïndexeerd zijn; daarnaast worden ook abstracts (samenvattingen) van boeken doorzocht, die door uitgevers zijn aangeleverd.

In Google Book Search vindt u boeken over alle mogelijke onderwerpen, niet alleen wetenschappelijke publicaties. Er zijn erbij waar geen copyright meer voor geldt, en die kunt u in hun geheel doorbladeren, maar als er nog copyright voor geldt, of als niet zeker is dat dat niet zo is, krijgt u maar een beperkt gedeelte van de tekst te zien, afhankelijk van wat de uitgever met Google is overeengekomen. De zoekmachine is bedoeld om fulltext in een grote collectie te kunnen zoeken, niet om ze in hun geheel leesbaar te maken. Uitgevers en bibliotheken werken aan het project mee omdat bij de van hen afkomstige boeken die bij een zoekvraag gevonden worden, een link naar de uitgever (plus een aantal internetboekhandels), resp. de bibliotheken in Amerika wordt aangeboden.

Als u geen gebruik maakt van de mogelijkheden om uw zoekvraag met een prefix (zie onder) te beperken, zoekt u fulltext in het hele boekenbestand. Nadat u een of meer boeken hebt gevonden die in eerste instantie niet de gezochte informatie blijken te bieden (op de getoonde pagina), maar waarvan de titel en de inhoudsopgave, eventueel ook het register, doen vermoeden dat ze die informatie wel bevatten, kunt u vervolgens met een aangepaste vraagstelling in elk boek afzonderlijk zoeken ('Search within this book'). Bij elk boek waarvan een bladzijde getoond wordt, wordt onder 'About this book' een beknopte bibliografische beschrijving gegeven, een link naar de copyrightpagina, en een naar de inhoudsopgave.

Een zeer beperkt aantal boeken is in andere talen dan het Engels: Duits, Frans, Italiaans, Nederlands, Portugees en Spaans. Ze zijn met de natte vinger te identificeren aan de combinatie van wat lidwoorden of voorzetsels, in het Nederlands bijvoorbeeld de het een.

Gegeven het materiaal biedt Google Book Search wat andere zoekmogelijkheden, met bijbehorende prefixen, dan Google's webzoekmachine. Naast de prefixen `intitle:` en `allintitle:` kunnen ook `author:`, `isbn:` en `inpublisher:` worden gebruikt. (ISBN staat voor International Standard Book Number en is een tiencijferige code die de editie van een boek identificeert.). Andere prefixen, bijvoorbeeld `numrange:`, werken in Google Book Search niet, phrases en de OR-operator wel, maar de tilde (~) voor synoniemen weer niet.

Zie ook: About Google Print (<http://print.google.com/googleprint/about.html>)

Greg R. Notess, Searching Books Between the Covers, Online 29, no. 3 (May/June 2005)

(<http://www.infotoday.com/online/may05/OnTheNet.shtml>).

Michael Liedtke, Google Opens Digital Library

(http://news.yahoo.com/news?tmpl=story&u=/ap/20050901/ap_on_hi_te/google_books)

4.2.2.9.4. Google Labs (<http://labs.google.com/>)

Google Labs is de speeltuin van Google, zoals Yahoo! Next en MSN Sandbox die van Yahoo! en Microsoft zijn. De apparaten in die speeltuin – de glijbaan, de trampoline, het klimrek – vinden we links op het scherm in de gedaante van een aantal zoekmachines. Rechts op het scherm vindt u een aantal experimenten vermeld die kennelijk zijn uitontwikkeld, hoewel de meeste ervan nog als Bèta genoteerd staan. Drie interessante experimenten die nog gaande zijn, zijn – naast Google Scholar (zie blz. 135 e.v.), en Google Desktop Search – Google Suggest en Google Sets.

[Google Suggest](#) lijkt mij een veelbelovend experiment. Het 'suggest' bestaat erin dat Google u terwijl u een string intypt, de termen laat zien die met die string overeenkomen. Op elk punt waarop u de string intypt, laat Google Suggest u dus de alfabetische omgeving van de string tot op dat punt zien, en een van de functies daarvan zou kunnen zijn dat u zoektermen aan de hand worden gedaan die u anders ontgaan zouden zijn. Interessanter zijn echter de phrases die Google bij een uitgetypte zoekterm serveert. Deze kunnen de gebruiker helpen tot zeer precieze zoekacties te komen. Ze blijven ook geenszins tot phrases van twee of drie woorden beperkt, maar kunnen ook een string als 'Once upon a midnight dreary while I pondered weak and weary' (de beginregel van Poe's gedicht 'The Raven') omvatten.

Google's suggesties bevatten spelfouten, zo laat de string 'brt' al twee verkeerd gespelde versies van de naam Britney Spears (met bijbehorende aantallen treffers) zien, dus de suggesties zullen wel gebaseerd zijn op de miljarden vragen die inmiddels aan Googles zoekmachine gesteld zijn. Daarmee is in overeenstemming het feit dat de suggesties niet precies in alfabetische volgorde worden gepresenteerd, zodat bij doortypen nog steeds weer nieuwe suggesties verschijnen.



As you type, Google will offer suggestions. Use the arrow keys to navigate the results. [Learn more](#)

Resultaat van het intypen van de string 'ik gi' in het zoekvenster van Google Suggest.

Officieel werkt Google Suggest alleen voor het Engels maar 'Berend Botje' blijkt er toch ook al uit te komen. Als je deze oplossing vergelijkt met de suggesties die andere zoekmachines bij te brede vraagformuleringen doen, eveneens op basis van eerdere zoekacties, is ze minder voor de hand liggend, maar eigenlijk inventiever.

[Google Sets](#) is een zoekmachine die items in een verzameling probeert te voorspellen, uitgaande van een paar zoektermen die andere items in die verzameling benoemen. Een heel simpel voorbeeld is de verzameling: kleuren. Afzonderlijke kleuren in die verzameling zijn rood, oranje, geel, blauw enz. Door Google Sets red, orange, blue op te geven proberen we andere kleuren te achterhalen (Sets werkt alleen in het Engels). Je kunt de vraag die zo aan Google Sets wordt gesteld vergelijken met de vraag "the various colors are" in de gewone Google in de hoop daarmee een lijst van kleuren te

vinden.



Labs.google.com, Google's technology playground.

Google labs showcases a few of our favorite ideas that aren't quite ready for prime time. Your feedback can help us improve them. Please play with these prototypes and send your comments directly to the Googlers who developed them.

New! [Google Related Links](#)

Provide your website's visitors with links to useful information
4/3/06 - [Give us feedback](#) - [Discuss with others](#)

New! [Google Mars](#)

View some of the most detailed maps of Mars created by NASA scientists
3/13/06 - [Give us feedback](#) - [Discuss with others](#)

New! [Google Page Creator](#)

Create your own web pages, quickly and easily
2/23/06 - [Give us feedback](#) - [Discuss with others](#)

[Google Dashboard Widgets for Mac](#)

Quickly access Google products from within Dashboard
2/23/06 - [Give us feedback](#) - [Discuss with others](#)

[Google Extensions for Firefox \(Two new extensions\)](#)

Add powerful new tools to your Mozilla Firefox browser
12/13/05 - [Give us feedback](#) - [Discuss with others](#)

[Google Transit](#)

Plan trips using public transportation (currently in Portland only)
12/7/05 - [Give us feedback](#) - [Discuss with others](#)

[Google Reader](#)

Use Google's web-based feed reader to keep up with what's important to you
10/7/05 - [Give us feedback](#) - [Discuss with others](#)

[Google Video](#)

Search TV programs and videos
9/27/05 - [Give us feedback](#) - [Discuss with others](#)

[Google Web Accelerator](#)

Save time online by loading web pages faster
5/4/05 - [Give us feedback](#) - [Discuss with others](#)

[Google Ride Finder](#)

Find a taxi, limousine or shuttle using real time position of vehicles
3/30/05 - [Give us feedback](#) - [Discuss with others](#)

[Google???????????](#) (Google Suggest in Japanese)

??
3/8/05 - [???](#) - [???????????](#)

[Google Suggest](#)

As you type your search, Google offers keyword suggestions in real time
12/9/04 - [Give us feedback](#) - [Discuss with others](#)

[Site-Flavored Google Search Box](#) Improved!

Get a search box that customizes results based on your website
6/17/04 - [Give us feedback](#) - [Discuss with others](#)

[Froogle Mobile US](#) | [Froogle Mobile UK](#) Improved!

Search for products from your mobile phone using Froogle
2/24/04 - [Give us feedback](#) - [Discuss with others](#)

[Google Compute](#)

Donate your computer's idle time to help scientific research
3/26/03 - [Give us feedback](#) - [Discuss with others](#)

[Google Sets](#)

Automatically create sets of items from a few examples
5/20/02 - [Give us feedback](#) - [Discuss with others](#)

Graduates of Labs

- [Personalized Search](#)**
Get the search results most relevant to you
- [Personalize Your Homepage](#)**
See information you care about on your Google homepage
- [Google Maps](#)**
View maps, get driving directions, and search for local businesses and services.
- [Google Scholar](#)**
Search through journal articles, abstracts and other scholarly literature
- [Google SMS \(US\)](#) | [Google SMS \(UK\)](#)**
Get precise answers to specialized queries from your mobile phone or device
- [Google Desktop](#)**
Find all your information, whether it's on the web or stored on your computer, from one convenient location
- [Google Groups 2](#)**
Create and join searchable discussion groups and mailing lists
- [Google Deskbar](#)**
Search using Google without opening your browser
- [Web Alerts](#)**
Find out about new web pages on a topic of interest
- [Search by Location](#)**
Restrict your search to a particular geographic area
- [Google Glossary](#)**
Find definitions for words, phrases and acronyms
- [Google News Alerts](#)**
Specify a topic and receive email updates when news breaks

Links: Google Labs, Google's speeltuin.

Google Sets kan bij allerlei vragen worden gebruikt waarbij men vanuit een aantal bekende items van een verzameling hetzij het overkoepelende begrip hetzij andere items in dezelfde verzameling probeert te vinden. In plaats van de vraag "the European capitals are" in Google in te geven in de hoop daarmee de lijst van Europese hoofdsteden te vinden, voeren we Google Sets Amsterdam, Brussels, Helsinki (liever de minder bekende met bij voorkeur weinig andere gemeenschappelijke kenmerken dan het obligate rijtje Bonn, Londen, Parijs); in Google zelf kan dat natuurlijk ook – het is een van mijn favoriete zoekmethodes – maar Google Sets is gespecialiseerd in het vinden en rapporteren van de andere items in de verzameling.

[Feedback](#) [Discuss](#) [Terms of Use](#)



Predicted Items
Violin
Violoncello
Viola
Clarinet
Flute
Piano
Oboe
Guitar
Bassoon
Percussion
Saxophone
Harp
pianoforte

Een voorbeeld van een zoekactie met Google Sets. Hier is begonnen met de twee zoektermen *violin* en *violoncello*. Google kan een kleine set maken met minder dan 15 verwante termen, of, via *Grow Set*, een uitgebreide set, die in dit geval 47 instrumenten omvat.

Google Sets kan ook voor lastiger vragen worden gebruikt, in combinatie met de algemene Google. Bijvoorbeeld: u wilt weten welk aanvankelijk goedgekeurd antibioticum in de VS een jaar of tien geleden weer uit de markt is genomen. De naam van het geneesmiddel herinnert u zich niet, maar de laatste jaren zijn twee geneesmiddelen waarvan u de namen nog wel weet teruggetrokken: Vioxx en Baycol. Google Sets verschaft u met die twee namen een lijst van geneesmiddelen waarmee problemen zijn geweest. Met die lijst kunt u in Google (of een andere zoekmachine) op zoek naar geneesmiddelen die echt van de markt zijn gehaald. Daarbij zult u lijstjes en overzichten tegenkomen die er steeds meer bevatten, totdat u – eventueel met behulp van `numrange`: voor jaartallen binnen een reeks van jaren – een overzicht vindt van alle geneesmiddelen die over een langere periode zijn teruggetrokken. Zoeken met items in een verzameling is in het algemeen een bevredigende zoekmethode als niet ondubbelzinnig duidelijk is hoe de verzameling benoemd moet worden.

4.2.3. Yahoo!

4.2.3.1. In kort bestek

URL (Basic Search): <http://www.search.yahoo.com>, plus – onder Preferences – een groot aantal nationale versies van de zoekmachine die naast mondiaal ook nationaal zoeken (op taal en binnen een land) mogelijk maken.

Omvang : volgens eigen opgave: ruim 19 miljard; volgens een studie van Gulli en Signorini (<http://www.cs.uiowa.edu/~asignori/web-size/>) indexeert Yahoo! ca. 8 miljard pagina's, iets meer dan 69% van het zichtbare Web.

Indexeert tot: ca. 150 K (HTML), ca. 33 K (PDF) [?] ([voorbeeld](#); vgl. [deze phrase](#) op blz. 5 van hetzelfde document)

Maximale omvang zoekargument in het zoekvenster: 100 tekens; **NB:** deze beperking kan worden omzeild door via de URL te zoeken of door een korte zoekvraag vanuit het resultatenscherm uit te breiden.

Indexeert andere formaten dan HTML: ja (zie ook:

<http://www.faganfinder.com/filetype/>)

Toont: nooit meer dan 1000 treffers

Updating: enkele minuten (News) – enkele maanden

Help en info: <http://help.yahoo.com/help/us/ysearch/>

Advanced Search: <http://search.yahoo.com/web/advanced?ei=UTF-8>

Directory: <http://dir.yahoo.com/>

Speciale zoekmachines: onder meer: <http://search.yahoo.com/images?&ei=UTF-8&p>

(afbeeldingen), <http://video.search.yahoo.com/?&ei=UTF-8&p> (video),

<http://search.yahoo.com/news?&ei=UTF-8&p> (nieuws), <http://ask.yahoo.com/ask/> (Ask

Yahoo!), <http://answers.yahoo.com> (Yahoo! Answers)

Testomgeving: <http://next.yahoo.com/>

Blog: <http://www.ysearchblog.com/>

Meer over Yahoo!:

~Inktomi's Search Syntax~: <http://www.searchlores.org/inktomi.htm>

Yahoo! Review on Search Engine Showdown

<http://www.searchengineshowdown.com/features/yahoo/review.html>

Google – Yahoo Comparison: Compare the Search Results

<http://www.googleguy.de/google-yahoo/>

Abondance : Fiche descriptive du moteur Yahoo! Search Technology

<http://outils.abondance.com/yahoo-moteur.html>

Overzicht

Soort zoekactie	Voorbeelden	Vindt:
Natuurlijke taal (alleen Engels!)	Can I use natural language in Yahoo! searching?	Probeer!
<i>Prefixes</i>		
intitle: (komt overeen met "Show results with the ... in the title of the page" in de Advanced Search)	intitle:databases religion religion intitle:databases intitle:once intitle:upon intitle:a intitle:midnight	documenten met de opgegeven zoekterm of phrase in de titel van de pagina (en eventuele andere zoektermen op een willekeurige plaats in de pagina) NB: Yahoo! rekent ook termen die in de META keywords en description voorkomen tot de titel! documenten met de opgegeven phrase ("Once upon a midnight") in de titel; NB: let op de bijzondere syntax: het prefix intitle moet voor elke term in de phrase worden herhaald en het geheel moet tussen dubbele aanhalingstekens worden gezet
TITLEWORDS (...) ¹⁾	TITLEWORDS (Remkes AIVD) TITLEWORDS (Remkes AIVD) onzorgvuldige	documenten met de tussen de haakjes opgegeven zoektermen in de titel; kan worden gecombineerd met een term die elders in het document voorkomt Opmerking: deze qua syntax vreemde eend in Yahoo!'s bijt (de hoofdletters en de haken zijn verplicht) biedt dus meer mogelijkheden dan Google's allintitle:
inurl: (komt overeen met "Show results with the ... in the URL of the page" in de Advanced Search)	studiefinanciering inurl:minocw notulen inurl:minbuza notulen originurlextension:pdf site:hva.nl inurl:handleiding intitle:database insects inurl:search	documenten die de opgegeven zoekterm in de (hele) URL bevatten (en eventuele andere zoektermen op een willekeurige plaats in de pagina); voor grote precisie te combineren met site:

Soort zoekactie	Voorbeelden	Vindt:
link:	link:http://www.searchability.com link:http://www.nfdb.nl/ AND link:http://www.danskefilm.dk/ maritiem link:http://www.museumgids.nl/	documenten met een link naar het opgegeven adres documenten die een verwijzing naar zowel een Nederlandse als een Deense filmdatabase bevatten documenten die een verwijzing naar de Museumgids bevatten, en in de tekst de term maritiem link: verzamelt de verwijzingen naar een bepaalde pagina, niet naar een hele site; gebruik voor dat laatste linkdomain: NB: http:// verplicht Opmerking: Yahoo!'s zoekvenster accepteert in eerste instantie niet meer dan 100 tekens, maar doet dat in tweede instantie (dus vanaf een resultatenschermbild); een zoekactie naar een document dat naar twee of meer opgegeven URL's verwijst, bij Google niet mogelijk, kan dus bij Yahoo! worden uitgevoerd door Yahoo! eerst een opdracht met één link-prefix te laten uitvoeren en daar dan andere link-prefixes aan toe te voegen.
linkdomain: ¹⁾	linkdomain:www.hva.nl Vgl. bijvoorbeeld: linkdomain:uva.nl (732.000) linkdomain:leidenuniv.nl (378.000) linkdomain:uu.nl (797.000) linkdomain:uu.nl -site:uu.nl (272.000)	documenten met een link naar een willekeurige pagina van de opgegeven site; linkdomain kan worden verbonden met andere opdrachten, bijvoorbeeld site: linkdomain:... -site:... geeft alle links naar pagina's van de Universiteit Utrecht met uitzondering van die welke in pagina's van de Universiteit Utrecht zelf voorkomen ('zelfcitaties')
site: domain: ¹⁾ (komt overeen met "Site/Domain" in de Advanced Search)	site:nl site:hva.nl site:ie.hva.nl site:htsa.ie.hva.nl handleiding site:ie.hva.nl handleiding AND (site:htsa.ie.hva.nl OR site:home.ie.hva.nl) handleiding AND (site:ie.hva.nl OR site:itv.hva.nl) NOT site:htsa.ie.hva.nl	documenten met de opgegeven term(en) bij de opgegeven site (domain), m.a.w. beperkt de zoekactie tot het opgegeven domain (inclusief subdomains)
hostname:	hostname:ub.leidenuniv.nl handleiding	documenten met de opgegeven term(en) op de opgegeven computer, m.a.w. beperkt de zoekactie tot de opgegeven computer; voor 'hostname' zie: http://manual.liquidweb.com/chapter3/glossary.html
path: ¹⁾ originurlpath: ¹⁾	path:help intitle:database path:full	documenten met een URL waarin de opgegeven zoekterm in het "pad" staat (dus noch in de sitenaam, noch in de bestandsnaam); geeft te weinig treffers
url: originurl: ¹⁾	url:http://www.hva.nl	de pagina met het opgegeven adres; NB: http:// verplicht!
linkextension: ¹⁾	linkextension:pdf linkextension:avi beethoven linkextension:avi	pagina's met een of meer links naar (een) document(en) met de opgegeven extensie

Soort zoekactie	Voorbeelden	Vindt:
outgoingurltype: ¹⁾	outgoingurltype:text/plain (komt ongeveer overeen met linkextension:txt)	pagina's met een of meer links naar (een) document(en) met de opgegeven MIME type (voor MIME types, zie http://www.datatrendsoftware.com/automail_lite_mimeypes.html)
originurlextension: ¹⁾ (komt overeen met "File Format" in de Advanced Search)	"Mary Vetsera" originurlextension:pdf	(tekst)bestanden of webpagina's met een bepaalde extensie, onder meer de extensie PDF (Acrobat), maar ook b.v. TXT, XLS, PPT, DOC, XML; kan worden gecombineerd met andere zoektermen en syntax (zie ook: http://www.faganfinder.com/filetype/)
depth: ¹⁾	depth:2 depth:4 -depth:0 -dept:1 -depth2 -depth:3 "Google Scholar" site:uu.nl depth:0	alle documenten met een niveau van tot en met 2 in de directorystructuur van een site, gerekend vanaf de root (depth:0) documenten met een niveau van 4 of dieper (4 is de maximaal op te geven diepte en staat voor 4 en dieper) documenten met de opgeven zoekterm met het opgegeven niveau in de directorystructuur van een site

Soort zoekactie	Voorbeelden	Vindt:
feature: ¹⁾	<p>feature:acrobat feature:acrobat -linkextension:pdf</p> <p>feature:activex</p> <p>feature:applet</p> <p>feature:applet title:play title:chess feature:audio paganini feature:audio</p> <p>feature:flash feature:form</p> <p>feature:form title:dictionary Nederlands translate feature:frame</p> <p>feature:homepage</p> <p>feature:homepage publicaties Wouden</p> <p>feature:image</p> <p>"Mila Georgieva" feature:image</p> <p>feature:index feature:index intitle:"Roman empire"</p> <p>feature:javascript</p> <p>feature:meta feature:script</p> <p>feature:shockwave</p> <p>feature:table</p> <p>feature:title</p>	<p>webpagina's met (een link naar) een bepaald kenmerk</p> <p>een PDF-bestand niet door Yahoo! geïndexeerde PDF-bestanden</p> <p>een ActiveX component, d.w.z. code in de vorm <object ... classid="clsid:... > in HTML</p> <p>een applet, d.w.z. een <applet ...> tag in HTML (Yahoo! vindt alleen applets met deze tag, dus niet met de tags <object...> en <embed...></p> <p>een online schaakpartij</p> <p>een link naar een geluidsfile, binnen de tags en <area href=...>, maar niet "embedded" met de tags: <embed...>, <bgsound...> en <META refresh></p> <p>bestanden met de extensie .fla, .spl of .swf</p> <p>een form, d.w.z. een vak waarin de bezoeker gegevens (b.v. een vraag) kan invullen</p> <p>online woordenboeken Nederlands/ Engels en Engels/Nederlands</p> <p>webpagina's met frames (en dus waarschijnlijk een menuframe)</p> <p>persoonlijke pagina's (met een ~ in het pad)</p> <p>persoonlijke pagina met publicatielijst van Ton van der Wouden bij de Universiteit Groningen</p> <p>webpagina's met een verwijzing naar een afbeelding in de tag in HTML</p> <p>webpagina's met de afbeelding van mijn favoriete violiste Mila Georgieva</p> <p>de ingangspagina's tot hele sites</p> <p>sites die de opgegeven zoekterm of phrase in de titel hebben en waarschijnlijk geheel aan dit onderwerp gewijd zijn</p> <p>webpagina's met de <script ...> tag, met daarin als attribuut "JavaScript"; Javascript in andere tags wordt niet herkend; hetzelfde geldt voor links naar bestanden met de extensie .js</p> <p>webpagina's met META-tags</p> <p>webpagina met andere scripts (tag <script...>) dan Javascript</p> <p>webpagina's met links naar bestanden met de extensie dcr, dir, fla, spl of swf</p> <p>webpagina's met daarin tabellen (tag <table...>, vaak gebruikt voor opmaak documenten met – zoals het hoort – een titel; andere dan HTML-</p>

Soort zoekactie	Voorbeelden	Vindt:
	<p>site:hva.nl -feature:index -feature:title -originurlextension:doc -originurlextension:ppt -originurlextension:pdf -originurlextension:txt -originurlextension:xls -originurlextension:cdm -originurlextension:js feature:video</p> <p>feature:vrml</p>	<p>bestanden hebben in de regel niet een titel die door zoekmachines als zodanig geïndexeerd wordt webpagina's binnen een bepaalde site die geen titel hebben</p> <p>webpagina's met (een) link(s) naar (een) videobestand(en) met de extensie avi, mpg, mpeg, mov, etc.; pagina's met de tag <embed src=...> of worden niet gevonden webpagina's met (een) link(s) naar (een) bestand(en) met de extensie wrl, wrz, vrml</p>
region: ¹⁾	<p>region:africa region:asia region:centralamerica region:downunder region:europa region:mediterranean region:mideast region:northamerica region:southamerica region:southeastasia</p>	<p>documenten met een domein (ccTLD) dat behoort tot een van de genoemde regio's; welke landen tot de opgesomde regio's behoren, vindt u in een inmiddels gearchiveerd hulpbestand: http://web.archive.org/web/20001217082300/dpsearchno.spray.net/hjalp.html Denk erom dat region alleen de ccTLD's van de betreffende landen omvat, niet de generieke domeinen (bijvoorbeeld .com) die ook onder die landen vallen en die men in de Advanced Search met de keuze Country kan doorzoeken (samen met het betreffende ccTLD). Het gevolg is dat de zoekactie "Máxima Zorreguieta" region:southamerica minder treffers oplevert dan "Máxima Zorreguieta" met Country Argentina in de Advanced Search.</p>
<i>Operatoren</i>		
term1 term2 (komt overeen met "Show results with all of these words" in de Advanced Search)	Verdi Requiem	de ene EN de andere zoekterm (standaard)
term1 -term2 (komt overeen met "Show results with none of these words" in de Advanced Search)	<p>cancer -lung</p> <p>Bach -"Johann Sebastian"</p> <p>intitle:"Roman empire" -site:com</p>	<p>de ene maar niet de andere term (of phrase); kan ook worden gecombineerd met een prefix/limit</p> <p>documenten met de opgegeven zoekterm of -phrase in de titel, maar afkomstig van sites die niet de TLD com hebben NB: als u – gebruikt kunt u niet in dezelfde zoekexpressie Booleaanse operatoren gebruiken!</p>
AND	cancer AND nutrition	de ene EN de andere zoekterm

Soort zoekactie	Voorbeelden	Vindt:
OR (komt overeen met "Show results with any of these words" in de Advanced Search)	nutrition OR diet	de ene OF de andere zoekterm (of beide)
NOT ¹⁾ AND NOT ¹⁾	cancer NOT lung intitle:"Roman empire" -site:com	de ene maar niet de andere term documenten met de opgegeven zoekterm of -phrase in de titel, maar afkomstig van sites die niet de TLD com hebben
nesting ¹⁾	(cancer OR tumor) AND (breast OR lung OR intestinal OR intestines) AND (nutrition OR diet) intitle:"Roman empire" (site:edu OR site:org) intitle:"Roman empire" AND (site:edu OR site:org)	termen in zowel EN- als OF-relaties in één en hetzelfde zoekargument; door het tussen haakjes plaatsen van de door OR verbonden termen krijgt de OR-bewerking prioriteit boven de AND-bewerking Om onverklaarbare redenen werkt de vraag zonder de operator AND hier wel, en met AND niet. In andere gevallen – zie hierboven bij site: werkt de combinatie AND en OR wel. NB: laat u niet verleiden om zoekacties als deze via de Advanced Search van Yahoo! uit te voeren. Show results with all of these words: kanker voeding EN any of these words: groenten fruit levert de volgende zoekexpressie op (met het bijbehorende aantal treffers): kanker voeding groente OR fruit, d.w.z. documenten die over kanker, voeding en groenten gaan, en documenten over fruit. (Google doet dit wel correct, ook al ziet de zoekexpressie er net zo uit.)
<i>Nabijheid</i>		
phrase ("...") (komt overeen met "Show results with the exact phrase" in de Advanced Search)	"Willem van Hanegem" "part time" "nothing of the sort"	de opgegeven naam of zinsnede als reeks van opeenvolgende woorden, met tussen elke twee woorden een spatie of een ander leesteken NB: begin september 2005 en eind maart 2006 vond deze phrase ook treffers waarin noch de phrase, noch de afzonderlijke woorden 'nothing' en/of 'sort' voorkwamen
* (wildcard)	"computer * design"	berichten met de opgegeven phrase, waarin de wildcard * voor 0 of 1 woord staat (in dit geval zowel "computer design" als "computer-aided design" enz.)
<i>Overige</i>		
Number of results ²⁾	10/15/20/30/40/100	kies het aantal treffers per pagina van de resultatenlijst
Updated ²⁾	anytime/ within the past 3 months / within the past 6 months / within a year	documenten met als laatste wijzigingsdatum een datum binnen een van de opgegeven periodes

Soort zoekactie	Voorbeelden	Vindt:
Language ²⁾	Persian Thai	kies een van de 37 talen in het menu onder Any language in de Advanced search , de Preferences of de Language Tools ; de voorbeelden betreffen talen die niet door Google worden aangeboden)
Country ²⁾	Argentina	kies een van de 24 landen in het menu 'any country' in de Advanced Search of gebruik de Language Tools (waar u land en taal met elkaar kunt combineren) NB: deze wijze van zoeken verschilt van die met de beperking site:ar (voorbeeld: Argentinië) en levert een aanzienlijk groter resultaat.
SafeSearch ²⁾	Filter out adult Web search results / Do not filter Web results	schakelt het Family Filter van Yahoo! uit / aan Standaard is Moderate filtering; gebruik de mogelijkheden van Preferences of de Advanced Search om die instelling te veranderen
Creative Commons Search ²⁾	Search only for Creative Commons licensed content / Find content I can use for commercial purposes / Find content I can modify, adapt, or build upon	materiaal dat door de auteur/maker beschikbaar is gesteld onder minder stringente voorwaarden dan vastligt in het Amerikaanse auteursrecht; gebruik Advanced Search om naar materiaal te zoeken dat onder zo'n licentie beschikbaar is
subscriptions ²⁾	"Steve Ballmer"	als eerste treffers records die Yahoo! bij de zoekvraag heeft gevonden in databases voor het verdere gebruik waarvan een subscription (abonnement) benodigd is
accenten	é, ï, ö, û etc. electricite electricité électricité lange Länge ode öde blöd site:nl oölogie oologie site:nl "Munchausen by proxy" site:edu "Münchausen by proxy" site:edu ("Munchausen by proxy" OR "Münchausen by proxy") site:edu	Yahoo! maakt een onderscheid tussen wel en niet geaccentueerde letters, met dien verstande dat de ongeaccentueerde zoekterm meestal ook de geaccentueerde vormen vindt, maar niet andersom: vindt electricite, electricité, électricité electricité électricité Dit gaat echter niet altijd op: lange, maar niet Länge Länge, maar niet lange ode, maar niet öde öde, oede en øde, maar niet ode; merkwaardig genoeg ook Hasseröder en Oederquart naast blöd een beperkt aantal treffers op onder meer bloedverlies, bloedgeheimen treffers met oölogie maar niet met oologie 2 treffers op oologie 555 treffers 2 treffers 560 treffers [?]
hoofdletters/ kleine letters (case-sensitivity)	next NeXT	Yahoo! maakt geen onderscheid tussen hoofdletters en kleine letters

Soort zoekactie	Voorbeelden	Vindt:
leestekens	<p>-,/_&+#\$</p> <p>part-time part-time NOT "part time" "part time" parttime e-merge c+ c++ &c single_user a# c# f# g# a&b&c</p> <p>a_b_c \$5000 €5000 \$134.50 I/O #2 >2</p>	<p>Yahoo! verwaarloost de meeste lees- en niet-lettertekens, of vat ze op als spaties (en legt dus een EN-relatie); punten en komma's in getallen worden als spaties opgevat, zodat getallen ook als deel van een groter getal kunnen worden gevonden (mits voorafgegaan of gevolgd door een punt of een komma)</p> <p>part AND time, maar niet parttime part AND time, maar niet als phrase "part time", inclusief part-time parttime e-merge c, inclusief c+ en c++, maar ook CNN, CIA etc. c++ zelfde resultaat als c "single user" a# c# f# g# a&b&c, maar ook a&b=c, A&B AND C etc. "a b c" 5000 5000 134 AND 50 I AND O 2 2</p>
getallen (zie ook: Codes onder Shortcuts)	<p>10000 10,000 14.668</p> <p>25.32 25,32 "14.668" "25,32"</p>	<p>10000, maar niet 10,000 10,000, maar ook 10 AND 000 14 AND 668 (inclusief 14,949 AND 78,669); Yahoo! leest komma's en punten in getallen als spaties 25 AND 32 25 AND 32 Los dit probleem (voor een deel) op door de getallen als een phrase te schrijven: "14,668", "25,32"</p>
stopwoorden	to be or not to be	Yahoo! kent, voor zover bekend, geen stopwoorden
(vertaling) [Web Page Translation]	"histoire de la France"	treffers in het Frans, Duits, Italiaans, Spaans en Portugees hebben als extra in de snippets een keuze Translate this page (automatische vertaling in het Engels)
(spellingcontrole) [Spell Checker]	<p>seperate (spelfout)</p> <p>swlling (typfout)</p> <p>Rottedam (typfout)</p> <p>"Berend Botje ging uit varen" "Berend Botje ging eens varen" "Berend Botje gong eens varen"</p>	<p>seperate is a common misspelling of separate. Results for seperate are shown below. Did you want results for seperate? (Yahoo! neemt dus zelf het initiatief tot de verbetering, maar biedt als alternatief de resultaten bij de opgegeven term.) Did you mean: swelling?</p> <p>Did you mean Rotterdam?</p> <p>geen spellingsuggestie (correct) geen treffers, geen suggesties geen treffers, geen suggesties</p>

Soort zoekactie	Voorbeelden	Vindt:
suggesties voor verder zoeken/categorisatie	potatoes	Also try: scalloped potatoes, baked potatoes, sweet potatoes, mashed potatoes More... (Bij het gebruik van één zoekterm biedt Yahoo! suggesties om de zoekactie te preciseren)
[term weighting]: volgorde zoektermen	kanker voeding voeding kanker nieuwe lichte rekruten lichte nieuwe rekruten nieuwe rekruten lichte	als eerste treffers in de resultatenlijst, afhankelijk van de volgorde van de zoektermen, documenten waarin het ene dan wel het andere aspect van de vraag overweegt (op basis van de door Yahoo! gebezigde statistische criteria) De volgorde van de zoektermen is ook van belang omdat Yahoo! automatisch naar phrases zoekt. NB: het effect is, te midden van alle andere rankingfactoren, gering
[term weighting] toekenning van gewichten aan zoektermen	cancer 0.1*nutrition 9.9*cancer nutrition nutrition 0.1*cancer 9.9*nutrition cancer 9.9*nutrition 0.1*cancer	documenten met als centraal onderwerp kanker en als bijkomstige term voeding documenten met als centraal onderwerp voeding en als bijkomstige term kanker gewichten worden toegekend met gehele of gebroken getallen met altijd een cijfer achter de punt. Het standaardgewicht is 1.0.
[term weighting]: herhaling van zoektermen	voeding voeding voeding voeding kanker kanker kanker kanker kanker voeding	als eerste treffers in de resultatenlijst, afhankelijk van de relatieve ranking en frequentie van de zoektermen, documenten waarin het ene dan wel het andere aspect van de vraag overweegt Frequentie zoektermen op eerste resultatenpagina van 10, resp.: 20 maal voeding / 40 maal kanker 31 maal voeding / 35 maal kanker Er is een effect, maar dat lijkt omgekeerd aan het beoogde.
truncatie	n/a	n/a
case sensitivity	n/a	n/a
[(automatic) stemming]; (automatische) toevoeging van zoektermen met dezelfde stam als de opgegeven zoektermen)	n/a	n/a
<i>Shortcuts</i>		
Woordenboek, encyclopedie, thesaurus		

Soort zoekactie	Voorbeelden	Vindt:
define (alleen Engels!)	define serendipity	als eerste treffer een omschrijving van de opgegeven term uit The American Heritage® Dictionary; NB: geen dubbele punt achter define
facts	Boer War facts	als eerste treffer een kort artikel over de opgegeven term uit de Columbia Electronic Encyclopedia; NB: facts komt na de zoekterm
synonym (alleen Engels!)	synonym beauty	als eerste treffer synoniemen van de opgegeven term uit Roget's II: The New Thesaurus; NB: geen dubbele punt achter synonym
Nieuws		
(Beurs) Typ: quote [ticker symbol]	quote YHOO ¹⁾	financiële informatie over de opgegeven ticker van Yahoo! Finance
(Bioscopen) Typ: showtimes [postcode of plaatsnaam]	showtimes alameda showtimes 94501	informatie over bioscoopvoorstellingen in de opgegeven plaats / postcodegebied
(Nieuws) Typ: news [onderwerp]	news Katrina	de laatste krantenkoppen over het opgegeven onderwerp
(Sportuitslagen) Typ: [team] scores	Red Sox scores	Sportuitslagen van het opgegeven team. NB: volgorde [team] scores
Lokale informatie		
(Benzineprijzen) [Gas Prices]; alleen VS Typ: gas [postcode]; of: [postcode] gas	40515 gas lexington gas gas lexington	als eerste treffer benzineprijzen voor de opgegeven postcode of plaats
(Files) [Traffic Reports]; alleen VS Typ: [plaatsnaam] traffic	San Francisco traffic	als eerste treffer de verkeerstoestand in de opgegeven stad
(Gouden gids) [Yellow pages]; alleen VS Typ: [stad] [staat][bedrijfstype]	lexington ky bookstore	als eerste hit bedrijven in de opgegeven plaats + staat (in dit geval Kentucky)
(Kaarten) [Map]; alleen VS Typ: [nummer] [straat] [plaatsnaam] [staat]	4175 24th Street San Francisco CA	als eerste treffer een kaart van de omgeving van dit adres

Soort zoekactie	Voorbeelden	Vindt:
(Kengetallen) [Area code]; alleen voor de VS en Canada	418	als eerste treffer het driecijferige kengetal van het Noord-Amerikaanse telefoonsysteem
(Postcodes) [Zip Codes]; alleen voor de VS Typ: zip code [plaatsnaam]	zip code Alameda	als eerste treffer de postcodes van de opgegeven plaats
(Weer) [Weather]; werkt wereldwijd Typ: [plaatsnaam] weather	Alameda weather Rotterdam weather	als eerste treffer informatie over de weersomstandigheden in de opgegeven plaats; met weervoorspelling; werkt mondiaal
Reisinformatie		
(Flight Tracker); Amerikaanse binnenlandse vluchten en internationale vluchten van en naar de VS, Canada en het Caribisch gebied Typ: [maatschappij] [vluchtnummer]	KLM 6037	als eerste treffer informatie over de opgegeven vlucht
(Hotel vinden) [Hotel finder]; werkt wereldwijd Typ: [stad] hotels	Pittsburgh hotels Amsterdam hotels	als eerste treffer informatie over hotels in de opgegeven stad
(Luchthaveninformatie) [Airport Information]; alleen VS en Canada Typ: [plaatsnaam of luchthavencode] airport	Kennedy airport JFK airport New Orleans airport	als eerste treffer diverse info (kaartjes, weer, vluchtinfo etc.) over de betreffende luchthaven Voor luchthavencodes, zie: http://www.world-airport-codes.com/
(Wisselkoersen) [Exchange Rates] Typ: convert [muntsoort] to [muntsoort]	convert euro to pound convert 4300 pound to euro	1.0000 Euro = 0.6900 British Pound
Overige		
(Vliegtuigregistratienummer) [Aircraft Registration]; alleen in de VS geregistreerde nummers, beginnend met N Typ: [registratienummer toestel]	N2921A	als eerste treffer het record van het vliegtuig met het betreffende staartnummer in de FAA registry
(Boekprijzenzoeker) [Book Price Finder]; alleen VS Typ: [ISBN]	0786868716	als eerste treffer de gegevens van het boek met het betreffende ISBN met daarbij een of meer leveranciers met hun prijzen

Soort zoekactie	Voorbeelden	Vindt:
(Postpakkettenvolger) [Package Tracker]; UPS, Fedex, USPS Typ: [volgcode] (UPS) fedex [volgcode] (Fedex) usps [volgcode] (USPS)	1z9999999999999999	als eerste treffer de status van het pakket met de opgegeven volgcode
(UPC code) [UPC Codes]; dit is een twaalfcijferige streepjescode op producten Typ: [UPC-code]	636920005506	als eerste treffer info over het artikel met de opgegeven code in de UPC database
(Amerikaanse octrooi code) [U.S. Patents] Typ: patent [nummer]	patent 6269361	als eerste treffer informatie over het octrooi met het opgegeven nummer in de USPTO patent database
(Voertuigidentificatienummer) [Vehicle Identification Number (VIN)] Typ: [VIN]	zffyt53a510122724	als eerste treffer informatie over de auto met de opgegeven VIN-code
(Tijdzones) [Time Zones] Typ: time in [stad]	time in Berlin	als eerste treffer de tijd in de opgegeven plaats (of plaatsen: er zijn nog vier Amsterdams in de VS en één in Zuid-Afrika)
(Omrekenen van maten en gewichten) [Weights and Measures Converter] Typ: convert [aantal] [eenheden] to [eenheden]	convert 6 pounds to kilograms	de ene maat of het ene gewicht omgerekend in de andere maat of het andere gewicht

1) Niet door Yahoo! zelf gedocumenteerde mogelijkheden in Yahoo!

Bron: ~ Inktomi's search syntax ~ (<http://www.searchlores.org/inktomi.htm>)

2) Alleen in Advanced Search onder de Preferences en/of de Language Tools

4.2.3.2. Basic Search

Yahoo!'s simpele zoekregel is vanaf diverse adressen te gebruiken, afhankelijk van ieders persoonlijke voorkeur: behalve vanuit het eigenlijke Basic zoekscherm (<http://search.yahoo.com>) vanuit de portal (<http://www.yahoo.com>), de webgids (<http://dir.yahoo.com>) en vanzelfsprekend vanuit diverse plaatsen in de My Yahoo!-omgeving. De webgids biedt de keus tussen zoeken in de gids en zoeken op het Web en zou om die reden de voorkeur kunnen genieten; maar de zoekpagina is (via [Edit]) aan de eigen wensen aan te passen, en dat heeft ook zo zijn voordelen. Beide pagina's liggen maar één muisklik van elkaar. Ik bespreek hier de zoekpagina (<http://search.yahoo.com>).

Yahoo!'s Basic Search lijkt veel op die van Google, en bevat onder de zoekregel enkele regelmatig wisselende elementen (op het moment dat ik dit schrijf een uitnodiging om My Web BETA te gebruiken en links naar de eigen mailbox en de nieuwszoekmachine). Rechtsboven is verder een uitnodiging om de Yahoo! audiozoekmachine (bèta) te proberen.

Boven de zoekregel staan een aantal links naar speciale zoekdiensten van Yahoo!:

Images (plaatjes), Video (bewegend beeld), Directory (de webgids van Yahoo!), Local (de zoekmachine van Yahoo! die je lokaal in de VS naar producten en diensten laat zoeken), News (de nieuwszoekmachine van Yahoo!, met die van Google een van de beste nieuwszoekers die er zijn) en Shopping (de winkel- en prijsvergelijkingszoekmachine van Yahoo!). Rechts daarvan staat een link naar [Edit], waaronder deze selectie kan worden aangepast en/of uitgebreid met Maps (kaarten), People Search, Travel en Jobs.

[Yahoo!](#)

[Yahoo! Audio Search Beta - Find podcasts, music...](#)

YAHOO! SEARCH

[Web](#) | [Images](#) | [Video](#) | [Directory](#) | [Local](#) | [News](#) | [Shopping](#) [Edit]

[Advanced Search](#)
[My Web BETA](#)

Search the Web

[[Show Mail](#) - [Show News](#)]

Copyright © 2005 Yahoo! Inc. All rights reserved. [Privacy Policy](#) - [Terms of Service](#) - [Copyright/IP Policy](#) - [Submit Your Site](#) - [Job Openings](#)

U kunt in de zoekregel van de Basic Search van Yahoo! zoveel zoektermen invullen als u wilt, in Booleaanse zoekexpressies en met gebruikmaking van `site:` en `inurl-limits` (en alle andere mogelijkheden die de krachtige interactietaal van Yahoo! biedt). Rechts van de zoekregel vindt u links naar de Advanced Search en (momenteel My Web Bèta, de gepersonaliseerde dienst van Yahoo! die u een eigen kopie van interessante webpagina's op de servers van Yahoo! laat opbouwen, een zoekgeschiedenis laat bewaren en een community met anderen met dezelfde belangstellingen laat vormen. Preferences en Language Tools, die Yahoo! net als Google ook heeft zijn daardoor op deze plaats (tijdelijk) buiten beeld geraakt. Vanzelfsprekend behandel ik ze hier toch.

4.2.3.3. Language Tools (Taalhulpmiddelen, <http://tools.search.yahoo.com/language/>)

Onder de Language Tools biedt Yahoo! de volgende hulpmiddelen.

1. In de eerste plaats vindt u hier de mogelijkheid pagina's in een taal en in een land te zoeken, desgewenst gecombineerd. Het is hier dus bij voorbeeld mogelijk naar Nederlandstalige pagina's in Singapore te zoeken, die niet noodzakelijk de domain name `sg` hebben (zie het voorbeeld in de tabellen). Zoeken in een taal en in een land, en in een combinatie van beide, is ook in de Advanced Search beschikbaar.
2. De tweede mogelijkheid is een lokale versie van Yahoo! te gebruiken. Voor de zoekresultaten heeft dat gebruik in sommige gevallen een andere sortering van het zoekresultaat tot gevolg.
3. In de derde plaats kunt u Yahoo! stukken tekst (die u in een venster typt of plakt) in een andere taal laten vertalen, of de URL van een pagina opgeven waarmee dat moet gebeuren. Beschikbaar zijn de combinaties:
 - Engels naar Chinees, Nederlands, Frans, Duits, Grieks, Italiaans, Japans, Koreaans, Portugees, Russisch, Spaans en omgekeerd
 - Frans naar Nederlands, Duits, Grieks, Italiaans, Portugees, Spaans en omgekeerd

4. Ten slotte vindt u hier Yahoo! in 26 nationale en regionale versies van hetzij de zoek-, hetzij de portalpagina. Bij veel daarvan betekent 'nationaal' dat u (naast de keuze: het hele Web doorzoeken) de keus krijgt tussen zoeken naar pagina's in het land en pagina's in de taal. Bij andere is alleen de taal als keuze naast alle talen beschikbaar. Bij de regionale Yahoo! Asia kan in het hele Web worden gezocht, dan wel in Indonesische, Maleisische, Pakistaanse, Filippijnse, Thaise of Vietnamese pagina's; daarnaast vindt u er links naar nationale portalversies voor Indonesië, Maleisië, de Filippijnen en Thailand. Deze komen niet in de lijst van 26 voor.

NB Een aantal van de nationale zoekversies heeft ook een link naar een lokale zoekmachine van Yahoo!, die dus een aanvulling vormt op de uitsluitend Amerikaanse lokale zoekmachine van de Amerikaans/internationale Yahoo!

4.2.3.4. Preferences (Voorkeuren; <http://search.yahoo.com/search/preferences>)

Hier kiest u achtereenvolgens:

- of de resultaten waarop u klikt al dan niet in een nieuw venster moeten verschijnen. Als u ervoor kiest dat ze dat doen, kunt u ze bij browsers die met tabs werken, zoals Opera en Firefox, onder een nieuwe tab laten verschijnen door de instellingen van uw browser aan te passen. Overigens is de keuze om tressers in een nieuw venster te tonen ook op bij iedere treffer op een resultatenpagina beschikbaar (onder een icoontje achter de titel van de treffers)
- het aantal treffers per pagina: 10/15/20/30/40/100; kies minimaal 20
- SafeSearch Filtering strikt (standaard) / gematigd (alleen voor expliciete afbeeldingen) / geen filter; dit is Yahoo!'s Family Filter
- SafeSearch Lock: als geregistreerd Yahoo!-gebruiker kunt u hiermee de instellingen gematigd en strikt voor uw apparaat vastleggen
- de taal van de documenten waarin uw zoektermen moeten voorkomen; de keuze is hier alle talen (standaard) of een of meer geselecteerde talen. Er is geen limiet aan het aantal talen die u kunt aanvinken.

4.2.3.5. Advanced Search (<http://search.yahoo.com/web/advanced?ei=UTF-8>)

Dit scherm is verdeeld in negen secties, die ieder een aantal opties bieden om een zoekactie te specificeren. Alle opties kunnen gelijktijdig worden gebruikt, met uitzondering van Creative Common Search en Subscriptions.

1. een sectie waarin u Yahoo! gebruikt om zoektermen en/of 'phrases' in te vullen en daartussen relaties aan te brengen, het deel van het document waarin de zoektermen worden gezocht te beperken, en de 'versheid' van het document te specificeren. De keuzes voor de zoektermen zijn:
 - all of these words (EN-relatie)
 - the phrase (string van opeenvolgende woorden)
 - any of these words (OF-relatie)
 - none of these words (NIET-relatie)

De opties voor het documentgedeelte waarin de zoektermen moeten voorkomen, zijn:

- hele document (standaard)
- in de titel
- in de URL

De opties voor de actualiteit van de te vinden documenten zijn ongelimiteerd (anytime, standaard) / laatste drie maanden / laatste zes maanden / afgelopen jaar.

2. In de tweede sectie kunt u de zoekactie beperken tot een domain of site. Mogelijkheden naast de vier meest gebruikte waarvoor met de keuzerondjes kan worden gekozen zijn alle generieke TLD's (dus ook net, info etc.) en ccTLD's, en verder specifieke sites: nl, hva.nl, ie.hva.nl etc.
3. Een voorziening in bèta is de derde sectie (Creative Commons), die het mogelijk

Advanced Web Search

You can use the options on this page to create a very specific search. Just fill in the fields you need for your current search.

Yahoo! Search

Show results with

all of these words any part of the page
 the exact phrase any part of the page
 any of these words any part of the page
 none of these words any part of the page

Tip: Use these options to look for an exact phrase or to exclude pages containing certain words. You can also limit your search to certain parts of pages.

Updated

Site/Domain

Any domain
 Only .com domains Only .edu domains
 Only .gov domains Only .org domains

only search in this domain/site:

Tip: You can search for results in a specific website (e.g. yahoo.com) or top-level domains (e.g. .com, .org, .gov).

Creative Commons Search
BETA

Search only for **Creative Commons** licensed content
 Find content I can use for commercial purposes
 Find content I can modify, adapt, or build upon



Tip: This special Yahoo! Search finds pages that have content with a Creative Commons license. [Learn more...](#)

File Format Only find results that are:

SafeSearch Filter Applies to this search only:

Filter out adult Web search results - SafeSearch On
 Do not filter Web results (results may include adult content) - SafeSearch Off

Advisory: Yahoo! SafeSearch is designed to filter out explicit, adult-oriented content from Yahoo! Search results. However, Yahoo! cannot guarantee that all explicit content will be filtered out.

[Learn more](#) about protecting children online.

Tip: If you'd like to block explicit content for every search, you can set this in [preferences](#). Keep in mind that this filter may not block all offensive content.

Country

Language Search only for pages written in:

any language
or
 one or more of the following languages (select as many as you want).

<input type="checkbox"/> Arabic	<input type="checkbox"/> German	<input type="checkbox"/> Polish
<input type="checkbox"/> Bulgarian	<input type="checkbox"/> Greek	<input type="checkbox"/> Portuguese
<input type="checkbox"/> Catalan	<input type="checkbox"/> Hebrew	<input type="checkbox"/> Romanian
<input type="checkbox"/> Chinese (Simplified)	<input type="checkbox"/> Hungarian	<input type="checkbox"/> Russian
<input type="checkbox"/> Chinese (Traditional)	<input type="checkbox"/> Icelandic	<input type="checkbox"/> Serbian
<input type="checkbox"/> Croatian	<input type="checkbox"/> Indonesian	<input type="checkbox"/> Slovak
<input type="checkbox"/> Czech	<input type="checkbox"/> Italian	<input type="checkbox"/> Slovenian
<input type="checkbox"/> Danish	<input type="checkbox"/> Japanese	<input type="checkbox"/> Spanish
<input type="checkbox"/> Dutch	<input type="checkbox"/> Korean	<input type="checkbox"/> Swedish
<input type="checkbox"/> English	<input type="checkbox"/> Latvian	<input type="checkbox"/> Thai
<input type="checkbox"/> Estonian	<input type="checkbox"/> Lithuanian	<input type="checkbox"/> Turkish
<input type="checkbox"/> Finnish	<input type="checkbox"/> Norwegian	
<input type="checkbox"/> French	<input type="checkbox"/> Persian	

Subscriptions
BETA Select from the sites below to include their content in your web search results.

To view content you will need a subscription to the site(s) below. For some sites you will need to pay for the subscription.

[Consumer Reports](#) [Forrester Research](#) [New England Journal of Medicine](#)
 [FT.com \(60 days\)](#) [IEEE publications](#) [TheStreet.com](#)
 [Factiva](#) [LexisNexis](#) [Wall Street Journal \(30 days\)](#)

[Learn more](#) about Yahoo! Search Subscriptions

Number of Results Display per page.

Yahoo! Search

maakt de zoekactie tot materiaal te beperken dat door de auteur/maker ter beschikking wordt gesteld tegen minder strikte voorwaarden dan in het Amerikaanse auteursrecht gelden. Zoek bijvoorbeeld naar vrij beluisterbare muziek van Beethoven met: `Beethoven feature:audio`

4. In de vierde sectie beperkt u de zoekactie tot een van de volgende bestandstypen:

- HTML (extensies: .htm, .html)
- Adobe PDF (extensie: .pdf)
- Microsoft Excel (extensie: .xls)
- Microsoft PowerPoint (extensie: .ppt)
- Microsoft Word (extensie: .word)
- RSS/XML (extensie: .xml)
- tekst (extensie: .txt)

Standaard is: alle bestandstypen.

5. Per sessie kan verder Yahoo's Family Filter worden ingesteld. Op een meer permanente basis kan dat ook in de Preferences, waar bovendien de instelling kan worden vastgelegd als u een geregistreerde gebruiker van Yahoo!-diensten bent. Standaard is dat alleen mogelijk aanstootgevend beeldmateriaal wordt uitgefilterd.

6. Vervolgens kunt u in sectie 6 uw zoekactie tot een bepaald land,

7. en in sectie 7 tot een of meer talen beperken. In het laatste geval kunt u zoveel talen aankruisen als u wilt. Zie ook bij Preferences. Standaard voor 6 en 7 is: alle landen en alle talen.

8. Subscriptions Beta biedt u de mogelijkheid Yahoo! als zoekmachine te gebruiken voor diensten zoals Factiva en LexisNexis waarop u geabonneerd bent. Of u vervolgens toegang krijgt tot zo'n dienst hangt ervanaf of u op de plaats waar die toegang wordt geboden als abonnee wordt herkend. Subscriptions Beta is een van Yahoo!'s oplossingen voor het bestaan van het Invisible Web (zie blz. 233 vv.).

9. Ten slotte kan het aantal treffers per pagina op 10/15/20/30/40/100 worden ingesteld. Hier per sessie, onder Preferences op een meer permanente basis.

Ook bij Yahoo! geldt dat enkele instellingen uit de Advanced Search niet in prefixen of limits in de Basic Search te vangen zijn. Daarbij gaat het – naast de beperking tot Creative Commons-materiaal en die tot Subscriptions – om de Last Modified-datum, om het zoeken in een land of een taal en om de SafeSearch-instelling. Yahoo! lost drie van de vier problemen elegant op door deze opties ook onder Preferences (taal, SafeSearch) en onder Language Tools (taal en land) beschikbaar te stellen. Alleen de beperking van de Last-Modified-datum tot de laatste 3/6/12 maanden is alleen in de Advanced Search beschikbaar.

Wie de moeite neemt zich Yahoo!'s interactietaal eigen te maken – wat niet al te moeilijk moet zijn, gezien het feit dat Yahoo! opdrachten gedaan in de Advanced Search als commandoregel in de Basic Search terugmeldt – en die in de Basic Search te gebruiken, wordt daarvoor ruimschoots beloond. Enkele voordelen zijn:

- Booleaanse zoekmogelijkheden (met nesting)
- herhaling van prefixen/limits mogelijk, bijvoorbeeld `originurlextension:` voor een bestandsformaat
- de beschikbaarheid van een heel arsenaal aan prefixen/limits die mogelijkheden bieden die ver uitgaan boven wat de Advanced Search biedt

4.2.3.6. Yahoo!'s resultatenlijst

Evenals Google toont Yahoo! in de URL van zijn resultatenlijst zowel de vraag zoals die gesteld is en de daarbij gebruikte instellingen, als ook de plaats waar men zich bij het doorbladeren van de lijst bevindt. Als die URL wordt bewaard en later opnieuw naar Yahoo! opgestuurd, voert de zoekmachine de vraag weer met dezelfde instellingen uit en verschijnt dezelfde resultatenlijst, vanaf de opgegeven treffer. Bij Yahoo! is dat het element `&b=#` (`#` staat voor het nummer van de eerste treffer die moet worden getoond). Dit en een aantal andere elementen in de URL van de resultatenlijst van Yahoo! kan net zoals bij Google worden gemanipuleerd, maar de URL is een stuk onleesbaarder dan die van Google, en het is slechts in een enkel geval de moeite waard hem te ontleden.

De eigenlijke resultatenlijst opent met de zoekregel met daarin de vraag zoals u die in de Basic Search of via de Advanced Search gesteld hebt. Daaronder meldt Yahoo! in enkele regels welke instellingen gebruikt zijn in hetzij de Advanced Search, hetzij de Preferences. (Welke instelling waar is gekozen, staat er niet bij, dus als u zonder enige beperking wilt zoeken is het verstandig vooraf nog eens goed naar uw Preferences te kijken.) Als u een zoekactie vanuit de Language Tools hebt ondernomen, bijvoorbeeld met als taal Nederlands en als land België, krijgt u daarvan op het resultatenscherf geen melding en kunt u ook niet met die instelling verder blijven zoeken.

Direct daaronder staat de resultatenbalk, waarin Yahoo! links de doorzochte database vermeldt (Web, Images etc.) en rechts de resultatenscore met de nummers van de getoonde treffers, de zoektermen van de vraag en de tijd die Yahoo! aan de vraag besteed heeft. De zoektermen zijn aanklikbaar en leiden u dan naar Yahoo! Education, een verzameling naslagwerken (woordenboek, encyclopedie, thesaurus, World Factbook etc.) waar u standaard in het woordenboek belandt. Naast de spellingsuggesties die Yahoo! ook op de resultatenpagina toont, vindt u hier nog meer van zulke suggesties (alleen Engels). De door Yahoo! gebruikte naslagwerken zijn van goede kwaliteit, maar zijn iets minder webgeoriënteerd dan het door Google gebruikte Answers.com.

Wat nu volgt is afhankelijk van uw vraagstelling. Als die heel algemeen is en Yahoo! kan specifiekere zoekacties tonen, dan suggereert het die hier bij 'Also try'.

Als Yahoo! de zoekterm in zijn nieuwsdatabase of als shortcut heeft gevonden, volgen nu eerst enkele links daarnaar; daarna volgen eerst, tegen een lichtgekleurde achtergrond, enige 'sponsored results', en daarna de reguliere treffers ('natural results'). 'Sponsored results' vindt u ook aan de rechter- en onderkant, het laatste wederom tegen een lichtgekleurde achtergrond.

The screenshot shows the Yahoo! search interface. At the top, there are navigation links like 'Yahoo!', 'My Yahoo!', 'Mail', and 'Welcome, Guest [Sign In]'. Below that is a search bar with the text 'gender language' and a 'Search' button. The search results are displayed below the bar, showing two main results. The first result is 'Writing with Gender-Fair Language' with a brief description and a link to 'www.rpi.edu/dept/llc/writecenter/web/genderfair.html'. The second result is 'LEO Gender-free Writing' with a description and a link to 'leo.stcloudstate.edu/style/genderbias.html'. On the right side, there are 'SPONSOR RESULTS' for 'Language And Gender at Amazon.com'.

Een treffer uit de directory als eerste resultaat bij de vraag: gender language.

Bij de 'natural results' geeft Yahoo! een prominente plaats aan treffers die in zijn directory zijn opgenomen. Dat is zeker het geval als een of meer zoektermen in het directorypad van een categorie voorkomen. Bij zulke treffers toont Yahoo! die categorie met een link en is het verstandig daar ook eens in te kijken. In deze gevallen toont Yahoo! ook niet de snippet, maar de beschrijving die de redacteur van de webgidsrubriek van de site heeft gegeven.

De eerste regel van een treffer geeft de titel van het document weer, of als de spider van Yahoo! die niet heeft kunnen vinden, de URL. Deze eerste regel van een treffer is altijd gelinkt naar het achterliggende document. Erachter staat een icoontje dat twee venstertjes voorstelt. Als u hierop klikt in plaats van op de link, krijgt u de pagina die u opent in een nieuw venster, ongeacht of u dit onder Preferences zo hebt ingesteld. Zoektermen die Yahoo! in de titel heeft gevonden zijn daarin vet gemarkeerd, evenals in de snippets onder de titel. Bij Yahoo! bestaan die snippets uit (een aantal van) de zoektermen plus wat context, maar de lengte ervan is beperkt, dus als er veel zoektermen zijn, zult u alleen de eerste te zien krijgen.

Achter de titel kan nog het File Format vermeld staan als dat geen HTML is, en in veel gevallen ook een uitnodiging het document als HTML te bekijken. Voor een snelle check valt dit aan te bevelen boven het volgen van de link onder de titel, die naar het document in zijn oorspronkelijke formaat (.pdf, .doc etc.) maar voor het tonen waarvan het bijbehorende programma aanwezig moet zijn en moet worden gestart. Bij documenten in een formaat voor het lezen waarvan een Microsoft-programma nodig is, vermeldt Yahoo! welk programma dat is.

Onder de snippets volgt dan nog een regel met de URL van de treffer, de grootte van het document – ook geen onbelangrijke informatie –, een link naar de cacheversie en een naar More pages from this site. Anders dan Google toont Yahoo! maar één pagina per site, en deze link voert naar alle overige treffers die Yahoo! bij dezelfde site gevonden heeft.

Als alternatief kunt u op de laatste resultatenpagina gebruik maken van de mogelijkheid Yahoo! de zoekactie te laten herhalen inclusief de weggelaten zoekresultaten.

Sommige treffers kunnen nog een extra regel tussen de snippet en de slotregel met de URL etc. hebben, namelijk: 'RSS: View as XML - Add to My Yahoo!' Het gaat daarbij om RSS feeds, berichtgeving in een speciaal formaat dat door een speciaal programma op uw eigen computer of online gelezen kan worden en u dan de koppen van elektronische nieuwsbrieven, weblogs, nieuwsberichten van websites en dergelijke laat zien (met natuurlijk een link naar de berichten zelf). Kies 'View as XML' als u de URL van zo'n feed wilt zien en in een RSS reader op uw computer wilt invoeren. (Of klik bij Firefox met de rechter muisknop op de link en kies 'Copy link location'.) De andere keuze, 'Add to My Yahoo!' kunt u gebruiken als zich bij Yahoo! als gebruiker geregistreerd hebt en een eigen My Yahoo! pagina hebt samengesteld. Daarover hieronder meer.

Onderin de resultatenpagina vindt u Googles zoekregel herhaald, met de vraag erin, zodat u van hieruit een nieuwe zoekactie kunt starten of de zoekexpressie kunt veranderen.

4.2.3.7. Bespreking Yahoo!

Te midden van zijn concurrenten Google en MSN Search is Yahoo! een speler op de zoekmarkt met heel wat ijzers in het vuur. Een bedrijf bovendien dat al heel wat concurrenten heeft zien komen en gaan, want Yahoo! als gids van het Web bestaat al sinds 1994.

In de jaren sindsdien heeft Yahoo! 'properties' verworven op vrijwel elk terrein dat met informatie en communicatie te maken heeft en biedt diensten aan die variëren van zoekinstrumenten tot mailboxen, van een Ask an Expert-dienst via een bookmark

manager tot hosting van groepen. Deze kleine greep uit alles wat Yahoo! te bieden kan nog worden aangevuld met een al jaren lang bestaand en steeds verder uitgebreid personalisatieprogramma, My Yahoo!, dat het de gebruiker mogelijk maakt een vrijwel compleet pakket aan Yahoo! diensten samen te stellen. Wie zich bij My Yahoo! heeft aangemeld – en dat kan op vrijwel elke pagina van elke dienst in het Yahoo!-pakket – kan vervolgens een eigen pagina met RSS-feeds, bookmarks, mail, zoekresultaten enz. samenstellen. De laatste loot aan de Yahoo!-stam is een bèta van My Web, die het – met behulp van de Yahoo! toolbar, die in de browser wordt geïnstalleerd – mogelijk maakt een eigen bestand van kopieën van hele pagina's in de My Yahoo!-omgeving aan te leggen (en vervolgens te doorzoeken). Dit alles wordt aangeboden in een gebruiksvriendelijke verpakking die eveneens stoelt op jaren van ervaring. Fraai is dat vanuit My Yahoo! alle diensten bereikbaar zijn en aan de My Yahoo!-omgeving kunnen worden toegevoegd, terwijl My Yahoo! ook als men nog niet is ingelogd op vrijwel elke pagina van elke dienst binnen vingerbereik is.

Een overzicht van alle Yahoo!-diensten biedt Everything Yahoo! (<http://docs.yahoo.com/docs/family/more/>), en het is beslist de moeite waard die pagina eens wat nader te bekijken, want ze bevat het een en ander waarmee de zoeker zijn voordeel kan doen. Zo zijn er diverse naslagwerken, zoals een woordenboek, enkele encyclopedieën, een thesaurus en het World Book of Facts, goed bekend staan verder Yahoo!'s People Search (alleen VS), Yahoo! News, Yahoo! Local en Yahoo! Finance.

Eigenlijk zou de eigenlijke zoekmachine van Yahoo! binnen de context van het hele Yahoo!-aanbod besproken moeten worden, maar het is nu eenmaal een feit dat veel mensen net als ik een lappendeken aan voorzieningen voor hun internetbehoeften hebben: zoeken bij Google, een mailbox bij Hotmail, de bookmarks bij Furl, de RSS-feeds via een RSS-reader op het eigen apparaat enz. U als zoeker weet beter dan ik wat goed voor u is, en daarom beperk ik mij bij de bespreking van de zoekmachine van Yahoo! tot de zoekmachine sec, met wat extra's die daarmee nauw verbonden zijn. (My Yahoo! Search, Y!Q)

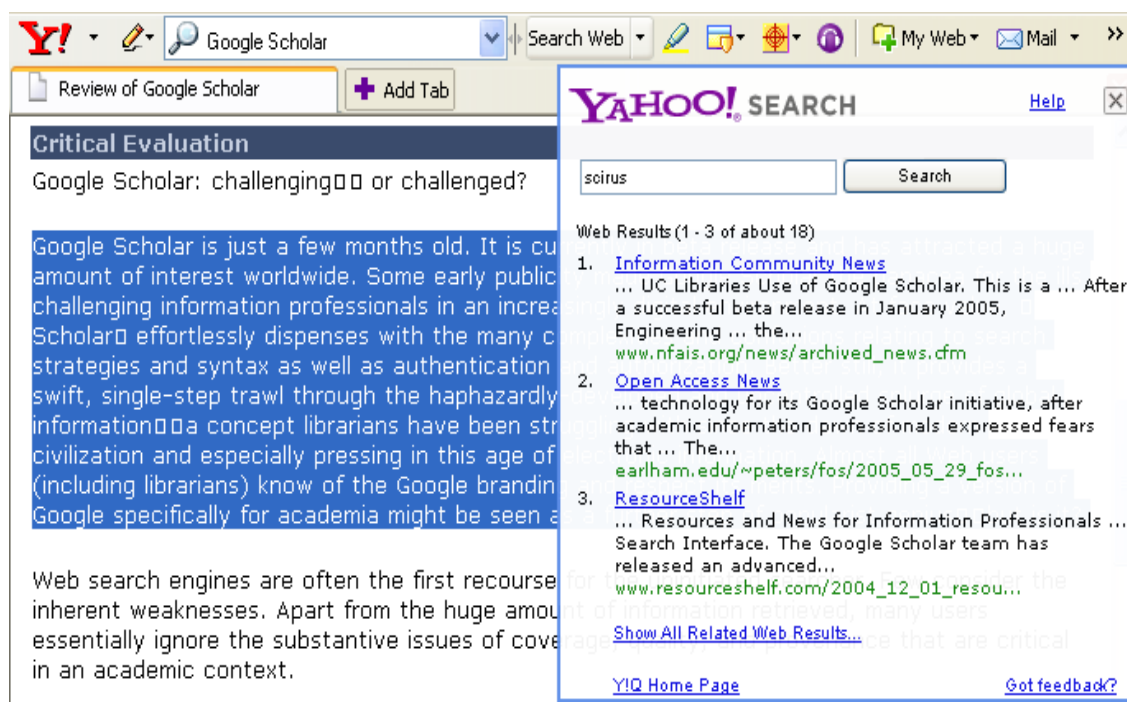
De ontwikkeling van Yahoo! tot een volwaardige zoekmachine is van recente datum, want nog maar een paar geleden was Yahoo! als zoekinstrument vooral een directory of webgids, die bij gebruik van de zoekregel naast resultaten uit de directory een aantal zoekresultaten van Google leverde. In die tijd waren AltaVista, AlltheWeb en Inktomi nog aparte bedrijven, maar inmiddels heeft Yahoo! die allemaal gekocht en daaruit een database en een zoekmachine gebouwd die als enige de vergelijking met Google kan doorstaan.

Over de hele linie blijft Yahoo! als zoekmachine iets bij zijn populaire concurrent achter. Zo is de database iets kleiner, zijn de zoekmogelijkheden iets geringer en is het oog voor detail bij Yahoo! iets minder ontwikkeld: Yahoo! heeft vermoedelijk ca. 8 miljard documenten in zijn database, ondersteunt minder bestandstypen en talen, het heeft bij voorbeeld geen numrange-prefix en het gaat minder zorgvuldig om met accenten, getallen en bijzondere en niet-lettertekens dan Google doet. Maar daar staat tegenover dat het zorgvuldiger omgaat met de melding van het aantal treffers, dat het vaak daadwerkelijk meer treffers toont dan Google uiteindelijk blijkt te doen, dat het geavanceerde Booleaanse zoekacties kan uitvoeren, dat het `link:`-zoekacties met andere prefixen kan combineren (en ook `link:... link:...`) en dat het met `linkdomain:` en `features:` over prefixen beschikt die tekortkomingen van Google kunnen compenseren. Niet onbelangrijk is ook dat Yahoo! bij langere documenten een groter deel indexeert en dus doorzoekbaar maakt dan Google, en dat het veel langere zoekexpressies toelaat.

4.2.3.8. Y!Q (<http://yq.search.yahoo.com/>)

Een van de sterke punten van de zoekmachine van Yahoo! is het schijnbaar onbeperkte aantal zoektermen dat hij aankan. Tegelijkertijd zal het u misschien zijn opgevallen dat Yahoo! in zijn resultatenlijst niet zoals Google een link naar 'Similar pages' heeft, en ook geen prefix *related:* kent. Die twee feiten gecombineerd vragen als het ware om een dienst zoals Yahoo! die in Y!Q heeft vormgegeven en die het aanduidt met de fraai klinkende benaming 'contextual search technology'.

Wat Y!Q doet is het volgende: u typt in natuurlijke taal (Engels) uw vraag in het zoekvenster, waarbij u erop let zoveel mogelijk specifieke termen te gebruiken die het onderwerp van uw vraag omschrijven; of u knipt een stuk tekst van een relevante webpagina en plakt dat in het zoekvenster. Als u dat als vraag opstuurt, analyseert Y!Q de tekst, haalt daar de (statistisch gezien) meest betekenisvolle termen en begrippen (phrases) uit en zoekt daarmee naar aan uw vraag of de geplakte tekst gerelateerde documenten. Het aardige is dat Y!Q de gekozen zoektermen en phrases ook toont, en u de minder relevante laat verwijderen: op het moment dat u een keuzevakje ontvinkt, krijgt u een nieuw resultaat. Tegelijkertijd is het ook mogelijk zoektermen aan uw vraag toe te voegen: bovenin het resultatscherm is het zoekvenster van Yahoo! beschikbaar, en als u daarin een of meer zoektermen invult en op Related Search klikt, wordt de zoekactie nogmaals uitgevoerd, met de nog aangevinkte zoektermen plus de in de zoekregel ingevulde. Jammer genoeg verschijnen de in de zoekregel ingevulde extra zoektermen niet als uitvinkbare termen in de resultatenlijst, dus experimenteren daarmee lukt niet.





Het handigst laat Y!Q zich gebruiken door de Yahoo! Toolbar te downloaden. Bovenin de afbeelding de toolbar zoals die in Internet Explorer verschijnt. Nadat met de zoekvraag Google Scholar is gezocht, is de link van een van de treffers gevolgd. In de tekst van het betreffende artikel is vervolgens een paragraaf gemarkeerd, waarbij op het scherm een Yahoo!-icoontje verscheen. Door hierin te klikken en "Search the Web for 'Google Scholar...'" te kiezen wordt vervolgens het Yahoo! Search venster opgeroepen, dat 3 van de 18 treffers bevat die Yahoo! op basis van de geselecteerde tekst heeft gevonden. De zoekregel gebruiken we om de vervolzoekterm *scirus* in te vullen.

scirus

Search Results Results 1 - 10 of about 29 for scirus - 0.04 sec.

 Results automagically refined using the following terms
 from: <http://www.charlestonco.com/review.cfm?id=225>
 google scholar beta release information professionals panacea

-  [News Results for google scholar](#)
[Google Health Next Niche Channel After Finance?](#) - Search Engine Journal
 - Mar 21 5:10 AM
[A feather in Skico's cap?](#) - The Aspen Times - 11 hours ago
[Scholars' Attack on Pro-Israel Lobby Met With Silence](#) - Forward - Mar 22
 11:45 PM
[Yahoo! Shortcut](#) - [About](#)

1. [Google Scholar Versus Metasearch Systems](#) 
 High Energy Physics Libraries Webzine. HEP Libraries Webzine. Issue 12 / February 2006. Google Scholar Versus Metasearch Systems. Tamar Sadeh, Ex Libris(*) 11/01/2006. Abstract ... However, a year after its beta release, Google Scholar is still facing a number of challenges that cause ... federated systems before Google Scholar. **Scirus** [5] is the first system that ...
library.cern.ch/HEPLW/12/papers/1 - 54k - [Cached](#) - [More from this site](#) - [More like this](#)

Het vervolg van de op de vorige pagina begonnen zoekactie. Yahoo! laat nu zien met welke termen en phrases uit de gemarkeerde tekst het de nieuwe zoekterm scirus gecombineerd heeft. Ieder van de door Yahoo! gekozen termen en phrases kan uit de selectie worden verwijderd en in de zoekregel kunnen nieuwe termen worden toegevoegd. Met 'Turn off related search' keert men van Y!Q naar Yahoo! Search terug.

Anders dan bij het gebruik van Yahoo!'s zoekpagina hebben treffers in de resultatenlijst van Y!Q wel een link naar eerdere pagina's, hier [More like this...](#) genaamd. Ook hier informeert Y!Q u over de zoektermen op basis waarvan het met zijn resultaten komt, en kunnen een of meer van die zoektermen worden uitgevinkt (met een onmiddellijk resultaat), dan wel nieuwe zoektermen in de zoekregel toegevoegd. Is Y!Q beter dan de Similar Pages van Google en soortgelijke functies in andere zoekmachines? Daar valt in zijn algemeenheid niets over te zeggen. Wel is Y!Q transparanter en flexibeler en kan het worden gebruikt om in een langere tekst razendsnel relevante termen en phrases voor een vervolgactie op te sporen.

4.2.3.9. My Yahoo!, MyYahoo! Search en MyWeb!

Zoekers hebben in toenemende mate behoefte aan een eigen plek op het Web waar ze hun spullen kunnen verzamelen en bewaren. Centraal staan daarin hun zoekinstrumenten, maar ook (de formulering van) zoekacties, de resultaten daarvan, bookmarks en RSS feeds (zie blz. 284) kunnen daarvan deel uitmaken, en, nog wat verder uit het centrum, een agenda, losse aantekeningen, hun post etc. Daarbij komt dat internet en de eigen computer als opslagmedia voor persoonlijke informatie steeds verder verweven raken, en als we nog een stapje verder gaan, zouden we sommige delen van die persoonlijke omgeving ook nog eens graag met een select gezelschap van anderen willen delen (op basis van voor wat hoort wat natuurlijk) en andere misschien met alle gebruikers van internet. Wat we hier in beeld krijgen zijn twee van de buzzwords van de huidige ontwikkeling van internet: *personalization*, de afbakening van een persoonlijke informatieruimte, en *collaboration*, het gezamenlijk genereren en

gebruik van informatie. Twee andere, even schijnbaar tegengesteld, zijn *localization*, de levering van informatie op de plaats waar ze gebruikt kan worden, en *global access*, de toegang tot informatie waar ook ter wereld.

Voorbeelden van deze algemene trends zijn er in de wereld van het zoeken te over.

Zonder er al teveel ruimte aan te willen besteden noem ik:

- de [Personalize Your Homepage](#) en [Personalized Web Search](#) in Google Labs, die gebruikers resp. in staat stellen een eigen pagina van Google-diensten samen te stellen, een eigen zoekgeschiedenis op te slaan, en een persoonlijk interesseprofiel samen te stellen waarmee zoekresultaten kunnen worden gemanipuleerd;
- [Findory](#), een nieuwsdienst die de persoonlijke nieuwsvoorziening aanpast aan keuzes die de gebruiker maakt bij het aanklikken van berichten;
- [My Stuff](#) van Yahoo!'s concurrent Ask, dat u bezochte pagina's laat bewaren;
- [Eurekster](#), een zoekmachine op basis van Yahoo! die gebruikers in staat stelt eigen zoekmachines (Swicki's) samen te stellen
- de lokale 'sponsored results' die Google presenteert, en de diverse variëteiten van Yahoo!'s Local search (zie boven);
- en Google Scholar, dat behalve een op één speciale gebruikersgroep gerichte zoekmachine ook een poging is die gebruikers buiten de reguliere, door wachtwoorden afgeschermd en aan licenties gebonden kanalen om, plaatsonafhankelijk bij wetenschappelijke informatie te brengen. (Dat voor de uiteindelijke toegang dan vaak toch nog moet worden betaald is in dit verband niet relevant; de informatie zelf verschilt niet van andere producten waarvoor betaald moet worden.)

In deze wereld is Yahoo! een aan de weg timmerende, ervaren speler; al jaren geleden beschikte het over alle diensten die voor overleving in de 'search engine wars' van belang zijn: een toonaangevende eigen directory, in de VS nog altijd goed voor een flink aandeel in de zoekmarkt, eigen groepen, een Ask Yahoo!-dienst, mail, een bookmark manager, Yahoo! Reference, en naast de tegenwoordig eigen algemene zoekmachine een hele reeks van speciale zoekmachines, waaronder die waarin lokalisatie wordt toegepast. En dan, niet te vergeten, is er nog My Yahoo!

My Yahoo! is een dienst aan geregistreerde gebruikers van Yahoo! die hen in staat stelt op de servers van Yahoo! een eigen persoonlijke mix van Yahoo!-diensten samen te stellen.

Registreren is gratis en gaat gemakkelijk. Ga naar <http://www.yahoo.com> en klik op MyYahoo! Op de pagina die nu opent ziet u naast het logo van MyYahoo! staan: New User? Sign up. Volg die link en klik op de volgende pagina nogmaals op Sign Up. Vul nu op het volgende scherm de gevraagde gegevens in. Nadat u met de 'terms of service' akkoord bent gegaan, krijgt u een bevestiging op uw opgegeven andere emailadres (u hebt er nu ook een bij Yahoo!), waarop u moet reageren door op een URL te klikken.

Nu kunt u inloggen en beginnen uw persoonlijke Yahoo! met content te vullen. Trek u nog even niets aan van de standaardspullen die op uw pagina staan, u kunt ze later verwijderen. Klik op 'Add Content', en kies op het volgende scherm de diensten die u op uw pagina wilt hebben. Daarbij kunt u Zoeken of Browsen. Om een idee te krijgen van wat My Yahoo! te bieden heeft, kunt u om te beginnen het beste browsen.

We beginnen met de Yahoo! Services in de rechter kolom. Op de volgende schermen ziet u onder meer Mail Preview, Bookmarks, Address Book, Email Address Finder. Klik op de ADD knop links van de items die u op uw pagina wilt hebben. De tekst van de knop verandert in Remove, en later zult u de geselecteerde items weer van uw MyYahoo!-pagina kunnen verwijderen.

Ga terug naar het Search/Browse-scherm en typ in het zoekvenster: search. U krijgt een resultatenlijst te zien waarvan de eerste en de derde treffer de moeite waard zijn: Yahoo!

Search en Saved Searches. Voeg ook deze aan uw MyYahoo!-pagina toe.

Bij Saved Searches ziet u staan: 'More content: Personal Information Management, Web Search'. Klik eerst de rubriek Personal Information Management aan en kies daaruit de diensten die u nog meer op uw pagina ter beschikking wilt hebben. Ettelijke hebt u al in uw lijstje staan, maar Notepad en Messageboards zijn twee diensten die het overwegen waard zijn.

Ga terug naar het Search/Browse-scherm en klik nu op Web Search. Kijk hier nog even of er nog iets van uw gading is, ga weer terug en doe nu een zoekactie naar weblogs (of meer in het algemeen: RSS en Atom feeds) over onderwerpen die u interesseren. Als dat postzegels verzamelen en Perzische tapijten zijn, zoek dan eerst met "stamp collecting" en daarna met "Persian rugs" en voeg de weblogs die u met die zoektermen vindt toe aan uw persoonlijke pagina.

Als u hier weinig of niets kunt vinden over de onderwerpen die u interesseren, geen nood. Rechts van de zoekregel ziet u twee links staan naar pagina's die u verder zullen helpen bij het vinden van RSS feeds.

Maar bekijk voordat u daarmee verder gaat eerst eens hoe uw MyYahoo!-pagina eruit ziet. Als u op 'Finished' klikt, krijgt u het eerste resultaat van uw inspanningen te zien: links een lijst van standaard nieuwsbronnen die Yahoo! op elke persoonlijke pagina zet – maar die u kunt verwijderen door op het kruisje achterin de kolom te klikken – gevolgd door de weblogs die u zelf hebt toegevoegd, en daaronder of ertussendoor een paar andere diensten: Address Book, Mail Preview, Notepad, Email Address Finder en Message Boards. Als het een rommeltje is, geen nood: met 'edit' achter in de kolom kunt u elke dienst een andere plaats in de kolom geven. Helaas is het niet mogelijk individuele diensten van de ene kolom naar de andere over te brengen. Wel kan via 'Change Layout' een derde kolom worden toegevoegd en kan men de kolommen van plaats laten verwisselen.

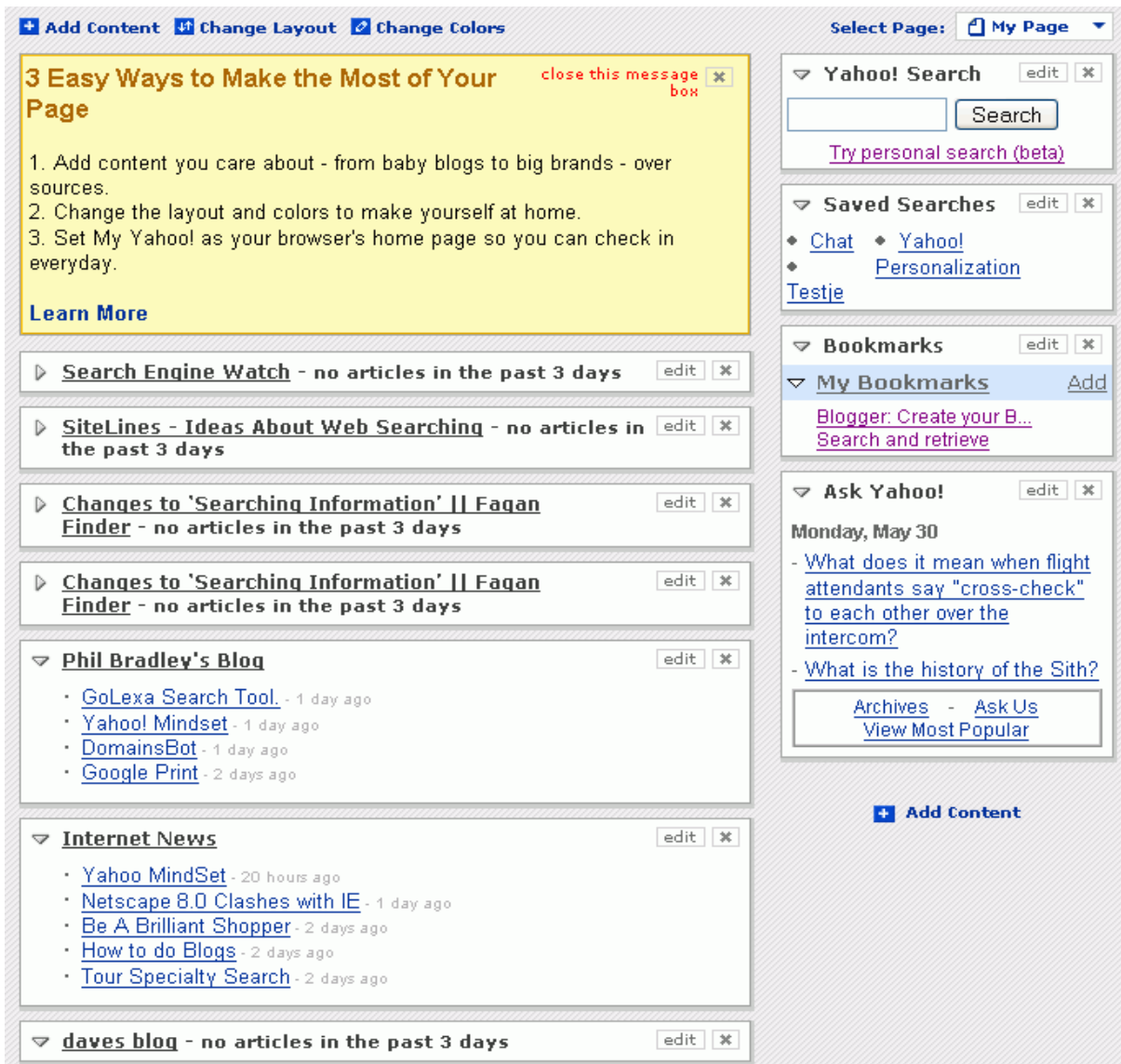
Rechts (zie volgende bladzijde) vindt u de andere diensten die u hebt opgegeven, zoals Bookmarks, Ask Yahoo!, Yahoo Search en Saved Searches.

U hebt nu een persoonlijke pagina op de Yahoo! server waar u een aantal voor uw informatiebeheer belangrijke spullen bij elkaar hebt, variërend van een adresboek en een schrijfblok tot uw bookmarks en opgeslagen zoekacties.

Bookmarks kunt u toevoegen via 'edit' – Edit Content – Add Bookmark. U zult daarbij zien dat u ook uw bestaande bookmarks in de Bookmark manager kunt importeren.

Op dezelfde manier kunt u via Saved Searches zoekacties formuleren, die als zodanig bewaard blijven. In de linker kolom werken diensten die u daar hebt toegevoegd maar die nog niet met content gevuld zijn, net zo. De RSS Feeds (weblogs) die u hier hebt geselecteerd lopen vanzelf en u zult er regelmatig nieuwe berichten zien verschijnen. Bij de andere diensten kunt u hetzij direct content toevoegen (Notepad), hetzij via 'edit' (bijvoorbeeld Address Book, Message Boards).

Enigszins verwarrend is dat Yahoo! momenteel aan het experimenteren is met een dienst die gedeeltelijk overlapt met MyYahoo!, namelijk MyWeb. Daarvoor moeten we terug naar de rechter kolom in de MyYahoo!-pagina. In het vakje Yahoo! Search ziet u een link Try Personal Search Beta. Volg die link en u komt op een zoekpagina van Yahoo! uit die er iets anders uitziet dan wat u gewend bent. De naam is MyYahoo!



Een vlot in elkaar geknutselde My Yahoo!-pagina.

Search Beta en onder de zoekregel vindt u naast de knop 'Search the Web' een knop 'MyWeb (beta)'. Voer nu een zoekactie uit en u zult zien dat de treffers twee opties extra hebben gekregen: Save en Block. Als u nu op Save klikt wordt de treffer (desgewenst met een noot) toegevoegd aan een MyWeb-folder, Saved genaamd; op dit moment hebt u ook de keus de pagina aan een andere folder toe te voegen, die u ook direct nieuw kunt aan maken. Klik nu op MyWeb (links boven in het resultatenscher) en u ziet een MyWeb-pagina met in de linker kolom een menu de Saved- folder, en daaronder nog twee andere, History en Blocked. Als u niet hebt aangegeven dat de zojuist bewaarde treffer in een andere folder dan Saved moest komen, zit hij daar nu in, en ziet u hem ook bovenaan in het rechter deel van het scherm. In de folders History en Blocked komen resp. treffers terecht die u bij uw zoekactie bezocht hebt door ze aan te klikken zonder ze te 'saven', en de treffers waarbij u in het resultatenscher op 'Blocked' hebt geklikt. Het effect van het laatste is dat u die treffers bij een volgende zoekactie die u als geregistreerd gebruiker met de zoekmachine onderneemt, niet meer in het zoekresultaat zult tegenkomen; daartegenover staat dat als het zoekresultaat treffers bevat die u eerder naar MyWeb gesaved hebt, daarbij vermeld staat dat u dat gedaan hebt, in welke folder en met welke noot.

Yahoo! My Yahoo! Mail Welcome, **mrjh1940** [Sign Out, My Account] Help Search the Web Search

YAHOO! SEARCH Features/FAQ Community Import Bookmarks Install Toolbar

My Web **BETA** Browsing items 1 to 17 of 17 items in "Saved" - 19 items saved in My Web

My Web (beta) Search the Web

Saved [My Search History ON - Stop](#) | [Show details](#)

Folders	<input type="checkbox"/>	Email	IM	Delete	Move to	Sort By: Date
<ul style="list-style-type: none"> Saved (17) History [Clear] Blocked (1) My Folders [Add - Edit] <ul style="list-style-type: none"> Search strategies (2) Shared (0) My Web Tips... Share your saved pages with friends. Learn more... 						
		May 30, 2005				
	<input type="checkbox"/>	After many months of completely and totally ignoring this si...				May 30, 2005
		Note: Edit http://www.google.com/ - My Saved Copy				
	<input type="checkbox"/>	Google Guide				May 30, 2005
		Note: Interactieve gids - Edit http://www.googleguide.com/ - My Saved Copy				
	<input type="checkbox"/>	main search engines' lore, usage and comparison				May 30, 2005
		Note: Searching, general - Edit http://www.searchlores.org/main.htm - My Saved Copy				
	<input type="checkbox"/>	Fravia's web-searching lore: Main entrance				May 30, 2005
		Note: Searching, general - Edit http://www.searchlores.org/ - My Saved Copy				

Gedeelte van de My Web- pagina bij Yahoo!, met bovenin de zoekfunctie, waarmee ofwel het hele Web, ofwel de opgeslagen pagina's in My Web kunnen worden doorzocht. Links het menu met de diverse folders, en rechts de in de folder 'Saved' opgeslagen adressen. Hier heeft men de keus naar de originele site door te klikken of de eigen kopie te bekijken.

In het menu aan de linkerkant vindt u onder My Folders de door u zelf gemaakte folders (die u hier ook kunt bewerken of waaraan u nieuwe folders kunt toevoegen, en een folder met de naam Shared. Pagina's die u in die folder onderbrengt, kunt u aan geïnteresseerden mailen of in een publieke folder bij [Yahoo! 360°](#) zetten. Dit laatste is een nog in een bètaversie verkerende community site van Yahoo!.

MyWeb en MyYahoo! staan niet los van elkaar. Hierboven heb ik beschreven hoe ik vanaf mijn MyYahoo!-pagina via Try Personal Search (beta) op de MyWeb-pagina ben uitgekomen. Andersom blijkt in de linker kolom van mijn MyYahoo!-pagina ook een rubriek Yahoo! My Web - All Saved Pages te zijn aangemaakt, die mij daarheen voert. Een directe link van de MyWeb-pagina naar de MyYahoo!-pagina ontbreekt nog, maar voor het overige biedt de combinatie de gebruiker op een gebruikersvriendelijke manier een schat aan mogelijkheden.

De hier beschreven situatie dateert van mei 2005. Inmiddels is de bètaversie van My Web 1.0 alweer opgevolgd door de bètaversie van My Web 2.0, die community-eigenschappen en tags aan My Web 1.0 toevoegt. Dat betekent: meer delen van resources met anderen, meer mogelijkheden tot contact met gelijk geïnteresseerden en de mogelijkheid om door het toekennen van tags (trefwoorden) tot een gezamenlijke onderwerpsontsluiting van de verzamelde resources te komen. Uitleg vindt u [hier](#).

4.2.4. MSN Search

4.2.4.1. In kort bestek

URL (Basic Search): <http://search.msn.com/> plus een groot aantal nationale versies van de zoekmachine die naast mondiaal ook nationaal zoeken (op taal en binnen een land) mogelijk maken; deze nationale versies bieden bij mondiaal zoeken anders gerangschikte en doorgaans minder resultaten dan de internationale/Amerikaanse versie

Omvang : ca. 5 miljard (volgens opgave van MSN); volgens een studie van Gulli en Signorini (<http://www.cs.uiowa.edu/~asignori/web-size/>) indexeert MSN Search ruim 7 miljard pagina's, bijna 62% van het zichtbare Web; naar mijn indruk (in maart 2006) is de database van MSN sinds begin 2005 kleiner geworden)

Indexeert tot: HTML: ca. 1 Mb (voorbeeld); PDF: ca. 1 MB (voorbeeld); TXT: > 750 K (voorbeeld)

Maximale omvang zoekargument in het zoekvenster: onbeperkt, maar bij invoer van meer dan 19 zoektermen in een EN-relatie vindt MSN het document niet meer; voor OR en prefer: gelden kennelijk andere regels

Indexeert andere formaten dan HTML: ja (vgl.: <http://www.faganfinder.com/filetype/>)

Toont: maximaal ca. 250 treffers

Updating: enkele minuten – enkele weken

Help en info: <http://search.msn.com/docs/default.aspx?FORM=HLHP;>

<http://search.msn.com/docs/help.aspx?FORM=HLDD3;>

http://search.msn.com/docs/help.aspx?t=SEARCH_REF_AdvSrchOperators.htm

Advanced Search: n/a (sommige geavanceerde mogelijkheden zijn ondergebracht in de Search Builder)

Directory: n/a

Speciale zoekmachines: <http://search.msn.com/news/results.aspx?FORM=NRDD>

(nieuws), <http://search.msn.com/images/results.aspx?FORM=IRDD> (images),

<http://search.msn.com/local/results.aspx?q=&FORM=PBHP2> (local),

<http://entertainment.msn.com/search/movie/?ss=> (movies), <http://shopping.msn.com/>

(shopping), <http://music.msn.com/> (music), <http://radio.msn.com/> (radio),

<http://tv.msn.com/> (tv), <http://music.msn.com/musicvideos/> (video)

Testomgeving: <http://sandbox.msn.com/>

Blog: <http://blogs.msdn.com/msnsearch>

Meer over MSN Search: <http://searchengineshowdown.com/features/msn/review.html>

<http://channel9.msdn.com/wiki/default.aspx/Channel9.MsnSearchFeedback>

A Guide to MSN® Search and the MSN Search Toolbar

(<http://download.microsoft.com/download/c/2/9/c2935f83-1a10-4e4a-a137-c1db829637f5/MSNSearch2005RG.doc>)

Overzicht

Soort zoekactie	Voorbeelden	Vindt:
Natuurlijke taal (alleen Engels!)	What is Albert Einstein known for? How deep is the Pacific Who won Wimbledon in 2003?	Probeer! De zoekmachine vindt antwoord op feitelijke vragen in Encarta Answers (http://encarta.msn.com/encnet/support/EncartaAnswers.aspx?page=home); succes hangt af van de vraagstelling en of Encarta een antwoord heeft

Soort zoekactie	Voorbeelden	Vindt:
<i>Prefixes</i>		
intitle:	intitle:Egyptology intitle:"Chinese cooking" intitle:dreaming "sleep walking" intitle:ireland intitle:history intitle:postal	<p>documenten met de opgegeven zoekterm of phrase in de titel van de pagina (en eventuele andere zoektermen op een willekeurige plaats in de pagina)</p> <p>documenten met de opgegeven termen in de titel; in werkelijkheid bevat het resultaat documenten die niet in het resultaat thuishoren</p>
inurl:	inurl:leidenuniv inurl:leidenuniv kunstgeschiedenis site:uu.nl inurl:handleidingen inurl:leidenuniv.nl inurl:frans site:leidenuniv.nl inurl:frans	<p>documenten met de opgegeven zoekterm in de url (host, pad of bestandsnaam) (en eventuele andere zoektermen op een willekeurige plaats in de pagina); voor grote precisie te combineren met site: of intitle: het prefix inurl: kan niet worden herhaald: het resultaat wordt dan onbetrouwbaar; gebruik waar mogelijk site: inurl:</p>
inanchor:	inanchor:"in vitro fertilisatie" inanchor:"in vitro fertilisatie" intitle:"in vitro fertilisatie" inanchor:"in vitro fertilisatie" -intitle:"in vitro fertilisatie" (intitle:"in vitro fertilisatie" OR inanchor:"in vitro fertilisatie") AND micromanipulatie	<p>documenten waarnaar met de opgegeven term of phrase wordt verwezen in een anker van andere documenten</p> <p>documenten met in de titel (intitle) de opgegeven term of phrase, waarnaar in een anker (inanchor) van een of meer andere documenten wordt verwezen met de opgegeven term of phrase (in dit geval is de phrase in titel en anker identiek)</p> <p>documenten waarnaar in een anker van andere documenten wordt verwezen met de opgegeven term of phrase, maar die de (in dit geval identieke) term of phrase niet zelf in de titel hebben</p> <p>documenten met de opgegeven term of phrase in de titel OF in het anker van een verwijzend document, EN met de tweede opgegeven term of phrase op een willekeurige plaats</p>

Soort zoekactie	Voorbeelden	Vindt:
inbody:	inbody:oneiromancy fraud -inbody:fraud -inurl:fraud -intitle:fraud references inbody:"G Hofstede"	de opgegeven term of phrase in de 'body' (zie blz. 66) van een document webpagina's die de opgegeven zoekterm zelf niet bevatten maar waarbij deze alleen in een anchor van naar de betreffende pagina verwijzende andere pagina's voorkomt NB: Deze melding heeft alleen betrekking op de gecacheerde versie van de pagina waarnaar wordt verwezen; de huidige versie kan de zoektermen inmiddels wel bevatten! de opgegeven term of phrase in de 'body' van de tekst, in combinatie met een andere term op een willekeurige plaats in het document; in het voorbeeld zijn we op zoek naar documenten waarin de naam G. Hofstede in een literatuurlijst voorkomt
site: (komt overeen met de optie Site/Domain in de Search Builder)	site:nl site:hva.nl site:ie.hva.nl site:htsa.ie.hva.nl handleiding site:ie.hva.nl handleiding (site:htsa.ie.hva.nl OR site:home.ie.hva.nl) handleiding (site:itv.hva.nl OR site:ie.hva.nl) -site:htsa.ie.hva.nl	documenten uit het opgegeven domein (of domeinen); in de Search Builder geeft u meer domeinen op door die aan de vraagstelling te blijven toevoegen; de OR-relatie en de – (NIET)relaties worden dan door de Search Builder gelegd NB: de combinatie van OR en – (voor NOT) in de zoekexpressie wordt door de Search Builder aangebracht
ip:	ip:132.229.7.180	sites die worden gehost door het opgegeven IP-adres
link: (komt overeen met de optie Links to in de Search Builder)	link:www.searchability.com link:www.nfdb.nl/ AND link:www.danskefilm.dk/ maritiem link:www.museumgids.nl/	documenten met een link naar het opgegeven adres documenten die een verwijzing naar zowel een Nederlandse als een Deense filmdatabase bevatten documenten die een verwijzing naar de Museumgids bevatten, en in de tekst de term maritiem Opmerking: evenals Yahoo! kan MSN Search documenten of sites vinden waarnaar door twee of meer andere sites wordt verwezen
linkdomain:	linkdomain:www.hva.nl Vgl. bijvoorbeeld: linkdomain:uva.nl (299.238) linkdomain:leidenuniv.nl (337.444) linkdomain:uu.nl (385.515) linkdomain:uu.nl -site:uu.nl (217.893)	documenten met een link naar een willekeurige pagina van de opgegeven site; linkdomain kan worden verbonden met andere opdrachten, bijvoorbeeld site: linkdomain:... -site:... geeft alle links naar pagina's van de Universiteit Utrecht met uitzondering van die welke in pagina's van de Umiversiteit Utrecht zelf voorkomen ('zelfcitaties')

Soort zoekactie	Voorbeelden	Vindt:
filetype:	filetype:xls begroting filetype:xls	bestanden met de opgegeven extensie, naast html-bestanden ook: TXT, PDF, DOC, RTF, XLS en PPT kan ook worden gebruikt in combinatie met (een) zoekterm(en)
contains:	contains:wav leeuwerik contains:wav	bestanden met een hyperlink naar een of meer bestanden met de opgegeven extensie; kan worden gecombineerd met een zoekterm, om bijvoorbeeld een geluidsbestand bij die term op te sporen
loc: location: (komt overeen met de optie Country/Region in de Search Builder)	loc:NL geneesmiddelenwijzer loc:BE	documenten uit het opgegeven land/regio (ccTLD's en generieke TLD's) Belgische documenten met de zoekterm geneesmiddelwijzer
language: (komt overeen met de optie Language in de Search Builder; ook standaard in te stellen onder Settings)	language:nl link:http://searchability.com/ (language:nl OR language:es)	documenten in de opgegeven taal documenten in een (of meer) van de opgegeven talen met een link naar de opgegeven webpagina
{frsh=[0-100]} (komt overeen met de optie Results ranking: Static/Updated recently in de Search Builder)	{frsh=100} Openbaar Ministerie” {frsh=100}	documenten waarin actualiteit (in de zin van recent gemaakt of bijgewerkt) een belangrijke rol speelt; het sorteercriterium moet worden gecombineerd met ten minste één zoekterm documenten met daarin de opgegeven phrase, gesorteerd naar actualiteit (= Last Modified date)
{popl=[0-100]} (komt overeen met de optie Results ranking: Less popular/Very popular recently in de Search Builder)	{popl=20} Openbaar Ministerie” {popl=100}	documenten waarvan de 'link popularity' minder belangrijk is dan actualiteit en 'approximateness'; het sorteercriterium moet worden gecombineerd met ten minste één zoekterm documenten met daarin de opgegeven phrase, gesorteerd naar “link popularity”

Soort zoekactie	Voorbeelden	Vindt:
<p>{mtch=[0-100] (komt overeen met de optie Results ranking: Exact match/Approximate match in de Search Builder)</p>	<p>{mtch=30}</p> <p>Openbaar Ministerie” {mtch=100}</p>	<p>webpagina's waarin de frequentie van de zoektermen betrekkelijk hoog is NB: de schaal gaat hier van Exact match (0) tot Approximate match (100); dat op zichzelf is al vreemd: men zou verwachten dat de mate van 'exactheid' zou toenemen, niet die van 'approximateness'; nog vreemder wordt het als blijkt dat het laatst getoonde (250e)document bij {mtch=100}, d.w.z. met de hoogste mate van 'approximateness', de hoogste frequentie van zoektermen vertoont, en het laatst getoonde (250e) document bij {mtch=0} de laagste frequentie; het sorteercriterium moet worden gecombineerd met ten minste één zoekterm</p> <p>documenten met daarin de opgegeven phrase, gesorteerd naar de mate van overeenkomst tussen zoekvraag en document (volgens het algoritme van de zoekmachine)</p>
url:	url:www.hva.nl	de pagina met het opgegeven adres
prefer:	<p>football prefer:organization</p> <p>site:nl prefer:site:leidenuniv.nl</p> <p>prefer:site:uu.nl sitemap</p>	<p>pagina's met een verplichte en een facultatieve (prefer:) term</p> <p>pagina's met de opgegeven zoekterm (sitemap) in het opgegeven domain (.nl), maar bij voorkeur (d.w.z. voorop in de ranking) bij een van de domains (leidenuniv.nl en uu.nl) die met het prefix prefer: aan de vraagstelling zijn toegevoegd</p> <p>prefer: is MSN's vertaling van 'best match' (zie blz. 53); werkt het best in combinatie met een ander prefix, zoals site: of intitle:</p>
feed:	<p>site:bbc.co.uk feed: bbc.co.uk</p> <p>site:bbc.co.uk feed: bbc.co.uk review</p>	<p>feeds (xml-bestanden) op de site van de BBC; voor uitleg en gebruik zie blz. 284 e.v.</p> <p>feeds met daarin de term review op de site van de BBC</p> <p>NB: achter het prefix feed: moet een spatie worden gezet</p>
hasfeed:	<p>site:bbc.co.uk hasfeed: bbc.co.uk</p> <p>site:bbc.co.uk hasfeed: bbc.co.uk review</p>	<p>pagina's op de site van de BBC die een RSS- of een Atom-feed hebben; voor uitleg en gebruik zie blz. 284</p> <p>pagina's op de site van de BBC met daarin de term review en met een RSS- of een Atom-feed</p> <p>NB: achter het prefix hasfeed: moet een spatie worden gezet</p>

Soort zoekactie	Voorbeelden	Vindt:
<i>Operatoren</i>		
term1 term2 (komt overeen met 'Show results with all of these words' in de Search Builder)	Verdi Requiem	de ene EN de andere zoekterm (standaard)
+term	defenders +for freedom	stopwoorden buiten phrases; zonder te plussen vindt MSN stopwoorden niet, tenzij ze in phrases voorkomen
term1 -term2 (komt overeen met 'Show results with none of these words' in de Search Builder)	cancer -lung Bach -"Johann Sebastian" "Roman empire" -site:com Nederland AND Duitsland AND betrekkingen NOT "economische betrekkingen"	de ene maar niet de andere term (of phrase) documenten met de opgegeven zoekterm of phrase, maar niet afkomstig van sites met de opgegeven TLD NB: gebruik in een zoekexpressie met AND en/of OR liever NOT dan – om een term uit te sluiten.
AND (&) (komt overeen met 'Show results with all of these words' in de Search Builder)	cancer AND nutrition	de ene EN de andere zoekterm
OR () (komt overeen met 'Show results with any of these words' in de Search Builder)	nutrition OR diet	de ene OF de andere zoekterm (of beide)
NOT	cancer NOT lung "Roman empire" NOT site:com	de ene maar niet de andere term documenten met de opgegeven zoekterm of phrase, maar niet afkomstig van sites met de opgegeven TLD
nesting	(kanker OR tumor) AND (dieet OR voeding) AND site:be "Roman empire" AND (site:edu OR site:org)	termen in zowel EN- als OF-relaties in één en hetzelfde zoekargument; door het tussen haakjes plaatsen van de door OR verbonden termen krijgt de OR-bewerking prioriteit boven de AND-bewerking
<i>Nabijheid</i>		
phrase ("...") (komt overeen met "Show results with the exact phrase" in de Advanced Search)	"Willem van Hanegem" "part time" "nothing of the sort"	de opgegeven naam of zinsnede als reeks van opeenvolgende woorden, met tussen elke twee woorden een spatie of een ander leesteken part time en part-time, maar niet parttime evenals Yahoo! vindt MSN Search bij deze vraag documenten waarvan de cacheversie noch de phrase, noch de woorden die er deel van uitmaken, bevatten

Soort zoekactie	Voorbeelden	Vindt:
<i>Overige</i>		
Display ¹⁾	Danish / Dutch / English etc.	stel onder Settings de taal in waarin de zoekmachine wordt gepresenteerd NB: dit is niet hetzelfde als het gebruik van de lokale versie van de zoekmachine op http://search.msn.nl
Number of results ¹⁾	10/15/30/50	kies onder Settings het aantal treffers per pagina van de resultatenlijst
[localisatie] Location ¹⁾		wie in de Amerikaanse versie van MSN Search onder Settings een Location heeft opgegeven, krijgt naast de Search knop een Near Me-knop gepresenteerd, die toegang beoogt te geven tot treffers in de opgegeven stad ; dan verschijnt in de resultatenlijst ook een menu waaruit andere steden kunnen worden gekozen (werkt momenteel alleen voor Noord-Amerika)
SafeSearch ¹⁾	Strict / Moderate / Off	schakelt het Family Filter van MSN Search uit / aan Standaard is Moderate filtering; gebruik de mogelijkheden van Settings om die instelling te veranderen
accenten	é, ï, ö, û etc. electricite électricité electricité lange Länge ode öde blöd site:nl oölogie oologie site:nl Munchausen by proxy syndrome" AND site:org "Münchausen by proxy syndrome" AND site:org ("Münchausen by proxy syndrome" OR "Munchausen by proxy syndrome") AND site:org	MSN Search maakt soms wel en soms geen onderscheid tussen geaccentueerde en niet geaccentueerde letters. vindt electricite, electricité, électricité electricite, electricité, électricité electricite, electricité, électricité Dit gaat echter niet altijd op: lange, maar niet Länge Länge, maar niet lange ode, maar niet öde öde, oede, øde, maar niet ode blöd, bloed treffers met oölogie maar niet met oologie één (Duitstalige) treffer 214 treffers 1 treffer 219 treffers (?)
hoofdletters/ kleine letters (case-sensitivity)	next NeXT	MSN Search maakt geen onderscheid tussen hoofdletters en kleine letters

Soort zoekactie	Voorbeelden	Vindt:
leestekens	<p>-,./_&+#\$</p> <p>part-time</p> <p>part-time NOT "part time"</p> <p>"part time"</p> <p>parttime</p> <p>c+</p> <p>c++</p> <p>&c</p> <p>single_user</p> <p>a#</p> <p>c#</p> <p>f#</p> <p>g#</p> <p>a&b&c</p> <p>a_b_c</p> <p>\$5000</p> <p>€5000</p> <p>\$134.50</p> <p>I/O</p> <p>#2</p> <p>>2</p>	<p>MSN Search verwaarloost de meeste lees- en niet-lettertekens, of vat ze op als spaties binnen een phrase; punten en komma's in getallen worden als jokers binnen het getal opgevat, zodat getallen met een punt of komma zowel in de in Nederland en België gebruikte als de Angelsaksische notatie worden gevonden</p> <p>part-time, "part time", maar niet parttime</p> <p>geen treffers</p> <p>"part time", inclusief part-time</p> <p>parttime</p> <p>vindt hetzelfde als c</p> <p>c++</p> <p>zelfde resultaat als c</p> <p>"single user"</p> <p>a, inclusief a#</p> <p>c#</p> <p>f, maar niet f#</p> <p>g, maar niet g#</p> <p>"a b c"</p> <p>"a b c"</p> <p>5000</p> <p>5000</p> <p>"134 50"</p> <p>"I O"</p> <p>2</p> <p>2</p>
getallen	<p>10000</p> <p>10.000</p> <p>14.668</p> <p>25.32</p> <p>25,32</p>	<p>10000, maar niet 10.000 of 10,000</p> <p>"10 000", inclusief 10,000 en 10.000</p> <p>"14 668", inclusief 14.668, 14-668 etc.</p> <p>"25 32", inclusief 25.32</p> <p>"25 32", inclusief 25,32</p>
stopwoorden	<p>to be or not to be</p> <p>to be or not to be"</p>	<p>MSN Search lijkt alleen not te vinden "to be or not to be"; willen stopwoorden gevonden worden, dan moeten ze worden geplust of in een phrase worden gebruikt</p>
(spellingcontrole) [Spell Checker]	<p>catagory (spelfout)</p> <p>swlling (typfout)</p> <p>Rotterdam (typfout)</p> <p>"Berend Botje ging uit varen"</p> <p>"Berend Botje ging eens varen"</p> <p>"Berend Botje gong eens varen"</p>	<p>de resultatenlijst bevat geen suggestie voor verbetering</p> <p>Were you looking for <i>selling</i>?</p> <p>(gevolgd door documenten met het woord in de verkeerde spelling)</p> <p>Were you looking for <i>Rotterdam</i>?</p> <p>(idem)</p> <p>Were you looking for <i>"Bernd Bothering unit varner"</i> (gevolgd door 37 treffers)</p> <p>Consider <i>"Bernd Bothering evens verne"</i> (1 treffer)</p> <p>geen suggesties, geen treffers</p>
suggesties voor verder zoeken/categorisatie	n/a	n/a

Soort zoekactie	Voorbeelden	Vindt:
[term weighting]: volgorde zoektermen	kanker voeding voeding kanker nieuwe lichting rekruten lichting nieuwe rekruten nieuwe rekruten lichting	als eerste treffers in de resultatenlijst, afhankelijk van de volgorde van de zoektermen, documenten waarin het ene dan wel het andere aspect van de vraag overweegt (op basis van de door MSN Search gebezigde statistische criteria) De volgorde van de zoektermen is ook van belang omdat MSN Search meestal automatisch naar phrases zoekt. NB: er is een effect, maar het is gering, en het is twijfelachtig of het betekenis heeft
[term weighting]: herhaling van zoektermen	voeding voeding voeding voeding kanker kanker kanker kanker kanker voeding	als eerste treffers in de resultatenlijst, afhankelijk van de relatieve ranking en frequentie van de zoektermen, documenten waarin het ene dan wel het andere aspect van de vraag overweegt . Frequentie zoektermen op eerste resultatenpagina van 10, resp.: 47 maal voeding / 31 maal kanker 26 maal voeding / 63 maal kanker Er is een duidelijk effect.
truncatie	n/a	n/a
[(automatic) stemming]; (automatische) toevoeging van zoektermen met dezelfde stam als de opgegeven zoektermen)	n/a	n/a
<i>Shortcuts</i>		
Woordenboek, encyclopedie, thesaurus		
define (alleen Engels!)	define serendipity define United Nations	als eerste treffer een omschrijving van de opgegeven term uit The American Heritage Dictionary; NB: geen dubbele punt achter define
Overige		
(Omrekenen van maten en gewichten) [Conversion tool] Typ: (how many) [eenheden] in [aantal] [eenheden]	how many days in 17 years meters in 386 feet	de ene maat of het ene gewicht omgerekend in de andere maat of het andere gewicht

1) alleen standaard instelbaar via Settings

4.2.4.2. Zoekscherf

MSN Search heeft één zoekscherf, waarop – afgezien van enkele instellingen onder Settings – alle mogelijkheden van de zoekmachine kunnen worden gebruikt. Het bestaat uit een zoekvenster met vijf databases waartussen geswitcht kan worden (Web, Desktop, News, Images, Local). U vindt deze (m.u.v. Desktop) ook terug onder het aanklikbare pijltje achter de Search-knop, met nog enkele andere mogelijkheden of toegangen: Look up word, Stock Quotes, Find Movies, Shopping en Music. Een andere

keuze dan Web in dit drop-downmenu laat u de zoekvraag in andere database uitvoeren. (In de Nederlandse versie zijn de keuzes hier: Web, Nieuws, Afbeeldingen, Shopping, Telefoongids en Bedrijvengids.)

Onder de zoekregel vindt u links naar de Search Builder, MSN's geavanceerde zoekopties, Settings, waar u enkele standaardinstellingen aan uw eigen wensen kunt aanpassen, Helpteksten en de Spaanstalige versie van de pagina.

Op de Nederlandse versie van de pagina ontbreken de link naar de Spaanstalige versie en de opties Desktop en Encarta. Wel vindt u hier aankruisvakjes voor Websites in het Nederlands en Alleen Websites in Nederland.

In de zoekregel typt u uw zoektermen, desgewenst verbonden met de operatoren AND (&), OR (|) en NOT (-). Nesting is mogelijk. MSN laat u meer dan negentien zoektermen in een EN-relatie invoeren, maar vindt dan – zonder melding van de oorzaak – niets meer. (Voor OR en prefer: gelden echter kennelijk andere regels.)



We couldn't find any results containing **denkend aan Holland brede rivieren traag oneindig * laagland ondenkbaar ijle populieren pluimen einder geweldige diepte verzonken boerderijen verspreid land kerken.**

Search tips:

Try similar words. For example, some sites may use "film" instead of "movie".

Make your search more general. For example, instead of using a specific product name, try typing the generic product category or manufacturer.

If you cannot find a page that you know exists, [send the address to us](#).

[Get more search tips](#).

Achter het zoekvenster vindt u de Search-knop met de al eerste genoemde pijl omlaag voor andere databases dan het Web, en eventueel een Near Me-knop, waarmee u in de VS gelokaliseerde zoekresultaten kunt opvragen. Hierover zo dadelijk meer.

4.2.4.3. Search Builder

Een deel van MSN Search' geavanceerde zoekfuncties wordt in een menu getoond als u de Search Builder aanklikt. Onder de zoekrechthoek verschijnt nu een tweede rechthoek met zes knoppen:

Search terms

Site/Domain

Links to

Country/Region

Language

Results ranking

Elk van de knoppen opent een invulvakje met of zonder extra opties, of, in het geval van de Results ranking, een venster met drie schuifpanelen. Door op de knop Add to search te klikken of na het verschuiven van een schuifregelaar de muisknop los te laten bevestigt u een toevoeging, instelling of keuze. In alle gevallen leidt dit tot opdrachten in het zoekvenster, die u, na enige ervaring te hebben opgedaan, ook gewoon zelf in de zoekopdracht kunt intypen of wijzigen.

De extra mogelijkheden die de Search Builder biedt kunnen in combinatie met het zoekvenster worden gebruikt, maar ook apart; het is dus niet nodig (een) zoekterm(en) in het zoekvenster ingetypt te hebben om de Search Builder te kunnen gebruiken. Maar als u termen in het zoekvenster hebt ingetypt, worden die standaard in een EN-relatie

verbonden met alles wat u in de Search Builder doet.

Met *Search terms* voegt u termen aan uw zoekactie toe. Opties die hierbij worden aangeboden zijn: All of these terms (EN-relatie tussen de termen die u hier toevoegt), Any of these terms (OF-relatie tussen de termen die u toevoegt), Exact phrase (termen onmiddellijk op elkaar volgend) en None of these terms (termen die niet in het zoekresultaat mogen voorkomen. In de zoekregel worden deze opties vertaald in AND, OR, "... " en -. Als u hier twee termen in een OR-relatie aan de zoekvraag toevoegt, wordt dit deel van de zoekexpressie automatisch tussen ronde haken gezet ('nesting'; dat ligt ook voor de hand aangezien de term(en) die u zelf in de zoekregel hebt ingetypt in een EN-relatie met de nieuwe termen wordt (worden) verbonden.

Als u liever zelf uw opdrachten intypt dan de Search Builder te gebruiken, kunt u in plaats van AND ook & gebruiken, in plaats van OR ook | en in plaats van de minoperator ook NOT.

Met *Site/Domain* beperkt u uw zoekactie tot een bepaalde site of een bepaald domein.

Als u dat tot meer dan een site en/of domein wilt doen, kunt u de handeling Add to search herhalen (dus elke keer niet meer dan één site/domain in het invulvak typen en dan op Add to search klikken). De Search Builder verbindt de toegevoegde sites/domains dan in een OR-relatie, zoals in het zoekvenster te zien is. Via de keuzerondjes 'Find only ...' en 'Exclude...' beperkt u de zoekactie tot de opgegeven sites/domains of sluit u die juist van de zoekactie uit. 'Find only...' wordt vertaald als impliciet AND, 'Exclude...' als -.

De domains omvatten zowel de generieke domains (.com, .edu etc.) als de ccTLD's (.nl, .be etc.). De sites kunt u links zo lang maken als u wilt (hva.nl, ie.hva.nl, ht.sa.ie.hva.nl), maar rechts niet langer dan de root (dus tot de eerste /). Een prefix `inurl:` heeft MSN Search ook, het is dus mogelijk een zoekactie te beperken tot een deel van een site.

Met *Links to* kunt u op zoek naar documenten of sites die naar een of meer andere documenten of sites linken. Sites die u zo aan de zoekactie toevoegt, worden in de opdrachtregel voorafgegaan door het prefix `link:`. Ook hier geldt: toevoegen gaat een voor een, en levert een OF-relatie op; u kunt die in de zoekregel zelf in een EN-relatie veranderen. *Links to* is goed met *Site/domain* te combineren om 'zelfcitaties' (links van de ene pagina binnen een site naar een andere pagina binnen dezelfde site) uit te



Het zoekscherm van MSN Search, met onder 'Results ranking' de schuifregelaars waarmee de zoekresultaten op resp. actualiteit, populariteit en relevantie kunnen worden gesorteerd.

sluiten: [link:www.hva.nl -site:hva.nl](#). De combinatie van `link:[site1] AND link:[site2]` is een bruikbare methode om 'hubs' op te sporen, en 'meer van hetzelfde'. Gebruik: [databases link:www.addall.com link:www.bookfinder.com](#) om de pagina Online Book Databases te vinden, die naar deze twee boekzoekmachines en vele andere linkt; met dezelfde zoekactie vindt u dan ook ILAB-LILA, waarvan u de naam vergeten was.

Gebruik *Country/Region* om uw zoekactie tot een of meer landen te beperken.

Verschijnt in de zoekopdracht als: `loc`: Voeg diverse landen in een (onderlinge) OF-relatie aan de vraag toe door meermalen een land te kiezen en op Add to search te klikken. Kan ook weer gecombineerd worden met andere prefixen/limit, bijvoorbeeld in de vraag: [link:www.aivd.nl loc:SE](#) (wie linken er vanuit Zweden naar de site van de AIVD). Als u `loc:...` direct in de opdrachtregel typt, hoeft u zich niet te beperken tot het korte lijstje van landen in de Search Builder. Voor een complete lijst, zie:

<http://www.iana.org/cctld/cctld-whois.htm>.

Op dezelfde manier kan de optie *Language* worden gebruikt om de zoekactie tot een of meer talen te beperken. Verschijnt in de zoekactie als: `language`: Voorbeeld: [koran site:NL language:ar](#) voor sites in het Nederlandse domein met de Arabische tekst van de koran.

Results ranking ten slotte bevat drie schuifregelaars waarmee de sortering van zoekresultaten kan worden gemanipuleerd. Door een of meer regelaars naar boven of naar beneden te trekken kan respectievelijk op freshness (laatste datum van wijziging), (link) popularity en exactness (overeenkomst tussen zoektermen en zoekresultaat), of een mix daarvan worden gesorteerd.

Als u na het verslepen van de regelaar de muisknop loslaat, verschijnt de gekozen ranking als onderdeel van de opdracht in het zoekvenster. De (stilzwijgende) standaardinstelling is `{frsh=50} {popl=50} {mtch=50}` en de stand van elk ervan varieert tussen 0 en 100; als u een van de regelaars op een andere stand zet, blijven de andere regelaars stilzwijgend op 50 staan. Meer hierover bij de bespreking van MSN Search.

Denk erom dat MSN Search over een heel aantal prefixen/begrenzers beschikt die geen plaats hebben gekregen in de Search Builder. Deze moeten dus direct in de zoekregel worden ingetypt. Het betreft hier de prefixen `contains:`, `filetype:`, `inanchor:`, `inbody:`, `intitle:`, `inurl:`, `ip:`, `linkdomain:`, `prefer:`, `feed:`, `hasfeed:`, en `url:`.

4.2.4.4. Settings

Hier kunt u enige wijzigingen aanbrengen in de standaardinstellingen van MSN Search. *Display*: wijzig de interfacetaal. Standaard is Engels (Nederlands voor de Nederlandse versie).

Results: stel het aantal treffers per pagina in op 10/15/30/50. Vink 'Group results from the same site' uit als u treffers van dezelfde site niet bij elkaar wilt zien; iedere pagina van een site krijgt dan zijn eigen ranking. Als u 'Group results from the same site' aan laat staan, stel de clustering van zoekresultaten per site dan in op 1/2/3. Vink 'Open links in a new browser window' aan als u de zoekresultaten in een nieuw venster of onder een nieuwe tab te zien wilt krijgen

SafeSearch: de standaardinstelling hier is dat MSN Search u geen afbeeldingen laat zien die niet geschikt worden geacht voor jeugdige internetgebruikers. De andere opties zijn dat u MSN ook sites met 'adult' teksten laat wegfilteren, dan wel dat u het filter geheel uitzet. Geen van de opties is geheel waterdicht.

Location: het kan zijn dat uw standaardlocatie hier al ingevuld staat; MSN Search herkent kennelijk de server die de verbinding met de zoekmachine legt. Desgewenst kunt u hier ook een andere plaatsnaam opgeven. Als dat een plaats in de Verenigde

Staten is, zult u resultaten Near Me voor die plaats kunnen opvragen.

Search language: beperk hier uw zoekactie tot documenten in een of meer talen.

Standaard is dat u documenten in alle pagina's zoekt, maar u wijzigt dit door het keuzerondje 'Limit my searches to pages written in the following languages' aan te klikken en vervolgens de talen aan te klikken waarin de documenten gesteld mogen zijn die u wilt vinden.

Op incidentele basis is deze optie ook beschikbaar in de Search Builder.

4.2.4.5. Het resultatenschermb

Het resultatenschermb toont het complete zoekscherm, zoals hierboven beschreven, met in het zoekvenster de vraag zoals die is uitgevoerd en daaronder de zoekresultaten. Als MSN Search een spelfout vermoedt, komt het met de suggestie voor een andere spelling, maar toont tevens de treffers die het met de mogelijk verkeerd gespelde term wel gevonden heeft. Erg verfijnd is de spellingcontrole niet: catagory wordt niet als een verkeerd gespelde term herkend, pasttime en indentiteit daarentegen wel.

De zoekresultaten openen met 1-3 sponsored sites, in een wat kleiner corps tegen een licht gekleurde achtergrond. Ook rechts en onder op elke pagina van de zoekresultaten staan deze advertenties, die gevonden worden omdat een van de door u opgegeven zoektermen overeenkomt met een term waarop een bedrijf een bod heeft gedaan: zo'n sponsored site is een advertentie die op uw scherm verschijnt als u met de betreffende term gezocht heeft en waarvoor het bedrijf een bepaald bedrag aan MSN (of Google of Yahoo!) betaalt iedere keer als u de link naar zo'n sponsored site volgt.

Als MSN Search in uw vraag de naam van een artiest, een album of een song uit zijn muziekdatabase herkent, toont het onder de Sponsored sites een plaatje met een link naar het betreffende record in MSN Music. Rechts vindt u links naar de TOP DOWNLOADS van de betrokken artiest, met samples die u kunt beluisteren.

Iets dergelijks gebeurt als MSN Search uw vraag als een nieuwsvraag herkent.

Voorbeeld: Oscars. MSN gebruikt voor zijn nieuwsresultaten de nieuwszoekmachine Moreover.

Web Results

Page 1 of 1,853,055 results containing **Lewis black** (0.34 seconds) (with SafeSearch: [Moderate](#))

Lewis Black Tickets At StubHub - www.stubhub.com SPONSORED SITES		
Buy Lewis Black tickets. Large selection of tickets with a wide range of prices. Fast, easy, and 100% guaranteed on StubHub.com.		
Lewis Black Items on eBay - www.ebay.com		
Find Lewis Black items at low prices. Shop for all kinds of unique products on eBay - The World's Online Marketplace.		
Lewis Black MP3s at eMusic - www.emusic.com		
Buy Lewis Black MP3s for less. Free 2 week trial. Get 50 free MP3s. Just \$9.99/month after free trial.		
» Artist: Lewis Black	TOP DOWNLOADS	SPONSORED SITES
Satirist and standup comic Lewis Black rose to prominence in the late '90s with regular appearances as a commentator on Comedy Central's The Daily Show. Obsessed with human...	Greed (LP Version) sample	Lewis Black at Amazon.com
MSN Music	A Sense Of Humor (LP Version)... sample	Save up to 35% on new and top selling DVD and videos at Amazon.com....
	Who's F**king Who (LP Version)... sample	www.amazon.com
The Official Website for Comedian Lewis Black		Lewis Black Tickets
Looking to book Lewis ?		Lewis Black tickets on Stubhub. Great prices. Wide selection. Big...
www.lewisblack.net Cached page		tickets.stubhub.com

De phrase "Lewis Black" levert als eerste treffers drie Sponsored sites en een record uit MSN Music op.

Tussen de sponsored sites in staan de eigenlijke treffers: een titel (als de spider die heeft kunnen vinden, anders de URL), gevolgd door een snippet, een URL, een link naar de cacheversie van het betreffende document en bij onlangs gespiderde documenten de datum waarop het document het laatst gespiderd is. Opvallend bij zoekacties bij MSN Search is het grote aantal 'redirects' waarop je stuit. Dat betekent dat MSN nog moeite heeft 'search engine spam' buiten de deur te houden.

Spammers gebruiken een heel scala aan technieken om zoekmachines en hun gebruikers te misleiden, variërend van onzichtbare (via de kleur) of onleesbare (via de lettergrootte) tekst op hun pagina's tot 'artificial link farms', netwerken van spammers die naar elkaars pagina's linken met het doel de 'link popularity' van alle deelnemers aan het netwerk kunstmatig te verhogen.

Zoekmachines reageren op dit soort praktijken door 'penalizing', waarbij de schuldigen (via het rankingalgoritme of op een andere manier) van de database worden uitgesloten of heel laag in de ranking worden gezet.

Voor meer over dit onderwerp, zie: <http://www.webmasterscene.com/article/41/>.

MSN heeft geen eigen directory en toont ook geen directoryresultaten. Wel treft men bij de eerste zoekresultaten wel eens treffers aan die MSN bij zijn concurrenten (Google, Ask Jeeves) vandaan heeft gehaald, iets wat ik nog bij geen andere zoekmachine heb aangetroffen. (Wel resultaten uit andere databases overigens.)

Bij documenten in andere formaten dan HTML kan de cachekopie als een HTML-tekst worden bekeken. 'Similar pages' kent MSN Search niet.

Onderaan de resultatenlijst kan naar volgende pagina's worden doorgebladerd en vindt u nog een zoekvenster. Hierin start u een nieuwe zoekactie, geen zoekactie binnen het resultaat dat u op dat moment bekijkt (al kunt u natuurlijk wel termen aan de bestaande zoekvraag toevoegen).

4.2.4.6. Commandoversie

Alle opdrachten en instellingen in de Search Builder kunnen ook in de opdrachtregel worden geformuleerd. Daarnaast beschikt MSN Search over een hele reeks prefixen die alleen via de zoekregel kunnen worden ingegeven; zie boven, 4.2.4.3. Alleen instellingen via de Settings kunnen niet in de vorm van een opdracht worden ingegeven. **NB:** Als u andere dan de standaardinstellingen hebt gekozen en opgeslagen, wordt u daar bij volgende zoekacties niet over geïnformeerd!

4.2.4.7. Bespreking

In november 2004 kwam Microsoft na twee jaar onderzoek met een bètaversie van een eigen zoekmachine voor de dag, en omdat het bedrijf overal waar het zich manifesteert direct als een grote speler geldt, waren de verwachtingen toen al geruime tijd hooggespannen. Niet alleen de buitenwereld gaf daar overigens uiting aan, Microsoft zelf heeft er nooit een geheim van gemaakt dat het ernaar streefde een nieuw hoofdstuk in de geschiedenis van de zoekmachines te schrijven, en presenteert zich ook nu het bètastadium verlaten is als een speler op de zoekmarkt die Google en Yahoo! – de huidige protagonisten, die mondiaal samen bijna drie kwart van de zoekacties afhandelen – ernstige concurrentie kan aandoen (voor aandelen in de zoekmarkt van eind 2005, zie: [Nielsen NetRatings Search Engine Ratings](#)).

De vraag of dat zo is valt natuurlijk niet te beantwoorden door alleen maar naar de kwaliteit van de zoekmachine, de grootte van zijn database en de relevantie van zijn resultaten te kijken. Maar als die factoren het enige criterium zouden vormen waardoor het aandeel van een zoekmachine bepaald werd, zou MSN Search in zijn huidige gedaante een tweederangsspeler blijven.

<i>Vraag</i>	<i>Google</i>	<i>MSN Search</i>	<i>Yahoo!</i>	<i>ASK</i>	<i>Exalead</i>	<i>Gigablast</i>
"quantum gravity" "boson stars"	954	78	321	243	10	25
caravaggio clair- obscur rembrandt	722	69	448	60	109	46
badminton "world champions" 1999 "mixed doubles"	283	80	314	190	54	5
"global warming" statistics contrails aviation	837	187	314	183	31	6
Ugric Dravidian loan Iranian	518	134	250	131	66	4
Kovacs Beethoven "violin sonatas"	405	116	353	273	460	108
"Ignaz Paderewski" premier	156	26	85	0	91	9
maxima amalia leontien	226	12	370	10	33	30
verdonk staatsveiligheid	585	61	530	27	118	69
"Ashura Day" killed	553	168	434	211	76	53
oorlog venetië casanova	542	154	186	14	64	39
"relevance ranking" recall precision Boolean "information retrieval"	900	223	297	248	35	45

De resultaten van de zes grote zoekmachines met een eigen database voor 12 vragen. De vragen zijn zo gekozen dat bij alle zoekmachines het aantal treffers kon worden genoteerd dat ze ook werkelijk tonen. Alleen bij Ask, dat zijn zoekresultaten op een andere manier clustert dan de andere zoekmachines, zijn de aantallen vermeld die de zoekmachine zelf opgeeft. Er zijn dus geen vragen bij die voor MSN Search een aantal treffers opleverden dat de 250 te boven ging. De resultaten van Exalead zijn met automatic stemming 'off'. De getallen dateren van 24-3-2006.

Ik begin met een opsomming van de kenmerken waarmee MSN Search aan de weg timmert.

- MSN Search heeft een onafhankelijke grote database van documenten in formaten die naast HTML- in elk geval ook PFD- en DOC-bestanden omvatten. Over de omvang van die database valt niets met zekerheid te zeggen: hij is aanzienlijk geringer dan die van Google, en vermoedelijk ook flink wat kleiner dan die van Yahoo! die in een testje van 12 vragen in 9 gevallen een groter resultaat blijkt op te leveren, en slechts in 3 een kleiner.
- MSN Search indexeert HTML-documenten tot 1 Mb, dat is meer dan welke andere zoekmachine ook.
- De meest opvallende vernieuwing waarmee MSN Search bezoekers wil trekken, is dat het de mogelijkheid biedt zoekresultaten op drie verschillende manieren te sorteren: freshness, popularity en exactness. Dit gebeurt via de zogeheten Search Builder (in het Nederlands: Geavanceerd zoeken), die onder meer de optie Results Ranking (Rangorde resultaten) biedt. Deze geeft toegang tot drie schuifregelaars die apart en in combinatie tussen een waarde van 0 en 100 kunnen worden fijngeregeld. De freshness van een document wordt gemeten naar wat MSN de Document date noemt, dat is de datum waarop de meest recente versie van het document op de server is opgeslagen. Zoals ik hierboven in 'What's in a date?' (blz. 128 e.v.) al heb uiteengezet, zegt die zogeheten 'Last Modified'-datum heel weinig over de actualiteit van de verstrekte informatie. Wel blijkt bij een testje van deze sortering dat het criterium freshness een bepaald type documenten naar voren haalt: blogs en fora. Dat ligt ook voor de hand, want het zijn juist die webpagina's die voortdurend ververst worden. Maar het betekent tegelijkertijd dat het zoekcriterium freshness op zich alleen kan worden gebruikt als men op zoek is naar het soort informatie dat in blogs en fora de boventoon voert, namelijk opinies en nieuws. En het betekent ook dat MSN Search wat dit zoekacties betreft de concurrentie moet aangaan niet met andere grote mondiale zoekmachines, maar met zoekmachines als BlogPulse en Feedster, die zich op blogs en nieuws concentreren.

Hieronder volgt een staatje van de eerste twintig treffers op de vraag

massavernietigingswapens {frsh=100}{popl=0} {mtch=0} loc:NL.

Massavernietigingswapens is de zoekterm, {frsh=100}{popl=0} {mtch=0} zijn de sorteerinstellingen en loc:NL beperkt de zoekactie tot Nederland.

Achtereenvolgens staan vermeld: de datum van de cacheversie, de datum die de server opgeeft voor de huidige versie van het document, indien beschikbaar een verschijningsdatum die de auteur in het document vermeld heeft en ten slotte een eventuele Last Modified-datum in het document zelf. Om een idee te geven van het karakter van de documenten waar het hier om gaat, staat in de laatste kolom het type aangegeven. Het testje werd gedaan op 24 februari 2005.

<i>Hit</i>	<i>Cache</i>	<i>Server</i>	<i>Publicatiedatum</i>	<i>Last Modified in document</i>	<i>Type publicatie</i>
1	14-02-05	24-02-05	25-01-05	23-02-05	blog
2	24-02-05	24-02-05			database
3	22-02-05	24-02-05			database
4	24-02-05	24-02-05	22-02-05		forum

<i>Hit</i>	<i>Cache</i>	<i>Server</i>	<i>Publicatiedatum</i>	<i>Last Modified in document</i>	<i>Type publicatie</i>
5	24-02-05	24-02-05	14-01-05		forum
6	23-02-05	24-02-05	10-02-05		blog
7	24-02-05	24-02-05	20-02-05		blog
8	22-02-05	24-02-05	24-01-05		blog
9	21-02-05	24-02-05	18-10-04		database
10	23-02-05	24-02-05	16-07-03		blogarchief
11	21-02-05	24-02-05	17-02-05		blog
12	24-02-05	24-02-05	18-01-03		archieff nieuwsblad 2003
13	24-02-05	24-02-05	22-01-04		archieff nieuwssite
14	24-02-05	27-02-04	2003		archieff serie columns
15	24-02-05	24-02-05	11-12-02		forum
16	24-02-05	24-02-05	13-06-04		blog
17	24-02-05	24-02-05	2003		forum
18	23-02-05	24-02-05	22-02-05		forum
19	19-02-05	24-02-05			Website
20	19-02-05	24-02-05	18-03-04		nieuwssite

Uit bovenstaande tabel blijkt inderdaad dat serverdatum en publicatiedatum nog wel eens verschillen en dat freshness van de 'document date' geen garantie is voor actualiteit van de informatie. Verder zou een vraag naar bijvoorbeeld wat de goegemeente van een onlangs verschenen boek vindt deze kwaliteit van MSN Search beter tot haar recht doen komen dan die naar informatie over massavernietigingswapens.

Bij de 'popularity ranking' gebeuren vreemde dingen. Zoals te verwachten, levert dezelfde vraag als hierboven gesteld, maar nu met de instelling {popl=100}, een andere volgorde van de treffers op dan met {frsh=100}. Volgens de Help is deze instelling bedoeld 'to add emphasis to sites by the number of other sites that link to them'. Je zou dus verwachten dat deze sortering documenten zou tonen met een aflopende reeks van aantallen andere documenten die daarnaar verwijzen. Maar als we dat met behulp van het prefix `link:` controleren, blijkt die reeks volgens een schijnbaar willekeurig patroon te verlopen.

In de tabel hieronder staan de URLs van de eerste 12 treffers voor de vraag:

"alternatieve geneeswijzen" {frsh=0} {mtch=0} {popl=100} (loc:NL) met de bijbehorende aantallen links voor MSN. (De instellingen {frsh=0} {mtch=0} zijn toegevoegd om eventuele versturende invloeden van andere rankingfactoren uit te schakelen.)

<i>URL</i>	<i>MSN</i>
www.klik.nl/start/index.htm	6
www.infolijn-ag.nl	150
www.loket.net/boeken/alternatieve.html	33
www.antenna.nl/altned	255
www.stelling.nl/simpos/kwakzalv.htm	176
www.gevonden.nl/gezondheid/alternatieve_geneeswijzen/alternatieve_geneeswijzen.html	13
www.soulwork.nl/person_list.asp?k=alternatieve_geneeswijzen	14
www.noag.org/geneeswijze/geneeswijze.htm	13
www.amerigo.nl/ow/geneeswijzen/geneeswijzen.html	15
alternatievegeneeswijzen.pagina.nl	1063
www.geneeswijzer.org/static/main.html	18
alternatieve-geneeswijzen.startkabel.nl	73

Ook als we in aanmerking nemen dat deze getallen maar bij benadering de werkelijke aantallen inkomende links voorstellen en dat de zoekmachine van MSN misschien net als Google en Ask bij het bepalen van de link popularity van een pagina of site niet alleen naar aantallen inkomende links kijkt, maar ook naar populariteit van de sites die de links leggen, is dit een bijzonder vreemd resultaat. Je zou verwachten met de instelling {popl=100} 'hubs' en 'authorities' te vinden, maar het lijkt er niet op.

De derde sorteermogelijkheid van MSN Search is die op exactheid. De schuifregelaar gaat hier van 'Exact match' onderaan naar 'Approximate match' bovenaan. Volgens de tekst van de helpfile moet de schuifregelaar voor dit sorteercriterium als volgt worden gebruikt: 'To put the most emphasis on the match between your exact search words and your results, move the slider on the right down. This will de-emphasize the two other slider rankings.' Dit betekent dus dat waar de regelaar bij 'Freshness' en 'Popularity' omhoog moest worden getrokken, dat bij deze derde schuifregelaar naar omlaag moet. Het effect daarvan zal zijn dat de invloed van de andere twee schuifregelaars afneemt, en, mag je verwachten, de 'exactness', anders gezegd de relevantie, van de eerst getoonde treffers toe. Verwarrend blijft dan wel dat naarmate regelaar omhoog wordt getrokken de waarde in {mtch=} toeneemt. 'Exact match' is dus 0, 'Approximate match' = 100. En die verwarring wordt er niet minder op als we in een op de site van Microsoft

beschikbare ['Guide to MSN Search and the MSN Search Toolbar'](#) (versie mei 2005) lezen dat: “Het op en neer bewegen van deze schuifregelaar bepaalt of zoekresultaten strikter dan wel minder strikt overeenkomen met de ingegeven zoektermen. Als klanten zeker zijn van de formulering van hun zoektermen en de manier waarop woorden met elkaar verbonden worden, zullen ze specifiekere en minder willekeurige resultaten krijgen door deze instelling dichter naar 'exact match' bij te stellen. Maar als ze er niet zeker van zijn dat ze naar de juiste termen op zoek zijn, zal een instelling 'approximate match' waarschijnlijker de resultaten opleveren waarnaar ze op zoek zijn.”

Het beste lijkt mij er maar van uit te gaan dat MSN Search u bij de instelling van de 'Exact match'-regelaar op 0 een zoekresultaat zal leveren dat uitsluitend of zoveel mogelijk wordt bepaald door MSN's relevantie-algoritmen, en niet of minder door de actualiteit en/of de populariteit van de treffers. Wat precies het effect is, heb ik niet kunnen vaststellen. Diverse proefjes met vragen waarbij de resultaten bij de instelling {mtch=0} en {mtch=100} met elkaar werden vergeleken laten wel verschillen zien, maar die beperken zich bij de eerste twintig resultaten tot wat stuivertje-wisselen van treffers. Het aantal treffers blijft gelijk, wat erop wijst dat 'Approximate match' niet inhoudt dat er treffers worden gevonden die niet aan alle zoekcriteria voldoen of die met variante woordvormen worden gevonden. Verder staan treffers die met {mtch=100} worden gevonden soms hoger en soms lager in de resultatenlijst dan dezelfde treffers in de lijst bij {mtch=0}; dat betekent dat er van een consequent rankingmechanisme op het punt van 'exactheid' geen sprake is.

{frsh=0} {popl=0} {mtch=0}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
{frsh=0} {popl=0} {mtch=100}	1	2	5	3	4	6	7	10	14	8	9	25	13	12	28	11	[23]	[22]	15	16

De volgorde van de eerste twintig treffers voor de vraag `child passenger safety technical encyclopedia` bij de instelling {mtch=0} vergeleken met de volgorde van dezelfde treffers bij de instelling {mtch=100}.

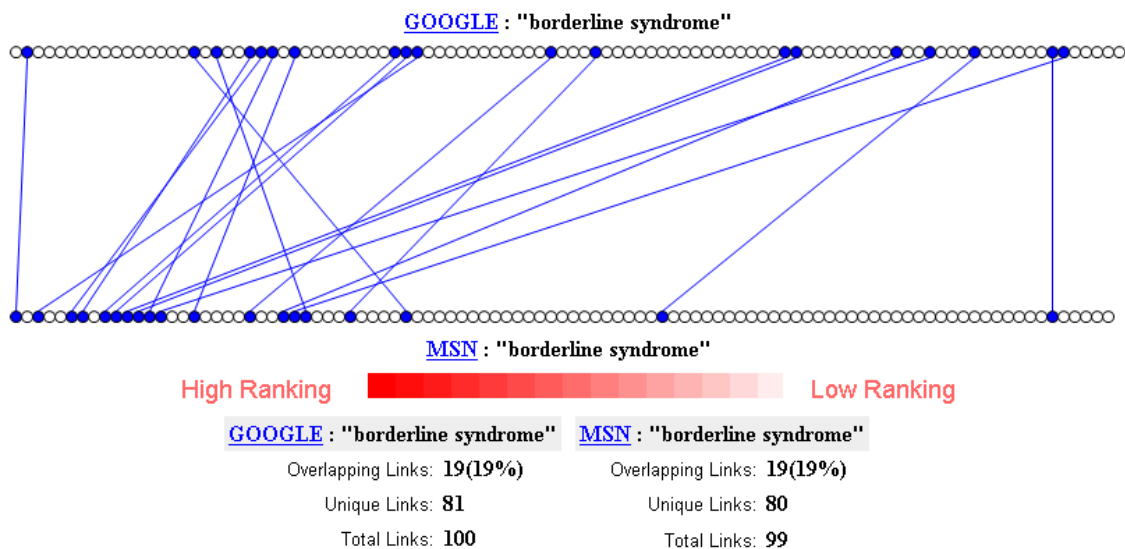
De rankingnummers tussen vierkante haken betreffen pagina's die niet helemaal identiek zijn aan die in de bovenste rij, maar wel tot dezelfde site behoren. De test is gedaan met de instelling: 1 document per site.

Wat bij inspectie van de eerste en de 250e treffer van diverse zoekacties met {mtch=100} wel opvalt is dat de frequentie van de zoektermen in het eerst getoonde document groter is dan die in het 250e, een aanwijzing dat 'exactness' onder meer iets met woordfrequenties te maken heeft.

Al met al heb ik niet de indruk dat dit gedeelte van de Search Builder nu een bijzonder nuttige functie van MSN Search is. Voor recente berichtgeving in de media, blogs en fora zijn meer gespecialiseerde zoekmachines beschikbaar, de popularity slider doet niet wat je ervan zou mogen verwachten en de werking van de Exact/Approximate Match slider blijft duister; en zolang we niet weten wat we ermee kunnen, kunnen we er niets mee.

Met zijn standaardinstellingen van {frsh=50} {popl=50} en {mtch=50} laat MSN Search zijn zoekresultaten in veel gevallen in een ranking zien die minder overeenkomsten met die van Google of Yahoo! vertoont dan ze onderling hebben. De overige functies van de Search Builder zijn tamelijk elementair. Prettig is dat opdrachten die via de Search Builder worden gedaan, kunnen worden herhaald. Dit leidt er onder meer toe dat op deze manier limits als `(site:nl OR site:be)` of `(loc:NL OR loc:BE)` aan een vraag kunnen worden toegevoegd, en vragen als [\(link:www.searchamovie.com/ OR link:www.bollywood.de/\)](http://www.searchamovie.com/) AND

[link:http://www.imdb.com](http://www.imdb.com) kunnen worden gesteld (dit laatste kan niet via de Search Builder, maar wel via wat aanpassingen in het zoekvenster).



Overlap van de eerste honderd treffers van Google en MSN Search voor "borderline syndrome".

- Encarta. Een aantrekkelijk extraatje bij het gebruik van MSN Search is dat u gedurende twee uur per vraag Encarta mag gebruiken, de verzameling van online naslagwerken van Microsoft. Het klapstuk van de verzameling, de encyclopedie, staat goed bekend en bevat ca. 40.000 artikelen. Dat zijn er wat minder dan de Columbia Encyclopedia die Google en Yahoo! in de aanbieding hebben, maar daar staat tegenover dat de artikelen iets minder beknopt zijn. Noch qua aantal noch qua lengte van de artikelen kan Encarta concurreren met Wikipedia dat in Google's pakket van Answers.com zit, maar concept en kwaliteit van deze door vrijwilligers samengestelde encyclopedie zijn omstreden.
Het gebruik van Encarta wordt vergemakkelijkt door MSN's Direct Answers, dat de mogelijkheid biedt om in natuurlijke taal vragen te stellen over onderwerpen die in Encarta worden bestreken. Direct Answers kan antwoord geven op vragen naar de betekenis van woorden (What is amaretto?), feitelijke vragen (How long is the Amazon River?), herkent soms vragen over muziek (Justin Timberlake) en voert u dan in één klik naar de winkel waar MSN Music online muziek verkoopt, en heeft een rekenmachine met conversiemogelijkheden.
- Near me. Wie vanuit Nederland MSN Search gebruikt zal doorgaans niet in aanraking komen met een toevoeging waar MSN flink mee aan de weg timmert: localization, een van de aspecten van personalisatie die in de huidige ontwikkeling van zoekmachines een prominente rol spelen. MSN Search is de eerste van de grote mondiale zoekmachines die zijn gebruikers de keuze geeft tussen mondiale en lokale zoekresultaten; in het laatste geval, als de gebruiker met de knop Near Me zoekt, past de zoekmachine de resultaten aan aan het IP-adres waar de bezoeker vandaan komt. U kunt dit zelf testen door een zogeheten 'public' of 'anonymous proxy server' te gebruiken; een voorbeeld is 'the Cloak', <http://www.the-cloak.com/login.html>. (Typ: <http://search.msn.com> in de regel boven de knop 'Start Surfing' en klik op de knop. Wacht tot het zoekscherm van MSN Search opstaat en kijk onder 'Settings'. U zult zien dat MSN Search u nu in Houston, Texas lokaliseert.)
- RSS. Handig is ten slotte dat zoekacties in MSN Search als een RSS-feed in XML-formaat in een RSS Reader kunnen worden ingelezen, waarna de zoekactie

automatisch periodiek kan worden herhaald en nieuwe treffers door de RSS Reader worden gemeld. Daarbij is MSN Search de enige zoekmachine die aparte prefixen heeft om feeds en pagina's met een feed op te sporen.

Naar mijn mening leggen van bovengenoemde punten waarop MSN Search de concurrentie met andere grote mondiale zoekmachines als Google en Yahoo! wil aangaan, eigenlijk alleen de hoge indexerlimiet van 1 Mb, de RSS-feed en de mogelijkheden om feeds te vinden gewicht in de schaal. Het is natuurlijk prettig dat we de beschikking hebben gekregen over nog een grote, zelfstandige zoekmachine die het hele Web bestrijkt, maar MSN's ambitie is een alternatief voor Google en Yahoo! te bieden, niet een aanvulling. En om redenen die ik zo dadelijk zal noemen, zou, als ik naast Google een tweede zoekmachine zou willen gebruiken, Yahoo! mijn eerste keus zijn en niet MSN. Over de sliders, MSN's innovatieve troef, heb ik al het nodige gezegd, en dat we gratis Encarta mogen gebruiken is natuurlijk fijn, maar voor de meeste vragen geen reden om op MSN over te stappen. De localization die MSN Search bij Near Me-zoekacties op de resultaten toepast, zou dat voor bepaalde vragen wel kunnen zijn als dit extraatje ook onze contreien bereikt.

De reden dat MSN Search voorlopig niet mijn eerste of tweede keuze onder de zoekmachines zal zijn is niet dat MSN Search onvoldoende zoekmogelijkheden biedt. Weliswaar is de functionaliteit van vde Search Builder nogal beperkt, maar sinds MSN in juni 2005 een hele reeks prefixen aan zijn arsenaal toe heeft gevoegd kan het zich in dit opzicht in elk geval meten met Yahoo!. Alleen de joker en het numrange:-prefix van Google ontbreken nog. Nee, de voornaamste reden waarom MSN Search voorlopig niet tot mijn favoriete zoekmachines zal behoren zijn de kwaliteit van de ranking en die van de snippets. Daarvan hieronder een voorbeeld.

Web Results

Page 1 of 18,411 results containing **le blanc vivaldi** (0.23 seconds)

[Le Blanc Travel Guide](#) - [travel.yahoo.com](#)

Traveling to Le Blanc? Start with Yahoo! Travel Guides and save up to 70% on hotels. Find user reviews, star ratings, maps,...

SPONSORED SITES

[Find Books by Le Blanc](#) - [www.allbris.com](#)

Buy used, new, rare and out-of-print books by Le Blanc. Millions of books from thousands of booksellers worldwide – all in...

[Vivaldi's Pasta and Grill - Wine List](#)

... and pasta dishes. Chardonnay/Pinot Grigio '**Le Rime**' (Banfi) (1) - A blend of rich Chardonnay and ... pairing with any fish dish. L'Acadie **Blanc** (Jost, Nova Scotia), Canada (1) - The L'Acadie ...

[www.vivaldis.ca/wine/index.html](#) [Cached page](#)

[Le Warwick Hotel Geneva - Latest Offers](#)

... Saisons Restaurant, La Bonne Brasserie and **Le Vivaldi** Piano Bar all of which are within the hotel. The ... Rex Boutique Hotel Hotel Balzac Residence Mont **Blanc** Hotel Les Nations Carlton Other Links Singles ...

[www.calle.com/world/CH/geneva/le-warwick.html](#) [Cached page](#)

[catherine redding :: soprano](#)

... Hamilton, Karina Gauvin, and Suzie **Le Blanc**. She has studied intensively with Nancy ... Requiem, Dvorak's Mass in D and **Vivaldi's** Gloria, in such prestigious venues as ...

[www.catherineredding.com/bio.html](#) [Cached page](#)

[catherine redding :: soprano](#)

... 2000 Stuart Hamilton, Karina Gauvin, Suzie **Le Blanc** Performing Experience Oratorio/Concert ... Bucks Choral Society/Festival di musica: **Vivaldi** Gloria (Jul. '03) Okanagan Festival ...

[www.catherineredding.com/resume.html](#) [Cached page](#)

[Show more results from "www.catherineredding.com".](#)

[ArkivMusic | Histoires / Claremont Duo](#)

... being realized by this combination (**Vivaldi**); the program is finally complemented by a ... Histoires: no 2, **Le** petit âne **blanc** by Jacques Ibert Performer: Maxine Neuman (Cello ...

[www.archivmusic.com/classical/album.jsp?album_id=74136](#) [Cached page](#)

In dit voorbeeld ben ik op zoek naar een CD-opname die de sopraan Suzie (of was het Susanne?) Le Blanc heeft gemaakt van geestelijke werken van Vivaldi. Erop vertrouwend dat mijn zoekmachine een phrase wel zal herkennen als hij er een ziet voer ik de zoektermen waarvan ik zeker ben op een simpele manier in: le blanc Vivaldi. Zoals u ziet begint het resultatenscherm van MSN Search met enkele 'sponsored results' waarin de phrase "Le Blanc" aanwezig is, maar de andere zoekterm, Vivaldi, vermoedelijk niet. Des te vreemder is het dat bij de eerste vijf algoritmische treffers er twee zijn waarin le noch met blanc, noch met Vivaldi verbonden is; een derde treffer bevat de phrase "le Vivaldi", ook raar, aangezien de phrase "le blanc" in MSN's database ca. 300 maal zo frequent voorkomt als "le Vivaldi". Hoe dan ook, van de CD is hier en verderop in MSN's resultatenlijst geen spoor te bekennen. Dat het anders kan, laat Yahoo! zien:

The screenshot shows a Yahoo! Search interface. At the top, there are navigation links for Web, Images, Video, Directory, Local, News, and Shopping. The search bar contains the text 'le blanc vivaldi' and a 'Search' button. Below the search bar, there are links for 'My Web BETA', 'Subscriptions (New)', 'Shortcuts', 'Advanced Search', and 'Preferences'. The search results section shows 'Results 1 - 100 of about 90,000 for le blanc vivaldi - 0.25 sec. (About this page)'. A yellow warning icon indicates a suggestion: 'Did you mean: leblanc vivaldi'. The first five results are listed below:

1. [Suzie Le Blanc Performs Sacred Works by Vivaldi - Music at Yahoo! Shopping](#)
Yahoo! Shopping is the best place to comparison shop for Suzie **Le Blanc** Performs Sacred Works by **Vivaldi** - Music. Compare products, compare prices, read reviews and merchant ratings.
[shopping.yahoo.com/p:Suzie Le Blanc Performs Sacred Works by Vivaldi... - More from this site - Save - Block](#)
2. [Vivaldi: Sacred Works / Le Blanc, Stubbs, Teatro Lirico - Overstock.com](#)
Only \$12.41 for this CD (\$1.40 shipping). Save big on Classical Music. Find **Vivaldi**: Sacred Works / **Le Blanc**, Stubbs, Teatro Lirico and the latest titles at clearance prices at
[overstock.com/cgi-bin/d2.cgi?PAGE=PRODUCT&PROD_ID=596797&cid=64666&... - More from this site - Save - Block](#)
3. [chatons chartreux](#) - Translate this page
... portée née le 3 mai- parents SIMBAD et TABBATAH ... **Vivaldi** est le premier à venir vers vous et à vous regarder dans **le blanc** des yeux en ronronnant, bien ...
[cartoonl.club.fr/chatons chartreux/chatons chartreux4.htm - 6k - Cached - More from this site - Save - Block](#)
4. [The Music Show - 11/11/2000: Saturday 11th NOVember, 2000](#)
... SALVE REGINA, RV 617: EXCERPT. (**VIVALDI**) Suzie **Le Blanc**, soprano ...
Director: Stephen Stubbs. **VIVALDI**: SUZIE LE BLANC. VANGUARD CLASSICS 99157 ...
[www.abc.net.au/rn/music/mshow/s210851.htm - 21k - Cached - More from this site - Save - Block](#)
5. [Amazon.fr: Musique: Suzie Le Blanc \[IMPORT\]](#) - Translate this page
... **Le** bouche à oreille ... avec **le** Programme Partenaires | Suzie **Le Blanc** [IMPORT] Antonio **Vivaldi** (Compositeur), Suzie **Le Blanc** & Stephen Stubbs ...
[www.amazon.fr/exec/obidos/ASIN/B00005YRF6 - 33k - Cached - More from this site - Save - Block](#)

Hier geen storende reclames en vier van de vijf eerste treffers in de roos. Niet bij alle treffers is de snippet daarover even duidelijk, maar in combinatie met elkaar bieden ze de gebruiker in één oogopslag die informatie die hij nodig heeft om te kunnen constateren dat hij heeft gevonden wat hij zocht.

Natuurlijk vindt MSN Search de juiste treffers ook, maar daar heeft het wel meer informatie van de gebruiker voor nodig. "[le Blanc](#)" [Vivaldi](#) levert pas op de plaatsen 9, 10 en 11 relevante treffers op, en pas bij de vraag "[suzie le blanc](#)"

[Vivaldi](#) komt MSN Search met een relevante treffer op de eerste plaats. Let ook op de kwaliteit van de snippet:

Web Results

Page 1 of 109 results containing "suzie le blanc" vivaldi (0.11 seconds)

[Le Blanc Travel Guide](#) - [travel.yahoo.com](#)

Traveling to Le Blanc? Start with Yahoo! Travel Guides and save up to 70% on hotels. Find user reviews, star ratings, maps,...

SPONSORED SITES

[Find Books by Le Blanc](#) - [www.alibris.com](#)

Buy used, new, rare and out-of-print books by Le Blanc. Millions of books from thousands of booksellers worldwide – all in...

[Suzie Le Blanc Performs Sacred Works by Vivaldi](#)

Suzie Le Blanc Performs Sacred Works by **Vivaldi** Suzie,Le,Blanc,Performs,Sacred,Works,by,Vivaldi, **Suzie Le Blanc** Performs Sacred Works by **Vivaldi** **Suzie Le Blanc** Performs Sacred Works by **Vivaldi** suzie ...

[www.allvcd.co.uk/Suzie-Le-Blanc-Performs-Sacred-Works-by-Vivaldi-B88-YRF-880-2.html](#)

Ik weet het, één voorbeeld is geen voorbeeld. Maar mijn indruk dat MSN Search niet de meest relevante resultaten van de grote zoekmachines levert, wordt door veel anderen gedeeld. Zie met name The Search Engine Relevancy Challenge at Rustybrick (<http://www.rustybrick.com/rustysearch-results.php>).

4.2.5. Ask (voorheen Ask Jeeves/Teoma)

4.2.5.1. In kort bestek

URL (Basic Search): <http://search.ask.com/>, <http://www.ask.com>

Omvang : volgens de meeste schattingen 2,5 à 3 miljard; volgens een studie van Gulli en Signorini (<http://www.cs.uiowa.edu/~assignori/web-size/>) indexeert Ask ongeveer 6 miljard pagina's, iets minder dan 58% van het zichtbare Web.

Indexeert tot: ca. 80 K (HTML)

Maximale omvang zoekargument 10 zoektermen, rest wordt genegeerd

Indexeert andere formaten dan HTML: ja (PDF; zie ook:

<http://www.faganfinder.com/filetype/>)

Toont: nooit meer dan ca. 200 treffers

Updating: enkele minuten – enkele weken

About Ask.com: <http://about.ask.com/en/docs/about/index.shtml>

Help en info: http://about.ask.com/en/docs/about/help_central.shtml;

<http://about.ask.com/en/docs/about/webmasters.shtml>

Advanced Search: <http://www.ask.com/webadvanced>

Blog: <http://blog.ask.com/>

Meer over Ask (Teoma):

Mapping the semantics of Web text and links:

http://www.informatics.indiana.edu/research/publications/maps_ir.pdf

Google – Teoma Comparison: Compare Search Results:

<http://www.googleguy.de/google-teoma/>

Overig:

<http://dc.internet.com/news/print.php/1002061>

<http://www.abondance.com/ouils/teoma.html>

<http://www.searchengineshowdown.com/features/teoma/review.html>

<http://www.at-web.de/suchboard/teoma.htm>

Overzicht

Soort zoekactie:	Voorbeelden:	Vindt:
Natuurlijke taal (alleen Engels!)	Can I use natural language in Ask searching?	Probeer!
<i>Prefixes</i>		
intitle: (komt overeen met de optie In page title onder 'Anywhere on page, page title, or URL' in de Advanced Search)	intitle:hadith flu intitle:homeopathy intitle:"stamp collecting" faq faq intitle:"stamp collecting" intitle:orchids site:edu intitle:database geoloc:africa intitle:computers lang:NL intitle:freeware last:week	documenten met de opgegeven zoekterm(en) of phrase in de titel van de pagina, ongeacht waar ze in de zoekexpressie staan (vgl. Google allintitle:) NB: Ask neemt intitle: letterlijk: u vindt hiermee geen documenten op zoektermen die in de META keywords en description (die door Ask wel worden geïndexeerd) voorkomen NB: intitle: kan niet in één vraag worden gecombineerd met inurl:; het kan wel met andere prefixen gecombineerd
inurl: (komt overeen met de optie In URL onder 'Anywhere on page, page title, or URL' in de Advanced Search)	inurl:minvws vaccinatie vaccinatie inurl:minvws welzijn inurl:minvws inurl:nieuws site:uva.nl inurl:search geoloc:europa inurl:search lang:NL inurl:occasions site:nl last:3months	documenten die de opgegeven zoekterm(en) in de (hele) URL bevatten, ongeacht waar ze in de zoekexpressie staan (vgl. Google allinurl:) geen treffers: Ask kent geen documenten met zowel minvws als welzijn in de URL NB: inurl: kan niet in één vraag worden gecombineerd met intitle:; het kan wel met andere prefixen worden gecombineerd
site: (komt overeen met de optie 'Domain or site' in de Advanced Search)	handleiding site:uu.nl handleiding site:let.uu.nl handleiding site:www2.let.uu.nl dierenwelzijn site:www.minlnv.nl site:www.minlnv.nl dierenwelzijn site:tv lang:NL de	documenten met de opgegeven term(en) bij de opgegeven site (domain of subdomain), m.a.w. beperkt de zoekactie tot het opgegeven domain (inclusief subdomains) of subdomain NB: alleen in combinatie met een zoekterm elders! site: kan niet in één vraag worden gecombineerd met geoloc: en kan niet worden herhaald; wel kan het met andere prefixen worden gecombineerd documenten in de Nederlandse taal die het woord 'de' bevatten en met de ccTLD tv (Tuvalu, een eilandstaatje dat zijn domain name te gelde maakt)

geoloc: (komt overeen met de opties onder 'Geographic region' in de Advanced Search)	laws geoloc:northamerica (of: NA) laws geoloc:europa (of EU) laws geoloc:southeastasia (of SEA) laws geoloc:indiaasia (of IA) laws geoloc:southamerica (of SA) laws geoloc:oceania (of OC) laws geoloc:afrika (of AF) laws geoloc:middleeast (of ME) laws geoloc:centralamerica (of CA)	documenten met een domein (ccTLD) dat behoort tot een van de genoemde regio's NB: alleen in combinatie met een zoekterm elders! geoloc: kan niet in één vraag worden gecombineerd met site: en kan niet worden herhaald; wel kan het met andere prefixen worden gecombineerd
lang: (komt overeen met de opties onder 'Language' in de Advanced Search)	Festschrift lang:DA kindergarten lang:DE kindergarten lang:EN gestalt lang:ES onze lang:FR agenda lang:IT Battus lang:NL dagen lang:NO pais lang:PT barn lang:SV	documenten in resp.: het Deens (DA), Duits (DE), Engels (EN), Spaans (ES), Frans (FR), Italiaans (IT), Nederlands (NL), Noors (NO), Portugees (PT) en Zweeds (SV) NB: alleen in combinatie met een zoekterm elders! Kan niet met andere prefixen worden gecombineerd.
last: (komt overeen met een van de opties onder 'Date page was modified' in de Advanced Search)	last:week last:2weeks last:month last:3months last:6months last:year last:2years "prinses Máxima" last:5weeks	documenten die in de opgegeven periode zijn gemaakt of bijgewerkt; last:3weeks e.d. kan ook NB: alleen in combinatie met een zoekterm elders ! last: kan niet worden gecombineerd met afterdate:, beforedate: of betweendate: berichtgeving over prinses Máxima in documenten die gedurende de afgelopen vijf weken op het Web zijn gezet of bijgewerkt
afterdate:[jjjjmmdd] (komt overeen met een van de opties onder 'Date page was modified' in de Advanced Search)	"Bin Laden" afterdate:20060301	documenten, gemaakt of gewijzigd na de opgegeven datum Het voorbeeld betreft de berichtgeving over Bin Laden (in documenten met een Last Modified datum) na 5 september 2005 NB: alleen in combinatie met zoekterm elders! afterdate: kan niet worden gecombineerd met last:, beforedate: of betweendate:
beforedate:[jjjjmmdd] (komt overeen met een van de opties onder 'Date page was modified' in de Advanced Search)	"Bin Laden" beforedate:20010911	documenten, gemaakt of gewijzigd vóór de opgegeven datum Het voorbeeld betreft de berichtgeving over Bin Laden vóór 11 september 2001 NB: alleen in combinatie met zoekterm elders! beforedate: kan niet worden gecombineerd met last:, afterdate: of betweendate:
betweendate:[jjjjmmdd, jjjjmmdd] (komt overeen met een van de opties onder 'Date page was modified' in de Advanced Search)	"Bin Laden" betweendate:20010910,20010911	documenten gemaakt of gewijzigd in de periode tussen de data aan weerszijden van de komma Het voorbeeld betreft de berichtgeving over Bin Laden op 10 september 2001 NB: alleen in combinatie met zoekterm elders! betweendate: kan niet worden gecombineerd met last:, afterdate: of beforedate:

<i>Operatoren</i>		
term1 term2 (komt overeen met de optie 'All the words' in Advanced Search)	Verdi Requiem	de ene EN de andere zoekterm (standaard)
-term (komt overeen met de optie Must not have onder 'Include or exclude words or phrases' in de Advanced Search)	Bach -Johann Bach -"Johann Sebastian"	de ene maar niet de andere term of phrase NB: - werkt niet met prefixen (wordt genegeerd)!
OR (komt overeen met de optie Should have onder 'Include or exclude words or phrases' in de Advanced Search)	Beethoven OR Mozart "Oost-Indische Compagnie" OR "West-Indische Compagnie" "prinses Maxima" OR "prinses Máxima" RVD "prinses Maxima" RVD OR "prinses Máxima" RVD "prinses Laurentien" OR "prinses Mabel" -prins "prinses Laurentien" -prins OR "prinses Mabel "prinses Mabel" -prins OR "prinses Laurentien" -prins "Rita Verdonk" OR "Ayaan Hirsi Ali" site:be	de ene OF de andere zoekterm, OF beide; OR verbindt twee woorden of twee phrases NB: hoewel de Advanced Search anders suggereert, is het in Ask niet mogelijk twee onderling met OR verbonden termen <i>tegelijk</i> in een AND-relatie met een derde term te verbinden (nesten); anders gezegd: EN heeft altijd prioriteit boven OF; ook toevoeging van een NOT-relatie met -term gaat maar ten koste van een van beide met OR verbonden zoektermen. documenten met daarin de phrase "prinses Maxima", plus documenten met daarin de combinatie van de phrase "prinses Máxima" met de term RVD documenten met daarin de phrase "prinses Maxima" in combinatie met de term RVD, plus documenten met daarin de phrase "prinses Máxima" in combinatie met de term RVD documenten met daarin de phrase "prinses Laurentien" (en eventueel de term prins), plus documenten met de phrase "prinses Mabel", waarin de term prins niet voorkomt documenten met daarin de phrase "prinses Laurentien" waarin de term prins niet voorkomt, plus documenten met daarin de phrase "prinses Mabel" waarin de term prins wel kan voorkomen documenten met daarin de phrase "prinses Mabel" en/of de phrase "prinses Laurentien", waarin de term Alexander niet voorkomt Wel kunnen termen in een OF-relatie worden verbonden met de diverse limits waarover Ask beschikt: uit België afkomstige documenten met de phrase "prinses Maxima" of "prinses Máxima"

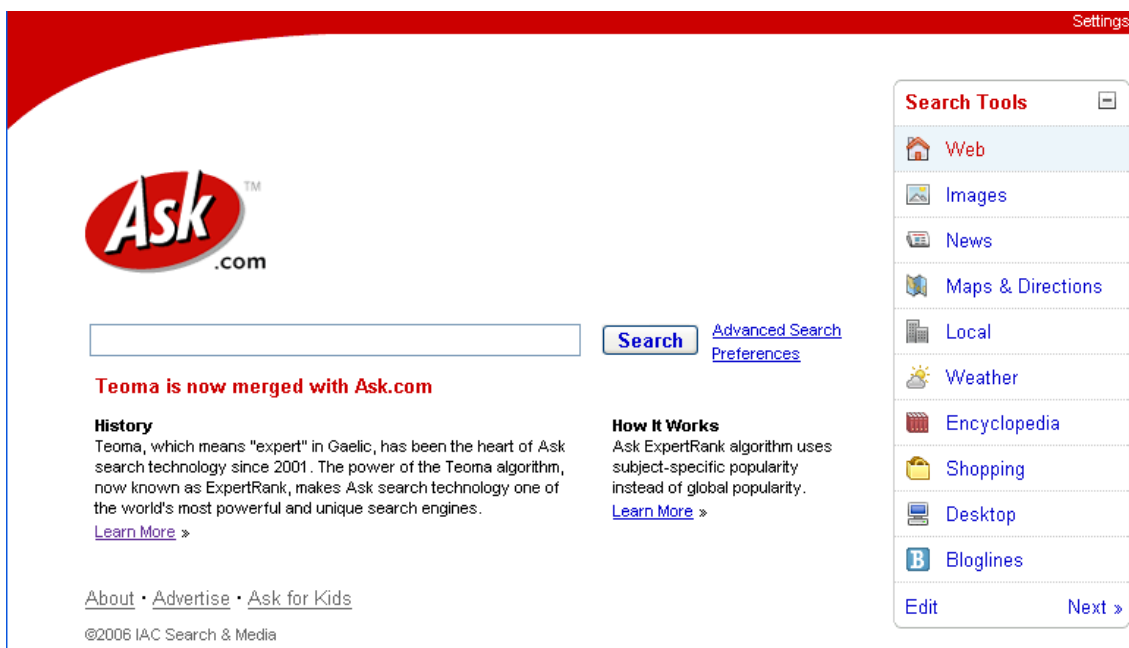
<i>Nabijheid</i>		
phrase ("...") (komt overeen met 'with the exact phrase' in de Advanced Search)	"Willem van Hanegem" "part time" "Willem van Hanegem" "Marco van Basten"	de opgegeven naam of zinsnede als reeks van opeenvolgende woorden, met tussen elke twee woorden een spatie of een ander leesteken documenten met de twee opgegeven namen
<i>Overige</i>		
.. results ¹⁾	10/20/30/50/100	kies het aantal treffers per pagina van de resultatenlijst
accenten	é, ï, ö, û etc. electricite électricité electricité électricité oölogie oologie oologie OR oölogie ode öde blöd site.nl	Ask maakt een strikt onderscheid tussen geaccentueerde en ongeaccentueerde vormen electricite électricité electricité électricité électricité oölogie, maar niet oologie oologie, maar niet oölogie oologie plus oölogie ode öde blöd, maar niet bloed
hoofdletters/ kleine letters (case-sensitivity)	next NeXT	Ask maakt geen onderscheid tussen hoofdletters en kleine letters
lees- en andere niet-lettertekens	-,/_&+#\$ part-time parttime c+ c++ &c single_user a# c# f# g# +a# c# f# g# a&b&c a_b_c \$5000 €5000 £5000 I/O #2 >2	Ask verwaarloost de meeste lees- en niet-lettertekens, of vat ze op als spaties binnen een phrase; maar in een aantal gevallen worden ze mee geïndexeerd en zijn als losse tekens of als onderdeel van een string van tekens vindbaar; probeer en let op de presentatie van de zoekresultaten (vet en evt. onderstreept wil zeggen: gevonden) part-time, "part time", maar niet parttime alleen parttime, maar niet part-time c, inclusief c+, c++ en &c c++ c, inclusief, &tc, c+ en c++ "single user", inclusief single_user, single-user enz. c EN f EN g (a = stopwoord) a EN c EN f EN g "a b c" abc 5000 5 EN 000, inclusief 5000 5 EN 000, inclusief 5000 "I O", inclusief I/O, i.o. 2 2 .

getallen	10000 10,000 10.000 25.32 25,32	10000, maar niet 10,000, 10.000 10 000, dus 10 EN 000 (bijv. 10 EN 5,000) "10 000", inclusief 10.000, 10,000, "10 000" "25 32", inclusief 25:32, 25,32, 25-32 25 32 (dus 25 EN 32, inclusief 25 EN 32,000)
stopwoorden	from Amsterdam to Warsaw +from Amsterdam +to Warsaw " from Amsterdam to Warsaw " " from Amsterdam " " to Warsaw "	Ask heeft wel stopwoorden, maar die kunnen toch worden gevonden door ze te 'plussen'; ook binnen een phrase zijn ze, al dan niet 'geplust', vindbaar Amsterdam [EN] Warsaw from [EN] Amsterdam [EN] to [EN] Warsaw "from Amsterdam to Warsaw" "from Amsterdam" [EN] "to Warsaw"
(spellingcontrole) [Spell Checker]	catagory (spelfout) wife-swlling (typfout) Rotterdam (typfout) " Berend Botje ging uit varen " " Berend Botje ging eens varen " " Berend Botje gong eens varen "	Spelling suggestion: category Spelling suggestion: wife-selling Spelling suggestion: Rotterdam? geen spellingsuggestie (correct) geen spellingsuggestie; geen treffers geen spellingsuggestie, geen treffers
suggesties voor verder zoeken: categorisatie: Narrow your search (in beperkte mate ook Nederlands)	history of china " history of china " regering	Chinese History, Ancient China, China Timeline, Chinese Culture, Dynasties of China, etc.; geen suggesties voor china = porselein Chinese History, Ancient China, China Timeline, Chinese Culture, Dynasties of China, etc.; geen suggesties voor china = porselein Nederlandse regering
suggesties voor verder zoeken: gerelateerde zoektermen: Expand your search (in beperkte mate ook Nederlands)	" Italian Renaissance " belasting	Ask zoekt pagina's met veel outgoing links bij het onderwerp van de opgegeven zoektermen Protestant Reformation Hundred Years War Humanism Sistine Chapel etc. Www Belastingdienst NI Belastingdienst Dutch Customs Dutch Tax Customs In Netherlands Netherlands Ministerie Van Financien
Suggesties voor verder zoeken: gerelateerde namen	" Italian Renaissance "	Michelangelo Mona Lisa Raphael Machiavelli Donatello, etc.
Suggesties voor verder zoeken: [term weighting]: volgorde zoektermen	kanker voeding voeding kanker hollandse nieuwe uitleg nieuwe uitleg hollandse nieuwe hollandse uitleg	als eerste treffers in de resultatenlijst, afhankelijk van de volgorde van de zoektermen, documenten waarin het ene dan wel het andere aspect van de vraag overweegt (op basis van de door Ask gebezigde statistische criteria) De volgorde van de zoektermen is ook van belang omdat Ask automatisch naar phrases zoekt.

[term weighting]: relatieve ranking en frequentie van de zoektermen	voeding voeding voeding voeding kanker kanker kanker kanker kanker voeding	als eerste treffers in de resultatenlijst, afhankelijk van de relatieve ranking en frequentie van de zoektermen, documenten waarin het ene dan wel het andere aspect van de vraag overweegt
truncatie	n/a	trunceren van termen is bij Ask niet mogelijk
[(automatic) stemming]; (automatische) toevoeging van zoektermen met dezelfde stam als de opgegeven zoektermen; voor de voorwaarden waaronder dit gebeurt, zie de volgende kolommen	child bicycle helmet	n/a

4.2.5.2. Basic Search

Ask's Basic Search-scherm is het toppunt van soberheid: Centraal staan de zoekregel en de zoekknop, met daaronder uitleg over de geschiedenis en de zoektechnologie van het bedrijf. Rechts vindt u een overzicht van de eerste tien van de in totaal 20 zoekinstrumenten waarmee Ask u laat werken; met Edit is dit overzicht naar eigen voorkeur aan te passen. Verder zijn er onder meer links naar de Advanced Search, Preferences en Settings en informatie over de werking van de zoekmachine en het bedrijf.



In de zoekregel kunt u uw zoektermen intypen, desgewenst met de operatoren OR en -, en met alle mogelijkheden van de Advanced search, vertaald in prefixes/limits, behalve het aantal treffers per pagina. De zoekregel accepteert meer dan tien zoektermen, maar verwaarloost de elfde en volgende. Stopwoorden tellen niet mee, tenzij 'geplust'

4.2.5.3. Preferences en Settings

Ik noem hier meteen ook de Preferences en Settings, waarmee een aantal voorkeursinstellingen kan worden opgegeven. Beide zien er hetzelfde uit en doen hetzelfde. Wel moeten de Settings die u hier instelt worden onderscheiden van de Settings die u onderaan het Advanced Search-scherm (zie 4.2.5.4) kunt saven. De

laatste hebben alleen betrekking op keuzes die de Advanced Search betreffen. De Preferences/Settings omvatten de keuze voor een of meer Locations (alleen van belang op het Noord-Amerikaanse continent) het aantal treffers per pagina (10-100) en de keuze of u de zoekresultaten in een nieuw venster ter zien wilt krijgen, een 'Family Filter', de mogelijkheid om Ask zoekacties te laten uitvoeren die u in de adresregel van uw browser typt, en de keuze voor een (inter)nationale versie van de zoekmachine. Als u hier voor Nederland kiest krijgt u de eerstvolgende keer dat u het adres <http://www.ask.com> bezoekt, de Nederlandse versie van de zoekmachine met als extra opties zoeken naar Nederlandse pagina's en pagina's in het Nederlands gepresenteerd.

4.2.5.4. Advanced Search

Het Advanced search-scherm opent met links naar de Settings (zie boven) en de Advanced Search Tips. Daarna volgt de hoofdzoekregel, waarin u een of meer zoektermen kunt intypen. Meer dan een zoekterm kunnen hier op twee manieren worden verbonden, met EN ('All of the words', de standaardverbinding) of als phrase. Desgewenst kunt u hier wel een of meer woorden met een of meer phrases verbinden door met 'All of the words te zoeken' en de phrase(s) zelf binnen dubbele aanhalingstekens te plaatsen.

The screenshot shows the Ask.com Advanced Search interface. At the top left is the Ask.com logo. To the right are links for 'Settings' and 'Advanced Search Tips'. A red banner below the logo reads 'Ask.com - Advanced Search'. The main search area contains a 'Find' section with a dropdown menu set to 'All the words' and a text input field containing 'child bicycle helmet', followed by a 'Search' button. Below this is the 'Include or exclude words or phrases' section, which has three rows for 'Should have', 'Must have', and 'Must not have', each with a dropdown menu and an empty text input field. Below these rows are buttons for '+ Add An Entry' and '- Delete An Entry'. The 'Location of words or phrases' section has a dropdown menu set to 'Anywhere on page'. The 'Language' section has a dropdown menu set to 'Any language'. The 'Domain or site' section has a radio button and an empty text input field. The 'Geographic region' section has a radio button and a dropdown menu set to 'Any geographic region'. The 'Date page was modified' section has a radio button and a dropdown menu set to 'Anytime', and two other radio buttons for 'Before' and 'Between' filters, each with dropdown menus for 'Month' and 'Year' and a text input field for a number. At the bottom of the form are buttons for 'Save Settings', 'Restore Previously Saved', and 'Restore Defaults'. A 'Search' button is also located at the bottom right of the main search area.

Het Advanced Search-scherm van Ask.

De gebruikelijke optie 'Any of the words' ontbreekt bij de hoofdzoekregel; de OF-relaties die daardoor tussen de ingegeven zoektermen worden aangebracht, moeten bij Ask worden gerealiseerd door in de volgende sectie van het zoekscherm termen bij 'Should have' in te vullen.

In de volgende sectie van het scherm kunt u zoektermen of phrases toevoegen, één per vakje. (U kunt deze sectie ook zelfstandig gebruiken, dus zonder iets in de hoofdzoekregel te hebben ingevuld.) De opties zijn hier Must have (standaard; EN-relatie), Must not have (NIET-relatie) en Should have (OF-relatie). De termen of phrases die u hier intypt worden met de geselecteerde relaties met elkaar en met de termen in de hoofdzoekregel verbonden. Denk erom dat bij Ask de EN-relatie altijd prevaleert: in de hoofdzoekregel *All of the words: Alexander Maxima* gecombineerd met *Should have: Máxima* resulteert in de vraag: *Alexander Maxima OR Máxima*, die door Ask als volgt wordt geïnterpreteerd: zoek de pagina's met daarin de termen Alexander EN Maxima, plus pagina's met daarin de term Máxima. Vandaar dat de zoekactie [Alexander Máxima OR Maxima](#) een veel groter resultaat oplevert dan [Alexander Maxima OR Máxima](#).

Denk er ook om dat Ask aan term weighting doet: de eerste term die u intypt krijgt een hoger gewicht dan de tweede en volgende. De zoekactie [Maxima Alexander OR Máxima](#) levert wel een zelfde aantal treffers op als [Alexander Maxima OR Máxima](#), maar anders geordend.

De NIET-relatie die u hier via 'Must not have' met andere termen aanbrengt, gaat altijd ten koste van de dichtstbijzijnde term. Zie de tabel voor een voorbeeld.

Desgewenst kunt u met 'Add an entry' meer zoekvakjes toevoegen, of met 'Delete an entry' het aantal zoekregels verminderen.

In de derde sectie kiest u waar uw zoektermen in de te vinden documenten moeten voorkomen: Anywhere on page (standaard), In page title (prefix: `intitle:`) of In URL (prefix: `inurl:`). Denk erom dat de keuzes In page title en In URL betrekking hebben op alle termen die u hebt opgegeven; `inurl:` en `intitle:` kunnen niet met een zoekterm elders worden gecombineerd en ook niet met elkaar. Wel is het mogelijk beide keuzes of hun prefixen met andere keuzes of prefixen (`site:`, `geoloc:`, `lang:`, `afterdate` etc.) te combineren.

In de volgende sectie kunt u uw zoekactie tot documenten in een bepaalde taal beperken. De keuzes zijn hier: Deens (`lang:DA`), Nederlands (`lang:NL`), Engels (`lang:EN`), Frans (`lang:FR`), Duits (`lang:GE`), Italiaans (`lang:IT`), Noors (`lang:NO`), Portugees (`lang:PT`), Spaans (`lang:ES`) Zweeds (`lang:SV`).

Standaard is Any language, wat natuurlijk nog veel meer talen omvat dan deze selectie.

Met Domain or site in de volgende sectie beperkt u de zoekactie tot een generiek domain of een ccTLD, of tot een bepaalde site. Alle niveaus zijn hier toegestaan: `site:uk`, `site:ac.uk`, `site:htsa.ie.hva.nl` en alles wat ertussen in zit. Per vraag kunt u maar één site of domain opgeven, dus ook verbinden met OR geeft niet het resultaat wat u ervan zou verwachten (de eerste site/domain in zo'n OF-relatie wordt door Ask genegeerd).

De andere keuze in deze zelfde sectie is Geographic region, waar Any geographic region de standaard is. Daarnaast kan men kiezen uit Noord-Amerika (`geoloc:NA`), Europa (`geoloc:EU`), Zuidoost-Azië (`geoloc:SEA`), India of Azië (`geoloc:IA`), Zuid-Amerika (`geoloc:SA`), Oceanië (`geoloc:OC`), Afrika (`geoloc:AF`), het Midden-Oosten (`geoloc:ME`) en Centraal-Amerika (`geoloc:CA`). Denk erom dat deze regio's enkel de landTLD's van de in die regio's gelegen landen bestrijken, niet de vaak talrijke generieke TLD's die ook in die landen zijn uitgegeven. In de regio Zuid-Amerika vind je dus wel pagina's over Máxima Zorreguieta met de landcode *ar* maar niet die met de TLD *com* die ook tot Argentinië behoren.

Ten slotte biedt de Advanced Search een aantal mogelijkheden om de te vinden documenten op datum te filteren. Dat kan op drie manieren:

1. terugrekenend van af de datum van vandaag met een periode van een week, twee weken, een maand, drie maanden, zes maanden, een en twee jaar. Dit staat gelijk met de syntax `last:1week`, `last:2weeks` enz. in de Basic Search-regel. Voordeel van die syntax is dat ze ook andere tijdsbestekken toelaat, bijvoorbeeld `last:2months` of `last:8weeks`.
2. voor of na een bepaalde datum
3. tussen twee data.

Met alle kanttekeningen die bij de datering van documenten op het Web moeten worden gezet, is dit een netjes ingericht zoekfilter van Ask, heel wat gebruikersvriendelijker dan wat Google op dit punt aan mogelijkheden biedt. Vergeet alleen niet het juiste keuzerondje voor het door u gekozen datumfilter aan te klikken, anders gaat het toch nog mis. Controleer ook in het resultatenscherf hoe Ask uw vraagstelling in een opdracht heeft vertaald, daarmee vermijdt u fouten en u leert ervan.

Helemaal onderaan het Advanced zoekscherm treft u de mogelijkheid aan de filters in de onderste vier secties van het venster (bijvoorbeeld Language en Geographic region) die u tijdens uw sessie hebt ingesteld te bewaren. De eerstvolgende keer dat u de Advanced Search gebruikt, vindt u ze dan ook weer zo ingesteld terug. Andere keuzes zijn 'Restore previously saved' of 'Restore defaults', voor als u tijdens de sessie (opnieuw) filters gewijzigd hebt.

4.2.5.5. Het resultatenscherf

Ask's resultatenscherf opent met de Basic Search-regel en de daarbij behorende links naar Advanced Search en Settings. Daaronder volgt, als u een tyfout hebt gemaakt, een spellingsuggestie (alleen Engels; probeer een van de 598 spellingsvarianten van 'Britney Spears' die Google herkent, zie <http://www.google.com/jobs/britney.html>).

Daarna volgen 1-3 Sponsored links, die door een gekleurde achtergrond en het kopje Sponsored Results van de 'natural' treffers zijn onderscheiden. Deze laatste bestaan uit een titel of een URL, een korte snippet met zoveel mogelijk van de zoektermen vet weergegeven, de URL en meestal een link naar een cachekopie, en ten slotte een link met de tekst Save. Deze laatste biedt de mogelijkheid geselecteerde treffers in de Bookmark Manager van Ask op te slaan, waarna daarmee de voor dit soort programma's gebruikelijke acties kunnen worden ondernomen. Zie [About Ask.com: My Stuff](#), en wat in paragraaf 4.2.3.9. over My Yahoo! is opgemerkt.

Handig is dat bij veel treffers een preview-mogelijkheid wordt geboden in de vorm van een verrekijker-icoontje vóór de URL. Door de muisaanwijzer daarop te zetten krijgt u een verkleinde weergave van de betreffende pagina te zien, klikken op het icoontje opent een nieuw venster met de betreffende pagina.

Ask toont maximaal twee pagina's per site; zijn er meer dan eindigt de tweede treffer met een link naar More results from... Daarnaast verschijnt bij een aantal treffers een link [Related Pages]. Als je daarop klikt lijkt het enige zichtbare effect te zijn dat de Sponsored Links verdwijnen en dat het aantal treffers per pagina tot tien wordt gereduceerd.

De rechter kolom van Ask's resultatenlijst is gereserveerd voor drie van de vijf specialiteiten die Ask moeten onderscheiden van zijn naaste concurrenten. Dit zijn, onder het hoofdje 'Narrow your search', termen en phrases waarmee breed opgezette vraagstellingen kunnen worden toegespitst (categorisatie dus), onder het hoofdje 'Expand Your Search' suggesties voor gerelateerde zoektermen en phrases, en, onder 'Related Names', persoonsnamen die met het onderwerp van de vraag in verband kunnen worden gebracht. Bij elk van de drie lijstjes zijn onder 'More >>' meer termen of namen te vinden. Bij het volgen van een link in een van de lijstjes wordt de



[Dutch Master Painters](#)

from the year 1800 till 1950 for sale in stock.
www.holland-art.com

[Painters](#)

Top 3 Sites on Painters
Best3Websites.com

[Painters](#)

Free Painters information straight from bloggers.
www.BlogHog.biz

Sponsored Results

[Dutch Artists: Artcyclopedia](#)

Artists by Nationality: Dutch Artists Chronological Listing Use ctrl-F (PC) or command-F (Mac) to search for a name

www.artcyclopedia.com/nationalities/Dutch.html · Cached · Save

[Dutch Painters - Artists from the Netherlands](#)

Dutch painters and Dutch artists from the Netherlands at the Linkism art directory for artists.

www.linkism.com/visual_artists/painting/netherlands-artists.htm · Save

[Dutch Museum Shop](#)

Kies uw taal: Nederlands Choose your language: English PlexiPhoto. Dutch Museum Amsterdam...

www.dutchmuseumshop.com/ · Cached · Save

[Bookshop of Dutch Painters of the Bookshop of Dutch Painters II](#)

Bookshop of **Dutch painters** of the Golden Age part II...

www.essentialvermeer.com/books/books_dutch_art_II.htm · Save

[Bookshop of Dutch Painters of the Bookshop of Dutch Painters](#)

Bookshop of **Dutch painters** of the Golden Age...

www.essentialvermeer.com/books/books_dutch_art.htm · Cached · Save

[More Results from www.essentialvermeer.com](#)

[Narrow Your Search](#)

[Famous Dutch Painters](#)

[17th Century Dutch Artists](#)

[Dutch Oil Painters](#)

[List of Dutch Artists](#)

[Dutch Painters 1600](#)

[17th Century Dutch Paintings](#)

[Dutch Landscape Painter](#)

[Dutch Painter Van](#)

[More »](#)

[Expand Your Search](#)

[Flemish Painters](#)

[Mark Twain's Jumping Frog](#)

[Vanitas Paintings](#)

[American Painters](#)

[Related Names](#)

[Vermeer](#)

[Jan Vermeer](#)

[Edith Cavell](#)

Een resultatenlijst van Ask.

oorspronkelijke vraagstelling vervangen door de tekst van de link. Dit gebeurt niet als phrase, dus aanpassing kan gewenst zijn.

Geheel onderaan de resultatenpagina vindt u nogmaals de zoekregel, waar een nieuwe zoekactie kan worden gestart of de bestaande gewijzigd.

4.2.5.6. Bespreking Ask

Tot voor kort was Ask in twee uitvoeringen beschikbaar: als de betrekkelijke nieuwkomer in de wereld van de zoekmachines Teoma en als oudgediende met een eigen karakter, Ask Jeeves; daarnaast betreft ook Lycos, een andere bekende, zijn resultaten bij Ask Jeeves /Teoma. Teoma, destijds een nog jonge zoekmachine met een geruchtmakende zoektechnologie werd in september 2001 door Ask Jeeves overgenomen. Sindsdien maakten beide zoekmachines gebruik van dezelfde database en waren de zoekresultaten en hun ranking eender. De verschillen waren dat Ask Jeeves zich presenteerde als een zoekmachine waar men zijn vragen in natuurlijke taal kon stellen (en die gebruikers vervolgens verder hielp), terwijl Teoma bij zijn resultaten ook 'Link collections by experts and enthusiasts' presenteerde, d.w.z. zogeheten 'hubs', waar de gebruiker sites over het onderwerp met veel outgoing links kon vinden. Nu zijn beide zoekmachines dus samengevoegd.



Source

[George Washington](#) | [Save](#)

George Washington is often called "the father of his country" for his crucial role in fighting for, creating and leading the United States of America in its earliest days. Washington was a surveyor, farmer and soldier who rose to command the colonial forces in the... [More](#) »

Search For: [Images](#)

Go To: [Official Site](#) | [Encyclopedia](#)

[Narrow Your Search](#)

[First President of United States](#)

[Who Was the First President of Mexico?](#)

[Who Was the 43rd President of US](#)

[Who Elects the President of the U.s.](#)

[US Presidents](#)

[All Names of United States Presidents](#)

[Who was the first president of the US?](#) [\[Web Answer\]](#)

George Washington was the first president of the US. Only unusual thing make news. The entire section of his Christian beliefs can be summed up by saying, Spencer has written in support of such...

en.wikipedia.org/wiki/Talk:Robert_Spence... · [Save](#) · [See 14 more Web Answers](#)

[John Hanson, First President](#)

John Hanson, American Patriot and **First President** of the United States (1715-1783)

www.marshallhall.org/hanson.html · [Save](#)

Ask als antwoordmachine.

Ask, de kleinste onder de groten, heeft vijf 'claims to fame':

- van Ask Jeeves heeft het geërfd dat de gebruiker vragen in 'natuurlijke taal' kan stellen. Voor de antwoorden put Ask uit een eigen, in de loop der jaren opgebouwde database van websites waar antwoorden op veel gestelde vragen te vinden zijn.
- van Teoma heeft het een ranking algoritme geërfd dat op het punt van de 'link popularity' beter zou zijn dan dat van Google
- de categorisatiemogelijkheden onder 'Narrow Your Search'
- de suggesties voor gerelateerde zoektermen onder 'Expand Your Search'
- de 'Related Names'

Het eerste kenmerk oogt fraai, en wie nog niet de truc had ontdekt dat je op het Web je vragen als antwoorden moet inkleden (The first American president was...”), zal bij dit type vragen aan Ask een betere zoekmachine hebben dan aan Google, Yahoo! en ook MSN Search (dat hiervoor een beroep doet op Encarta). Ook Ask's rankingalgoritme doet solide aan. Zoals bekend speelt in Google's rankingalgoritme 'link popularity' een zekere rol; simpel gezegd komt dit erop neer dat Google bij zijn ranking van zoekresultaten mede let op hoeveel incoming (of 'backward') links de gevonden pagina's trekken, en hoe groot de eigen 'link popularity' is van de pagina's waar die links vandaan komen. Bij Pagerank 'stemmen' dus alle andere webpagina's over de populariteit, en dus de ranking, van een treffer mee. De verbetering die Ask op dit concept aanbrengt is dat het niet alle verwijzende pagina's over de populariteit van een treffer laat mee 'stemmen', maar alleen die welke zelf ook tot het zoekresultaat behoren. (Ask noemt dit 'Subject Specific Popularity' (SSP).) Dat maakt het gezelschap natuurlijk selecter en dus de kwaliteit van het oordeel hoger. De vraag is alleen of het gezelschap niet te select is geworden: alleen die sites stemmen immers mee die aan uw zoektermen voldoen, en natuurlijk zijn er daarnaast nog heel wat die dat niet doen maar wel relevant voor het onderwerp zijn.

In de praktijk is moeilijk vast te stellen of Ask's opvatting van 'link popularity' beter werkt dan die van Google; 'link popularity' is immers maar een van de vele factoren van het totale rankingalgoritme bij beide zoekmachines. Verschillende vragen tonen bij de eerste honderd treffers een uiteenlopende maar constant geringe overlap (<25%) met de

resultaten van zowel Google als Yahoo!, maar dat ligt alleen al door de uiteenlopende grootte van de drie databases voor de hand. Experts raden het gebruik van Ask als derde zoekmachine naast Google en Yahoo! aan, en gezien juist de verschillen tussen de drie lijkt dat een verstandige raad.

Voor wie de resultaten van diverse zoekmachines met elkaar wil vergelijken, zijn er mogelijkheden te over. Ik noem er een paar:

- metazoekmachines. In principe zijn deze voor dit doel geschikt, mits ze aan twee voorwaarden voldoen: 1. Resultaten van de belangrijkste zoekmachines, met name Google, worden opgenomen; en 2. Bij elke treffer wordt per zoekmachine gemeld welke plaats die treffer in de resultatenlijst van de betreffende zoekmachine inneemt. De enige metazoekmachine waarvan ik eind maart 2006 heb kunnen vinden dat hij aan deze twee voorwaarden voldeed, is Jux2 (<http://www.jux2.com/>), dat de resultaten van Google, Yahoo! en MSN Search verzamelt.
- springplankzoekmachines. Deze laten u vanuit één zoekregel de vraag aan diverse zoekmachines stellen. Voordelen: je hoeft de vraag maar één keer te formuleren en kunt hem aanpassen aan de gebruikte zoekmachine. Nadeel: je krijgt de resultaten na, niet naast elkaar te zien. Een voorbeeld is TurboScout (<http://www.turboscout.com>).
- Resultaten van Google en Yahoo! naast elkaar bieden Google – Yahoo Comparison: Compare the Search Results (<http://www.googleguy.de/google-yahoo/>) en Twingine (<http://twingine.com>); hetzelfde kan met Google en Ask (Teoma): <http://www.googleguy.de/google-teoma/>. Het meest compleet hier zijn Compare Search Engines (<http://ranks.nl/tools/compare.html>), dat onder meer de nationale versies van Google en MSN Search, en Nederlandse zoekmachines kan vergelijken, en Stereosearch (<http://www.radio.xampled.com/projects/hizmo/stereosearch.asp>), dat ook de resultaten van blogzoekmachines (zie blz. 278 e.v.) naast elkaar zet.
- Heel informatief zijn ook de vergelijkende schema's van de (maximaal) eerste honderd treffers van verschillende zoekmachines die Langreiter (<http://www.langreiter.com/exec/yahoo-vs-google.html>) en Thumbshots (<http://ranking.thumbshots.com>) tonen. Langreiter doet dat voor Google en Yahoo!, Thumbshots voor – per tweetal – AlltheWeb, AltaVista, Google, MSN (de huidige versie), Ask (Teoma), WiseNut en Yahoo!. In beide gevallen omvat het schema twee evenwijdige horizontale assen met ieder maximaal honderd rondjes die de maximaal eerste honderd treffers voor een vraag in elk van beide zoekmachines voorstellen. Deze rondjes zijn door lijnen met elkaar verbonden waar treffers in de ene en in de andere zoekmachine met elkaar overeenkomen. Klikken op een rondje leidt u naar het resultaat. Voor een voorbeeld, zie blz. 197.

Een bezwaar van Ask, juist als complementaire zoekmachine, vind ik dan wel weer de eigenaardigheden die Ask bij het gebruik van prefixen, phrases en OF- en NIET-relaties vertoont. Diverse prefixen kunnen niet worden herhaald of met elkaar gecombineerd, met phrases wordt gesjoemeld als ze geen resultaat opleveren en OR en - (voor NOT) werken anders dan we van Google, Yahoo! en MSN Search gewend zijn. Voor de geavanceerde zoeker geen aanbeveling om Ask te gebruiken!

Ask's 'Narrow Your Search' en 'Expand Your Search'-functies laten – al naar gelang de zoekvragen – uiteenlopende indrukken achter; maar dat geldt voor alle zoekmachines en metzoekers die met categorisatie en suggesties voor nieuwe zoektermen werken. De zoekterm "[electronic publishing](#)" biedt een mooi voorbeeld van Ask's capaciteiten op deze terreinen. Voordelen, definitie, geschiedenis, invloed, metadata en rechten zijn maar enkele van de meer toegespitste onderwerpen die onder het hoofdje 'Narrow Your Search' bij dit verschijnsel worden genoemd, en onder 'Expand Your

Search' worden Word Processing en How Does a Manuscript Become a Book? gesuggereerd.

'Narrow Your Search' en 'Expand Your Search' leveren de beste resultaten bij een algemene vraagstelling, bij voorkeur met één zoekterm. Dus liever `pianists` dan `women pianists`, en liever `identity` dan `identity theft`.

4.2.6. *Andere zoekmachines*

Google, Yahoo!, MSN Search en Ask zijn de grote vier die u in ieder document over de grote mondiale zoekmachines genoemd vindt. Ze vormen in die zin een klasse apart dat de omvang van hun linkdatabases groter is dan die van de zoekmachines van het tweede echelon. Om die reden kan de verleiding groot zijn om, als een vraag zich ertoe leent met een grote mondiale zoekmachine te worden opgelost, die 'tweederangs' zoekmachines maar links te laten liggen. Toch zou men zich daarmee ernstig te kort doen. Onder deze categorie zoekmachines zijn er minstens vijf die een database van meer dan een miljard sites hebben: AlltheWeb, AltaVista, Gigablast, A9 en Exalead. Daarvan zijn de laatste twee het interessantst.

4.2.6.1. A9 (<http://www.a9.com>)

A9 is een zoekmachine van de online boekhandel Amazon. Voor zijn webresultaten maakt deze zoekmachine gebruik van een beperkte versie van de database van Google, net zoals bijvoorbeeld AOL en Netscape dat doen; daarnaast gebruikt het ook resultaten van Alexa, een database die wordt opgebouwd doordat de bewegingen van geregistreerde gebruikers op het Web worden gevolgd. Tot zover weinig nieuws dus. Maar A9 heeft meer te bieden:

1. In de eerste plaats toont A9 standaard de resultaten uit zijn webdatabase en die in zijn imagesdatabase in twee kolommen naast elkaar, wat in veel gevallen voordelen biedt boven de resultaten uit elk van die databases apart. Stel u maar een zoekactie met de zoekterm 'Borobudur' voor.
2. In een aparte kolom bovenin het zoek- en het resultatenscherf toont A9 een aantal selectievakjes waaronder het toegang geeft tot nog andere databases. De interessantste daarvan is 'Books'. Als u daarop klikt nadat u een vraag in de web- en de imagesdatabase gesteld hebt, toont A9 in een derde (in feite de tweede) kolom ook de zoekresultaten uit meer dan 100.000 gedigitaliseerde boeken waarover het dank zij overeenkomsten met een groot aantal uitgevers kan beschikken. Ongeregistreerde gebruikers krijgen alleen de titels van boeken te zien die aan de zoektermen voldoen of waarin deze zoektermen voorkomen; in dat laatste geval wordt ook een paginanummer gegeven en een zoekterm met een klein stukje context. Geregistreerde gebruikers daarentegen krijgen toegang tot de pagina's zelf, en kunnen ook andere pagina's van een boek opzoeken waar de opgegeven zoektermen, of andere, voorkomen. Aangezien het heel goed kan voorkomen dat een zoekactie op het Web onbetrouwbare of onbevredigende resultaten oplevert, is dit een belangrijke uitbreiding van het arsenaal van de zoeker.
3. Drie andere databases die tot het standaard menu van de A9 zoekmachine behoren, zijn Movies, Reference en Yellow Pages. Klikken op elk van die menukeuzes levert een nieuwe kolom met mogelijke resultaten voor de opgegeven zoekvraag op, zodat er meer dan tien kolommen met zoekresultaten naast elkaar kunnen komen te staan; de zevende kolom bevat de menukeuzes. De databases Movies, Reference en Yellow Pages bevatten resultaten van respectievelijk de Internet Movie Database (helaas niet de IMDbPro Professional Details, waartoe alleen abonnees op die database toegang hebben), Answers.com (zie bij Google) en een eigen Gouden Gids.
4. Naast deze standaardverzameling van databases heeft A9 nog meer dan 200 andere databases te bieden die in één zoekactie kunnen worden doorzocht. Daartoe behoren

onder meer zoekmachines voor blogs en RSS feeds (zie blz. 284), de PubMed database van de National Library of Medicine (zie bij Google Scholar) en de website van de British Library. Uit deze lijst kunnen extra databases en zoekmachines worden geselecteerd en aan de persoonlijke lijst toegevoegd. Wie niet geregistreerd en ingelogd is kan steeds één zoekmachine aan het standaardlijstje toevoegen; wie dat wel is, kan dat voor toekomstige zoekacties met meer zoekmachines en databases doen. Dit wordt vergemakkelijkt doordat de verschillende 'columns' ('verticale' of speciale zoekmachines), zoals ze heten, van trefwoorden zijn voorzien waarop je kunt zoeken en doorklikken. Dat is ook wel nodig want de tweehonderdzoveel databases die tot dusver bij A9 als 'columns' zijn aangemeld vormen een buitengewoon divers allegaartje. Sommige hebben zelfs een recordstructuur die zich absoluut niet leent voor de zoektaal van Google.

- Overigens ziet de databaselijst er wel netjes uit. Bij elke database zijn Try it!, Add, en More info-knoppen of links beschikbaar, en geselecteerde databases/zoekmachines worden gemarkeerd en krijgen in plaats van de Add- en Remove-knop.

The screenshot shows the Amazon A9 search interface. At the top, the search bar contains "Victor Acier" Meissen with a "GO" button. Navigation links for Home, Prefs, Toolbar, and Sign Out are visible. A personalized greeting says "Hello M.R.J. Hofstede. [Click here](#) if this is not you." Below the search bar, there are "Hide Column Choices" and a grid of checkboxes for various content types: Web, Books, Images, Movies, Reference, Yellow Pages, Your History, Your Bookmarks, Your Diary, Furl, jobs.feedster, Top Blogs, blogs.feedster, Feedster, Movalog, Blogdigger, Blogwise, SEO Blog, and IT-weblog. A "More Choices..." dropdown is also present.

The results are displayed in three columns:

- Web Results** (Showing 1 - 10 of about 23):
 - International Auctioneers**: 56, aus der Serie Gärtnerkinder von Michael **Victor Acier** 1778 ... Höhe 18,5 cm, Hutrand, wenige Blüten und Blätter minimal restauriert **Meißen**, ... http://www.internationalauctioneers.com/int/auction_lots.asp?AucID=6252&first=251 - 24k Cached [Site Info](#)
 - International Auctioneers**: Michael **Victor Acier** um 1775, aus der Serie kleine Amoretten, Malernr. 27 (Ru) ... Höhe 11,5 cm **Meißen**, unterglasurblaue Schwertermarke 2. Hälfte 20. ... http://www.internationalauctioneers.com/int/auction_lots.asp?AucID=6252&first=276 - 25k Cached [Site Info](#)
 - Suchergebnis**: **MEISSEN** um 1860, 1.Wahl Entwurf von Michel **Victor Acier** 1775, ... **MEISSEN** 1.Hälfte 19.Jahrhundert, 1.Wahl Modell von Michel **Victor Acier** 1772, Modell-No. ... http://www.auction-bergmann.de/ger/browse.asp?nachverkauf=False&T1_ID=1&T2_ID=104 - 28k Cached [Site Info](#)
- Book Results** (Showing 1 - 1 of about 1):
 - The Arcanum : The Extraordinary True Story** by Janet Gleeson (15 January, 2000) [page 240](#) : "... was forced to take work on a large estate near **Meissen** in order to provide for her sons until they were ... " See more [references to "Victor Acier" Meissen](#) in this book.
- Image Results** (Showing 1 - 1 of about 1):
 - Four small thumbnail images of porcelain figurines, including a seated figure and a group of figures.

Gedeelte van een resultatenpagina van de zoekmachine A9 van Amazon. Bovenin een kolom met aanklikbare databases, daaronder in drie kolommen kolommen naast elkaar achtereenvolgens treffers van het Web (Google), boeken en afbeeldingen.

6. Registratie brengt nog meer voordelen met zich mee. U krijgt een gepersonaliseerde omgeving tot uw beschikking waarin ook uw zoekvragen bewaard blijven. Zo kunt u steeds terugkeren naar de juiste formulering van vragen die u in het verleden aan A9 gesteld hebt. Deze zoekvragen verschijnen op het zoekscherm als u A9 gebruikt, maar zitten ook onder een knop in het menu die Your History heet; daarnaast zijn er nog drie andere gepersonaliseerde knoppen, Your bookmarks, Your Diary en Furl. De eerste daarvan is een online bookmark manager, die wordt gecompleteerd door Your Diary, de mogelijkheid om annotaties die u bij websites heeft gemaakt, te doorzoeken. Terwijl u uw History opbouwt vanuit het zoekscherm van A9, worden de online bookmarks en de history-teksten aangemaakt als u een toolbar van A9 gebruikt (zie het hoofdstuk Browser buttons en toolbars, blz. 264 e.v.). Furl ten slotte is een archief van webpagina's die door gebruikers van Furl op de servers van die dienst worden opgeslagen en – ten dele – publiek toegankelijk worden gemaakt (vgl. MyWeb van Yahoo!, zie blz. 177). Het is een soort kruising tussen een online bookmark manager – met als verschil dat in plaats van bookmarks hele pagina's worden opgeslagen – en het Internet Archive – met als verschil dat alleen die pagina's worden opgeslagen die daadwerkelijk door gebruikers bezocht worden. Furl is verder een voorbeeld van een 'collaborative service', in zoverre als gebruikers van de dienst eerst pagina's verzamelen en de adressen vervolgens aan anderen beschikbaar stellen – maar dan verrijkt met de informatie dat ze tot de verzameling van een gebruiker behoren. Hier speelt hetzelfde mechanisme als bij de informatie die Amazon bij boeken geeft: 'Customers who bought this book also bought...'

Zoals u ziet biedt A9 u een mix van diensten, die ten dele te maken hebben met zoeken, en ten dele met informatiebeheer, zoals we die ook al bij Yahoo! zijn tegengekomen. De informatie die u in de 100.000+ boeken kunt vinden vormt zonder twijfel een verrijking van uw zoekmogelijkheden op het zichtbare en onzichtbare Web. Dat kan een reden zijn om u bij A9 te registreren. Dat betekent echter niet automatisch dat het aanbod van de overige diensten van A9 – de op een subset van Google gebaseerde webdatabase, de opgeslagen zoekvragen en de overige informatiebeheermogelijkheden – voor u de best mogelijke optie zijn.

Mocht u zich bij A9 registreren, verzuim dan niet van de mogelijkheden gebruik te maken die onder [Prefs](#) te vinden zijn.

Zie ook: Greg R. Notess, Searching Books Between the Covers, Online 29, no. 3 (May/June 2005) (<http://www.infotoday.com/online/may05/OnTheNet.shtml>); A9 > Company > What's New? (<http://a9.com/-/company/whatsNew.jsp>)

4.2.6.2. Exalead (<http://www.exalead.com>)

Een betrekkelijk nieuwe verschijning in de zoekmachinewereld is Exalead, een van origine Franse zoekmachine die als intranet-zoekmachine is begonnen en onder meer de websites van Danone, Cap Gemini Ernst & Young, enkele grote banken en verzekeringsmaatschappijen en het Franse Ministerie van Onderzoek en Nieuwe Technologie bedient. Misschien wel de belangrijkste klant is echter de Franse tak van America Online (AOL), want het is die klandizie die ertoe heeft geleid dat Exalead als een soort uithangbord voor zijn technologie ook een webzoekmachine onderhoudt die met meer dan vier miljard pagina's in zijn database inmiddels tot de grotere spelers behoort.

Dat op zich zou misschien niet voldoende reden zijn er hier uitgebreider aandacht aan te besteden, maar Exalead beschikt over een aantal bijzondere eigenschappen en mogelijkheden die het tot een unieke en voor sommige vragen onvermijdelijke zoekmachine maken. Deze eigenschappen en mogelijkheden betreffen de zoek- en andere faciliteiten die vanaf de resultatenpagina worden geboden.

Ik zal hier geen complete bespreking aan Exalead wijden, die komt bij de volgende update van dit boek, maar de belangrijkste eigenschappen wil ik hier wel alvast noemen.

Om te beginnen kunnen er een aantal overeenkomsten met de andere grote zoekmachines in heel kort bestek worden genoemd.

- de standaard (verzwegen) operator tussen meer zoektermen is AND
- Exalead ondersteunt Booleaans OR en NOT, en als alternatief voor het laatste het minteken onmiddellijk voor de uit te sluiten zoekterm
- net als Google past Exalead bij twee of meer zoektermen automatisch stemming toe. Een verschil met Google is echter dat dat niet standaard gebeurt, maar alleen als men het in de Advanced Search of in de Preferences zo heeft ingesteld
- eveneens net als bij Google wordt het +-teken gebruikt om in een vraag met twee of meer zoektermen automatic stemming uit te schakelen. Om precies te zijn: als u automatic stemming heeft ingeschakeld voor een vraag met twee of meer zoektermen, geldt die automatic stemming voor alle zoektermen; door er een of meer te plussen maakt u de automatic stemming voor die zoektermen ongedaan
- Exalead ondersteunt phrases ("...")
- als text en document limiters ondersteunt Exalead `title:`, `site:`, `date:` (`date>=jjjjmddd` en `date<=jjjjmddd`, beide kunnen in één vraag worden gebruikt), en `language:`

<i>Termen</i>	<i>Google</i>	<i>Exalead</i>
woman book child illustration	woman, book, child, illustration, books, children, illustrations, illustrator, illustrators (maar niet women)	woman, book, child. illustration, illustrations (maar niet women)
children book	children, book, child, books	children book
children corner	children, corner	children, corner
crown jewel	crown, jewel	crown, jewel, crowne, crowned
internal affair	internal, affair, affairs	internal, affair, affairs
multinational affair	multinational, affair	multinational, affair, affairs, affaire, affaires, multinationale, multinationales multinationals
international affair	international, affair, affairs	international, affair, affairs, affaire, affaires, internationale, internationales
corporate affair	corporate, affair	corporate, affair, affairs, affaire, affaires (maar niet corporatif, corporation)
market player	market, player, markets, players	market, player, markets, players, playing. played (maar niet play, marketing)

<i>Termen</i>	<i>Google</i>	<i>Exalead</i>
market party	market, party	market, party, markets, parti (onder meer als string in Parties als benaming voor een ODP-categorie) (maar niet parties)
market players	market, players	market (ook als string in Marketing als benaming van een ODP categorie), players, markets, player, playing (maar niet marketing)
movie star	movie, star	movie, star, movies, moving (maar niet stars)
movies stars	movies, stars, star	movies, stars, movie, moving, start, starts (maar niet star)
bungee jump	bungee, jump, jumping (maar niet jumps, bungy)	bungee, jump, bungy (maar niet jumps, jumping)

(Automatic) stemming, de (automatische) toevoeging van variante woordvormen aan een zoekvraag, verhoogt de recall (zie blz. 75) van een zoekactie. Probleem is echter dat zoekmachines die daarmee werken, daar heel uiteenlopende invullingen aan kunnen geven. In de tabel hierboven staat een aantal zoektermen met daarbij voor Google en Exalead de woordvormen die bij de zoekactie op die zoektermen zijn meegenomen.

Interessanter dan de overeenkomsten zijn de bijzonderheden van Exalead. Achtereenvolgens zijn dat:

A) Zoekmogelijkheden

- als aanvulling op zijn (functioneel beperkte) spellingchecker heeft Exalead in zijn Advanced Search – naast Exact search en Automatic stemming – de mogelijkheid van een Phonetic search (zie blz. 59) en een Approximate spelling (zie blz. 59; = 'fuzzy' zoeken) search. Het praktisch nut van beide lijkt mij dubieus.
 - Exalead biedt de mogelijkheid met (rechts) getrunceerde zoektermen te zoeken: `catalog*` vindt `catalog`, `catalogs`, `catalogi`, `cataloged`, maar merkwaardig genoeg niet `catalogus` en `catalogue`.
 - daarnaast ondersteunt Exalead zoeken met 'regular expressions'. Deze bieden vooral mogelijkheden om te maskeren en om ranges aan te geven. Vgl. Google: `numrange:`, maar hier kan ook met alfanumerieke tekens worden gewerkt: regular expressions zijn patronen die reeksen numerieke of alfanumerieke tekens voorstellen. Ze staan altijd tussen slashes (/.../).en maken gebruik van de volgende tekens:
 - een punt (.) voor één teken
 - | voor OF
 - [en] om tekengroepen te omsluiten
 - - voor een range in een tekengroep (bijvoorbeeld [a-z] voor een willekeurige letter uit het alfabet)
 - ? om een tekengroep optioneel te maken
 - * om een teken of een tekengroep te herhalen
- Voorbeelden van 'regular expressions' zijn `/4[0-9]/` voor 40, 41...49, `/4[0-2]*/`

voor 4, 40, 41, 42, 400, 401, 402, 410, 411, 412, 420, 421, 422, 4000, 4001 enz.); en `/int[a-z]*tie/` voor interventie, integratie, introductie, intentie, intelligentie, internetpresentatie

Meer over regular expressions in <http://www.regular-expressions.info/tutorial.html>

- Exalead sorteert standaard op relevantie; in plaats daarvan kan al in de zoekopdracht een andere sorteervolgorde worden opgegeven, op Last Modified-datum: `sort:new` sorteert de treffers van nieuw naar oud, `sort:old` van oud naar nieuw
- naast MSN Search (`prefer:`) is Exalead de enige grote zoekmachine die het mogelijk maakt naar 'partial matches' te zoeken. Met de operator OPT is het mogelijk aan te geven dat de erop volgende zoekterm in het zoekresultaat mag maar niet hoeft voor te komen. (Vroeger had elke grote zoekmachine in zijn Advanced search een optie 'Should contain' of 'Should include', nu heeft alleen Ask die nog, maar legt daarmee alleen maar een OF-relatie.)
- naast de (impliciete) EN-operator heeft Exalead de meer specifieke operator NEAR, die een EN-relatie legt tussen zoektermen die maximaal 16 woorden uit elkaar mogen staan
- in plaats van met phrases kan ook met de NEXT-operator worden gewerkt

B) Resultatenpagina's

Zoals al eerder opgemerkt bezit Exalead een spellingchecker met een beperkte functionaliteit. Deze herkent 'catagory' bijvoorbeeld niet als een foutieve spelling van 'category'. Als de spellingchecker een fout wel herkent, verschijnt in de linker kolom van het resultatscherm een suggestie voor een juiste spelling.

Voor het overige zijn Exaleads resultatschermen de fraaiste en meest functionele van alle grote zoekmachines.

- Links op het scherm verschijnt een kolom met vier lijstjes, van respectievelijk Related terms, Related categories, Web site locations en Document types. De eerste twee hiervan bieden, op twee niveaus, suggesties voor verder zoeken.
- 'Related terms' kan suggesties voor een verdere verfijning van de zoekactie bieden (narrower terms), maar ook suggesties voor een verbreding van de zoekactie (broader en related terms). Zoeken met de term 'oneiromancy' (voorspellen op basis van dromen) bijvoorbeeld levert als suggesties onder meer 'crystal gazing' en 'tea leaf' (related terms), maar ook 'forms of divination' en 'means of divination' (broader terms) op.
- Onder 'Related categories' worden de rubrieken van het Open Directory Project opgesomd die zelf de zoekterm in hun benaming bevatten, ofwel sites in de naam of beschrijving waarvan de zoekterm voorkomt. Hoewel directories altijd een goed middel vormen om meer van hetzelfde te vinden, is de vormgeving hiervan bij Exalead helaas gebrekkig.
- 'Web site locations' en 'Document types' bevatten filters die achteraf op het zoekresultaat kunnen worden toegepast. Het eerste serveert sites uit het ODP; hierop is hetzelfde bezwaar van toepassing als hierboven genoemd. 'Document types' biedt de mogelijkheid de zoekactie tot een van de gevonden documenttypen te beperken.
- Nog meer filters verschijnen in een balk bovenin het scherm. Deze bieden toegang tot de audio- en videobestanden in uw zoekresultaat, en tot pagina's waarop een RSS-feed (zie blz. 284) beschikbaar is.
- Rechts van de zoekresultaten laat Exalead thumbnails van de gevonden pagina's zien. In de Preferences is die plaats te veranderen. Technisch heel fraai is ten slotte ook dat Exalead in de trefferlijst zelf, dus in het centrale deel van het venster een preview van de treffers verschaft. Daarvoor hoeft u alleen maar op een willekeurige plek in de treffer of het bijbehorende plaatje te klikken.

Een resultatenscherf van Exalead.

Exalead is niet een zoekmachine die u als eerste moet gebruiken. Daarvoor is hij (nog) te klein en te gammel. Maar voor sommige vragen, die voor andere zoekmachines een probleem zijn, biedt hij misschien een oplossing.

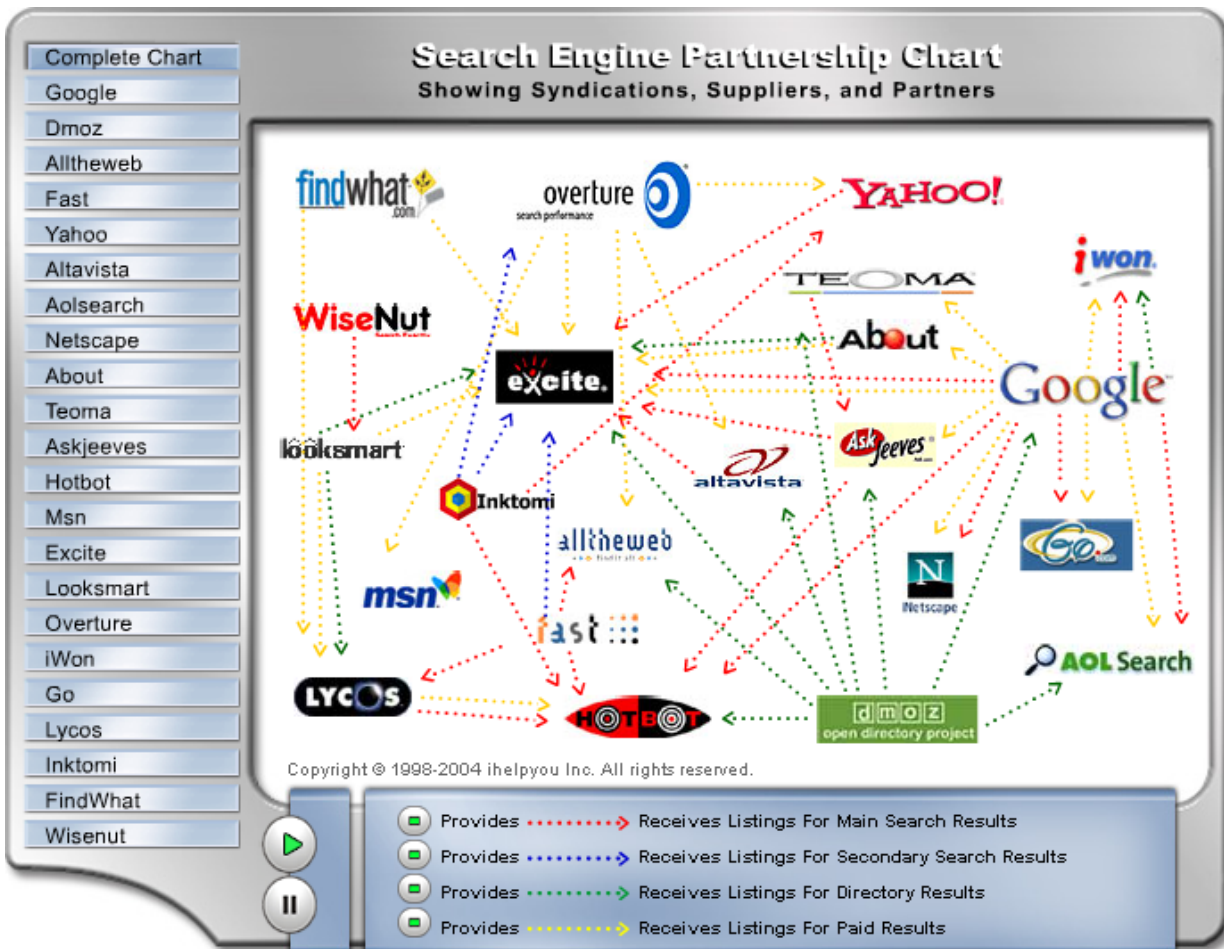
4.3. Metazoekmachines

In hoofdstuk 2.5.1. is al even ter sprake gekomen dat metazoekmachines aanzienlijke bezwaren hebben. De belangrijkste daarvan zijn dat ze beperkte zoekfunctionaliteit bieden, zelden of nooit toegang tot de complete Google-resultaten hebben, de treffers van normale zoekmachines vermengen met die van commerciële (die goede plaatsen in hun trefferlijsten verkopen) en nogal eens achterlopen bij actuele ontwikkelingen in de zoekmachinewereld. Daarbij komt dan vaak nog dat hun ranking gebaseerd is op dubbeltellingen van treffers van zoekmachines die – met dezelfde ranking – uit dezelfde database putten, en dat ze onstabiel zijn: het ene moment gebruiken ze de ene mix van zoekmachines, het volgende aan ander.

Voor mij persoonlijk zijn dit voldoende redenen om geen gebruik van metazoekmachines te maken tenzij dat onvermijdelijk is – bijvoorbeeld om hun resultaten te vergelijken. Ik sluit mij dan ook aan bij hetgeen hierover gezegd wordt inde gids van de University of California in Berkeley, <http://lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch.html>. Wie positiever over dit soort zoekinstrumenten denkt, kan ik op dit moment (maart 2006) alleen Jux2

1. [HISTORY OF MEXICO - CHAMELEON ADVENTURER: THE ASTONISHING CAREER OF ...](#) [new window] [frame] [preview] [clusters]
 History of Mexico - CHAMELEON ADVENTURER: THE ASTONISHING CAREER OF AGUSTIN DE **ITURBIDE**. By Jim Tuck, Author in Mexico Connect - Mexico's monthly ezine / magazine devoted to informing about and ...
Cache: <http://cc.msncache.com/cache.aspx?q=1860243706868:=en-US&FORM=CVRE4>
www.mexconnect.com/.../history/jtuck/jtaugustiniturbide.html - Wisenet 2, Lycos 3, Ask Jeeves 3, MSN Search 4, Looksmart 5, MSN 9
2. [Images of the Spirit: The Evocative Vision of Graciela Iturbide](#) [new window] [frame] [preview] [clusters]
 ... Arts & History: Graciela **Iturbide** Images of the Spirit: The Evocative Vision of Graciela **Iturbide** The Mexican photographer blends history, lyricism and portraiture to record cultures in transition. The ...
Cache: <http://cc.msncache.com/cache.aspx?q=1947582058192:=en-US&FORM=CVRE>
www.smithsonianmag.si.edu/.../issues98/jun98/iturbide.html - MSN Search 1, MSN 2, Looksmart 3, Lycos 6, Ask Jeeves 6
3. [Handbook of Texas Online - Agustín de Iturbide](#) [new window] [frame] [preview] [clusters]
 Encyclopedia entry furnishes a biographical sketch of the Mexican military figure who briefly ruled as Emperor of Mexico from May 1822 to March 1823.
www.tsha.utexas.edu/.../online/articles/view/ll/fit1.html - Lycos 4, Ask Jeeves 4, MSN 5, Wisenet 5, Looksmart 6

De eerste drie treffers op de zoekterm 'iturbide' bij de metazoekmachine Vivísimo, in 2005. Al bij deze treffers valt op dat Google en Yahoo! kennelijk niet tot de gebruikte zoekmachines behoren en dat Lycos en Ask Jeeves steeds dezelfde treffer op dezelfde plaats hebben. Dat wekt weinig vertrouwen in het rankingmechanisme van Vivísimo.



Schema van de relaties tussen een aantal bekende zoekmachines.

(<http://www.jux2.com>) en P.R.A.S.E. (<http://www.prase.us/>; P.R.A.S.E. staat voor Page Rank Assisted Search Engine) aanbevelen. Beide werken met alleen Google, Yahoo! en MSN Search; P.R.A.S.E. laat u Page Rank als filter op de resultaten van die drie zoekmachines toepassen.

4.4. Speciale zoekmachines

Zoekmachines als Google en AltaVista hebben iets van een kanon waarmee je op een mug schiet. Natuurlijk, ze zijn gemaakt om 'het Web' full-text te indexeren en doorzoekbaar te maken, en uw vraag is er een van de miljoenen soortgelijke die ze elke dag afhandelen. Maar het heeft toch iets raars dat u voor een vraag die zich misschien maar op een heel beperkt terrein afspeelt – u wilt iets weten over een Nederlands museum of over een medisch ongerief – een zoekmachine aan het werk zet die naast beide onderwerpen nog talloze andere bestrijkt.

Ze zijn minder bekend dan de algemene, mondiale zoekmachines, en ze kunnen daar niet mee concurreren wat de omvang van hun linkdatabase betreft, en vaak ook niet wat de verfijning van hun zoekfaciliteiten betreft, maar ze zijn er wel degelijk, de speciale zoekmachines, waarmee u in een veel beperktere omgeving zoekt. En afgezien van het feit dat met een kanon op een mug schieten altijd iets gênants heeft, bieden ze in sommige gevallen voordelen boven de hierboven genoemde krachtpaters:

- doordat je in een beperktere omgeving zoekt, is de kans op dubbelzinnigheid kleiner. Zoektermen krijgen een meer gerichte betekenis, zowel in die gevallen waarin termen in verschillende talen uiteenlopende betekenissen kunnen hebben en je dus beter binnen je eigen taalgebied kunt zoeken dan mondiaal, als in die waarin termen in een bepaald vakgebied een heel andere betekenis hebben dan daarbuiten.
- het kan zijn dat speciale zoekmachines delen van het Web ontsluiten waar de grote, mondiale zoekmachines niet komen. Dit is het 'invisible' of 'deep' Web dat op blz. 233 e.v. ter sprake zal komen.

Speciale zoekmachines zijn in twee hoofdgroepen te onderscheiden:

- die welke een beperkt geografisch gebied bestrijken, bij voorbeeld zoekmachines die alleen het Nederlandse domein of dat van de Benelux beslaan
- de welke een bepaald vakgebied of onderwerp bestrijken, bij voorbeeld de internationale betrekkingen, de wijsbegeerte, de geneeskunde of heel beperkte onderwerpen als beroemdheden en modelbouw.

Daarnaast bestaan ook nog speciale zoekmachines met gespecialiseerde functies, zoals de zoekmachines die prijzen vergelijken, en zoekmachines voor speciale materiaaltypen zoals beeld en geluid.

4.4.1. Nationale zoekmachines: Nederland

Alvorens een nationale zoekmachine te gaan gebruiken moeten we ons afvragen of dat nodig en nuttig is. Dat klinkt misschien vreemd na het pleidooi hierboven om de zoekruimte beperken, maar nationale zoekmachines zijn niet de enige waarmee dat kan. Ook de grote mondiale zoekmachines hebben immers zo hun mogelijkheden om – indien nodig – met een taal en/of binnen een domein te zoeken, en daarnaast is er alle reden je af te vragen of het enorme potentieel dat de grote nationale directories en startpagina's (zie boven bij Startnederland.nl) bieden voor de meeste vragen niet geschikter is dan een nationale zoekmachine. (In heel veel gevallen is dat zo!)

Nederlandse zoeksites

Overall ter wereld vind je portals die speciaal aan zoeken op het Web zijn gewijd. Nederland vormt daarop geen uitzondering. Twee Nederlandse sites waar u veel over zoeken kunt leren en op de hoogte kunt blijven van de ontwikkelingen bij zoekinstrumenten zijn:

- Zoeken makkelijk gemaakt: Zoekprof leert zoeken op internet (<http://www.zoekprof.nl>); en
- Voelspriet - zoekmachine, zoeknieuws, zoekhulp en zoekforum (<http://www.voelspriet.nl>)
- StartNederland.NL (<http://www.startnederland.nl/nl/z/zoek/index.html>) geeft onder zoeken een opsomming van een groot aantal startpagina's op zoekgebied.

Twee Nederlandstalige zoekhandleidingen zijn:

- Zoeken op het World Wide Web van drie medewerkers van de bibliotheek van de Universiteit van Tilburg (<http://www.uvt.nl/diensten/bibliotheek/instructie/www/onlinecursus/>).
- Cursus Zoeken op Internet van de School voor Informatie en Communicatie, Hanzehogeschool Groningen (http://www.schic.nl/studenten/cursus_zoeken/).
- Zie verder het overzicht van cursussen Leren zoeken op het web van Leren.nl: http://www.leren.nl/rubriek/computers_en_internet/internetten/zoeken_op_het_web/. Let wel op de datum van de laatste update.

Een grote mondiale zoekmachine gebruiken om binnen het Nederlandse domein en of naar documenten in de Nederlandse taal te zoeken ligt te meer voor de hand omdat de Nederlandse zoekmachines niet, zoals je zou verwachten, een groter deel van het Nederlandse Web dekken dan de grote mondiale. Het tegendeel is het geval: op de meeste simpele zoektermen die een controlebaar resultaat opleveren, scoren de grote mondiale zoekmachines meer treffers dan enkele bekende Nederlandse zoekmachines. Om dat te illustreren volgen hier (zie volgende bladzijde) de resultaten van een vergelijkend testje (van medio september 2005) met een dozijn niet al te frequent voorkomende zoektermen voor drie grote, mondiale en twee Nederlandse zoekmachines. Voor de laatste heb ik Vindex en Track als voorbeeld genomen omdat andere zoekmachines ofwel met dezelfde database werken als een van deze twee (Vindex ↔ Ilse, Track ↔ Kobala) ofwel metzoekers zijn (Zoeken.nl, Vinden.nl). In het algemeen gesproken kan de conclusie niet anders zijn dan dat 'autochtone' zoekmachines als Vindex en Track het in elk geval wat de omvang van hun linkdatabase betreft moeten afleggen tegen de grote mondiale zoekmachines als Google en Yahoo!, en eveneens tegen de Nederlandse versies van AltaVista (dat met een wat andere ranking uit dezelfde database put als Yahoo!). Voor zoekacties die bij Ilse of Track een heel gering of geen resultaat opleveren heeft het altijd zin het ook nog eens met een grote mondiale zoekmachine te proberen, waarbij overigens de mogelijkheid om een metzoekmachine te gebruiken ook zeker niet uit het oog moet worden verloren. Zo levert Vinden.nl voor de vraag naar de term tuitelig één treffer op die Google niet heeft.

Omvang is natuurlijk niet alles. Functionaliteit en gebruikersinterface van de zoekmachine zijn ook belangrijk, evenals andere extra's. Zo vindt Ilse ook pagina's met een generiek domein (.com, .org enzovoort) in Nederland, iets wat Google niet kan. Een bijzonderheid van Track.nl is de mogelijkheid die deze zoekmachine biedt per provincie te zoeken.

Vergelijking van de resultaten bij de grootste mondiale en de grootste Nederlandse zoekmachines voor pagina's in het Nederlands*)

	<u>Google</u>	<u>Yahoo!</u>	<u>MSN Search (.nl)</u>	<u>Vindex</u>	<u>Track</u>
uitvreters	<u>856</u>	<u>>1000</u>	<u>250</u> (548)	<u>141</u>	<u>175</u>
binnenvetter	<u>894</u>	<u>850</u>	<u>250</u> (575)	<u>154</u>	<u>264</u>
zitgebeuren	<u>23</u>	<u>9</u>	<u>1</u> (1)	<u>3</u>	<u>2</u>
smulbos	<u>>1000</u>	<u>592</u>	<u>250</u> (689)	<u>95</u>	<u>238</u>
exobiologie	<u>791</u>	<u>387</u>	<u>250</u> (529)	<u>37</u>	<u>37</u>
welvoeglijkheid	<u>399</u>	<u>132</u>	<u>115</u> (115)	<u>28</u>	<u>56</u>
beekpunge	<u>588</u>	<u>620</u>	<u>250</u> (464)	<u>101</u>	<u>152</u>
hannessen	<u>846</u>	<u>978</u>	<u>250</u> (648)	<u>219</u>	<u>196</u>
tuitelig	<u>29</u>	<u>13</u>	<u>17</u> (17)	<u>5</u>	<u>2</u>
roodkeelpieper	<u>522</u>	<u>355</u>	<u>250</u> (423)	<u>76</u>	<u>66</u>
hotsknotsbegonia-voetbal	<u>92</u>	<u>111</u>	<u>8</u> (5)	<u>22</u>	<u>2</u>

*) Bij de drie internationale zoekmachines werd gezocht binnen het Nederlands domein (site:nl) en met als taalinstelling Nederlands, en met het 'Family Filter' uit; de resultaten van Vindex en Track omvatten ook andere domeinen dan .nl (.com, .org e.d., en zelfs incidenteel .be en .de). en zijn dus geflatteerd.

Geteld zijn niet de door de zoekmachines opgegeven, maar de werkelijk getoonde resultaten, 'herhaald met de weggelaten resultaten' (repeated with the omitted results included). Aangezien MSN Search niet meer dan 250 treffers toont is bij deze zoekmachine ook het aantal treffers vermeld dat deze zoekmachine zelf opgeeft. Vermoedelijk is dat te hoog, net zoals dat bij de meeste andere zoekmachines het geval is. Track toont slechts 100 sites, met geclusterde resultaten. Omdat het opgegeven aantal treffers bij deze zoekmachine hoogstwaarschijnlijk juist is, is dat aantal vermeld.

Zinvoller dan het gebruik van algemene nationale zoekmachines lijkt mij het gebruik van onderwerpsspecifieke nationale zoekmachines zoals NL-menu (<http://zoeken.bibliotheek.nl/>, onder Websites) voor organisaties en de SURFnet Search Engine (<http://search.surfnet.nl/selectSearchMode.do>) die is toegesneden op het Nederlandse hoger onderwijs en onderzoek.

4.4.2. Nationale en regionale zoekmachines van andere landen

Natuurlijk hebben ook andere landen dan Nederland hun nationale, regionale en zelfs plaatselijke zoekmachines. Of het nodig en nuttig is die te gebruiken, hangt van de plaatselijke situatie af. Op te sporen zijn ze in elk geval voor het merendeel wel, want er bestaan een aantal uitgebreide overzichten van. Hieronder volgt een lijstje:

- Country based search engines (<http://www.philb.com/countryse.htm>): ongeannoteerd overzicht van 2086 zoekmachines voor 216 landen, territoria en regio's, die regelmatig wordt geactualiseerd
- Search Engines Worldwide: The 1-Hit Search Engines Directory (<http://www.1-hit.com/search-engines/>): geannoteerde lijst van 1405 nationale zoekmachines
- Beknopt, maar ook selectief: Search Engines – Beaucoup! (Geographically Specific) (<http://www.beaucoup.com/>)
- Uitgebreid overzicht van Franse regionale zoekmachines: Abondance : annuaire des outils de recherche régionaux en France (<http://www.abondance.com/annuaire/index.html>)
- Aanbevolen: ~ Regional search engines ~ (<http://www.searchlores.org/regional.htm>)

4.4.3. *Onderwerpsspecifieke zoekmachines*

Wat algemene mondiale zoekmachines nog wel kunnen als het erom gaat een zoekactie tot een bepaalde taal of een bepaald geografisch domein te beperken, kunnen ze niet of nauwelijks als het om zoekacties gaat die zich beperken tot een bepaald onderwerp of vakgebied. Dat betekent dat voor deze machines elke keer weer uit de vraag zelf moet blijken waarover de vraag eigenlijk gaat en in welk vakgebied hij thuishoort. Neem een toch nog redelijk specifieke term als hyperactiviteit. Een ouder of opvoeder zal daarover heel andere informatie willen krijgen dan een criminoloog die onderzoek doet naar het verband tussen hyperactiviteit en criminaliteit, een socioloog die het verschijnsel bestudeert als factor in de groepsdynamiek van een klas en een psychiater die geïnteresseerd is in de hersenfysiologie van het syndroom. Of neem een term als vervreemding: gaat het daarbij om een juridisch, een filosofisch of een psychologisch begrip? Natuurlijk kunt u een algemene mondiale zoekmachine wel duidelijk maken waar het u om gaat – door de zoekterm voldoende context mee te geven in de vorm van andere zoektermen –, maar het zou een stuk gemakkelijker zijn als dat niet hoefde doordat het onderwerpsgebied waarbinnen u zocht, veel beter afgegrensd was.

Nieuws en personen zijn twee onderwerpen waarbij het gebruik van een speciale zoekmachine bijna vanzelfsprekend is. Van de grote mondiale zoekmachines is voor nieuws vooral Yahoo! aan te raden, met zijn Yahoo! News Full Coverage (<http://fullcoverage.yahoo.com>), dat nieuws van de afgelopen dertig dagen dekt. Voorbeelden van internationale nieuwszoekmachines zijn NewsHub (<http://www.newshub.com/>), met elk half uur geactualiseerde headlines uit de internationale (vak)pers en een 'archief' van 24 uur, Daypop (<http://www.daypop.com>), dat nieuwssites en blogs doorzoekt met een actualiteit van minimaal een maal per dag en een 'archief' van twee weken, en RocketNews (<http://www.rocketnews.com>), met nieuws van nieuwssites en blogs (zie blz. 277 e.v.) van tot vier dagen terug. In Nederland heeft Nieuwstracker.nl (<http://www.nieuwstracker.nl/>) een eigen zoekmachine en een 'archief' van twee dagen; wie wat verder wil terugkijken, kan terecht bij Redactiestart.nl (<http://redactiestart.nl/>), een portal met onder meer nieuws en een aantal (meta)zoekmachines voor nieuws. Een overzicht van Nederlandse nieuwsbronnen biedt de nieuwspagina van Startkabel.nl (<http://www.startkabel.nl/k/nieuws/>). Over nieuws per e-mail, rss etc. meer in het hoofdstuk Bijblijven, blz. 273 e.v. Zoeken naar mensen op het Web is een onderwerp waaraan een heel boek gewijd zou kunnen worden. De in de hieronder genoemde 'Guide to Specialized Web Search Engines' vermelde verzamelingen van speciale zoekmachines besteden natuurlijk ook aan dit onderwerp aandacht, maar snelheidshalve worden hier alvast twee uitgebreide overzichten genoemd die de belangrijkste aspecten en hulpmiddelen van het zoeken naar personen opsommen: Cyndi's List (<http://www.cyndislist.com/finding.htm>) met een vooral genealogische inslag en een uitgebreide en goed bijgehouden Nederlandse afdeling (<http://www.cyndislist.com/nether.htm>) en de People Finder Guide (<http://www.virtualchase.com/people/index.html>). De belangrijkste zoekinstrumenten voor het zoeken naar personen, zijn hier vermeld. De Nederlandse zoeksite Voelspriet (<http://www.voelspriet.nl/personenzoeken.htm>) heeft een handige Personenzoeker, die een aantal zoekfaciliteiten bij elkaar brengt.

Zoekmachines die zich op een speciaal vakgebied of een speciaal onderwerp concentreren zijn er ook, en dat ligt ook voor de hand. Het programmeren van een zoekmachine is een kunst die talloze IT-ers verstaan, en bovendien, er zijn programma's voor te koop of zelfs gratis te verkrijgen of te huren. Soms verschijnt er zo'n zoekmachine voor een tijdje op het Web, om dan weer te verdwijnen: ten tijde van de

affaire-Lewinsky bestond er een half jaar een zoekmachine die GoMonica heette en waarmee je alle details van de affaire in een besloten omgeving kon opzoeken. Voor die zoekmachine bestond er maar één Bill en één Monica! Er zijn talloze zoekmachines die zich met (vooral Amerikaanse) celebrities bezighouden, niet alleen celebrities in het algemeen, maar ook individueel: [ElvisFind](#) en de [Anna Kournikova Search Engine](#) zijn maar enkele van de voorbeelden. Maar ook serieuze onderwerpen als de filosofie, de internationale betrekkingen, de letterkunde, de medicijnen, chocolade en sigaren hebben hun eigen zoekmachines.

Zulke zoekmachines zijn ruwweg in drie soorten in te delen:

- de mooiste zijn die welke net als de grote mondiale zoekmachines (delen van) servers op het Web spideren en indexeren, maar dat alleen bij die servers doen waar veel of uitsluitend documenten op een bepaald vakgebied of over een bepaald onderwerp te vinden zijn. Deze zoekmachines worden wel LASEs genoemd, Limited Area Search Engines. Er zijn er onder meer voor [het milieu](#) en de [internationale betrekkingen](#). Een variant op deze LASEs vormen zoekmachines die met een thesaurus of knowledgebase werken. Een voorbeeld hiervan is de zoekmachine [MedHunt](#), waarvan de spider met een woordenboek van 12.000 termen uit de medische literatuur het Web afzoekt naar documenten die die termen bevatten. Die documenten worden opgehaald en geïndexeerd, waarna men er met MedHunt naar kan zoeken.
- zoekmachines die een met de hand verzamelde hoeveelheid links met beschrijvingen op een bepaald vakgebied of over een onderwerp indexeren en toegankelijk maken. Dit is het merendeel. Ze gaan meestal vergezeld van een directory.
- databaseprogramma's. Dit zijn eigenlijk geen echte search engines, in de zin dat aan de ontsloten database geen direct toegankelijke webpagina's ten grondslag liggen. We komen hier op de grens van wat wel het 'Invisible Web' wordt genoemd, een deel van het Web dat achter de poorten van databases verborgen ligt.

Uit deze opsomming blijkt wel welke voordelen speciale zoekmachines kunnen hebben boven de grote mondiale. Bij de eerste categorie zoekt men binnen de afgebakende omgeving van gespecialiseerde servers, waarop documenten met een gespecialiseerd idioom staan. Bij de tweede categorie zijn mensen betrokken geweest bij het selecteren en beschrijven van resources, wat een garantie kan zijn voor kwaliteit; en bij de derde categorie gaat het om documenten die met een gewone zoekmachine misschien niet gevonden zouden worden.

Overigens vinden ook speciale zoekmachines die geen databaseprogramma's zijn maar net als de grote mondiale zoekmachines direct toegankelijke webpagina's spideren, nog wel eens documenten die deze zoekmachines missen. Zo vindt de LASE van het International Relations and Security Network bij de vraag `virus site:ecp.nl` [41 treffers](#), tegen Google [23](#), MSN Search [18](#) en Yahoo! [1](#).

Een ander voordeel dat speciale zoekmachines wel eens kunnen hebben is dat ze ook met metadata volgens standaarden als bijvoorbeeld [Dublin Core](#) kunnen werken. Een voorbeeld is de fraaie zoekmachine van SURFnet (<http://search.surfnet.nl/selectSearchMode.do>), die de sites van het Nederlands hoger onderwijs en onderzoek doorzoekt.

Er bestaan duizenden van deze speciale zoekmachines op het Web, en er is geen beginnen aan ze allemaal op te noemen. Daarom volgen hier een aantal adressen waar ze verzameld en soms ook geannoteerd worden.

- Kijk in de eerste plaats eens bij de grote mondiale zoekmachines: veel ervan hebben nog allerlei speciale zoekmachines die zich op een vakgebied of onderwerp

concentreren. De belangrijkste staan bij de hierboven genoemde algemene zoekmachines genoemd. Meestal gaat het daarbij om zoekmachines die in de behoefte aan veel gevraagd materiaal zoals muziek en plaatjes kunnen voldoen. Yahoo! en Google voor (internationaal) nieuws en voor plaatjes zijn hiervan voorbeelden. (Voor een thematisch overzicht van afbeeldingen en resources voor afbeeldingen op het Web, zie Paula Berinsteins Finding Images Online: Links to Image Resources (<http://www.berinsteinresearch.com/fiolinks.htm>).

- Dé plaats om te beginnen als u bij de specialiteiten van de grote mondiale zoekmachines niet vindt wat u zoekt, is Paula Dragutsky's 'Guides to Specialized Web Search Engines (with Descriptions)' (<http://www.searchability.com>). Deze site geeft een geannoteerde opsomming van verzamelingen van zoekmachines over populaire en academische onderwerpen, verzamelingen van regionale zoekmachines en verzamelingen van zoekmachines voor kinderen. Aparte hoofdstukken zijn gewijd aan de grootste verzamelingen, gespecialiseerde metazoekmachines (die niet de algemene mondiale, maar speciale zoekmachines voor zich aan het werk zetten) en de Booleaanse operatoren. Ook is er een op omvang geordend overzicht en een alfabetisch register van de behandelde verzamelingen.
- Een niet in Dragutsky's overzicht genoemde verzameling is het oorspronkelijk Spaanstalige, maar inmiddels ook in het Engels beschikbare Buscopio: Search of Searches (<http://www.buscopio.net/eng/>), met meer dan 3800 speciale zoekmachines.
- Net zoals gewone metazoekmachines de algemene mondiale zoekmachines voor zich laten werken, zo doen speciale metazoekmachines dat met speciale zoekmachines. Drie voorbeelden zijn CNET's Search.com (<http://www.search.com/>), Intelliseeks Profusion (<http://www.profusion.com>), en GoshMe (<http://www.goshme.com>).
- [Search.com](http://www.search.com) kan 800 speciale zoekmachines en andere zoekinstrumenten bevragen, niet allemaal tegelijk, maar in onderwerpsgroepjes van allemaal ongeveer dezelfde interesse (klik op Research by topic). Net als bij Yahoo! en andere directories zijn de onderwerpen verdeeld over een boomstructuur, met rubrieken ('channels') en subrubrieken. Alleen staat hier voor elke subrubriek een aankruisvakje waarin u kunt aankruisen of u de betreffende rubriek doorzocht wilt hebben. Via een link 'customize' kunt u verder aankruisen welke bestanden in een subrubriek u doorzocht wilt hebben, en door middel van een aankruisvakje 'Remember selections' kunt u die keuzes ook opslaan. Op die manier kunt u diverse rubrieken tegelijk doorzoeken nadat u in die subrubrieken de te doorzoeken bestanden geselecteerd hebt. De metazoekmachine zelf ondersteunt phrases en Booleaanse (AND, OR, AND NOT) en unitaire operatoren, maar aangezien niet alle gebruikte zoekmachines dat ook doen, blijven sommige ervan onbenut; dat kan ook gebeuren doordat een bevroegde zoekmachine niet snel genoeg antwoordt. De resultaten worden per zoekmachine gepresenteerd, niet altijd compleet, maar in dat geval wel met een link naar het complete resultaat bij de zoekmachine zelf. Soms kunnen resultaten ook op een andere manier worden gesorteerd.
- Behalve een metazoekmachine van speciale zoekmachines is Search.com ook een gewone metazoeker, maar de zoekmachines die het daarbij gebruikt vormen nogal een allegaartje.



Web | Downloads | Shopping | Games | more »

 Search

[preferences](#) • [help](#) • [research by topic](#)

find: [personals](#) • [email addresses](#) • [missing people](#) • [privacy](#) • [mailing lists](#)

Remember selections [[restore defaults](#)]

American Memory

<http://lweb2.loc.gov/ammem/mbdqeny.htm>

FamilySearch

<http://www.familysearch.com>

GenForum

<http://genforum.genealogy.com/>

GenSource

<http://www.gensource.com/>

Genealogy.com

<http://www.genealogy.com/>

RootsWeb

<http://worldconnect.genealogy.rootsweb.c>

Yahoo! Groups

<http://groups.yahoo.com/>

Choose another channel:

Search.com is een metazoekmachine van speciale zoekmachines. Een van de subrubrieken onder People is Genealogy, met daarin zeven genealogische zoekmachines (opgesomd onder de zoekregel). Zoeken met bijvoorbeeld de naam Zorreguieta levert bij twee daarvan een (genealogisch interessant) resultaat op. Dit is een voorbeeld van waar speciale (meta)zoekmachines goed in zijn. Zouden we met Google of Yahoo! hebben gezocht, dan zouden we van alles over Zorreguieta hebben kunnen vinden, maar zouden de hier gevonden vermeldingen, als ze al gevonden waren, waarschijnlijk verscholen zijn gebleven tussen de andere treffers.

- [Profusion](#) functioneert standaard als een gewone metazoekmachine die een aantal van de grote mondiale zoekmachines bevraagt, waartoe onder meer Yahoo! en MSN Search behoren, maar niet Google. Als algemene metazoeker is het dus niet zo interessant. Als gespecialiseerde metazoekmachine kan het honderden speciale zoekmachines bevragen, en biedt de gebruiker daarbij ook nog de mogelijkheid een eigen mix van gebruikte zoekmachines samen te stellen. Om deze laatste en enkele andere geavanceerde functies te kunnen gebruiken moet u zich registreren, waarbij onder meer een mailadres moet worden opgegeven waarheen 'Search alerts' kunnen worden gestuurd en enkele standaardinstellingen (bijvoorbeeld het aantal zoekresultaten per zoekmachine en per resultatenpagina) kunnen worden gewijzigd. (Het is daarbij verstandig de Search Timeout laag in te stellen!) Via de link [Click here for a directory of all "Specialized Searches" on ProFusion](#) bereikt u de lijst van onderwerpen waaruit u kunt kiezen, verdeeld in 16 categorieën en meer dan 130 subcategorieën. Eenmaal in een subcategorie aangeland, ziet u daar een lijstje van 1 tot meer dan 20 zoekmachines die vanuit de zoekregel kunnen worden bevraagd. In sommige gevallen gebruikt Profusion er daarvan standaard

maar een deel, maar u kunt ze ook allemaal gebruiken of een andere keuze maken door de aankruisvakjes voor de zoekmachines aan te vinken.

Onder de zoekregels staat een aantal opties ('Search Options') waarvoor Profusion een standaardinstelling heeft die per zoekactie, of door het wijzigen van de standaardinstellingen na registratie ook voor toekomstige zoekacties, kunnen worden veranderd. Dat wijst zich vanzelf.

Na een zoekactie laat Profusion je het verloop daarvan bij de diverse zoekmachines zien en komt ten slotte met een resultaat van standaard 10 treffers per zoekmachine, met per treffer de plaats die deze in de eerste resultaten van de gebruikte zoekmachines heeft gescoord. Die resultaten kunnen worden geprint of gemaïld, en er kan een Search Alert op de vraag worden gezet waarmee nieuwe resultaten dagelijks, wekelijks of maandelijks naar het door u bij de registratie opgegeven mailadres worden gestuurd.

De resultaten van een zoekactie worden standaard in volgorde van relevantie getoond, waarbij Profusion treffers ordent volgens een score die de resultante is van het aantal zoekmachines dat een treffer hoog in de lijst heeft staan en de plaats die de treffer op deze lijst inneemt. Daarnaast kunnen de resultaten ook per zoekmachine worden bekeken. Kiest u voor relevantie, dan worden nog drie andere

The screenshot shows the Profusion search interface. At the top left is the Profusion logo. To its right are navigation buttons for 'home', 'alerts', 'options', 'feedback', and 'help'. Below these is a 'Welcome Guest!' message with a 'sign in' link. The search bar contains the text 'bookmarklets' and is followed by 'in Home > Developer Toolbox' and a 'go' button. Below the search bar, it says 'Did you mean bookmark lets?'. The results section shows '27 Results were discovered (Search Analysis)'. Search terms are 'bookmarklets' and 'Results Per Engine: 10 Sources: Design Trends, Developer Resources, HTML...'. There are links for 'Set Search Alert', 'Print Results', and 'Email Results'. The view is set to 'Relevance' and 'Source'. Search options include 'Search More Sources', 'Turn Off Highlighting', and pagination controls showing '1 2 3 All >> >'. The results are sorted by 'Score | Title | URL | Date'. The first four results are listed:

- 1. Googling with Bookmarklets** (open window) [Set Page Alert]
Create interactive **bookmarklets** to perform Google functions from the comfort of your own browser.
Context: ... Googling with **Bookmarklets** Create interactive **bookmarklets** to perform Google functions from the comfort of your own browser. Contributed by: ...
URL: <http://hacks.oreilly.com/pub/h/148>[O'Reilly Network #1]
Status: OK Score: 63% Date: 17-Jan-2005 [Find Similar Results]
- 2. accessibility related bookmarklets from ...** (open window) [Set Page Alert]
... 5. <http://www.studioid.com/projects/bookmarklets/> Even EARL is in on the act ... Next in thread : Sarah Horton: "Re: accessibility related **bookmarklets**" ...lists.w3.org/Archives/Public/w3c-wai-eo/2002AprJun/0019.html - 8k -
URL: <http://lists.w3.org/Archives/Public/w3c-wai-eo/2002AprJun/0019.html> [W3C #1]
Status: OK Score: 61% [Find Similar Results]
- 3. Google with Bookmarklets** (open window) [Set Page Alert]
Create interactive **bookmarklets** to perform Google functions from the comfort of your own browser
Context: ... Google with **Bookmarklets** Create interactive **bookmarklets** to perform Google functions from the comfort of your own browser [Discuss (0) | Link to this...]
URL: <http://hacks.oreilly.com/pub/h/2724>[O'Reilly Network #2]
Status: OK Score: 60% Date: 17-Jan-2005 [Find Similar Results]
- 4. Bookmarklets** (open window) [Set Page Alert]
Home page for **bookmarklets**. www.devx.com/webdev/Link/17128 - 2k -
URL: <http://www.devx.com/webdev/Link/17128>[Dev X #1]
Status: OK Score: 59% [Find Similar Results]

Resultatenlijst voor een vraag naar 'bookmarklets' in de rubriek Developer Toolbox van Profusion.

sorteermogelijkheden getoond, nl. alfabetisch op titel, alfabetisch op URL en chronologisch op datum (dit laatste alleen als Profusion in de META-informatie van de gevonden documenten een datum heeft kunnen vinden).

Drie extra's die Profusion biedt zijn een controle op dode links, een zogeheten 'search analysis', en suggesties om (een andere subcategorie van) zoekmachines te gebruiken. De 'search analysis' geeft het aantal resultaten per bevroegde zoekmachine, het aantal dode links, de tijd die de zoekactie gekost heeft en de status: complete of incomplete. Verder is er een overzicht van geëlimineerde resultaten ('no match'), eveneens met de vermelding waarom ze geëlimineerd zijn: wegens een te lage relevantiescore, een gebroken link of een onbereikbare server. Een interessante nieuwigheid die Profusion voor heeft op alle andere (meta)zoekmachines is dat het alvorens een zoekresultaat te tonen zelf nog eens de actuele documenten kan ophalen om duplicaten en gebroken links te identificeren en te verwijderen.

Een laatste welkom extraatje is dat Profusion geregistreerde gebruikers bij het tonen van de resultatenlijst in de gelegenheid stelt bij de afzonderlijke treffers 'Page Alerts' op te geven, dat wil zeggen dat Profusion u per email meldt wanneer er iets aan de betreffende pagina verandert. (Zie ook het hoofdstuk Bijblijven, blz. 273 e.v.)

Profusion lijdt aan hetzelfde euvel als andere metazoekmachines, namelijk dat het alleen het alleen die zoekfuncties kan leveren die alle bevroegde zoekmachines gemeen hebben. Waar het meer pretendeert te kunnen – namelijk met de optie Search Type: Boolean – levert de metazoeker bij zoekmachines die geen (volledig) Booleaans zoeken ondersteunen, resultaten die niet aan de zoekcriteria voldoen; ook dat is een kenmerk dat Profusion met andere metzoekers gemeen heeft. Niettemin heeft deze aan de University of Kansas ontwikkelde metzoekmachine een aantal leuke innovatieve kenmerken. En in de wereld van speciale metzoekers is de concurrentie niet groot.

- Een potentieel interessante nieuweling in deze wereld is het Braziliaanse GoshMe – The Web Search Assistant. Net als Search.com en Profusion gebruikt GoshMe andere zoekmachines voor zijn resultaten, maar het toont geen eigen resultatenlijst. In plaats daarvan laat het per zoekmachine het aantal treffers zien dat die zoekmachine gevonden heeft. Wat GoshMe in verband met het onderwerp Speciale zoekmachines interessant maakt is dat het net als Search.com en Profusion niet alleen de grote mondiale zoekmachines gebruikt, maar ook een aantal directories (zoals ODP, Librarian's Internet Index en About.com) en een verzameling speciale zoekmachines, waarvan de laatste net als bij Search.com en Profusion in een aantal (17) onderwerpsgebieden is ingedeeld. Hier wordt de zoekactie tot maximaal 5 onderwerpsgebieden tegelijk beperkt, waarna GoshMe bij elk van de zoekmachines waarin resultaten worden gevonden het aantal treffers, met een link naar de resultatenlijst in de betreffende zoekmachine, vermeldt. Tot de bevroegde speciale zoekmachines behoren ook databases, zodat GoshMe ten dele ook een zoekmachine voor het Invisible Web (zie blz. 233 e.v.) is. Wat GoshMe minimaal doet, is u attenderen op speciale zoekmachines en databases waarin antwoorden op uw zoekvragen gevonden worden.
- Jammer is dat GoshMe geen phrases en operatoren ondersteunt. Dat leidt ertoe dat in zoekmachines of databases die standaard met een OF-relatie werken, bij meer zoektermen ten onrechte treffers worden gevonden.
- Speciale zoekmachines zijn vaak te vinden via de grote onderwerpsdirectories en portals op het Web. Deze vindt u als 'Web directories' in de directory van Yahoo!.
- En ten slotte kan zoeken naar speciale zoekmachines natuurlijk ook via de grote mondiale zoekmachines. Trefwoorden: special search engines, specialty search engines, specialized search engines, subject-specific search engines, niche search engines, topical search engines, vertical search engines.

wiles fermat theorem

Go Now

 All Words

Find the best search engines to your query

[choose up to 5 categories of interest]

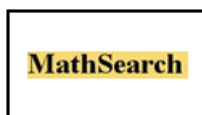
<input type="checkbox"/> Animals & Farming Pets, Agriculture, Veterinary, Zool...	<input type="checkbox"/> Arts & Social Sciences Architecture, Literature, Philosoph...	<input type="checkbox"/> Audio, Video & Images MP3, Enterviews, Broadcasting, Gene...
<input type="checkbox"/> Economics & Business Industry, Jobs, Yellow Pages, Busin...	<input type="checkbox"/> Environment & Society Government, Religion, Books, Public...	<input type="checkbox"/> Health Medicine, Nutrition, Multimedia, Bi...
<input type="checkbox"/> Home & Garden Home Decor, Cooking, Garden & Fores...	<input type="checkbox"/> Informatics & Internet Games, Technical, Software, Domain ...	<input type="checkbox"/> Law International Law, Criminal, Forms,...
<input type="checkbox"/> News Blogs, Latest, Regional Newspapers,...	<input type="checkbox"/> Recreation & Entertainment Movies, People, Fashion & Beauty, G...	<input type="checkbox"/> Reference Articles, Universities, Encyclopedi...
<input checked="" type="checkbox"/> Science Engineering, Chemistry, Mathematics...	<input type="checkbox"/> Shopping Comparison, Auctions, Arts, Classif...	<input type="checkbox"/> Sports Basketball, Cycling, Fishing, Socce...
<input type="checkbox"/> Tourism Hotels, Shopping, Maps, Informati	<input type="checkbox"/> Adults Video & Images, Hentai, WebCams, G	

De vraag wiles fermat theorem, uit te voeren in de speciale zoekmachines en databases in de categorie Science van GoshMe. Zoeken met phrases is helaas niet mogelijk.


[Scirus](#) ★ [Close preview](#)

Scirus is the most comprehensive science-specific search engine. It found **60** results in its search for **wiles fermat theorem**. Scirus searches over 200 million science-specific Web pages, enabling you to quickly pinpoint scientific, scholarly, technical and medical data on the Web.

Category: Science


[MathSearch](#) 🔍

Mathsearch: of over 200,000 documents on English-language mathematics and statistics servers across the Web. **285** results for **wiles fermat theorem**.

Category: Science >Mathematics


[Google Scholar - Science](#) 🔍

Google Scholar (Science) looks only for academic materials that can help you in yours studies. **666** articles and papers regarding **wiles fermat theorem**.

Category: Science

Gedeelte van het resultaat van de voorgaande zoekactie in GoshMe. Bij de eerste treffer is gebruik gemaakt van de 'preview'- mogelijkheid onder het loop-pictogram rechts van de snippet.

4.4.4.1. Een eigen zoekmachine

Speciale metazoekmachines als Search.com, Profusion en GoshMe hebben natuurlijk een paar belangrijke tekortkomingen: ze bestrijken maar een beperkt aantal onderwerpen en gebruiken voor de onderwerpen die ze bestrijken een beperkt aantal gespecialiseerde zoekmachines die u niet zelf kunt kiezen. Maar als u beroepshalve of voor uw hobby steeds dezelfde websites moet doorzoeken (of de bezoekers van uw site daartoe in staat zou willen stellen), zou het prettig zijn als u uw eigen LASE voor die sites zou kunnen hebben. Gesteld bijvoorbeeld dat u een site heeft over Chinees porselein en u wilt bezoekers van die site niet alleen op uw eigen site laten zoeken maar ook op enkele tientallen verwante sites die u in het ODP of in de directory van Google of Yahoo! heeft gevonden, kan dat dan zonder dat u een duur programma koopt en een ook niet goedkope IT-specialist inhuurt? Ja dat kan, als we voor dat doel een grote mondiale zoekmachine, bijvoorbeeld Yahoo!) gebruiken die we zo inrichten dat hij alleen zoekt op de sites die wij hem opgeven. Daarvoor hebben we in de eerste plaats een stukje HTML-code nodig dat een zoekvenster op onze webpagina plaatst van waaruit de ingetypte zoektermen naar de zoekmachine worden opgestuurd, waarna deze het resultaat op een nieuwe pagina of in een nieuw venster toont:

```
<FORM METHOD="GET" ACTION="http://search.yahoo.com/bin/search" "target=V_blank">  
<INPUT SIZE="30" NAME="p" VALUE="Typ hier de vraag"><INPUT TYPE="submit"  
VALUE="Zoek met Yahoo!"></FORM>
```

Als wijzelf of onze bezoekers nu op de plaats waar in het zoekvenster nu Typ hier de vraag staat onze zoektermen intypen en op de knop Zoek met Yahoo! drukken, stuurt Yahoo! ons het resultaat in een nieuw venster toe.

Het enige wat nu nog moet gebeuren is dat we in het hierboven getoonde stukje HTML-code van tevoren opgeven tot welke sites Yahoo! de zoekactie moet beperken. Dat doen we door de tekst 'Typ hier de vraag' te laten volgen door een aantal spaties en de volgende tekst: *AND (site:aaa.com OR site:bbb.edu OR site:ccc.nl OR site:ddd.org)* etc. De spaties zetten we erin om dit deel van de vraag buiten het zoekvenster te laten vallen. De zoektermen die nu in het zoekvenster worden ingetypt worden door de eerste AND met het door ons opgegeven restant van de vraag verbonden.

Als we de ingevoerde zoekvragen niet tot hele sites willen beperken, kan dat ook, en wel door naast *site:* ook *inurl:* of *path:* te gebruiken. De tekst die we opgeven komt er dan ongeveer zo uit te zien: *AND (site:aaa.com AND inurl:china) OR (site:bbb.edu AND path:shops) OR (site:ccc.nl AND inurl:stores) OR site:ddd.org)*. Deze simpele eigen zoekmachine is natuurlijk wel gebonden aan de beperkingen van Yahoo! zelf. Wat Yahoo! niet heeft geïndexeerd, zal het ook niet vinden, of het nu om pagina's van uw eigen site of die van andere sites gaat. Dat zou iemand op de gedachte kunnen brengen dat je hetzelfde beter met Google kunt doen, dat immers een grotere linkdatabase heeft. Maar helaas, met Google gaat dit maar in bescheiden mate, omdat Google niet meer dan 32 zoektermen combineert, zodat het aantal te doorzoeken sites dat je kunt opgeven nogal beperkt is; bovendien is Booleaans zoeken bij Google maar beperkt mogelijk en kan een vraagbeperking zoals *AND (site:aaa.com AND inurl:china) OR (site:bbb.edu AND inurl:chinese)* etc. in Google niet worden toegepast. Een andere grote zoekmachine die de mogelijkheid van een eigen zoekmachine wel uitdrukkelijk aanbiedt en u ook uitlegt hoe u dat op een eigen webpagina moet realiseren is Gigablast met zijn zogeheten [Custom Topic Search](#). Maar met zijn database van ruim 2 miljard pagina's is Gigablast een stuk kleiner dan Yahoo! en hoewel Gigablasts documentatie vermeldt dat sites inclusief een pad kunnen worden opgegeven, is mijn indruk dat die sites vervolgens niet door Gigablast worden doorzocht. Wie voorbeelden wil zien van eigen zoekmachines met Yahoo! en Gigablast, vindt die in het

artikel Yahoo's No-Limit Query Limit As Opposed to Google's – So What? in Tara Calishains ResearchBuzz van 19 augustus 2004 (http://www.researchbuzz.org/2004/08/yahoos_nolimit_query_limit_as.shtml), op de site van Pandia (Pandia Search Engine Topic Search, <http://www.pandia.com/searchworld/gigablast.html>; probeer en bekijk het resultaat) en in de eerste twee berichten van mijn weblog Search and Retrieve (<http://searchretrieve.blogspot.com/>). Op Yahoo! gebaseerde gepersonaliseerde zoekmachines zijn verder te vinden in de 'searchrolls' van Rollyo (<http://www.rollyo.com>) en, nog fraaier, de 'swicki's' van Eurekster (<http://swicki.eurekster.com/>). Ten slotte biedt ook MSN Search sinds kort de mogelijkheid op basis van Microsofts zoektechnologie (en de linkdatabase van MSN Search) 'search macro's' op te zetten die hetzelfde doen als hierboven beschreven. Zie MSN Search's Weblog: Build Your Own Search Engine (<http://blogs.msdn.com/msnsearch/archive/2006/03/28/563482.aspx>) en voor voorbeelden: Microsoft Gadgets (<http://microsoftgadgets.com/Macros>). Het bovenstaande is natuurlijk alleen bedoeld voor degenen die een paar man's' oplossing wensen voor het specifieke probleem hoe je een zoekactie tot een aantal goede, relevante sites kunt beperken.

The screenshot shows the Microsoft Gadgets Search Macro Gallery interface. At the top, there is a navigation bar with links for Gallery, Blog, Forums, Build, Macros, and Login. Below this, a search bar is visible with the text "Search Macro Gallery » A macro that helps locate corporate financial documents" and a "Search" button. The main content area displays a macro titled "research.companyfinancials" with a five-star rating. The macro's description is "A macro that helps locate corporate financial documents". There is a "Live.com" button and a "Comment" button. The macro's default scope is set to "web". The definition of the macro is: "(filetype:pdf | filetype:doc | filetype:ppt | contains:pdf | contains:doc | contains:ppt) ("annual report" | "balance sheet" | "income statement" | "sec filing" | "sec filings" | "investor relations" | "quarterly report" | "quarterly statement" | "earnings per share")". At the bottom of the page, there is a copyright notice: "© Copyright 2005 Microsoft Corporation." and links for "Terms of Use", "Privacy Statement", and "Contact Us". A footer note states "This site is hosted for Microsoft by ORCSWEB".

Voorbeeld van een zoekmacro van Microsoft, die – in combinatie met eigen zoektermen – kan worden uitgevoerd op [Windows Live](http://windowslive.com).

5. Op zoek naar het Invisible Web

Het Invisible Web is onzichtbaar voor zoekmachines. Maar er zijn wel manieren om het te vinden. In dit hoofdstuk gaan we ernaar op zoek.

Wie naar de aantallen documenten kijkt die de grote mondiale zoekmachines zeggen te indexeren, en daarbij ook nog eens bedenkt dat volgens de experts de overlap tussen hun linkdatabases betrekkelijk gering is, kan zich misschien nauwelijks voorstellen dat het 'zichtbare Web' dat de spiders van zoekmachines zien, maar een klein gedeelte van het hele Web is, en dat daarbuiten nog een heel 'invisible Web' bestaat. En toch is dat zo. Om dit te illustreren twee voorbeelden.

Het eerste betreft de ervaringen van de bibliotheek van de University of Buffalo in de VS, die haar catalogusrecords als HTML-bestanden op het Web plaatste en vervolgens afwachtte wat Google daarmee ging doen. Het aantal bestanden, verdeeld over 80.000 folders, bedroeg ruim 2 miljoen, maar toen Google's spider, Googlebot, was langs geweest bleek die er, na de 80.000 directories te hebben opgehaald en geïndexeerd, er na 20.000 catalogusrecords de brui aan te hebben gegeven. Kennelijk is er – onder omstandigheden – een grens aan het aantal bestanden dat Google van één site ophaalt. In de documentatie van zoekmachines wordt op dat feit ook wel gezinspeeld, omdat een uitgebreide spiderbeurt voor een site zeer belastend kan zijn (zie: Googling the Library Catalog (<http://www.theshiftedlibrarian.com/2003/02/03.html#a3569>), ook verschenen als Breaking Through the Invisible Web (<http://libraryjournal.reviewsnews.com/index.asp?layout=article&articleid=CA266430&publication=libraryjournal>)).

Mijn tweede voorbeeld is er een dat u zelf kunt zien. Bij de bibliotheek van de Technische Universiteit in Wenen heeft iemand een overzicht van gratis raadpleegbare bibliografische en andere bestanden op een server gezet. Voor elke letter van het alfabet is een aparte webpagina, 23 in totaal omdat er voor de letters q, x en y geen bestanden zijn (zie: <http://www.ub.tuwien.ac.at/ska/cdrom/hdstartalph.htm>). Deze menupagina is door geen van de grote mondiale zoekmachines geïndexeerd, net zomin als de index van de directory waarin de bestanden zijn opgeslagen. Maar deze laatste is wel te bekijken (<http://www.ub.tuwien.ac.at/ska/cdrom/>) en uit deze index blijkt dat de directory behalve de pagina's met de letters van het alfabet (hda.htm, hdb.htm ... htz.htm) nog honderden andere pagina's bevat.

Als we nu kijken wat de spiders van de grote zoekmachines met deze directory hebben gedaan, blijkt het resultaat bedroevend:

Zoekmachine	Zoekopdracht	Aantal treffers
Google	site:www.ub.tuwien.ac.at inurl:cdrom	6
Yahoo!	site:www.ub.tuwien.ac.at inurl:cdrom	7
MSN	site:www.ub.tuwien.ac.at inurl:cdrom	2
Ask	site:www.ub.tuwien.ac.at inurl:cdrom	1
Exalead	site:www.ub.tuwien.ac.at inurl:cdrom	23

Waarom alleen maar sommige pagina's in de folder geïndexeerd zijn, en waarom Google andere pagina's heeft geïndexeerd dan bijvoorbeeld Yahoo!, is een raadsel. Sommige pagina's hebben incoming links, bij andere zijn die er niet of niet meer. In het

laatste geval zijn ze misschien door de spiders van zoekmachines gevonden doordat ze zelf outgoing links hebben, die door spiders elders als incoming links worden aangetroffen en terug gevolgd (zulke incoming links worden bijvoorbeeld door tellers als NedStat als '[referrers](#)' geregistreerd).

Hoe dan ook, deze twee voorbeelden laten een van de twee typen beperkingen zien waaraan de spiders van zoekmachines onderhevig zijn: zoekmachines kunnen best miljoenen pagina's van dezelfde site spideren en ze kunnen best alle referrer links in logbestanden volgen, maar om praktische en economische redenen doen ze dat niet. Daarnaast zijn er ook technische redenen waarom zoekmachines niet in sommige delen van het Web kunnen doordringen.

5.1. Beleidsmatige beperkingen aan het door zoekmachines geïndexeerde Web

Om beleidsredenen behoren tot het Invisible Web:

- documenten die niet worden gespiderd omdat er een bepaalde limiet aan het aantal gespiderde bestanden per site is gesteld
- documenten die zo diep in de directorystructuur van een site verstopt zitten dat de spiders van search engines er niet komen; dat wil overigens niet zeggen dat die documenten niet met de search engine van de site of via het browsen door een menustructuur kunnen worden gevonden
- teksten in andere formaten dan die welke door de grote mondiale zoekmachines worden geïndexeerd, denk bijvoorbeeld aan de diverse gecomprimeerde formaten (ZIP, ARC, TAR etc.)
- sites, delen van sites of documenten waarvoor door webmasters in een bestand robots.txt of in de METAtags 'robots' met de inhoud 'noindex,nofollow' is aangegeven dat ze niet door zoekmachines moeten worden geïndexeerd
- een deel van de 'dynamisch gegenereerde' pagina's die door de talloze websites (die vanuit een database worden gecreëerd), databases en catalogi op het Web worden geproduceerd

In feite gaat het bij deze laatste categorie om twee verschillende, maar ten dele overlappende gevallen. In de eerste plaats kan een spider niet wat u en ik wel kunnen: zoektermen in een interactief formulier (een 'form', bijv. een zoekregel) invullen en op Enter drukken. Pagina's die alleen op die manier gevonden kunnen worden, worden door spiders niet gevonden (zie bij technische beperkingen). In de tweede plaats zijn spiders erop geprogrammeerd links naar URLs waarin een vraagteken voorkomt, niet of slechts in beperkte mate te volgen. Zulke URLs ontstaan als pagina's niet statisch zijn maar 'on the fly' uit een database worden gegenereerd, en dat is zowel het geval als er een zoekformulier wordt gebruikt en een 'site search engine' aan het werk wordt gezet, als wanneer de hele site vanuit een database wordt gegenereerd. In dat laatste geval bevatten alle links op zo'n site een script (met een vraagteken erin) dat de server ertoe brengt om de opgevraagde pagina's te tonen. Zulke links kunnen spiders zoals Googlebot nog wel volgen, maar alleen in beperkte mate en voor zover ze aan bepaalde voorwaarden – met niet meer dan één vraagteken en met een beperkt aantal 'query string parameters' van een beperkte lengte in het gedeelte van de URL na het vraagteken – voldoen.

Zie: Search engine crawlers and dynamic web pages

(<http://www.theallineed.com/webmasters/04092101.htm>); Make Your Dynamic Web Site Search Engine Friendly (<http://www.searchenginepromotionhelp.com/m/articles/search-engine-optimization/dynamic-content-promotion.php>); Google Information for Webmasters (<http://www.google.com/webmasters/2.html#A1>).

5.2. Technische beperkingen

Technische oorzaken spelen een rol bij het uit de boot vallen van:

- webpagina's zonder indexeerbare tekst, die alleen uit een plaatje, een audio- of een videobestand of een Flash-file bestaan (bijv. <http://www.tefal.de/aktuell/ingenio/ingenio.swf>)
- teksten die alleen als plaatje en niet als indexeerbare documenten beschikbaar zijn
- data op het Web die aan voortdurende verandering onderhevig zijn, zoals beurskoersen, vertrek- en aankomsttijden van vliegtuigen, weersberichten, filemeldingen enz.
- webpagina's die alleen door een zoekopdracht in een form en via een script kunnen worden geproduceerd (en dus niet via browsen door een menustructuur)
- webpagina's die niet bij zoekmachines zijn aangemeld en die niet door spiders kunnen worden ontdekt, enerzijds omdat er geen documenten zijn die ernaar linken of waarnaar ze zelf linken, en anderzijds omdat ze verstopt zitten op plaatsen in een directorystructuur waar spiders niet meer komen (bijvoorbeeld in een subdirectory van een lege directory)
- sites, delen van sites of documenten die door een wachtwoord worden beveiligd
- documenten in de talloze intranetten die deel uitmaken van het Web en die alleen vanaf bepaalde IP-adressen kunnen worden bekeken

Eigenlijk moeten hier nog twee types slechts gedeeltelijk zichtbare documenten aan worden toegevoegd:

- documenten die op iedere denkbare vraag zo laag in het zoekresultaat van zoekmachines ranken dat ze altijd buiten de boot van het aantal door zoekmachines getoonde treffers vallen
- documenten die zo lang zijn dat een deel van de tekst niet door zoekmachines wordt geïndexeerd. Met zoektermen die alleen in dat deel van het document voorkomen, wordt het document dan niet gevonden.

Tot het onzichtbare Web reken ik ook de resultaten van een functie waarover de grote mondiale zoekmachines niet, of alleen met een speciale zoekmachine (Froogle), beschikken, nl. prijsvergelijking. Een bekend voorbeeld is [Addall](#).

Zoekmachines als Google, Yahoo! en MSN Search vinden tegenwoordig ook treffers in grote databases zoals PubMed, de grote medische database van de Amerikaanse National Library of Medicine en WorldCat, de overkoepelende catalogus van de bibliotheekorganisatie OCLC. PubMed artikelen vindt u met Google door aan uw zoektermen *site:ncbi.nlm.nih.gov inurl:pubmed* toe te voegen. Titels in WorldCat vindt u met Google en Yahoo! door aan uw zoektermen *site:worldcatlibraries.org* toe te voegen. In het eerste geval gaat het om dynamisch gegenereerde pagina's, in het tweede om statische pagina's.

Overigens is het niet raadzaam Google en soortgelijke zoekmachines te gebruiken om in catalogi en databases te zoeken. In het algemeen is de databasesoftware zoals PubMed en WorldCat die gebruiken daarvoor veel beter geschikt, omdat de indexeer- en zoekfuncties van deze software beter zijn toegesneden op het soort gestructureerde informatie die databaserecords bevatten. Daarbij komt dat bij vergelijking van de aantallen treffers op willekeurige zoektermen in de titel tussen WorldCat en PubMed enerzijds en Google en Yahoo! anderzijds blijkt dat nog lang niet alle records uit beide databases door de zoekmachines gevonden worden.

Wat beeld betreft is het goed te beseffen dat zoekmachines als Google en Yahoo! weliswaar mogelijkheden bieden om via zoektermen (tekst dus) images op het Web te vinden, maar dat deze mogelijkheden maar heel beperkt zijn in vergelijking met tekst.

Zoekmachines kunnen zelf niet 'zien' waar een beeld over gaat; daarvoor zijn ze afhankelijk van de informatie die in de bestandsnaam (balkenende.gif) of in de ALT-tag en/of de hyperlink van een verwijzend tekstdocument wordt gegeven. En die informatie is doorgaans maar heel summier.

Speciale zoeksoftware, die doorgaans maar één database tegelijk ontsluit, kan wel meer. Terwijl de Advanced Search van Google Images als beeldontsluitingscriterium eigenlijk alleen de optie Coloration: any colors/black and white/greyscale/full color heeft, biedt de QBIC (= Query By Image Content, zie: <http://www.qbic.almaden.ibm.com/>) software van IBM musea zoals de Hermitage in St.-Petersburg de mogelijkheid hun collecties op kleurencombinaties en lay-out te ontsluiten. Zie: <http://www.hermitagemuseum.org/fcgi-bin/db2www/qbicSearch.mac/qbic?sellLang=English>. Voor meer over dit onderwerp, zie: A Survey on: Contents Based Search in Image Databases, <http://www.isy.liu.se/cvl/ScOut/TechRep/Papers/LITHISYR2215.pdf>.

Hetzelfde geldt voor geluid en bewegende beelden. Weliswaar bestaan daarvoor speciale zoekmachines die het Web in het algemeen of een deel ervan pretenderen te dekken, maar ook die baseren zich grotendeels op contextuele informatie die in gewone HTML-bestanden wordt verstrekt. Voorbeelden van zulke zoekmachines zijn SingingFish (<http://search.singingfish.com/>) voor audio en video (zoek: Jennifer Aniston, Paganini Accardo) en FindSounds (<http://www.findsounds.com/>) voor geluidseffecten (zoek: nightingale, applause). Alleen de SpeechBot van Hewlett Packard, die Amerikaanse radioprogramma's ontsloot, maakte gebruik van 'speech recognition technology' en 'indexeerde' dus de inhoud van de geluidsbestanden zelf.

5.3. Databases opsporen

Zoals ik hierboven heb uiteengezet, is het 'Invisible Web' dat gedeelte van het Web dat wel kan worden bekeken maar niet met de grote zoekmachines zoals Google en Yahoo! wordt geïndexeerd. Meestal gaat het hierbij om grotere of kleinere databases, die als zodanig wel via Google etc. vindbaar zijn, maar waarvan de inhoud achter een zoek- en/of een inlogformulier verborgen ligt. We bevinden ons dus hier in een ruimte waarin als het ware de voordeuren met het naambordje van de bewoner – catalogus van de Universiteitsbibliotheek van Amsterdam, het Octrooiregister van het Bureau voor de Industriële Eigendom, CBS StatLine, het archief van de Volkskrant – wel via de grote zoekmachines te vinden zijn, maar alles daarachter niet. Zo heel bezwaarlijk hoeft dat niet te zijn: als we maar enig idee hebben wat er achter de voordeur verscholen kan liggen en waar we de voordeur moeten zoeken, kunnen we uiteindelijk wel bij de informatie komen waarnaar we op zoek zijn.

Om informatie op het Invisible Web op te sporen hebben we de beschikking over vier verschillende methoden:

- A. via zoekinstrumenten op het Web die speciaal bedoeld zijn om ons de voordeuren van dit deel van het Web te helpen vinden;
- B. via zoekmogelijkheden die de grote mondiale zoekmachines en gespecialiseerde directories/webgidsen ons bieden om databases te vinden;
- C. door op die plaatsen op het Web te gaan kijken waar we mogen verwachten dat het type informatie waarnaar we op zoek zijn, aanwezig is;
- D. via databaserecords die bij zoekmachines terechtgekomen zijn.

5.3.1. Speciale instrumenten om databases te vinden

Verzamelingen van databases vindt u – onder meer – op de volgende plaatsen:

- ◆ direct search - Web Search Tools and Directories (featuring Invisible Web

Resources) (<http://www.freepint.com/gary/direct.htm>); de verzameling is afgesloten, maar voor nieuwe resources zie ResourceShelf (<http://www.resourceshelf.com/>).

- ◆ INTERNET INVISIBLE.com (<http://www.internetinvisible.com/>), een Spaanstalige, maar internationale verzameling van 1700 databases op diverse terreinen
- ◆ Datenbank-Infosystem Universitätsbibliothek Regensburg (http://www.bibliothek.uni-regensburg.de/dbinfo/?bib_id=ub_r). Zeer uitgebreid, thematisch ingedeeld overzicht van bibliografische en andere databases en databanken op het Web, waarvan de meeste vrij toegankelijk.
- ◆ Fachbibliographische Datenbanken (<http://new.heimat.de/home/bibdat/fachbib.htm>). Thematisch overzicht van bibliografische databases, met een aanduiding van het aantal records dat ze bevatten.
- ◆ Les bases de données gratuites (<http://dadi.enssib.fr/>). Thematisch ingedeelde verzameling van 716 vrij toegankelijke bibliografische en andere databases.
- ◆ Bases de données bibliographiques sur la Toile (<http://www-bu.univ-paris8.fr/Ref/DocBdd.html>). Bibliografische databases.
- ◆ Kostenfreie Literaturinformationen im Web (<http://www.ub.tuwien.ac.at/ska/cdrom/hdstartalph.htm>). Bibliografieën en bibliografische databases, woordenboeken, encyclopedieën etc.
- ◆ Informationen aus kostenlosen Internet-Datenbanken (<http://www.informationsmarkt.de/free.html>). Duits georiënteerd allegaartje, met een aantal nuttige verwijzingen.
- ◆ Répertoire des bases de données (<http://www.chu-rouen.fr/documed/tex.html>). Allegaartje van databases over diverse onderwerpen.
- ◆ Deep Web Research 2005, een overzicht van materiaal m.b.t. het Invisible Web door Marcus P. Zillman (<http://www.llrx.com/features/deepweb2005.htm>) en, van dezelfde auteur: Deep Web Research Subject Tracer Information Blog (<http://www.deepwebresearch.info/>) en Data Mining Subject Tracer Information Blog (<http://www.dataminingresources.info/>).

Daarnaast zijn er enkele zoekinstrumenten die zich zoekmachines van het Invisible Web noemen:

- ◆ IncyWincy: The Invisible Web Search Engine (<http://www.incywincy.com>); deze zoekmachine, die ook als metazoekmachine kan worden gebruikt, zoekt in eerste instantie in de directory van het Open Directory Project (ODP; zie 4.1.1.3), en dan speciaal naar die door die directory ontsloten pagina's die een eigen zoekformulier hebben. Dit wordt door IncyWincy zelf getoond waarna de gebruiker er verder mee kan zoeken. Helaas, de spider van IncyWincy mist nogal wat pagina's met een zoekregel, en bij wat hij wel vindt werkt het zoekvenster dat IncyWincy dan verschaft niet goed, is sprake van verdubbeling – zo kom je de Surfnets zoekmachine meermalen tegen, steeds voor een faculteit of een opleiding van een universiteit – of blijkt de zoekmachine van de site Google te zijn – geen Invisible Web dus (maart 2006 was niet duidelijk of de zoekmachine nog bestond).

Ook de directory van Google biedt mogelijkheden om databases te vinden die verder gaan dan het bestand van ca. 4 miljoen sites die de directory omvat. Anders dan het Open Directory Project zelf indexeert Google namelijk ook alles wat op de homepages van de sites in de directory voorkomt.

- ◆ CompletePlanet (<http://aip.completeplanet.com/>). Dit is een directory van meer dan 70.000 sites met een database of een speciale zoekmachine. De zoekpagina's van

deze sites zijn geïndexeerd en zijn met de zoekmachine van Complete Planet te doorzoeken; daarnaast kan door de directory worden gebladerd, en kunnen de summary's worden bekeken. Het resultaat is niet erg bruikbaar: in de eerste plaats is de toewijzing van de databases aan de diverse categorieën en subcategorieën weinig consistent en zit veel op de verkeerde plaats (zo zit een database over gambling literature onder literature en niet onder sociale wetenschappen); en in de tweede plaats geeft de tekst op de geïndexeerde zoekpagina's vaak maar een gebrekkige indruk van waarnaar eigenlijk gezocht kan worden. Interessanter is misschien het product Deep Query Manager (DQM) van het bedrijf dat Complete Planet onderhoudt, BrightPlanet. Met dit product kunnen sites uit de lijst van 70.000 databases worden geselecteerd en vervolgens in hun geheel gespiderd ('geharvest').

- ◆ Turbo10 Search Engine (<http://turbo10.com/>). Dit is een metazoekmachine voor ruim 3300 databases die tot het Invisible Web behoren. De zoekmachine heeft meer geavanceerde kenmerken, maar voor het zoeken in het Invisible Web zijn de belangrijkste dat de gebruiker een eigen selectie uit de bestaande collectie kan maken (die via een cookie op de eigen computer bewaard blijft) en daar ook via een op het oog eenvoudige procedure eigen databases aan toe kan voegen. Een poging om dat met de [BL Online](#) van de Koninklijke Bibliotheek te doen liep stuk op het feit dat die database met frames werkt.

Zie ook: Deep Web Search Tools

(<http://www.bhsu.edu/education/edfaculty/Itturner/Deep%20Web%20Search%20Tools1.htm>)

Deze benadering van het probleem hoe we het Invisible Web kunnen doorzoeken schiet naar mijn mening ernstig tekort. Ze miskent de verscheidenheid van het karakter, de omvang, de dekking en de inhoud van de databases, databanken en catalogi die op het Web te vinden zijn. En bovenal, ze is gecompliceerd en tegelijk intrinsiek incompleet.

5.3.2. *Met Google en webgidsen naar databases zoeken*

Databases op het Web kunnen gaan over alle mogelijke onderwerpen, van fruitvliegjes via orchideeën en Beethoven tot aan de filosofie en haar verschillende subdisciplines en richtingen. Een simpele manier om (een deel van) het Invisible Web bij een bepaald onderwerp te vinden is daarom er met Google of een andere grote zoekmachine naar te zoeken: vat het onderwerp waarover u informatie zoekt in een paar termen samen en voeg daar termen als 'database', 'data', 'dataset', 'archive', 'bibliography', 'index', 'statistics' (afhankelijk van het soort informatie waarnaar u op zoek bent) aan toe, om te beginnen met de prefixen `intitle:` en `inanchor:` (voor Google) en (eveneens voor Google) eventueel met het synoniemteken `~` (`intitle:~statistics`) eraan toegevoegd. Afhankelijk van het nationale of regionale belang van het onderwerp kunt u natuurlijk met de taal variëren.

In het algemeen geldt: hoe specifiek de database, des te meer materiaal hij over het onderwerp bevat. Zoek dus liever eerst met een term als 'orchids' in plaats van 'plants' of 'botany' als u een database met gegevens over orchideeën zoekt en liever met 'aria' dan met 'opera' of zelfs 'music' als het u speciaal om aria's gaat. Als een te specifieke zoekterm niets oplevert kunt u altijd nog met bredere termen werken.

Toevoeging van zoektermen als 'database' kan een goede methode zijn om databases op te sporen, maar leidt ons niet altijd naar het Invisible Web. Er zijn databases te over die met statische webpagina's werken, en de inhoud daarvan vinden Google, Yahoo! en andere zoekmachines ook. Maar we hebben meer pijlen op onze boog. Als we de URLs van dynamische webpagina's wat nader bekijken, komen we daarin elementen tegen die we als zoekterm kunnen gebruiken om databases op te sporen die zulke pagina's genereren. Dit zijn elementen als: `cgi`, `bin`, `search`, `asp`, `cfm`, `query` en `php`. Door deze elementen als zoekterm toe te voegen aan de termen die we gebruiken om het

onderwerp te omschrijven waarnaar we op zoek zijn, kunnen we databases over dat onderwerp op het spoor komen. Een paar voorbeelden:

Google: [mycology inurl:cfm](#) levert twee direct als zodanig identificeerbare databases op: de [Systematic Botany and Mycology Fungal Databases](#) en de [LGC Promochem: Fungi & Yeast Collection](#). Vervangen we `inurl:cfm` door `inurl:asp`, dan komt daar o.a. de [Bibliography of Systematic Mycology](#) nog bij.

Google: [hotels paris inurl:search](#) levert als een van de eerste treffers [France Hotels, Paris Hotels and the France Travel Guide](#) op.

Het gebruik van grote zoekmachines om het Invisible Web te ontdekken biedt nog andere mogelijkheden. Pagina's waarop zoektermen kunnen worden ingevuld bevatten een invulregel of een invulformulier dat in de HTML van de webpagina als een zogeheten 'form' staat aangegeven. De zoekmachine van Yahoo! maakt het mogelijk met allerlei bijzonderheden van webpagina's te zoeken en `feature:form` is daar een van. Omdat formulieren ook andere dan zoekformulieren kunnen zijn, is het raadzaam aan de zoekvraag de term `search / zoek OR zoeken / etc.` toe te voegen:

Yahoo!: [intitle:elvis feature:form search](#) zoekt de webpagina's met in de titel de term Elvis, ergens de term search en een invulformulier. Yahoo!: [site:kb.nl AND feature:form AND \(zoek OR zoeken\)](#): zoekt webpagina's van de Koninklijke Bibliotheek met een zoekformulier en ergens (in de URL, de titel of de tekst) het woord zoek of het woord zoeken (of beide)

Hoewel zoekmachines dus wel mogelijkheden bieden om databases op te sporen en daarmee toegang te krijgen tot een stukje van het Invisible Web, is dit werken met `inurl: cgi etc.` en `feature:form` een methode om dat te doen die niet erg precies is en nogal wat ruis (vooral in de vorm van zogeheten 'site search engines') oplevert. Aanbevolen als het echt niet op een andere manier lukt een database over een onderwerp te vinden, of bij heel gespecialiseerde onderwerpen.

Meer perspectief biedt de benadering via grote vakspecifieke directories, gateways en portals, die, zoals al eerder ter sprake is gekomen, zijn op te sporen via de directory van Yahoo! (met het trefwoord 'Web directories'); overigens heeft ook Google, met zijn rankingmechanisme dat sites beoordeelt die veel links trekken, de neiging zulke directories en portals bij zijn eerste resultaten te tonen. Als een database echt van belang is voor een bepaald vakgebied mag je verwachten dat hij in een vakspecifieke directory of overzicht van resources vermeld zal zijn.

Databases and Indexes

- ♦ **RAMBI**

"The Index of Articles on Jewish Studies" from the Jewish National and University Library.

<http://jnul.huji.ac.il/rambi>

- ♦ **THEOLDI**

"Theological Literature Documented in Innsbruck."

<http://bildi.uibk.ac.at/index-theoldi-en.html>

Databases waarvan de inhoud tot het onzichtbare Web behoort, zijn vaak op te sporen via portals of gateways van een vakgebied. Hier een gedeelte van The [ORU Library Gateway to Theological Resources on the Internet](#).

5.3.3. Databases zoeken op de plaatsen waar ze te verwachten zijn

Dit brengt me op de derde methode om databases op te sporen: door op het Web daar te gaan kijken waar het type informatie waarnaar we op zoek zijn naar te verwachten aanwezig is.

Om simpel te beginnen: niemand zal Google gebruiken om het huidige adres en telefoonnummer van de vertaler dr. C.M.E. Harrisson op te sporen, die een paar jaar geleden nog in Oegstgeest woonde, maar daar op het adres waar hij woonde nu niet meer te vinden is. Het beste begin je bij zo'n vraag in een database van het onzichtbare Web die zo vanzelfsprekend is dat je niet eens beseft met het onzichtbare Web te maken te hebben: een telefoonboek. En dat waarschijnlijk niet om de heer Harrisson op te zoeken – hoewel de twee meest gebruikte online telefoongidsen die er voor Nederland zijn beide zoeken in heel Nederland mogelijk maken – maar om de huidige bewoner van het adres waar hij woonde te vragen of de heer Harrisson een adres heeft achtergelaten. Om dat te kunnen vragen moet je even het zichtbare Web op, namelijk om een telefoongids te vinden die een naam en telefoonnummer bij een adres vindt. Iedere startpagina noemt er wel [een of meer](#).

Als de huidige bewoner dat niet kan of wil zeggen, volgt de volgende stap: je zoekt de heer Harrisson in een gewone telefoongids op zonder een plaats op te geven, dus in Nederland (dat kan in het Uitgebreid zoeken-schermbild van zowel [De Telefoongids](#) als de [Nationale Telefoongids](#)) en/of je zoekt in de 'Tovergids' van de branchevereniging waartoe de betrokkene mogelijk behoort, het Nederlands Genootschap van Tolken en Vertalers. Ook deze gids is een database die tot het onzichtbare Web behoort, evenals de email-adressenbestanden die aan de beurt komen als de heer Harrisson niet in de telefoongids of het adressenbestand van vertalers te vinden zou zijn.

Met de telefoongids en het adressenbestand van vertalers bevinden we ons in twee uithoeken van het onzichtbare Web die we bij zo'n vraag intuïtief opzoeken. Dat deze databanken bestaan en worden bijgehouden is nogal wies: de telefoonmaatschappijen verschaffen telefoongidsen omdat ze er belang bij hebben dat er zoveel mogelijk gebeld wordt, en als het om gouden gidsen gaat hebben bedrijven er belang bij dat men ze weet te vinden; en hetzelfde geldt voor een branche- of beroepsorganisatie. Anders gezegd, telefoonmaatschappijen en branche- en beroepsorganisaties zijn twee soorten bedrijven/organisaties waarvan we kunnen verwachten dat ze specifieke informatie leveren. Of nog anders, en algemener: informatie, op het zichtbare én op het onzichtbare Web, is daar te vinden waar mensen of organisaties er een belang bij – of plezier in – hebben die informatie te verschaffen.

Met dit uitgangspunt kunnen we heel wat vragen die uiteindelijk op het onzichtbare Web zullen moeten worden beantwoord, intuïtief proberen te benaderen. Natuurlijk, het is altijd mogelijk dat we het antwoord op een vraag direct bij Google of een andere grote zoekmachine vinden; maar als dat niet zo is, moeten we ons afvragen wie de informatie waarnaar we op zoek zijn zou kunnen leveren en misschien in een database heeft opgeslagen.

Ter illustratie van het bovenstaande volgt hieronder een lijst van 27 enigszins exotische vragen waarop het antwoord alleen of bij voorkeur op het Nederlandse onzichtbare Web moet worden gezocht. Met 'bij voorkeur' bedoel ik dat het antwoord mogelijk ook wel met een van de grote mondiale zoekmachines kan worden gevonden, maar dat de specialisatie en de organisatie (en de daaraan aangepaste zoekfuncties) van de database die over het onderwerp gaat, het gebruik van die database preferabel maken. (Een voorbeeld daarvan is de [Nederlandse Onderzoek Databank](#): de documenten in deze database zijn ook met Google te vinden, maar het [Uitgebreid Zoekschermbild](#) van de NOD zelf maakt inperken op onderzoek in uitvoering/afgesloten/beide en op promotie-onderzoek ja/nee mogelijk.)

Opvallend is dat vrijwel alle hieronder genoemde databases op onderwerp vindbaar zijn

via de Google webgids. In de vierde kolom vindt u in welke tak(ken) van de boom, zodat u zich een idee kunt vormen hoe u databases ook intuïtief of via een (min of meer) logische weg zou kunnen vinden.

Google indexeert de homepages van de sites die het in zijn directory heeft zitten. Daardoor zult u er ook databases vinden die niet in de directory zelf zijn opgenomen, maar wel op zo'n homepage staan. Waar dat het geval is heb ik de betreffende homepages met een extra pijltje aangegeven. Overigens bevat de oorspronkelijke directory van ODP een aantal van die databases wel (een voorbeeld is StatLine, dat slechts indirect via de directory van Google en de homepage van het CBS wordt gevonden, maar direct via het ODP).

<i>Vraag betreft</i>	<i>Database</i>	<i>Instelling</i>	<i>Vindbaar via</i>
naam en adres van een Nederlandse bloembollenexpert	World Taxonomist Database	Universiteit van Amsterdam, ETI - Expert Center for Taxonomic Identification	Google Directory → Science > Biology > Taxonomy
in Nederland van toepassing zijnde wetsartikelen over het onderwerp besnijdenis	Database Nederlandse wetgeving en seksualiteit	Rutgers Nisso Groep – Kenniscentrum seksualiteit	
literatuur over non-verbale communicatie	BL Online The bibliographical database of linguistics	Koninklijke Bibliotheek	
het telefoonnummer van de molenaar van de Maredijkmolen in Leiden	Nederlandse Molendatabase Het Molenbestand	Hippo WebDesign, Amsterdam Vereniging De Hollandsche Molen	Google Directory → World > Nederlands > Kunst > Architectuur > Geschiedenis > Gebouwtypen > Molens
het adres van een rozenkwekerij	Tuinlinks – Bloemen en planten startpagina	Groeninfo.nl	Google Directory → World > Nederlands > Huis en tuin > Tuin
de exacte plaats van de kustlijn op landkaarten	Kartografisch woordenboek	sectie Kartografie en Geovisualisatie van GIN, vereniging Geoinformatie Nederland	Google Directory → World > Nederlands > Naslag > Woordenboeken > Thema
calorische waarde van kippenmest	Phyllis, database for biomass and waste	Energieonderzoek Centrum Nederland	
alle personen in Nederland wier telefoonnummer begint met de combinatie 51215	Gebeld.nl (geavanceerd zoeken)		Google Directory → World > Nederlands > Naslag > Adresgegevens

<i>Vraag betreft</i>	<i>Database</i>	<i>Instelling</i>	<i>Vindbaar via</i>
de boete die een officier van justitie zal eisen voor het bespieden van iemand op een woonschip	Boetebase	Openbaar Ministerie	
het schip waarop Johannes Hofstede uit Den Haag in 1733 zijn laatste reis als matrozenhulpje maakte	Uitgevaren voor de Kamers van de VOC	Nationaal Archief en archiefdiensten van Amsterdam, Delft, Enkhuizen, Hoorn, Rotterdam en Zeeland	Google Directory → World > Nederlands > Maatschappij > Geschiedenis > Op Onderwerp > VOC ; <i>of:</i> Google Directory → World > Nederlands > Maatschappij > Genealogie > Hulpmiddelen > Doorzoekbare Databases
de ontwikkeling van de gemiddelde leeftijd van huwende mannen over de periode 1993-2004	StatLine	Centraal Bureau voor de Statistiek	Google Directory → Science > Social Sciences > Economics > Indicators and Statistics > Official Statistics > Europe → Netherlands - Centraal Bureau voor de Statistiek ; <i>of:</i> StartNederland.nl → Statistieken
een afbeelding van de oudste bewaard gebleven stoomlocomotief in Nederland	Het Nederlands Spoorwegmuseum	Het Nederlands Spoorwegmuseum	Google Directory → Reference > Museums > Transportation > Railroad > Europe
prijzvergelijking van direct ingaande lijfrentes	Independer.nl	Robin Hood N.V.	Google Directory → World > Nederlands > Zakelijk > Financieel > Hypotheken
foto van een markt in Atjeh, omstreeks 1900	fotocollectie van het Department of Archives & Images	Koninklijk Instituut voor Taal-, Land- en Volkenkunde	
kleurstoffen en pimenten die in Nederland aan diervoeding mogen worden toegevoegd	Databank toevoegingsmiddelen	Productschap Diervoeder	Google Directory → World > Nederlands > Zakelijk > Voedings- en Genotmiddelen > Diervoeding → Productschap Diervoeder

<i>Vraag betreft</i>	<i>Database</i>	<i>Instelling</i>	<i>Vindbaar via</i>
diogeneesmiddelen die in Nederland zijn toegestaan voor de behandeling van koliek	Diergeneesmiddelen Database	Bureau Registratie Diergeneesmiddelen	Google Directory → World > Nederlands > Zakelijk > Gezondheidszorg > Diergeneeskunde > Producten > → Diergeneesmiddelen Bureau Registratie Diergeneesmiddelen
literatuur over gallicismen in het Nederlands	Bibliografie van de Nederlandse taal- en literatuurwetenschap (BNTL)	Nederlands Instituut voor Wetenschappelijke Informatiediensten	Google Directory → World > Nederlands > Naslag > Wetenschappelijk → Nederlands Instituut voor Wetenschappelijke Informatiediensten; of: Google Directory → World > Nederlands > Wetenschap > Geesteswetenschappen > Taal- en Letterkunde > Talen > Nederlands → Taalunieversum
een octrooi voor een fiets en fietsslot dat in gesloten toestand zowel de stuurkolom als de trapas van de fiets blokkeert	esp@cenet	Bureau voor de Industriële Eigendom	Google Directory → World > Nederlands > Regionaal > Nederland > Overheid > Rijksoverheid > Ministeries > Ministerie van Economische Zaken
Large scale modeling of landsurface-atmosphere interactions	Nederlandse Onderzoek Databank	Nederlands Instituut voor Wetenschappelijke Informatiediensten	Google Directory → World > Nederlands > Naslag > Wetenschappelijk → Nederlands Instituut voor Wetenschappelijke Informatiediensten
geschiedenis van de kerk van de H. Catharina van Alexandrië te Hengstdijk, Zeeuwsch-Vlaanderen	Bedevaartplaatsen in Nederland	Meertens Instituut	Google Directory → World > Nederlands > Wetenschap > Geesteswetenschappen > Taal- en Letterkunde > Talen > Nederlands → Meertens Instituut

<i>Vraag betreft</i>	<i>Database</i>	<i>Instelling</i>	<i>Vindbaar via</i>
het hotel in Domburg waar minister H.A. van IJsselstein en zijn familie op 10 juli 1919 logeerden	Zeeuws Archief Isis	Zeeuws Archief	Google Directory → World > Nederlands > Maatschappij > Genealogie > Hulpmiddelen > Doorzoekbare Databases
de tekst van het herderlijk schrijven dat de aartsbisschop en de bisschoppen van Nederland op 21 februari 1943 in kerken en kapellen lieten voorlezen	Het "Illegale Parool"-archief 1940-1945	Stichting Democratie en Media, voorheen Stichting "Het Parool"	Google Directory → World > Nederlands > Maatschappij > Geschiedenis > Naar Tijdperk > Twintigste Eeuw > Tweede Wereldoorlog
de tekst van art. 366 van het Wetboek van Strafrecht	Overheid.nl – Wet- en regelgeving	Ministerie van Binnenlandse Zaken en Koninkrijksrelaties	Google Directory → World > Nederlands > Maatschappij > Recht > Juridische Informatie
de tekst van het octrooi, verleend aan de West-Indische Compagnie, van 3 juni 1621	Het Geheugen van Nederland – The Atlantic World	Koninklijke Bibliotheek	Google Directory → Regional > Europe > Netherlands > Guides and Directories → Portals to the world - The Netherlands; of: Google Directory → Kids and Teens > International > Nederlands > School > Geschiedenis
berichtgeving over de moord op Nienke Kleiss, 2000-heden	AD.nl Archief	Het Rotterdams Dagblad (tegenwoordig AD Archief)	Google Directory → World > Nederlands > Nieuws en Media > Dag- en Nieuwsbladen > Nederland
een baan als hypotheekadviseur	Werk.nl	Centrum voor Werk en Inkomen	Google Directory → World > Nederlands > Zakelijk > Banen > Loopbaan
ReVox B77 MK2 Bandrecorder	Marktplaats.nl	Ebay	Google Directory → World > Nederlands > Webwinkelen > Advertenties

Beperkt als hij is geeft deze lijst hopelijk wel een indruk van de diversiteit aan informatie die op het Invisible Web te vinden is. Tegelijk valt op dat de databases waarin die informatie gevonden kan worden van een beperkt aantal aanbieders afkomstig is. De overheid, op diverse niveaus, is een belangrijke aanbieder, evenals semi-overheden en zelfstandige bestuursorganen, toezichthouders en inspecties, (nationale) bibliotheken, (nationale) archieven, musea, wetenschappelijke instellingen, kranten en omroepen, bedrijven en branche- en bedrijfsorganisaties, verenigingen en clubs, hobbyisten. De belangrijkste aanbieders hebben over het algemeen de meeste en de grootste databases. Wie zich probeert voor te stellen waar op het Invisible Web informatie te vinden kan zijn moet zich rekenschap geven van:

- de aard van de informatie die hij/zij zoekt en het soort instelling/organisatie/persoon dat die informatie zou kunnen verschaffen
- de omvang van de database waarin de informatie te vinden zou kunnen zijn
- geografische (provincie/plaats) en chronologische (verleden/heden) aspecten van de vraag.

De hierboven gegeven voorbeelden hebben merendeels betrekking op Nederland en in Nederland aangeboden databases. Daarmee bevinden we ons in het Nederlandse deel van het onzichtbare Web. Wagen we ons daarbuiten, dan wordt het terrein oneindig veel groter en ook complexer, maar wie zich eenmaal een beeld heeft gevormd van de Nederlandse situatie, kan de blauwdruk daarvan ook op het buitenland proberen toe te passen. In veel gevallen vormt de Google Directory, met zijn zoekboom, daarvoor een handig hulpmiddel: als de Nederlandse site Overheid.nl te vinden is in de rubriek [World > Nederlands > Regionaal > Nederland > Overheid](#), dan zoeken we de Franse Service-Public.fr in de rubriek [World > Français > Régional > Europe > France > Etat et politique](#). Vervolgens kunnen we met Yahoo! en 2x link ([link:http://www.overheid.nl](http://www.overheid.nl) [link:http://www.service-public.fr](http://www.service-public.fr)) sites proberen te vinden die behalve naar de Nederlandse en de Franse overheidsportals nog naar andere overheidsportals linken. Of we kunnen meteen doorgaan naar het [Nations Online Project :: A Destination Guide to Countries and Nations of the World](#), waar bij elk land sites worden gegeven in de volgende sectoren: Official Sites – Map – News – Culture – Business – Country Guides – Education – Genealogy – History – Science & Research – Search – Additional Links.

5.3.4. *Databases opsporen via records in zoekmachines*

De vierde en laatste methode om het onzichtbare Web te vinden die ik hier wil noemen, maakt gebruik van het feit dat de grens tussen zichtbaar en onzichtbaar Web steeds minder scherp wordt doordat zoekmachines tegenwoordig in staat zijn ook dynamische pagina's te spideren. Tal van treffers die u bij een zoekactie kunt tegenkomen komen tegenwoordig uit databases, en dat betekent natuurlijk dat we een database – als die door een zoekmachine ontsloten wordt – ook kunnen opsporen via de informatie die we in zo'n database mogen verwachten aan te treffen. Een voorbeeld is een bibliografische database met literatuur over een bepaald onderwerp. Door de exacte bibliografische gegevens van een artikel (auteursnaam, woord(en) uit de titel van het artikel, woord(en) uit de titel van het tijdschrift) aan Google te voeren kunnen we de database(s) opsporen waarin de beschrijving van dat artikel zich bevindt – natuurlijk alleen als Google toegang tot die database heeft.

5.4. **Zoeken in databases: een stoomcursus**

Het aantal databases op het Web loopt in de honderdduizenden of meer, en binnen het bestek van dit boek is het onmogelijk anders dan in heel algemene termen aandacht te besteden aan de vraag hoe die moeten worden doorzocht. Daarvoor is de verscheidenheid in type – van simpele naam/adres/woonplaats/telefoonnummer-

bestanden via bibliotheekcatalogi tot octrooidatabases – te groot, en zijn de interfaces te verschillend. Wat ik hier kan behandelen, zijn de 'bare bones', de eigenschappen die databases en het zoeken daarin over de hele linie wel met elkaar gemeen hebben. Aan u als gebruiker van deze of gene database is de taak deze algemene eigenschappen door het matglas van hun gebruikersinterface te herkennen en te benutten.

5.4.1. *Structuurverschillen*

Allereerst moet worden opgemerkt dat het bij de databases waarvan hier sprake is, doorgaans niet om fulltext databases gaat, zoals de databases die zoekmachines ontsluiten. Webpagina's en andere documenten die door zoekmachines worden geïndexeerd onderscheiden zich door hun betrekkelijk grote rijkdom aan tekst en een betrekkelijk geringe mate van structuur; en waar die structuur er is, kan een zoekmachine daar weinig mee omdat er geen uniforme structuur voor documenten op het Web is. Zo zal de auteur van een wetenschappelijk artikel op het Web dat netjes indelen in titel, abstract, tekst (met kopjes en onderkopjes) en noten, maar is een roman op hetzelfde Web in hoofdstukken ingedeeld. Als gevolg van die variëteit aan structuren beperken zoekmachines zich bij hun indexing tot een onderscheid tussen die elementen die webdocumenten wel grotendeels met elkaar gemeen hebben, of die in elk geval waar ze voorkomen, uniform zijn: URL, titel, tekst, link, anchor, feature. Databases, zoals die hier besproken worden, bevatten doorgaans geen artikelen of webpagina's of andere objecten – in de breedste zin des woords --, maar (korte) beschrijvingen daarvan, en wel uniforme beschrijvingen op basis van een aantal vooraf vastgestelde kenmerken. Dat laatste brengt mee dat ze in hoge mate gestructureerd zijn en dat de zoekmachines van die databases heel gevarieerde deelverzamelingen (sets) van zulke beschrijvingen kunnen vinden op basis van identieke kenmerken. De hierboven geschetste verschillen tussen webdocumenten en de beschrijvingen (in het jargon: records) in databases brengen aanzienlijke verschillen in het zoeken met zich mee. Zo kan de gebruiker van een webzoekmachine, in overeenstemming met het karakter van de documenten waarnaar hij/zij op zoek is, wel met phrases zoeken, maar kan hij/zij niet aangeven dat een zoekterm in de literatuurreferentie van een artikel moet voorkomen; omgekeerd kan de zoeker in een bibliotheekcatalogus wel specificeren dat een zoekterm in het plaats-van-uitgave veld van de beschrijving van een boek moet staan, maar heeft zoeken met phrases veel minder zin omdat gebruikmaking van de structuur van een catalogusrecord veel beter werkt: auteur:Hermans AND titel:kamer werkt veel zekerder dan "de donkere kamer van Damokles" (of was het Damocles?). Daarbij komt dat databases vaak niet alleen maar formele beschrijvingen van objecten bevatten (naam, adres, woonplaats), maar dat die beschrijvingen, afhankelijk van het gebruiksdoel van de database, verrijkt zijn met informatie die bedoeld is om er binnen de database een gemeenschappelijk karakter aan te geven en de deelverzameling waartoe ze daarmee behoren van andere deelverzamelingen te onderscheiden. Zo kan de toevoeging van de informatie rijk/modaal/arm aan de formele beschrijving van huishoudens in een naam/adres/woonplaats database die database bruikbaar maken voor mailings, net zoals de toevoeging van trefwoorden of onderwerpscodes aan de bibliografische beschrijvingen van een bibliotheekcatalogus het mogelijk maakt deelverzamelingen op onderwerp of vakgebied te selecteren. De mogelijkheden die dit aspect van databases aan de zoeker biedt, ontbreken (vooralsnog) op het Web. Samengevat: databaserecords zijn doorgaans korter en meer gestructureerd dan webdocumenten en voorzien van informatie die deelverzamelingen in de hele database betreft; dat maakt ze gemakkelijker en op meer uiteenlopende kenmerken vindbaar, mits men maar van alle mogelijkheden van de zoeksoftware gebruik maakt.

5.4.2. Interfaces

Dat veel mensen niettemin een afkeer van het zoeken in databases hebben, komt door het grote aantal dat ervan bestaat (hoe kies ik de juiste database?) en de verscheidenheid van de interfaces waarmee ze zijn opgetuigd (hoe presenteert die database zich aan mij en zoek ik er op de juiste manier in?). Het aantal mondiale zoekmachines dat meer dan een miljard Webdocumenten in zijn database heeft, is op de vingers van twee handen te tellen; databases daarentegen zijn er bij de vleet, en ze dekken alle mogelijke verschillende (en, per beschreven soort object, vrijwel altijd overlappende) verzamelingen. En daarbij komt dan nog eens dat een groot deel ervan zich op volkomen verschillende manieren aan de gebruiker presenteert, zodat hij of zij het zicht op de essentialia van het zoeken in databases dreigt te verliezen.

Voor het interface van een database maakt het niet of nauwelijks uit wat voor soort objecten daarin worden beschreven. Alleen de objectkenmerken en hun benamingen verschillen: in een personendatabase kunnen die kenmerken geboortejaar, geslacht, lichaamslengte en kleur van de ogen zijn, in een woningendatabase adres, postcode, plaats, oppervlakte en vraagprijs, en in een bibliotheekcatalogus auteur, titel, uitgever, plaats van uitgave en jaar van uitgave. Het enige verschil dat er hier toe doet is dat tussen letters/woorden (geslacht, plaats, titel) en getallen (geboortejaar, vraagprijs, jaar van uitgave). Bij de laatste kan de databasesoftware het mogelijk maken niet alleen op = (standaard), maar ook op > (groter dan) en < (kleiner dan) te zoeken.

Heel in het algemeen kunt u bij databases vier verschillende interfaces tegenkomen:

- eenvoudig zoeken (Simple Search, Basic Search): in een zoekregel voert u een of meer termen in die in een standaardrelatie (EN of, soms, OF) met elkaar worden verbonden en in het hele databaserecord worden gezocht; daarnaast kan in een keuzemenuutje de mogelijkheid worden geboden preciezer naar afzonderlijke kenmerken in de objectbeschrijvingen te zoeken, bijvoorbeeld het adres in een naam/adres/woonplaats database. Kenmerkend voor dit soort interface is dat het gebruikers wel termen laat combineren en een kenmerk (attribuut) waarop ze willen zoeken laat kiezen, maar geen kenmerken laat combineren.

Eenvoudig zoeken

Zoekterm(en):

(Titel)woord: [help](#)

Onderwerp: [help](#)

Persoonsnaam: [help](#) [Index](#)

Koninklijke Bibliotheek - Nationale bibliotheek van Nederland
Nederlandse Taalunie

Voorbeeld van een Eenvoudig zoeken-scherm in een database. Bij het kenmerk of attribuut (veld) Persoonsnaam is ook het browsen van een index mogelijk gemaakt.

- zoeken via een index (browsen of scannen), zie hieronder: de gebruiker krijgt een lijst van alle termen in de database en van alle termen per kenmerk gepresenteerd en kan daaruit termen kiezen. Doorgaans is de indexterm gelinkt, zodat op die zoekterm direct een zoekactie kan worden uitgevoerd; soms vindt u daarbij ook nog een mogelijkheid om diverse indextermen aan te vinken en een zoekactie met de aangevinkte termen – doorgaans in een OF-relatie – uit te voeren.

Bladeren in een alfabetische index

Terug

Vul woord(en) in: Schmidt

Kies een index: Auteur

Zoek

Index: Auteurs

Aantal	Treffer
1	Schmidt, Anna
1	Schmidt, Anne (1951-)
3	Schmidt, Anne M
1	Schmidt, Annette (1956-)
1	Schmidt, Annette (1966-)
1	Schmidt, Annette (1968-)
10	Schmidt, Annie M. G
1	Schmidt, Anton
9	Schmidt, Anton Franz -
1	Schmidt, Anton Wilhelm

Vorige

Volgende

Zoeken via een index. Vóór de indextermen de aantallen records waarin ze voorkomen.

- geavanceerd zoeken (Uitgebreid zoeken, Advanced Search): dit gaat doorgaans via een zoekformulier waarin de diverse kenmerken in de beschrijving van objecten van elkaar worden onderscheiden. In de plaats van de standaard EN-relatie wordt verder vaak een keuze tussen EN, OF en NIET geboden. Een veel voorkomend extra kunnen secundaire kenmerken zijn, kenmerken waarop in de database enkel in combinatie met een primair kenmerk kan worden gezocht. Een voorbeeld van zo'n secundair kenmerk in een woningendatabase kan zijn: aantal kamers, of: wel of geen CV.

Formulieren voor geavanceerd zoeken bieden om praktische redenen nogal eens een beperkt aantal invoervelden aan. Bij sommige databases wordt de gebruiker uitgenodigd zelf velden toe te voegen als hij/zij dat nodig vindt. Meestal gebeurt dat op de Windows-manier met plustekens die een folder met extra termen of opties openen.

Search terms: [Clear all fields](#)

and	Person as subject	Chomsky, Noam	Get list
and	Title	grammar	Get list
and	Language (classification)		Get list
and	Journal		Get list
and	Publisher		Get list

and or not

publication: Books Articles Reviews

publication: All years Before After Equals

Between and

Fill in at least one field. Be sure to use the Get List button(s) to insert search items in their canonical form.

Search

Interface met de kenmerken van een Geavanceerd zoeken-scherm. Kenmerken (velden) kunnen met EN, OF en NIET worden verbonden. Ook hier is het raadplegen van een index ('Get list') mogelijk gemaakt.

- commandozoeken (Command Line Search): net als bij het scherm Eenvoudig zoeken hebben we hier één zoekregel, maar dan wel met alle mogelijkheden (en vaak nog meer) van het Geavanceerd zoeken-scherm. Op de diverse kenmerken kan worden

gezocht door ze – volgens een database-afhankelijke standaard – te benoemen: in een bibliotheekcatalogus bijvoorbeeld AUT = auteur, TRF = trefwoord. Zoektermen (en kenmerken) kunnen met elkaar worden verbonden door AND, OR en NOT, en nesting, het combineren van EN- en OF-relaties in één en dezelfde zoekvraag, is mogelijk. Vaak kan op meer kenmerken worden gezocht dan in het Geavanceerd zoeken-scherm, en kunnen daarbij ook secundaire kenmerken worden gebruikt. In de hierboven genoemde Woningendatabase zou dat bijvoorbeeld kunnen in de vorm van KAM=5 voor het secundaire kenmerk dat een huis minimaal 5 kamers moet hebben.



Electronic Research Archive for Mathematics

Jahrbuch Database

[Contact](#)

Search

Advanced Search

Command Search

Citation Checker

2005 PAM Division Award

News

- Update [may 2005](#)
- Years **1868-1942**
- 17 035** links to facsimiles

Services

- [Classification](#)
- [Journals](#)

Information

- [About the Jahrbuch Project](#)
- [Letters to the editor](#)
- [Mirror](#)



Command Search

[Help](#)

Please enter a query in the box below

Display: entries per page.

Format:

Search

Clear Form

HINTS FOR USAGE:

& = AND | = OR ^ = AND NOT

au = wiles,a* & ti = modular elliptic & py = 1995

cc = 62H30 ^ bi = cluster analysis

au = erdős, pa* & dt = book ^ cc = 05-06 & py = 1980:1990

au = ((bender,* ^ canfield,*) | (canfield,* ^ bender,*)) & py >= 1997

- an [Accession number](#): JFM item number
- au [Author name](#)
- bi [Basic index](#)
- cc [Mathematical subject classification code](#)
- dt [Document type](#)
- kw [Keywords](#)
- la [Language](#)
- py [Publication year](#)
- rv [Reviewer's name](#)
- so [Source \(Journal, Publisher...\)](#)
- sh [Subject Heading](#)
- ti [Title words](#)

Copyright Jahrbuch Project : © 2001 [European Mathematical Society](#).
[edbm/w3] Retrieval & display software : © 2001 [Cellule MathDoc](#), [UJF](#) & [CNRS](#).

Voorbeeld van een 'Command search'-interface. De vraag kan in één venster worden ingetypt, inclusief zoek sleutels (au, ti), operatoren (& voor AND, | voor OR, ^ voor NOT) en wildcards (*).

Deze basistypen van gebruikersinterfaces komen in allerlei combinaties en soms ook mengvormen voor. Zo kan het voorkomen dat het Eenvoudig zoeken-scherm tevens dient om commandogestuurde zoekacties mogelijk te maken, dat zoeken via een index apart beschikbaar is of met Eenvoudig zoeken en Geavanceerd zoeken geïntegreerd, of alleen met een van beide. Over wat Eenvoudig zoeken en wat Geavanceerd zoeken eigenlijk is, bestaan kennelijk grote verschillen van mening.

Naast deze vier zoekinterfaces bieden veel databases nog enkele extra mogelijkheden om verder te zoeken als men eenmaal een zoekactie heeft uitgevoerd en daarbij treffers heeft gevonden. Een van die mogelijkheden is dat men in de complete presentatie van een record dat men met een zoekactie heeft gevonden, een link volgt vanuit een term die men daar aantreft naar andere records die diezelfde term bevatten. Om een voorbeeld te noemen: iemand kan in de Bibliografie van de Nederlandse Taal- en Literatuurwetenschap (BNTL, <http://wwwoud.niwi.knaw.nl/cgi-bin/nph-bntl.pl>) beginnen met op zoek te gaan naar artikelen die in de titel de naam 'Piet Grijs' hebben, en dan in die artikelen het gelinkte trefwoord 'Hugo Brandt Corstius' tegenkomen. Als hij daarop klikt zoekt hij in feite verder met het trefwoord Hugo Brandt Corstius. Voor een illustratie, zie blz. 262.

Een andere mogelijkheid om verder te zoeken is deze. Tijdens het zoeken wordt bij elke zoekactie die treffers oplevert een zogeheten set gevormd, een deelverzameling uit de hele database. Alle records in die set voldoen aan de opgegeven zoekcriteria.

Verscheidene zoekacties achter elkaar tijdens één sessie leveren verscheidene sets op, en er zijn databases waarin die sets zolang de sessie duurt bewaard blijven. Net zoals dat met zoektermen gaat kunnen die sets dan met EN, OF en NIET worden gecombineerd.

The screenshot shows the BNTL search interface. At the top, there are navigation tabs: Zoeken, Resultaten, Geavanceerd, Mijn archief, and Help. The search bar contains 'zoeken [en]' and 'trefwoord (trf)'. The search criteria are 's5 en s6'. Below the search bar, there are radio buttons for 'Thesaurus ingangen', 'Publicaties', and 'Beide'. The user is logged in as 'BNTL'. Below the search bar, there are tabs for 'zoekgeschiedenis', 'titellijst', and 'titelgegevens'. The main content area shows an 'Overzicht zoekopdrachten' table with columns 'set', 'treffers', and 'actie'.

set	treffers	actie
▶ 7.	4	zoeken [en] s5 en s6
▶ 6.	53	zoeken [en] s3 of s4
▶ 5.	47	zoeken [en] s1 of s2
▶ 4.	16	zoeken [en] (trefwoord (trf)) Anglicismen
▶ 3.	38	zoeken [en] (alle woorden (all)) Engelse leenwoorden
▶ 2.	19	zoeken [en] (trefwoord (trf)) gallicismen
▶ 1.	28	zoeken [en] (alle woorden (all)) franse leenwoorden

U kunt sets combineren door setnummers in de commandobalk als zoekvraag op te nemen, bijvoorbeeld: 's1 en s2' of 's1 of s3'.

Voorbeeld van setcombinatie in de Bibliografie van de Nederlandse Taal- en Literatuurwetenschap (BNTL). Achtereenvolgens is gezocht met

- de termen Franse en leenwoorden in de basic index (alle woorden) → set 1
- met de term Gallicismen als trefwoord → set 2
- met de termen Engelse en leenwoorden in de basic index → set 3
- en met het trefwoord Anglicismen → set 4

Daarna zijn set 1 en set 2 in een OF-relatie met elkaar gecombineerd (→ set 5), en is hetzelfde gebeurd met set 3 en 4 (→ set 6). Ten slotte zijn set 5 en 6 in een EN-relatie gecombineerd, met als resultaat set 7, met daarin literatuurreferentie over Franse en Engelse leenwoorden (gallicismen en anglicismen).

Het kunnen combineren van sets is een geavanceerde zoekfunctie die niet door zo heel veel vrij toegankelijke databases op het Web wordt aangeboden. Waar dat wel gebeurt, zit het in de gebruikersinterface meestal onder Zoekgeschiedenis, History of een dergelijke optie. Soms wordt de zoekgeschiedenis direct op het resultatenschermbijgehouden.

5.4.3. *Records, velden, indexen, zoek sleutels*

Tot dusver heb ik het over objectbeschrijvingen en objectkenmerken gehad om het verschil met webzoekmachines te benadrukken, waar we over webpagina's/documenten en prefixen/limiters/qualifiers spreken. In databases hebben we te maken met beschrijvingen van objecten, bij (de databases van) zoekmachines om (kopieën van) de objecten zelf; alleen zijn die objecten in dit geval geen woningen of personen, maar uitsluitend elektronische documenten – teksten, plaatjes of geluidsfiles. In de taal van de databases praten we echter niet over objectbeschrijvingen en objectkenmerken, maar over records en velden. Een objectbeschrijving in een database noemen we een record, en het objectkenmerk dat onderdeel uitmaakt van zo'n beschrijving, een veld. Tegen velden kun je nog op een andere manier aankijken, en ze kunnen dan ook anders worden benoemd.

Net als bij zoekmachines worden bij databases records geïndexeerd. Indexen kun je je voorstellen als wat registers voor boeken zijn: lijsten van alle woorden die in de records van een bestand voorkomen, met een verwijzing (pointer) naar de records waarin ze voorkomen. Als een gebruiker in een simpele zoekregel een term intypt, zoekt de software als het ware die term eerst in het register op en slaat dan de bijbehorende pagina's open; dat gaat veel sneller dan wanneer het hele bestand op de zoekterm zou moeten worden doorgelopen. Bij twee zoektermen in een EN-relatie gaat het relatief nog sneller: nu hoeven alleen de pointers van zoekterm a en zoekterm b in het register maar even te worden vergeleken en dan alleen de pagina's (records) opgeslagen die overeenkomen.

In werkelijkheid gebeuren er veel ingewikkelder zaken, maar het gaat er hier maar om dat u zich een voorstelling kunt maken van wat er ongeveer gebeurt. Als we de vergelijking met het register van een boek nog even volhouden, kunt u zich voorstellen dat er verschillende soorten registers zijn. Een geschiedenisboek bijvoorbeeld kan een algemeen register bevatten, waarin de beschreven gebeurtenissen geïndexeerd zijn, maar daarnaast ook een register van alleen persoonsnamen. Zo kun je de bladzijden die in het boek aan de slag bij Waterloo zijn gewijd niet alleen in het algemene register onder 'Waterloo, slag bij' vinden, maar ook in het persoonsregister onder Wellington.

Achter wat u in een objectbeschrijving (record) als een objectkenmerk (veld) ziet, gaat in feite zo'n speciaal register, een veldindex schuil, een lijst van de termen die dat speciale kenmerk beschrijven. Naast het algemene register van een database, dat alle woorden van alle records in de database bevat, bestaan er speciale registers voor de afzonderlijke velden die alle woorden (en in sommige gevallen phrases, zie hieronder) in die velden bevatten. Het algemene register noemen we 'basic index', de speciale registers veldindexen. Deze komen dan weer overeen met wat u in databaseprogramma's als Access als tabellen kent.

Wat is het voordeel van verschillende indexen en velden boven één enkele index voor het hele record? Dat is vanzelfsprekend het onderscheidend vermogen. Stel u een database van persoonsbeschrijvingen voor waarin de namen van personen, de plaats waar ze wonen en de kleur van hun ogen vermeld staan. Niet alle personen die Hoozevee heten wonen in Hoozevee, en niet alle personen met de oogkleur blauw heten Blauw. Het is dus verstandig aparte indexen voor namen, woonplaatsen en oogkleuren te hebben.

Ook hier is de werkelijkheid meestal wat ingewikkelder dan hier is geschetst. Vaak komen de velden die u in een record ziet, niet overeen met één index, maar met verschillende indexen die meer gedetailleerde objectkenmerken 'registeren'. De voorloopteksten in records duiden aan om welk veld en welke index(en) het gaat. Soms zijn ze lang: Auteur: Titel: etc., soms kort: AU:, TI: enz. In het laatste geval komen ze soms overeen met de veld aanduidingen of veldlabels die u in de commandoversie van een gebruikersinterface gebruikt om aan te duiden in welk(e) veld(index) een term gezocht moet worden: AU=hermans AND TI=kamer. In een record bekeken wordt een index (of een combinatie van indexen) een veld. In een zoekformulier (of een daarmee overeenkomend afrolmenu in een Eenvoudig zoeken-scherm) is hetzelfde niet een veld of een index maar een zoeksleutel. We hebben dus ook de auteursleutel, de titelsleutel etc. als het om een bibliotheekcatalogus gaat.

5.4.4. Termen en operatoren

Als u in databases zoekt (en trouwens ook als u dat met een zoekmachine op het Web doet), doet u dat met zoektermen waarvan u hoopt dat ze overeenkomen ('matchen') met termen die in een of meer records in de database (of de documenten op het Web) voorkomen. De overeenkomst tussen zoeken op het Web en zoeken in databases is dat u in beide gevallen naar 'matches' zoekt. Het verschil is het verschil tussen de structuur en de inhoud van een document en de structuur en de inhoud van zijn beschrijving. Net als bij webzoekmachines kunnen zoektermen en velden in databases met elkaar worden verbonden door AND, OR en NOT. Ook hier is AND meestal de standaardrelatie die zowel zoektermen als (in Geavanceerd zoeken-schermen) de diverse velden met elkaar verbindt. Nesting – het combineren van EN- en OF-relaties in één zoekvraag – is meestal wel mogelijk. De geavanceerde structuur van databaserecords en de hoge eisen die aan het vinden van de juiste records in databases worden gesteld brengen mee dat de OF-relatie een wat belangrijker rol in databasezoeken vervult dan in webzoeken.

5.4.5. Trunceren en maskeren

Dat blijkt ook uit het feit dat de meeste databases trunceren en maskeren van zoektermen toelaten. Trunceren is het vervangen van een (of nul) of meer tekens in een woord of getal door een joker ('wild card'). Dat kan zich in drie vormen voordoen: rechts (dit komt het meeste voor), links en intern. Voorbeelden van getrunceerde termen zijn: psych* (voor psyche, psychiater, psychiatrie, psycholoog, psychologie enz.) en 199* (voor bijvoorbeeld 1990-1999), *kanker (voor darmkanker, longkanker, mondkanker enz.) en arch*logy (voor archeology, archaeology). Een voorbeeld van maskeren is organi?ation, waarbij het vraagteken één letter (in dit geval s of z) vervangt. De tekens die voor trunceren en maskeren worden gebruikt verschillen per database. Veel voorkomend zijn: *, ?, \$, # en %. Raadpleeg de Help van een database om te zien of trunceren en maskeren mogelijk zijn, wat de gebruikte tekens zijn en – ook belangrijk – voor hoeveel tekens die staan. Het verschil tussen een teken dat voor één letter staat en een teken dat voor 0 of 1 teken staat kan het verschil tussen vinden en niet vinden uitmaken.

In sommige databases hoeft u zich over trunceren geen zorgen te maken. Daar worden (in alle of alleen in bepaalde velden) rechtse truncaties automatisch uitgevoerd. Als dat gebeurt ziet u dat vanzelf wel aan het resultaat: hebt u met de term jacht gezocht en treffers met de term jachttafereel gevonden, dan heeft de software u geholpen door automatisch te trunceren.

Trunceren en maskeren zijn manieren om heel snel een aantal termen aan een vraag toe te voegen in een OF-relatie met elkaar verbonden zijn. Wat ze in feite doen is een vraagstelling verbreden. Ze verhogen de recall (zie blz. 75) van een zoekactie.

5.4.6. Nabijheidsoperatoren

Een verdere uitwerking van de EN-operator vormen de zogeheten nabijheidsoperatoren, die de maximale afstand tussen zoektermen in een EN-relatie specificeren. Daarbij kan dan nog onderscheid worden gemaakt tussen een nabijheidsoperator die ook de volgorde van de zoektermen specificiert (logica ADJ/3 wetten) en een nabijheidsoperator die dat niet doet (wetten NEAR/3 logica). De vorm en de precieze naam van de nabijheidsoperatoren verschillen per database. Naast NEAR en ADJ (voor adjacent) komen ook louter tekens (/3 enz.) en andere nabijheidsoperatoren als SAME en WITH voor. Raadpleeg de Help van een te doorzoeken database om te zien of zoeken met nabijheidsoperatoren mogelijk is, en zo ja, hoe dat dan moet.

De meest extreme vorm van nabijheidszoeken is natuurlijk het gebruik van de phrase of woordgroep. Zoeken hiernaar wordt in databases meestal wel mogelijk gemaakt, hetzij met nabijheidsoperatoren (bijvoorbeeld NEAR sec), hetzij door middel van aanhalingstekens of door een aan te kruisen keuzevakje.

Phrases en nabijheidsoperatoren leggen verbanden tussen zoektermen die het resultaat van een zoekactie verkleinen. Ze zijn een middel om de precision (zie blz. 75) van een zoekactie te verhogen.

5.4.7. Indexen en thesaurussen

Het kunnen bekijken van indexen in het algemeen is een belangrijk hulpmiddel bij het zoeken in databases omdat het een idee geeft van de alfabetische context van de termen waarmee u overweegt te gaan zoeken, en ook van het aantal malen dat die termen in het bestand voorkomen. Twee speciale doeleinden waarvoor u indexen kunt gebruiken zijn het opsporen van een exacte naam of code en het vinden van de termen die in een bepaalde zoekleutel worden geaccepteerd. Twee simpele voorbeelden uit een van de databases van het Rijksbureau voor Kunsthistorische Documentatie illustreren dit:

The screenshot shows the RKD search interface. At the top, there are navigation tabs: RKD collections & archives, RKD artists&, RKD images, RKD portraits, and RKD library. The main search area has a search bar with the text 'Roelofs' entered. Below the search bar are several filter options: Nationaliteit/school, Geboorteplaats, Sterfplaats, Plaats van werkzaamheid, Periode geboortedatum, Periode sterfdatum, Periode van werkzaamheid, Kwalificatie, and Onderwerp. A dropdown menu is open, displaying a list of artist names under the heading 'Kunstenaarsnamen'. The list includes: Roelofs, Adèle; Roelofs, Albert; Roelofs, Otto Willem Albertus; Roelofs, Alida; Bouwsema-Roelofs, Alida; Roelofs, Alide; Bouwsema-Roelofs, Alide; Roelofs, Anne; Roelofs, Anton (1922-1990); Zutphen, A. van; Roelofs, Antonie Christiaan; and Roelofs, Antonie Christiaan Roelofs, Anton (1922-1990). The interface also features a 'Zoek' button and a navigation bar at the bottom with links for 'wis', 'Uitgebreid zoeken', and 'Help'.

In dit voorbeeld was de opdracht: vind informatie over de kunstenaar Roelofs in de Artists database van het RKD. Omdat geen voornaam bekend was en het pictogram van een lijstje rechts van de zoekleutel Naam op een index wijst is eerst de naam Roelofs in de zoekregel ingetypt en daarna op het pictogram geklikt. In een apart venstertje rechtsonder ziet u nu het gedeelte van de index van kunstenaarsnamen in de omgeving van de naam Roelofs. Daaruit blijkt dat er heel wat Roelofsen geweest zijn, en het lijstje dat we nu zien, zal ons hopelijk helpen degene die we op het oog hebben te selecteren. Hieronder kom ik nog op dit voorbeeld terug, maar hier vestig ik uw aandacht even op het feit dat in deze database niet van elk veld een index te raadplegen is; achter de zoekleutel Naamsdeel (bijvoorbeeld: van) staat geen pictogram. Het andere voorbeeld, in dezelfde database, betreft het raadplegen van een index om te zien welke termen daarin eigenlijk worden geaccepteerd:

Rijksbureau voor Kunsthistorische Documentatie - databases - Mozilla Firefox

File Edit View Go Bookmarks Yahoo! Tools Help

http://www.rkd.nl/rkddb/default.asp?databa

RKD collections & archives RKD artists RKD images RKD portraits RKD library

Vul zoektermen in

Naam

Naamsdeel

Nationaliteit/school

Geboorteplaats

Sterfplaats

Plaats van werkzaamheid

Periode geboortedatum

Periode sterfdatum

Periode van werkzaamheid

Kwalificatie

Onderwerp

Zoek wis Uitgebreid zoeken Help

http://www.rkd.nl - Rijksb...

Periode

<input checked="" type="checkbox"/>	1600-1699	
<input checked="" type="checkbox"/>	1625-1649	
<input checked="" type="checkbox"/>	1650-1674	
<input checked="" type="checkbox"/>	1650-1699	
<input checked="" type="checkbox"/>	1675-1699	
<input checked="" type="checkbox"/>	1700-1724	
<input checked="" type="checkbox"/>	1700-1749	
<input checked="" type="checkbox"/>	1700-1799	
<input checked="" type="checkbox"/>	1725-1749	
<input checked="" type="checkbox"/>	175-199	
<input checked="" type="checkbox"/>	1750-1774	
<input checked="" type="checkbox"/>	1750-1799	
<input checked="" type="checkbox"/>	1775-1799	
<input checked="" type="checkbox"/>	1800-1824	
<input checked="" type="checkbox"/>	1800-1849	
<input checked="" type="checkbox"/>	1800-1899	
<input checked="" type="checkbox"/>	1825-1849	
<input checked="" type="checkbox"/>	1850-1874	
<input checked="" type="checkbox"/>	1850-1899	
<input checked="" type="checkbox"/>	1875-1899	

Hier was de vraag: zoek informatie over een schilder met de naam Roelofs die in 1874 geboren is. Bij de zoekleutel Periode geboortedatum staat 1874 ingevuld, maar omdat deze combinatie niets opleverde, is de index van dit veld opgeroepen, waar blijkt dat de periode die hier had moeten worden ingevuld 1850-1874 is. Als we die term met het pijltje naar links naar de zoekregel overbrengen, levert de zoekactie Naam: Roelofs en Periode geboortedatum wel een resultaat op van twee schilders met die naam die in 1874 geboren zijn.

Zoals uit het hierboven gegeven voorbeeld van de Periode geboortedatum blijkt, hoeft het feit dat een veld geïndexeerd is, niet te betekenen dat het woord voor woord geïndexeerd is. In het geval van termen als 'Derde Wereld', 'hoger onderwijs' is het evident dat er ook sprake kan zijn van 'phrase-indexering'. Anders dan bij phrase-indexering door zoekmachines op het Web gaat het daarbij doorgaans echter niet om de indexering van willekeurige stukken tekst, zoals dichtregels of citaten, maar om die van speciale velden, bijvoorbeeld auteursnamen of trefwoorden, in een beschrijvend record. Zulke velden bevatten samengestelde termen – 'Hermans, Willem Frederik', 'syndroom van Gilles de la Tourette' – die een persoon of onderwerp in een vaste vorm identificeren.

In het algemeen kan men onderscheid maken tussen drie vormen van indexering: woord-indexering, phrase-indexering en veld-indexering. In het laatste geval is de indexterm de hele inhoud van een veld.

Deze drie vormen van indexering sluiten elkaar niet uit; zo is het heel goed mogelijk dat termen die in het trefwoordenveld phrase-geïndexeerd zijn, in de basic index woord-geïndexeerd zijn. In het trefwoordenveld vind je het trefwoord 'Derde Wereld' dan alleen als 'Derde Wereld', terwijl je records met datzelfde trefwoord met Zoeken in alle velden (d.w.z. in de basic index) met derde EN wereld vindt.

In het laatste voorbeeld bevat de index van het veld Periode geboortedatum niet de werkelijke geboortedata van de kunstenaars, maar zogeheten 'genormaliseerde' data. De databasebouwer heeft er hier – terecht – voor gekozen gebruikers niet met een enkel, wellicht foutief, jaartal te laten zoeken maar meerdere jaartallen in periodes onder te brengen om de 'vangst' van zoekacties te verhogen. We komen hier op een belangrijk aspect van veel databases, dat van het gecontroleerde vocabulaire en van thesaurussen. De periode 1850-1874 voor alle tussenliggende jaartallen die gebruikers voor het geboortjaar van een kunstenaar zouden willen invullen is een voorbeeld van een gecontroleerd vocabulaire: alleen met deze term: 1850-1874 is het gezochte te vinden. Zo'n gecontroleerd vocabulaire kan behalve voor jaartallen ook voor taal bestaan, bijvoorbeeld als alle records in een database die over ouderdomsdementie, dementerende bejaarden, de ziekte van Alzheimer gaan, het trefwoord 'ziekte van Alzheimer' hebben. Dat trefwoord maakt dan deel uit van een gecontroleerd vocabulaire en heet een 'voorkeursterm'. Een ander voorbeeld van zo'n voorkeursterm is in een onderwijsdatabase de term adolescenten voor alle 12-18-jarigen.

Gecontroleerde of voorkeurstermen (het bekendste voorbeeld zijn trefwoorden) worden aan records toegevoegd om gebruikers in staat te stellen met één zoekterm mede alle records te vinden die die zoekterm in de formele objectbeschrijving niet bevatten, maar wel een synoniem van die term. Ze kunnen in diverse velden worden gebruikt, en in feite hebben we er hierboven al een voorbeeld van gezien: in dat van de kunstenaar met de naam Roelofs over wie informatie moest worden gevonden. Als u naar de bijbehorende illustratie terugbladert, zult u zien dat in de index van namen van Kunstenaars die daar in het venster getoond wordt, meermalen twee, soms drie namen bij elkaar zijn geplaatst. Een of twee daarvan worden voorafgegaan door een pictogram dat op een inrijverbodsbord lijkt, maar per groepje is er ook altijd één naam die zo'n

pictogram niet heeft. Dat is in deze database de genormaliseerde naam, waaronder alle records over deze schilder gevonden worden, ook al heeft hij soms andere namen gebruikt of staat hij ook onder andere naamsvormen bekend.

Een zoekleutel die zich bij uitstek voor toepassing van een gecontroleerd vocabulaire leent is Onderwerp. Hieronder ziet u een trefwoordenlijst die is opgeroepen bij de term jacht (om te kijken of jachtaferelen een goede onderwerpsterm zou zijn):

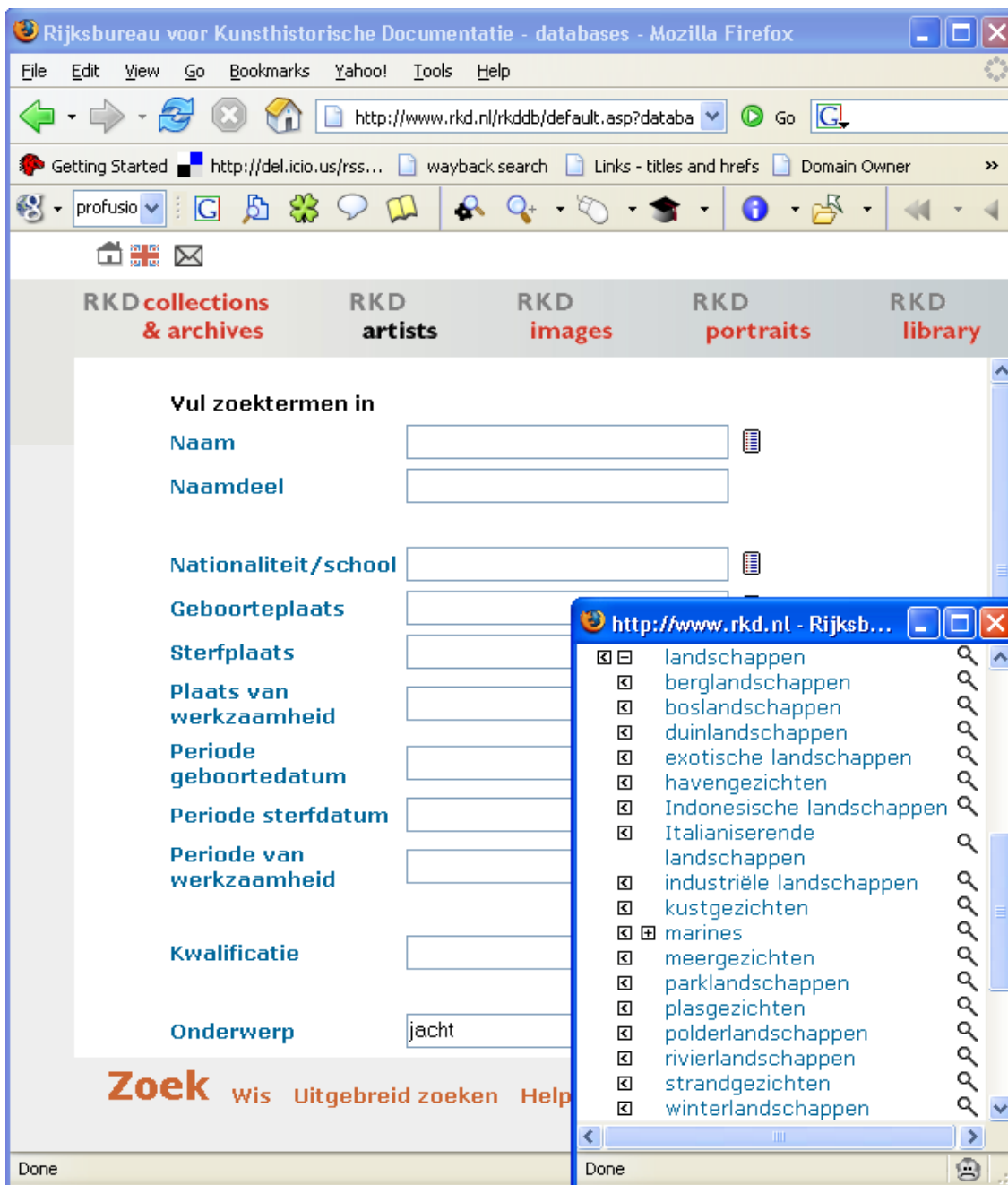
The screenshot shows a Mozilla Firefox browser window displaying the website 'Rijksbureau voor Kunsthistorische Documentatie - databases'. The search interface includes a search bar with the text 'jacht' and a dropdown menu titled 'Trefwoordenlijst' (Keyword list). The dropdown menu contains the following terms:

- jachtstillevens
- jachtvoorstellingen
- joods genre
- karikaturen
- katten
- poezen
- kerkinterieurs
- keukenstillevens
- kinderportretten
- klassiek genre
- koeien
- koppen (geen portretten)
- kozakken
- kunstgalerijen
- kustgezichten
- landschappen
- literatuur
- marines
- marktscènes
- marktstillevens

U ziet dat het antwoord nee is: de twee toegelaten termen zijn jachtstillevens en jachtvoorstellingen, maar u ziet ook hier een voorbeeld van een gecontroleerde term: poezen, met als niet-geprefereerde variant: katten.

Voor drie van de trefwoorden in dit lijstje (bijvoorbeeld: landschappen) ziet u een plusje staan, en dit brengt mij op een verdere verfijning van het fenomeen gecontroleerd vocabulaire. Heel wat databases beschikken niet alleen over zo'n gecontroleerd vocabulaire, maar ook over een thesaurus waarin de relaties tussen termen

geëxpliciteerd zijn. Een synoniem voor de term fiets is rijwiel, en dat kan de voorkeursterm zijn; maar er bestaan ook nog termen als tweewieler en vouwfiets, waarvan de eerste een meer algemene term is die het begrip rijwiel mede omvat, en de tweede een meer specifieke term die een bepaald type fiets aanduidt. In een thesaurus is het hele vocabulaire van trefwoorden voor een database ondergebracht, inclusief verwijzingen naar ruimere (broader, BT), specifiekere (narrower, NT) en verwante (related, RT) termen. Met het laatste worden geen synoniemen bedoeld, maar termen die tot een zelfde betekeniscluster behoren: voor tweewieler bijvoorbeeld driewieler en vierwieler; of een antoniem, zoals huilen voor lachen. In het voorbeeld van de RKD-trefwoorden hierboven vinden we alleen narrower terms, maar niettemin is dit ook een voorbeeld van een thesaurus:



De vorige afbeelding, maar nu met de thesaurus op landschappen 'opengeslagen'. Zoals u ziet zijn er heel wat 'narrower terms'.

Thesaurussen in databases kunnen een grote hulp zijn bij het zoeken. Wel moet u bedenken dat trefwoorden altijd maar op één niveau aan records worden toegevoegd: het is niet zo dat een record met het trefwoord boslandschappen ook nog eens het trefwoord landschappen meekrijgt; dat betekent dus dat u, als u een thesaurus gebruikt, zich goed rekenschap moet geven op welk conceptueel niveau u zoekt: als u naar boslandschappen zoekt, zult u daarbij de landschappen niet vinden, en andersom. Indexen, als ze kunnen worden gebruikt, worden meestal bij een zoekscherm (eenvoudig of geavanceerd) aangeboden; soms wordt browsen of zoeken via een index als een aparte zoekmogelijkheid gepresenteerd.

Thesaurussen vind je zelden zo in de gebruikersinterface verwerkt als we dat bij de RKD-database hebben gezien. Als er een thesaurus wordt gepresenteerd gebeurt dat meestal apart, bijvoorbeeld onder een tab thesaurus. Een andere mogelijkheid is dat in de resultatenlijst van een zoekactie op trefwoorden ook thesaurusrecords verschijnen, waarin een omschrijving van de term, eventuele synoniemen ('Used For'), en broader, narrower en related terms worden vermeld.

Dat laatste kan ook gebeuren als men in de complete presentatie van een record de link onder een trefwoord volgt. Bijzonder fraai is dat in de RKD-database geregeld: klik je hier op een term dan vind je daarbij een lijstje met de de gerelateerde thesaurustermen; klik je daarentegen op het loeppictogram achter de term, dan doorzoekt de software de database op die term en toont van die zoekactie de resultaten:

The screenshot shows the RKD database search results for 'Vermeer, Johannes'. The search criteria are 'Naam: Vermeer, Johannes' and 'Onderwerp: de database RKD artists' (1 hits). The results table lists various attributes such as 'Naam', 'Naamsvariant', 'Man/vrouw/groep', 'Nationaliteit/school', 'Geboorteplaats/datum', 'Sterfplaats/datum', 'Plaats/periode van werkzaamheid', 'Kwalificatie', and 'Onderwerp'. A pop-up window titled 'Thesaurus' is overlaid on the results, showing the term 'stadsgezichten' with its domain and description. The thesaurus window also lists related terms: 'architectuur (als genre)', 'vedute', 'dorpsgezichten', and 'topografische voorstellingen'.

Term	Domein	Beschrijving
stadsgezichten	onderwerpstype	Uitbeeldingen van een stad of deel van een stad (straat/straten, plein/pleinen en/of gebouw/gebouwen). De term stadsgezichten kan worden gebruikt voor voorstellingen van bekende en onbekende steden, straten, pleinen en gebouwen. Bekende stadsgezichten vallen ook onder topografische voorstellingen. (G. Krebber, mei 2003)
architectuur (als genre)		
vedute		
dorpsgezichten		
topografische voorstellingen		

Other related terms listed in the thesaurus window:

- Ruimere term
- Nauwere term
- Verwante term

Als resultaat van doorklikken op de term stadsgezichten in dit record over Johannes Vermeer in de RKD Artists database verschijnt een venster met uitleg over de term en ruimere, specifiekere en verwante termen.

Thesaurussen en indexen kunnen termen bevatten die niet in de database voorkomen waar zo'n thesaurus of index bij hoort. Dat komt doordat met name van thesaurussen standaardedities bestaan, die door databases worden gebruikt, liever dan dat men voor zo'n database een eigen thesaurus maakt. En aangezien bij elk nieuw record dat aan de database wordt toegevoegd de kans bestaat dat het 't eerste record is waaraan de thesaurusterm wordt toegekend, neemt men ook niet de moeite thesaurustermen die geen bijbehorende treffers hebben uit de database te verwijderen. De diverse thesaurussen van de RKD-databases zijn daar een voorbeeld van.

The screenshot shows the ERLWebSPIRS 5 search interface. At the top, there is a navigation bar with 'Search | Database Guide | Help | Logout'. Below this, a search bar contains 'PsychINFO(Weekly 2005/08 Week 1,2005/07,2005 Part B,2005 Part A,2004 Part B,2004 Part ...' and a 'Change Database(s)' button. A menu bar includes 'Search', 'Advanced', 'Thesaurus', 'Index', 'Find Citation', 'Search History', and 'How do I...?'. The main area is divided into two columns. The left column has a 'Term:' input field and a 'Go To Terms' button. The right column is titled 'Limit Search To:' and contains several checkboxes: 'Publication Year greater than 1997', 'Records with a URL field', 'PsycArticles Full Text', and 'All Journals'. There are 'More' and 'Clear' buttons below these options. At the bottom of the interface, there are buttons for 'Change Display', 'Clear Marked', 'Search Marked', 'Explode Marked', and 'Back'.

Subject: Anger

SCOPE NOTE: Year term introduced: 1967.

USED FOR: Rage

To search for the subject, or one or more terms, select the check boxes next to each one then click **Search Marked**.
Click on a term to see more information.

Subject

Anger

More Specific (narrower) terms

[Hostility](#)

Less Specific (broader) terms

[Emotional States](#)

Related terms

[Anger Control](#)

[Hate](#)

[Jealousy](#)

[Tantrums](#)

Een thesaurusrecord in PsycInfo, een database van literatuur over de psychologie. Het record geeft ook informatie over specifiekere, ruimere en verwante termen en biedt gebruikers de mogelijkheid de term zelf en onderliggende en verwante termen aan te vinken en daarmee vervolgens via Search Marked mee te zoeken. Search Marked betekent: zoeken met de aangevinkte termen, Explode Marked betekent: neem in de zoekactie ook termen mee die nog onder de aangevinkte termen verborgen liggen. 'Hostility', als meer specifieke term van 'anger', is daar een voorbeeld van.

5.4.8. 'Explode'-functie in thesaurussen en classificatiesystemen

In het voorbeeld van de RKD-database hierboven kwam al even ter sprake dat wie daar met de term landschappen zoekt, geen boslandschappen zal vinden. De simpele reden is dat objectbeschrijvingen maar met een beperkt aantal trefwoorden kunnen worden verrijkt en dat niet de hele hiërarchische structuur van een thesaurus aan ieder individueel record kan worden toegevoegd. In het geval van de vouwfiets zou dat immers betekenen dat aan een beschrijving die daarover ging ook de termen fiets, tweewieler, wielvoertuig en transportmiddel zouden moeten worden toegevoegd. Toch zijn er wel geavanceerde zoeksystemen die dit probleem oplossen: als u een term in de thesaurus gevonden hebt, laten ze u ook de narrower terms zien en daaruit die termen selecteren die u in uw zoekactie naar de bredere term wilt meenemen. Eventueel bestaande nog specifiekere termen kunnen ook in de zoekactie worden meegenomen via de zogeheten Explode- of Expand-functie (zie de illustratie op de linker bladzijde). Diezelfde functie – de mogelijkheid om sets te vormen die niet alleen de algemene zoekterm dekken maar ook alle termen die daaronder liggen, dus specifiek zijn – wordt ook aangeboden door classificatiesystemen. Daarbij wordt het hele gecontroleerde vocabulaire, inclusief de zie-verwijzingen vanuit de niet geprefereerde termen (katten, zie poezen) en de hiërarchische relaties tussen termen, in één systeem ondergebracht dat elke term, behalve die op het hoogste, meest algemene niveau, in een complete boomstructuur van termen onderbrengt die het hele vakgebied of onderwerp van een database dekt. Je kunt het vergelijken met wat directories zoals Yahoo! of ODP doen voor de websites die ze dekken, met als belangrijk verschil dat in de algemenere categorieën tevens de meer specifieke categorieën zijn inbegrepen. Simpel gezegd, aan de hand van de voorbeelden die ik hierboven heb gegeven: als je op landschappen zoekt vind je ook treffers met het trefwoord boslandschappen, en als je op tweewieler zoekt vind je zowel de fiets als de vouwfiets.

Een voorbeeld hiervan biedt de BL Online, een bibliografie op het gebied van de taalkunde die bij de Koninklijke Bibliotheek wordt bijgehouden. In onderstaande illustratie laat ik eerst een deel van het onderwerpsclassificatieschema zien:

- 9.4. Neurolinguistics and language disorders
 - 9.4.1. Neurolinguistics
 - 9.4.2. Language disorders
 - 9.4.2.1. Disorders of language development
 - 9.4.2.2. Aphasia
 - 9.4.2.3. Other language disorders
- 10. Sociolinguistics and dialectology
 - 10.1. Sociolinguistics
 - 10.1.1. Language attitudes
 - 10.1.2. Language policy and language planning
 - 10.1.3. Language for special purposes
 - 10.1.4. Language loss and maintenance
 - 10.2. Multilingualism, language contact
 - 10.2.1. Multilingualism
 - 10.2.2. Diglossia
 - 10.2.3. Language contact
 - 10.3. Linguistic geography
 - 10.4. Dialectology
 - 10.4.1. Dialectal lexicon
- 11. Historical and comparative linguistics
 - 11.1. Historical linguistics, language change and comparative philology
 - 11.2. Linguistic typology, universals of language
- 12. Mathematical and computational linguistics

Het hele schema omvat 13 categorieën, met daaronder 103 'subjects'. Deze zijn ondergebracht in een schema waarin de codes de hiërarchische relaties tussen categorieën en onderwerpen duidelijk maken. Zo is het subject *Dialectical lexicon* (10.4.1.) een sub-onderwerp van het onderwerp *Dialectology* (10.4) en behoort dit weer tot de algemenere categorie *Sociolinguistics and dialectology* (10). Deze classificatie maakt het bij deze database nu mogelijk een woordenboek van het Fries niet alleen op de zoektermen *Frisian* (als taal) en *Dialectical lexicon* (onderwerp) te vinden, maar ook in de combinatie van *Frisian* en *Dialectology* en in die van *Frisian* en *Sociolinguistics and dialectology*:

Searched: (object language = 'Frisian') and (subject heading = '*Dialectal lexicon*') and (publ.type = 'Books or Articles or Reviews') and (publ.year = all years).

Found: 11 **Select page:** [1](#) [New Search](#) [Go back](#)

Most recent publication on top

[B](#) [Visser, Willem](#) : [Dyk, Siebren](#) : Eilander wezzenbúek = Woordenboek van het Schiermonnikoogs --- Ljouwert : Fryske Akad., 2002, xliii, 844 p.. - (Fryske Akad., 934) , (Rige Fryske dialektstúdzjes, 10)

[B](#) [Perdok, Els](#) : Woordenboek Nederlands-Schiermonnikoogs --- Schiermonnikoog : Cultuurhistorische Vereniging 't Heer en Feer, 2001, ix, 410 p.

[B](#) [Duijff, Pieter](#) : Wurdlisten fan'e Fryske stedsdialekten (Transl: Wordlists of the Frisian city-dialects) --- Ljouwert : Fryske Akad., 1998, 177 p.. - (Fryske Akad., 860) , (Rige Fryske dialektstúdzjes, 8)

[A](#) [Fort, Marron C.](#) : Das romanische Lehnwort im Saterfriesischen --- NOWELE, 28-29, 1996, 141-149

Searched: (object language = 'Frisian') and (subject heading = '*Dialectology*') and (publ.type = 'Books or Articles or Reviews') and (publ.year = all years).

Found: 41 **Select page:** [1](#) [New Search](#) [Go back](#)

Most recent publication on top

[B](#) [Visser, Willem](#) : [Dyk, Siebren](#) : Eilander wezzenbúek = Woordenboek van het Schiermonnikoogs --- Ljouwert : Fryske Akad., 2002, xliii, 844 p.. - (Fryske Akad., 934) , (Rige Fryske dialektstúdzjes, 10)

[A](#) [Versloot, Arjen P.](#) : Die Wangeroogischen Diminutivendungen -*uuk* und -*ijk* --- AB&G, 56, 2002, 223-231

[B](#) [Perdok, Els](#) : Woordenboek Nederlands-Schiermonnikoogs --- Schiermonnikoog : Cultuurhistorische Vereniging 't Heer en Feer, 2001, ix, 410 p.

[A](#) [Kathol, Andreas](#) : Syntactic categories and positional shape alternations --- JCGL, 3/2, 2000 (2001), 59-96 | On inflected complementizers in German, Dutch and Frisian dialects

Searched: (object language = 'Frisian') and (subject heading = '*Sociolinguistics and dialectology*') and (publ.type = 'Books or Articles or Reviews') and (publ.year = all years).

Found: 128 **Select page:** [1](#) [2](#) [3](#) [New Search](#) [Go back](#)

Most recent publication on top

[A](#) [Versloot, Arjen P.](#) : Fries in Holland in de 17e eeuw --- TeT, 55/1, 2003, 1-40, 1 map | E. ab.

[B](#) [Visser, Willem](#) : [Dyk, Siebren](#) : Eilander wezzenbúek = Woordenboek van het Schiermonnikoogs --- Ljouwert : Fryske Akad., 2002, xliii, 844 p.. - (Fryske Akad., 934) , (Rige Fryske dialektstúdzjes, 10)

[A](#) [Versloot, Arjen P.](#) : Die Wangeroogischen Diminutivendungen -*uuk* und -*ijk* --- AB&G, 56, 2002, 223-231

[A](#) [Bree, Cor van](#) : Een fries substraat in de drentse woordenschat? --- Acta Neerlandica : bijdragen tot de Neerlandistiek : Debrecen, 2, 2002, 283-298

Zoals u in de illustraties hierboven ziet wordt het zoekresultaat steeds groter naarmate het subonderwerp *Dialectical lexicon* wordt verbreed, eerst tot *Dialectology* en daarna tot *Sociolinguistics and dialectology*, maar komt het woordenboek van het Schiermonnikoogs van Visser en Dyk in alle drie de zoekresultaten voor. Niettemin heeft het alleen het trefwoord *Dialectical lexicon* (= de code 10.4.1.) gekregen:

[New Search](#) [Go back](#)

Publication	Book
Classific.	Modern Frisian / Dialectal lexicon
Author	Visser, Willem
Author	Dyk, Siebren
Title (Frisian)	Eilander wezzenbúek
Title (Dutch)	Woordenboek van het Schiermonnikoogs
City	Ljouwert
Publisher	Fryske Akad.
Year	2002
Pages	xliii, 844 p.
Series	Fryske Akad. , 934
Series	Rige Fryske dialektstúdzjes , 10

De simpele notatie:

1

...

10

10.1

...

10.4

10.4.1.

maakt het dus mogelijk alle onderwerpen die onder 10 vallen in de zoekactie mee te nemen. In een commandogestuurde zoekactie zou ditzelfde worden bereikt door simpelweg achter 10. te trunceren: 10.*.

In de interface van de BL Online zitten de diverse classificatieschema's – er is er ook een voor de talen – verstopt onder Lists, samen met de genormaliseerde lijsten van (afkortingen van) talen waarin publicaties geschreven zijn, tijdschriften en uitgevers. Om ze te kunnen gebruiken moeten ze als zoek sleutel worden gekozen, waarna met 'Get List' een classificatieschema kan worden opgeroepen. Ook hier geldt weer dat de functie in de diverse interfaces op heel verschillende manieren gepresenteerd kan worden. Daarbij komt dat de Help-informatie vaak uiterst summier is. Voor alles wat niet woordgeïndexeerd is, met een gecontroleerd vocabulaire of een thesaurus werkt, of een classificatie gebruikt, geldt eigenlijk: ga ernaar op zoek als u in een database geen resultaten krijgt waar u die wel verwacht te krijgen.

5.4.9. De presentatie van zoekresultaten

De meeste databases presenteren hun zoekresultaten in minimaal twee formaten: kort, in de vorm van een lijst, en lang, voor de beschrijving van één object. Vanzelfsprekend

 Titel:	Battus/Piet Grijs / [door] Johan Diepstraten
Auteurs:	Johannes Josephus Maria Diepstraten (1951-1999)
In:	Diepstraten, Johan. De kunst van het schrijven
Uitgave:	Arnhem : Ellessy
Jaar:	2000
Omvang:	p. 63-65
Syst. code:	2.6.4.9.2 - per auteur - vanaf 1830 - beschrijvende studies - literatuurwetenschap 0.3.9 - besprekingsartikelen en recensies over meerdere werken
Trefwoorden:	Brandt Corstius, Hugo

De lange presentatie van een titel in de BNTL (zie blz. 250). Het record is gevonden op de titelwoorden Piet [EN] Grijs. De links in het record (alles wat blauw is weergegeven) voeren naar andere records met hetzelfde kenmerk, bijvoorbeeld andere publicaties van J.J.M. Diepstraten en andere publicaties over Piet Grijs (onder trefwoord: Brandt Corstius, Hugo).

voert de korte presentatie naar de lange en is er in de lange presentatie ook een link terug naar de lijst, maar ook hier laat de interface weer alle mogelijke varianten tussen onvindbaar en vanzelfsprekend zien. In de complete presentatie van een record kunnen – naast de menukeuzes – weer diverse links zichtbaar zijn. Die kunnen u dan naar andere objecten van dezelfde auteur (kunstenaar, uitvinder, verkoper op een veiling enz.) leiden, of naar andere objecten (kunstwerken, uitvindingen, artikelen enz.) van hetzelfde type.

5.5. Tot slot

Hierboven heb ik nogal wat aandacht besteed aan het deel van onzichtbare Web dat wel gratis openbaar toegankelijk is, maar niet of slechts ten dele door de grote mondiale zoekmachines ontsloten. Naast dat deel bestaat nog een uitgestrekt terrein van het onzichtbare Web dat alleen tegen betaling toegankelijk is. De zoekmachine Google Scholar ontsluit dat terrein voor een deel door ons toegang te geven tot abstracts, preprints etc. Maar de complete, officieel gepubliceerde literatuur zit meestal achter slot en grendel, en daartoe krijgen we slechts toegang als er betaald wordt. Bij de bespreking van Google Scholar heb ik al aangegeven hoe u dat probleem op een simpele en goederkope manier kunt oplossen: door u in te schrijven bij een grote wetenschappelijke bibliotheek krijgt u niet allen toegang tot het zichtbare en grote delen van het onzichtbare Web, maar ook nog eens tot de wereld van het gedrukte boek en de niet geringe kennis waarmee professionals u bij het zoeken naar informatie over welk onderwerp ook ter zijde willen staan.

Literatuur: The invisible web: uncovering sources search engines can't see

(http://www.findarticles.com/p/articles/mi_m1387/is_2_52/ai_112542832)

The Deep Web: Surfacing Hidden Value (<http://www.press.umich.edu/jep/07-01/bergman.html>)

Deep Web FAQ (http://brightplanet.com/deepcontent/deep_web_faq.asp)

Invisible Web: What it is, Why it exists, How to find it, and Its inherent ambiguity

(<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html>)

Deep Web Research Subject Tracer™ Information Blog (<http://www.deepwebresearch.info/>)

The Deep Web (<http://library.albany.edu/internet/deepweb.html>)

Those Dark Hiding Places: The Invisible Web Revealed

(http://library.rider.edu/scholarly/rlackie/Invisible/Inv_Web.html)

OSTI Communications: Deep Web Video (<http://www.osti.gov/media/DeepWebVideo.html>)

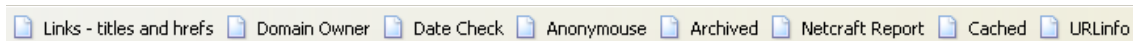
6. Browser buttons en toolbars

In dit hoofdstuk komt een paar snufjes ter sprake die het leven van de zoeker kunnen veraangenamen en vereenvoudigen.

6. 1. Browser buttons (bookmarklets, favelets)

Een van de redenen waarom ik als browsers de voorkeur geef aan Firefox en Opera boven Internet Explorer is dat deze browsers je veel meer informatie verschaffen over een webpagina dan Microsofts product dat doet. Klik bij Firefox met de linker muisknop in een pagina of een frame, of op een plaatje, en je krijgt de mogelijkheid aangeboden het adres van het plaatje te kopiëren, het frame in een nieuw venster te openen, de Last Modified-datum van een pagina te zien en alle links op een pagina in een lijstje te zien. Jammer is dat je ze niet ook allemaal in één keer kunt kopiëren, zoals dat met Opera wel kan. Maar voor dat soort kleine of soms ook grotere wensen bestaan 'bookmarklets', ook bekend als 'browser buttons' of 'favelets', kleine scriptjes die als een knop in een balk van je browser kunnen worden gezet en die dan de ontbrekende functie van de browser op die manier toch toegankelijk maken. In Opera is de balk waarin je 'bookmarklets' kunt onderbrengen de 'Personal bar', in Firefox is het de 'Bookmark bar'. Bij deze browsers sleep je de link of de knop waar de functie onder zit eenvoudig naar de juiste balk (klikken, slepen, loslaten). Internet Explorer heeft voor dit doel de 'Link bar' (Nederlands: Koppelingen). Slepen lukt mij hier niet, maar door rechts op de link of knop van de 'bookmarklet te klikken, in het daarop volgende menu 'Add to Favorites' ('Toevoegen aan favorieten') te kiezen en de 'bookmarklet' in de map Koppelingen te zetten kunt u ook hier uw doel bereiken.

De 'bookmarklet' die voor mij in Firefox de 'outgoing' links van een pagina toont en laat kopiëren, heet 'Links – titles and hrefs' en is te vinden op een site waar meer van dit soort snufjes worden aangeboden. U vindt ze onder de hierboven genoemde namen (bookmarklets etc.) met een zoekmachine, waarbij u eventueel ook nog een term kunt toevoegen die de functie omschrijft die u in de vorm van zo'n knop aan uw browser zou willen toevoegen, bijvoorbeeld: "[search bookmarklets](#)".



Zo maar een aantal bookmarklets in de Bookmark bar van Firefox.

Verzamelingen ervan vindt u meestal wel op de site van de browser die u gebruikt, en veel daarvan zijn ook browser-specifiek: ze kunnen alleen in een bepaalde browser worden gebruikt. Daarnaast bestaan er browseronafhankelijke verzamelingen. Ik geef hier een paar voorbeelden van bookmarklets die mij nuttig lijken. Tenzij anders vermeld, werken ze in de meest gangbare browsers.

<i>Naam</i>	<i>Adres</i>	<i>Functie</i>	<i>Voorbeeld/Opmerking</i>
Make Search Bookmarklet Tool	http://www.bookmarklets.com/mk.phtml	doorzoekt een site of database met zijn eigen zoekmachine op tekst die op een willekeurige plek gemarkeerd wordt	werkt alleen als de zoekterm in de URL van de resultatenlijst van de 'site search engine' verschijnt

<i>Naam</i>	<i>Adres</i>	<i>Functie</i>	<i>Voorbeeld/Opmerking</i>
FaganFinder URLInfo Bookmarklet Maker	http://www.faganfinder.com/urlinfo/?goto=http%3A//www.faganfinder.com/urlinfo/bookmarklet.php	maak zelf uw eigen bookmarklets voor een reeks van functies die met zoeken te maken hebben, en voor verschillende zoekmachines	voorbeeld: track met deze bookmarklet kunt u een (van vier) trackingdiensten activeren die veranderingen op webpagina's bijhouden en u daarover per e-mail informeren
wayback newest, resp. wayback search	http://www.squarefree.com/bookmarklets/search.html	toont de nieuwste, resp. alle oudere versies van de op uw scherm afgebeelde pagina in het Internet Archive (Wayback Machine)	zoek een oudere versie van een inmiddels verdwenen pagina (biedt meer mogelijkheden dan de cachekopie van Google)
google site search	http://www.squarefree.com/bookmarklets/search.html	doorzoekt een site met Google op tekst (een woord of een phrase) die op een van de pagina's van die site gemarkeerd wordt	gebruik Google als alternatief voor de site search engine
Google Definition Search	http://www.imilly.com/bm.htm	activeert Google's define: shortcut bij een door u gemarkeerde term	definities verschaffen context. Heb voor iedere eventuele zoekterm de definitie paraat
Answers.com Search	http://www.imilly.com/bm.htm	voert met een gemarkeerde term een zoekactie uit in een aantal 'naslagwerken' op het Web	de meest uitgebreide gratis 'ready reference' op het Web, met veel in de context ingebedde links; Google heeft op zijn resultatenpagina ook een link, maar dit gaat nog sneller
OneLook Search	http://www.imilly.com/bm.htm	voert met een gemarkeerde term een search uit in honderden woordenboeken	als aanvulling op Answers.com voor specialistische termen
URLinfo Yahoo backlinks	http://www.faganfinder.com/urlinfo/?goto=http%3A//www.faganfinder.com/urlinfo/bookmarklet.php	zoekt met Yahoo! naar de incoming links van de afgebeelde pagina	zoek naar documenten die met de afgebeelde pagina gerelateerd zijn doordat ze ernaar linken (dit is maar één van de vele bookmarklets die met FaganFinder kunnen worden gemaakt; een extra voordeel is dat het FaganFinder menu op de pagina verschijnt)
Check MSN backlinks for the current URL	http://www.researchbuzz.org/2004/11/some_bookmarklets_for_the_new.shtml	zoekt met MSN Search naar de incoming links van de afgebeelde pagina	zoek naar documenten die met de afgebeelde pagina gerelateerd zijn doordat ze ernaar linken

<i>Naam</i>	<i>Adres</i>	<i>Functie</i>	<i>Voorbeeld/Opmerking</i>
alltheweb external backlinks	http://www.squarefree.com/boobookmarklets/seo.html	zoekt met AlltheWeb naar de incoming links van de afgebeelde pagina	zoek naar documenten die met de afgebeelde pagina gerelateerd zijn doordat ze ernaar linken
google backlinks	http://www.squarefree.com/boobookmarklets/seo.html	zoekt met Google naar de incoming links van de afgebeelde pagina	zoek naar documenten die met de afgebeelde pagina gerelateerd zijn doordat ze ernaar linken
Links – titles and hrefs	http://www.accessify.com/tools-and-wizards/accessibility-tools/favelets/	toont anchors en (outgoing) links op de afgebeelde pagina	verzamel deze van verschillende pagina's en sorteer ze om te ontdebellen
Meta data	http://www.accessify.com/tools-and-wizards/accessibility-tools/favelets/	toont de meta tag data van de afgebeelde pagina	kan handig zijn voor als een zoekmachine een pagina presenteert waarin u de zoekterm(en) niet terug kunt vinden; welke metadata heeft een pagina die hoog scoort bij een zoekmachine die mede op basis van metadata rankt

6.2. Toolbars

Browser buttons hebben de charme van de eenvoud. Ze zijn gemakkelijk te downloaden – sleep ze naar een balk in uw browser –, te beheren – via de rechter muisknop kunt u ze hernoemen, ze bewerken (als u javascript kent), of verwijderen – en te gebruiken. Hun eenvoud is echter hun beperking. Anders dan een zoekactie in één zoekmachine uitvoeren, hetzij vanuit een zoekvenster of met een gemarkeerde term of phrase, doen ze niet. Als u meer wilt, moet u op zoek naar toolbars.

Toolbars zijn, zoals de term al zegt, gereedschapsbalken die u aan uw browser kunt toevoegen. Het meest elementaire dat ze doen is in zo'n balk een zoekregel tonen, waarin zoektermen kunnen worden ingevuld. Als u dan op de knop of de loep naast het invulvak klikt, wordt de vraag opgestuurd naar een zoekmachine, die vervolgens resultaten laat zien.

Een vorm van extra functionaliteit kan zijn dat het programma achter de toolbar de geschiedenis van uw zoekacties bewaart, u ziet dat dan wel aan het pijltje naast het zoekvak. Een andere extra functionaliteit kan zijn dat, met behoud van de zoektermen in het invulvak, een zoekactie achter elkaar in diverse zoekmachines kan worden uitgevoerd; daarbij kan men gebruik maken van alle zoekmachines in het pakket, of alleen van die welke men zelf heeft uitgekozen. Let wel: het gaat hier dus niet om een gelijktijdige zoekactie met diverse zoekmachines, maar om zoekacties die na elkaar worden uitgevoerd, net alsof u uw eigen sprinkplankzoekmachine in uw browser heeft. Wel is het natuurlijk mogelijk dat u de toolbar van een metazoekmachine zoals Vivísimo, of van een zoeksoftwareprogramma zoals Copernic hebt geïnstalleerd, dan zet die metazoekmachine of dat programma diverse zoekmachines voor u aan het werk en aggregereert de resultaten.

6.2.1. Toolbars speciaal voor Firefox

Voor de zoeker die naast de grote mondiale zoekmachines ook een select aantal gespecialiseerde zoekmachines in zijn toolbarpakket wil hebben – waarbij, zoals

gezegd, een en dezelfde zoekactie achtereenvolgens met verschillende zoekmachines vanuit de toolbar kan worden uitgevoerd – biedt de Firefox browser verreweg de meeste mogelijkheden, ook al geldt dat op zichzelf genomen veel toolbars niet voor andere browsers dan Internet Explorer beschikbaar zijn.

Ik noem een paar van de mogelijkheden die op de site van

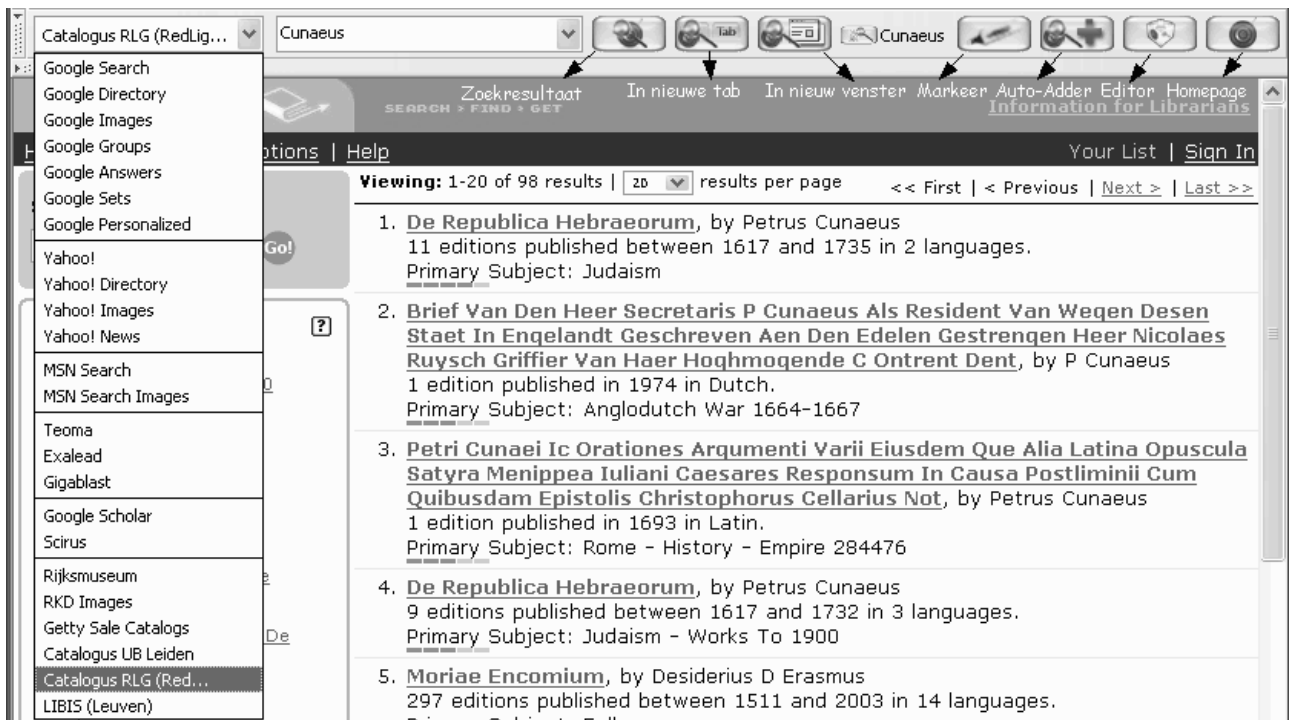
Firefox(<http://www.mozilla.org/products/firefox/central.html>) worden aangeboden:

- In de rechterbovenhoek van het scherm heeft Firefox standaard een zoekregel met een paar zoekmachines (Google, Yahoo!, Ebay) waaruit gekozen kan worden. Daaraan kan een hele batterij zoekmachines worden toegevoegd, en er kunnen natuurlijk ook zoekmachines worden verwijderd. Momenteel zijn er meer dan 3000 zoekmachines en databases beschikbaar, verdeeld over 32 categorieën en subcategorieën, en er komen er nog altijd bij. Het toevoegen van zoekmachines gebeurt door vrijwilligers, hetgeen betekent dat niet elke categorie even goed vertegenwoordigd is. Maar populaire rubrieken als Music, Shopping, File Sharing, Academic, Reference en Language Dictionary zijn ruimschoots gevuld.
- Hiernaast ziet u een selectie van zoekmachines en databases waarvoor plug-ins bestaan. Door de volgorde van downloaden van de plugins heb ik getracht enige orde te bewaren, wat niet helemaal is gelukt. Helaas is het momenteel nog niet mogelijk de zoekmachines in groepen van elkaar te onderscheiden. Niettemin is de algemene groepering hier wel duidelijk: algemene zoekmachines, wetenschappelijke zoekmachines, naslagwerken, sharewaresites, diversen.
- Een ander zoeksnuifje in Firefox en Mozilla is NeedleSearch, dat er iets minder gelikt uitziet, maar feitelijk meer mogelijkheden biedt. In de eerste plaats heeft deze toolbar een eenvoudige eigen editor, die u in staat stelt de zoekmachines die u kiest naar eigen inzicht te ordenen, groepen te benoemen en door een separator van elkaar te onderscheiden, maar bovendien bent u hier niet, zoals bij de andere mogelijkheid, gebonden aan de selectie van zoekmachines waarvoor anderen een script hebben gemaakt. Dat script maakt u zelf op een heel simpele manier:



1. zorg ervoor dat de zoekregel van de zoekmachine (ook site search engines komen in aanmerking) of database op uw scherm staat
2. typ daarin de zoekterm 'NeedleSearch' (met de hoofdletters!) en druk op een knop in de toolbar (de Auto-Adder); u krijgt een mededeling dat de zoekmachine is toegevoegd en dat u hem in de lijst kunt vinden onder zijn URL
3. start de editor door op de knop Launch the NeedleSearch editor te klikken, en pas de naam van de zoekmachine aan. Met de editor kunt u de zoekmachine ook ergens anders in de lijst neerzetten, een scheidingslijn tussen zoekmachines of groepen zoekmachines aanbrengen en groepen benoemen door een zoekmachine met alleen een naam (Zoekmachines: Kunst) toe te voegen.
4. bevestig uw keuze, sluit de editor en kies voortaan de zoekmachine die u wilt gebruiken uit de lijst die onder het pijltje naast de zoekregel in de toolbar zit.

Een van de aardige dingen van NeedleSearch is dat het ook zoekmachines aan kan die met meer velden werken. In zo'n geval nummert men de te doorzoeken velden door bij het toevoegen van een zoekmachine in plaats van 'NeedleSearch' in de enkele zoekregel,



De toolbar van NeedleSearch met een voorbeeldlijstje van zoekmachines. Bovenin de illustratie is met zwarte pijltjes aangegeven wat de knoppen doen. Zoekresultaat wil zeggen: zoekresultaat in zelfde venster tonen, In nieuwe tab en In nieuw venster wil zeggen: zoekresultaat onder nieuwe tab en in nieuw venster tonen; Markeer: markeer (highlight) de zoektermen in het zoekresultaat, Auto-Adder en Editor: zie boven, Homepage: naar de homepage van NeedleSearch.

in de verschillende zoekregels 'NeedleSearch1', 'NeedleSearch2' enz. in te vullen. Bij het zoeken kan men dan diverse zoektermen aan de diverse zoekregels toedelen door er een puntkomma tussen te zetten. Probleem is natuurlijk wel dat u zelf moet onthouden waar de velden voor stonden (auteur, titel enz.).

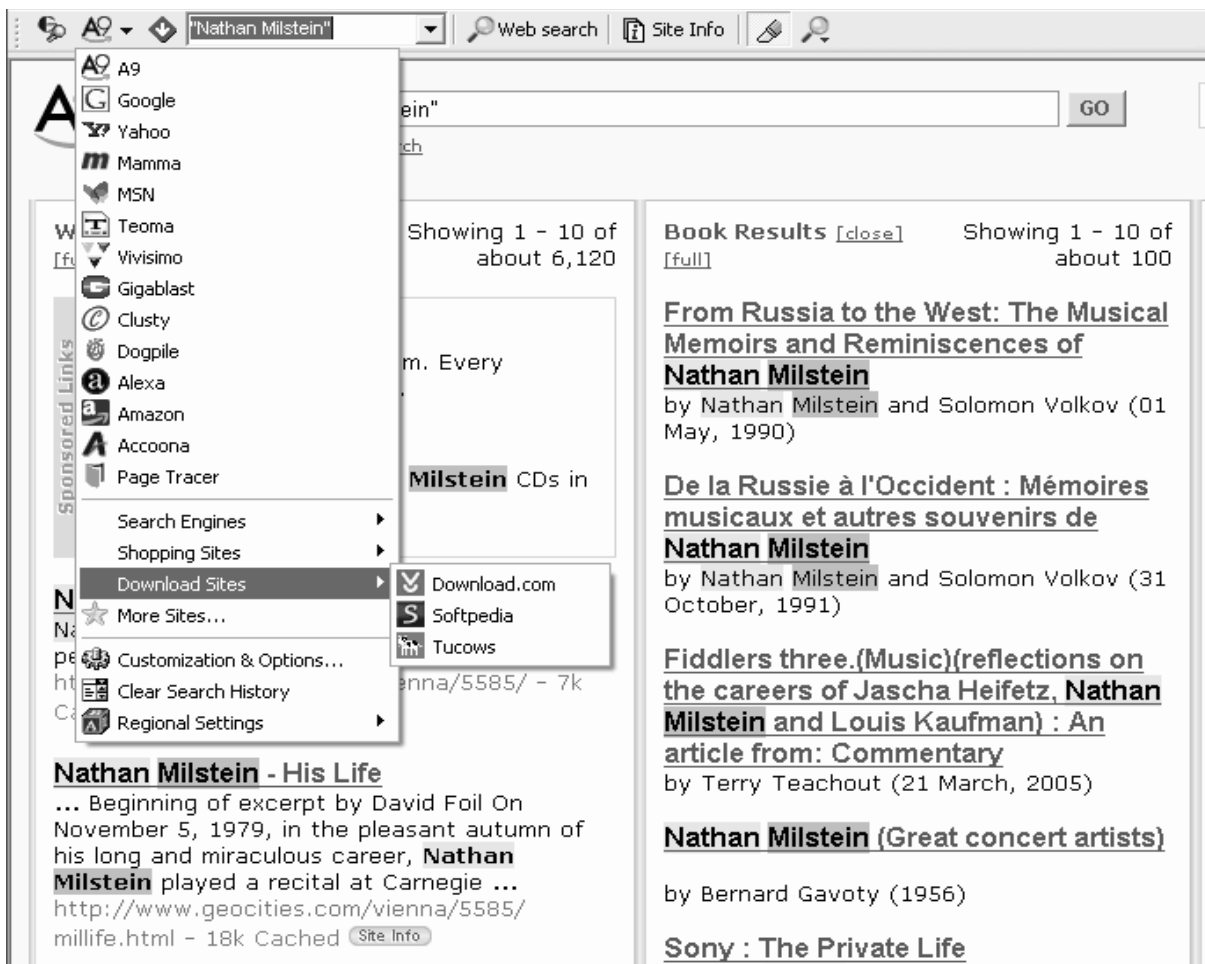
Needle Search werkt niet met zoekmachines die niet direct te benaderen zijn. In het voorbeeld werkten zoekacties in de RKD-database en in de Getty Sale Catalogs niet; in zulke gevallen werkt NeedleSearch dus als een simpel bookmarklijstje.

6.2.2. Andere search toolbars

De kracht van de hierboven genoemde toolbars is dat ze een snelle toegang tot een arsenaal aan zoekmachines en databases bieden, waarvan het merendeel ook via één zoekregel doorzoekbaar is. Meer hebben ze niet te bieden. Daarnaast bestaan nog allerlei andere zoektoolbars met extra functionaliteit. Deze zijn meestal afkomstig van zoekmachines en hebben klantenbinding ten doel. Een andere eigenschap die ze gemeen hebben is dat er meestal alleen een versie voor Internet Explorer van is; begrijpelijk als je bedenkt dat dit de verreweg meest gebruikte internetbrowser is.

Een zoektoolbar tot voor kort alleen voor IE beschikbaar was, maar nu ook voor Firefox is de Groowe Toolbar (<http://www.groowe.com>). Net als de hierboven besproken toolbars voor Firefox heeft deze toolbar los van het enkele invoervenster voor een reeks van zoekmachines weinig extra functies. Twee gemakken die deze toolbar echter bij alle zoekmachines biedt zijn de markering (highlighting) van zoektermen in een gevonden webpagina, en de mogelijkheid (onder een pictogram dat een loep voorstelt) om de zoekterm in de pagina te zoeken.

De Groowe toolbar geeft u toegang tot 18 zoekmachines, 2 winkels (Amazon en Barnes & Noble) en 3 downloadsites, maar dat is nog lang niet alles. Bij elk van de diensten



Een Zwitsers zakmes voor de zoeker: de Groowe toolbar.

worden ook een aantal belangrijke nevendiensten via de toolbar toegankelijk gemaakt; de toolbar verandert dus van gedaante als men van zoekmachine verandert. Zo zijn bij Google naast de webzoekmachine onder meer ook die voor News, Images, Groups en de Directory beschikbaar; als knoppen wel te verstaan, hetgeen betekent dat daarmee direct op de zoekterm(en) gezocht kan worden. Download.com daarentegen heeft onder meer Music en New releases. Het aantal mogelijkheden is dus aanzienlijk uitgebreider dan de hierboven genoemde getallen doen vermoeden. Daarbij komt dat er updates van de toolbar uitkomen, waarbij het aantal plug-ins verder wordt uitgebreid.

Bijzonder handig is een pijl naast het zoekvak waarmee men direct naar de volgende zoekmachine kan 'doorbladeren'.

De Groowe toolbar heeft beperkte personalisatie (customization) mogelijkheden. Zo kan de volgorde van de zoekmachines worden gemanipuleerd, kunnen per zoekmachine opties worden toegevoegd of (uit het zicht) verwijderd en kunnen zoekmachines uit de actuele lijst naar een reservelijst worden verplaatst (onder Customization & Options). Dat kan niet met allemaal, sommige zijn 'affiliated sites', die de gratis beschikbaarheid van het product mogelijk maken.

Een extra dat ook onder Customization te vinden is en niet geactiveerd, is een eenvoudige popup-blocker.

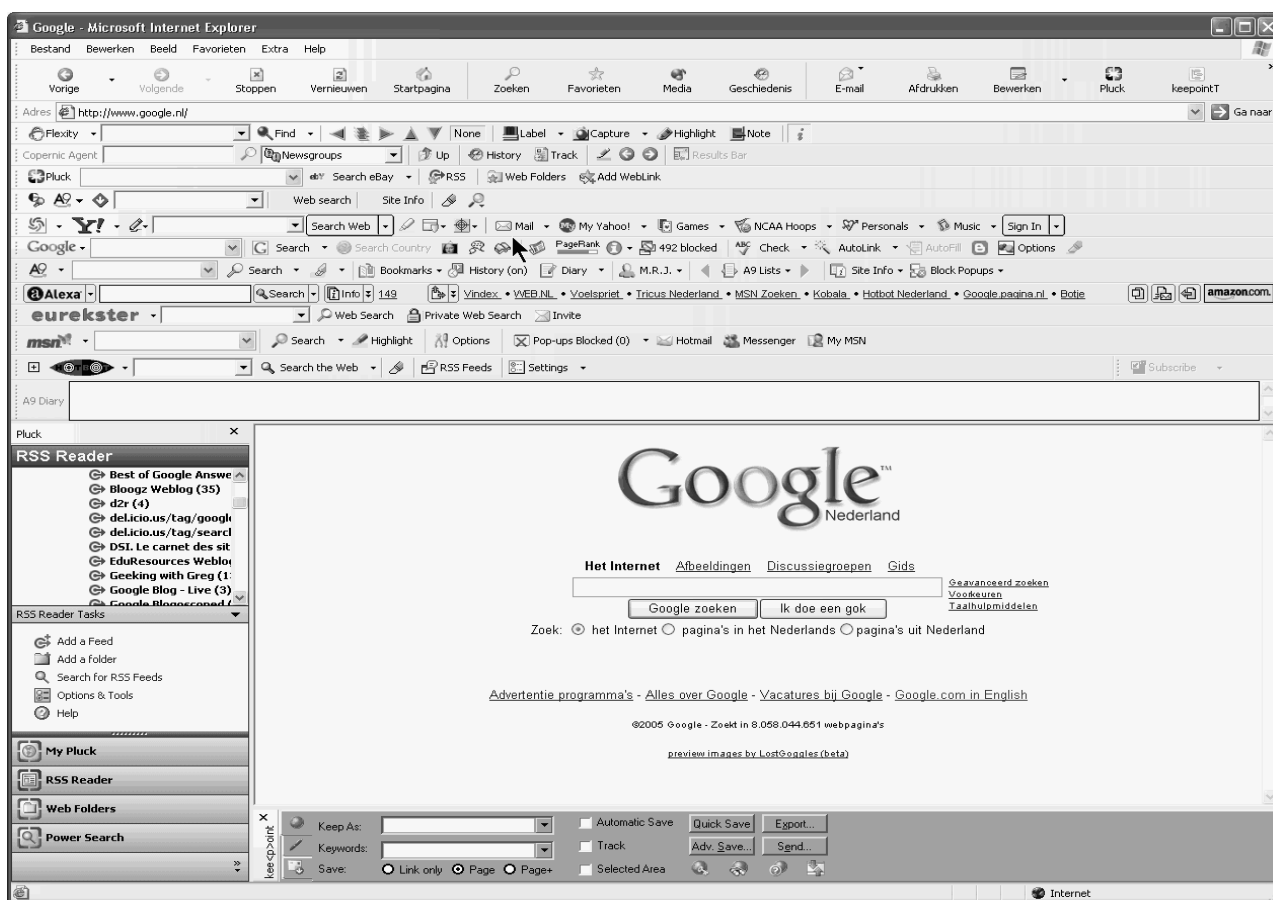
Anders dan de Groowe toolbar zijn de meeste andere toolbars zoekmachine-specifiek. Hoewel er niets op tegen is er een paar te installeren, raakt uw browser natuurlijk wel met elke balk meer steeds verder overladen en raakt u al gauw het overzicht over alle

verschillende functies kwijt. (Ja, ik weet dat je toolbars kunt verbergen, maar gebruik je ze dan nog?) Dit geldt te meer omdat u misschien ook nog wel andere toolbars wilt gebruiken, voor een online bookmark manager als Backflip of een information manager zoals Pluck. En die hebben dan ook weer functies die overlappen met die van zoekbalken. Zoals zo vaak hebben we hier weer een probleem van kiezen tussen recall en precision, tussen overdaad en soberheid.

Het aantal toolbars waarmee men zijn browser kan uitrusten loopt vermoedelijk in de tientallen, en gegeven de diversiteit van de behoeften en manieren van werken van hun gebruikers heeft het geen zin ze hier allemaal te bespreken. Ik volsta er daarom mee een aantal kenmerken te noemen die veel ervan met elkaar gemeen hebben, en vervolgens enkele bijzonderheden te noemen die karakteristiek zijn voor een of enkele van degene die ik onder ogen heb gehad.

Alle toolbars die hier ter sprake komen hebben een zoekvak waarin u de zoektermen van een vraag kunt intypen, die dan vervolgens in de bij de toolbar behorende zoekmachine wordt uitgevoerd. De enige uitzondering die ik ken is de Y!Q (zie blz. 172) demobar van Yahoo!, waarin u geen zoektermen intypt maar die als een browser button werkt: als u een stuk tekst in een document markeert en u klikt op de Y!Q loop, dan zoekt Yahoo! naar gerelateerde pagina's op basis van karakteristieke termen in de gemarkeerde tekst.

De zoekmachines die niet alleen een database van webpagina's, maar ook van nieuws, plaatjes e.d. hebben, en/of op hun website hulpmiddelen als woordenboeken en encyclopedieën (MSN's Encarta) beschikbaar stellen, bieden die natuurlijk ook via de toolbar aan



Een tikje overdreven? Internet Explorer op de computer van de auteur, kort voordat deze door nog onopgehelderde oorzaken crashte...

Frequent zijn verder een pop-up blocker, waarmee ongewenste pop-up vensters kunnen worden geweerd, een highlightfunctie waarmee zoektermen in een gevonden webpagina kunnen worden gemarkeerd (zoals dat bij Google in de cacheversie van een pagina gebeurt) en een woordzoeker om de zoektermen in lange pagina's op te sporen. Het laatste kan natuurlijk ook met de browser, maar als u niet gewend bent dat met de toetscombinatie Ctrl-F te doen, gaat het zo sneller. Een aardig extraatje van de Yahoo!-toolbar naast de popup-blocker is een spyware killer, die allerlei bestandjes en programma's op uw computer opspoorde en verwijderde die informatie over uw surfgedrag verzamelen en doorgeven. Als u daarnaast met een toolbar werkt – bijvoorbeeld met die van Amazon, Alexa en/of A9 – die zijn werk alleen kan doen door uw surfgedrag bij te houden, is het zaak op te passen dat u die niet buiten werking stelt door de trackingfuncties van die toolbar te elimineren.

Een aantal toolbars geeft 'site info' over de site waarvan u een pagina op het scherm hebt staan. Die kan bestaan uit een PageRank (Google) of een ander rankingmechanisme (Traffic Rank voor Alexa, Amazon en A9, HipRank voor SearchHippo), maar de informatie kan ook aanzienlijk uitgebreider zijn. Zo biedt de Alexa toolbar in handige venstertjes veelal naam, adres en telefoonnummer van de eigenaar van de site, informatie over bezoekersaantallen en welke sites door die bezoekers nog meer worden bezocht, de backlinks van de pagina en de categorie waarin de site eventueel in het Open Directory Project is ondergebracht; de toolbars van Amazon en A9 bieden die informatie op een wat andere manier ook.



Alexa heeft niet alleen een grote eigen zoekmachine, maar de toolbar heeft ook een aantal nuttige zoekfuncties, zoals een link naar een ODP-categorie en 'People also visit...'

Bedrijven die behalve een zoekmachine ook gepersonaliseerde diensten aanbieden, geven geregistreerde gebruikers daar veelal via de toolbar ook toegang toe. Een voorbeeld is e-mail: zowel Google als MSN en Yahoo! laten u via de toolbar in hun mailservers inloggen. Als een zoekmachine andere gepersonaliseerde diensten aanbiedt, zoals Yahoo! dat met My Yahoo! doet en MSN met MyMSN, biedt de toolbar geregistreerde gebruikers de mogelijkheid daarop in te loggen. Zoals bij de bespreking van Yahoo! al even ter sprake is gekomen, is Yahoo! naar mijn indruk de zoekmachine met het meest volledige en best geïntegreerde pakket aan zulke gepersonaliseerde diensten. Met de komst van Yahoo! 360 als blogging service wordt dit pakket verder gecompleteerd. Yahoo! als zoekmachine is een goede tweede na Google, met een breed pakket aan databases naast de webdatabase en een grote, gerenommeerde directory. Gegeven het feit dat het verstandig is minstens twee zoekmachines naast elkaar te gebruiken en in het licht van wat ik hierboven heb aangestipt inzake de wenselijkheid een beperkt aantal toolbars te kiezen die toegang geven tot een zo compleet mogelijk pakket aan diensten zou de keuze van de toolbars van Google en Yahoo! een aantrekkelijke optie zijn.

Naast alle hierboven genoemde gemakken bieden sommige toolbars nog andere aan, in de vorm van online vertalingen, een rekenmachine, een klok, een kalender, een woordenboek, weersberichten etc.

7. Bijblijven

Zoeken op het Web is eigenlijk een testimonium paupertatis, een bewijs van een tekort. Het feit dat je op zoek moet wijst erop dat je iets niet hebt wat je eigenlijk zou moeten hebben, of iets kwijt bent dat je eens gehad hebt en weer nodig hebt.

Ik geef een voorbeeld uit mijn eigen praktijk: een paar maanden geleden kwam ik erachter dat gratis en voor niets een fantastische catalogus op het Web te raadplegen is, die van de [Research Libraries Group](#) (RLG), een internationale organisatie van bibliotheken, archieven en musea. Deze catalogus bevat niet minder dan 130 miljoen beschrijvingen (meer dan 40 miljoen titels) en is, misschien op WorldCat na, de grootste in zijn soort.

Op het bestaan van deze catalogus stuitte ik toevallig in een mailtje van een nieuwsbrief waarop ik geabonneerd ben, maar bij nadere inspectie bleek de site allerm minst nieuw: de oudste versie in het Internet Archive dateert van december 2003 en er blijkt zelfs al in april 2003 over geschreven te zijn.

Nu zoek ik beroepshalve nogal eens naar oude boeken en gebruik daarvoor allerlei grotere en kleinere bibliotheekcatalogi. Omdat dat een karweitje is dat nogal wat tijd kost, begin ik bij de bibliotheken met de grootste collecties oude boeken of de nationale bibliotheek van het land waar zo'n boek is uitgegeven, en daal dan langzamerhand naar lagere regionen af. Had ik eerder van het bestaan van RedLightGreen af geweten, dan zou ik heel wat zoekacties in die catalogus zijn begonnen.

Hoe had ik het bestaan van RedLightGreen kunnen missen? Raadpleging van GoogleGroups leert dat de term RedLightGreen sinds 1981 32 keer in nieuwsgroepen gebruikt is. De oudste daarvan is een 'post' van David D. Dillard van begin oktober 2003 in de lijst Net-Happenings. Tja, dat ik die gemist heb, kan ik me wel voorstellen, want het aantal happenings in die tijd was nogal overvloedig. Een andere plek waar ik op het bestaan van RedLightGreen geattendeerd had kunnen worden was Gary Price's nieuwsbrief [ResourceShelf van 2 oktober 2003](#), maar ook daarin moet ik erover heen gelezen hebben. Heel toepasselijk staat de oorzaak er in een post waarmee de archiefmap van ResourceShelf van oktober 2003 opent, bij: het is infoglut, ook wel bekend als 'information overload'.

In dit geval was de informatie beschikbaar, en ook het kanaal waarlangs die informatie mij had kunnen bereiken. Dat ze dat niet heeft gedaan heeft een reden waarop ik straks nog terugkom, maar hier gaat het erom dat als ik niet op NetHappenings en Resourceshelf geabonneerd was geweest, de kans dat ik RedLightGreen zou hebben ontdekt niet erg groot zou zijn geweest. Dat ik deze waardevolle resource uiteindelijk in februari 2005 tegenkwam had ik ook niet aan zoeken te danken, maar aan weer een andere nieuwsbrief, de weblog van SearchEngineWatch, waarin Gary Price ook schrijft. Als je niet weet dat je niet weet, is de kans dat je te weten zult komen ook niet erg groot. Maar wacht eens even – het Web is toch niet voor niets het *Web*, waar alle punten door draden met elkaar verbonden zijn? Ik had RedLightGreen toch al surfend over de links kunnen tegenkomen? Jawel, en als er genoeg draden (links) zijn die naar een punt leiden, vind je dat op enig moment ook wel. Alleen, bij RedLightGreen ligt dat niet zo voor de hand, want Google kent maar 170 links naar een van zijn URL's (www.redlightgreen.com, www.redlightgreen.org, www.redlightgreen.net). Dat zijn er ongeveer net zoveel als naar de website van de gemeente Leiderdorp.

De moraal van dit verhaal is dat zoeken maar één kant van de medaille van kennisverwerving is. De andere is bijblijven, en ze horen inderdaad bij elkaar als twee kanten van een medaille. Dat u wel eens vergeet waar u iets gelezen hebt, is niet zo erg: naar iets waarvan u weet dat het moet bestaan, kunt u zoeken. Maar eerst moet u weten dat het bestaat, en daarvoor is het nodig dat u bijblijft.

In dit hoofdstuk behandel ik in het kort een aantal mogelijkheden op het Web – naast discussielijsten, fora en nieuwsgroepen – om bij te blijven, om u op de hoogte te houden en te laten houden van de ontwikkelingen die zich op uw interessegebied voordoen. De ervaring die ik hierboven heb beschreven, speelt daarbij in zoverre een rol dat infoglut altijd op de loer ligt. De middelen die u kiest om u op de hoogte te laten houden, bepalen voor een deel hoe efficiënt en effectief dat zal gebeuren.

7.1. Alerts

Ik begin met een dienst die door ettelijke zoekmachines maar ook door veel individuele websites wordt aangeboden in de vorm van 'Alerts'. Een voorbeeld van zo'n Alert-dienst van een zoekmachine is Google Alerts (<http://www.google.com/alerts>; niet te verwarren met de commerciële, niet met Google gelieerde dienst Google Alert, <http://www.googlealert.com/>). Deze dienst is gebaseerd op e-mail, wat natuurlijk betekent dat u zich moet registreren en wat vanzelfsprekend meteen twee dingen veronderstelt: 1. u leest uw e-mail en bent nog niet bezweken onder de vloedgolf van spam die dit kanaal om op de hoogte te blijven teistert; en 2. u beperkt de inhoud en frequentie van de alerts tot een hoeveelheid die u aankunt.

Google Alerts werkt heel simpel: net als bij een zoekactie op het Web geeft u de dienst een of meer zoekvragen op, vervolgens uit welke database (News, Web of beide) u treffers wenst te ontvangen, en ten slotte de frequentie waarmee dat moet gebeuren. Bij dit laatste hebt u de keus tussen eenmaal per dag, eenmaal per week en 'as it happens'. Bij deze keuze gaat het niet om hoe vaak u e-mails ontvangt, maar om hoe vaak Google de opgegeven vraag uitvoert; 'as it happens' betekent in dit verband dat Google dat voortdurend doet.

Waarop attendeert Google u? Ook dat is simpel: op nieuwsberichten en of websites die voor de vraag die u gesteld hebt tot de eerste tien, resp. twintig treffers in de resultatenlijst zijn doorgedrongen.

Verder hebt u nog de keuze tussen mails in HTML-formaat en in tekstformaat. Google Alerts laat u in eerste instantie maximaal tien zoekvragen opgeven. Nadat u gereageerd hebt op een e-mail waarin om bevestiging wordt gevraagd, kunt u er meer opgeven. Via de e-mails die u ontvangt en via de inlogpagina kunt u 'alerts' bewerken of verwijderen. Een andere (meta)zoekmachine die 'alerts' aanbiedt is Profusion (<http://www.profusion.com/index.htm>), een zoekmachine die we al in het hoofdstuk over het 'Invisible Web' zijn tegengekomen. Profusion kan twee typen 'alerts' voor u verzorgen, 'Page Alerts' en 'Search Alerts'. Dit tweede type lijkt op de dienst die Google biedt. Ook hier voert u een vraag in, die door Profusion periodiek wordt uitgevoerd en waarvan de resultaten naar u toe worden gemaïld, maar omdat Profusion een metazoekmachine is, moet die vraag eerst worden uitgevoerd, en daarvoor is nodig dat u Profusion duidelijk maakt in welke zoekmachines dat moet gebeuren. Hebt u dat gedaan (via: <http://www.profusion.com/nav>) en de vraag uitgevoerd, dan biedt het resultatscherm de mogelijkheid de vraag in de gekozen zoekmachines in een Search Alert op te slaan.

'Page Alerts' zijn waarschuwingen die betrekking hebben op veranderingen in een webpagina, waarvan u de URL opgeeft. Ook hier kunt u eerst een vraag uitvoeren en dan bij een door u geselecteerde treffer in de resultatenlijst een Page Alert aanbrengen, maar u kunt de URL van de pagina die u wilt volgen ook direct via de Alert Manager (<http://www.profusion.com/alertsmanager>) opgeven. Aangezien veel pagina's voortdurende wijzigingen ondergaan en lang niet al die wijzigingen interessant zijn, is het verstandig gebruik te maken van de optie om in de pagina termen op te geven waarop die veranderingen betrekking moeten hebben; dan krijgt u alleen 'alerts' bij veranderingen die die termen betreffen. Ook hier geeft u de frequentie op waarmee Profusion de gekozen zoekmachines de opgegeven pagina op veranderingen laat

controleren, plus het e-mail adres waarheen een rapport of een melding (voor een PDA of een mobiele telefoon) van die veranderingen moet worden moet worden verstuurd. Via de Alert Manager kunt u 'alerts' deactiveren/activeren (Disable), uitvoeren (Run), bewerken/wijzigen (Edit) en verwijderen (Delete). Ook kunt u de History van de 'alerts' opvragen en zelfs – in een bij Profusion bewaard archiefexemplaar van de gevolgde pagina – de veranderingen zien die Profusion op een bepaald moment in het verleden geconstateerd heeft.

Profusion heeft een categorie Web Search Engines waarin de belangrijkste grote mondiale zoekmachines zitten, met uitzondering van Google. Alerts uit deze categorie vormen dus een mooie aanvulling op Google Alerts. Daarnaast zijn er natuurlijk de speciale categorieën, en de eventueel door u zelf samengestelde My Search Groups, waarin u eigen combinaties van zoekmachines kunt maken.

Trefwoorden waarmee u dit soort diensten met een zoekmachine kunt vinden: alerts, "alerting services", "staying current", "current awareness services", "keeping up-to-date".

Ook individuele sites waar men aan het verzamelen van resources of het volgen van ontwikkelingen doet, bieden vaak de mogelijkheid u per e-mail van ontwikkelingen op de hoogte te houden. Een voorbeeld is Librarian's Index of the Internet (<http://lii.org/search/file/maillinglist>), waarbij u zich op een nieuwsbrief kunt abonneren die u attendeert op de nieuwe sites en pagina's die aan deze selectie van internetbronnen worden toegevoegd. Een mogelijkheid om er snel achter te komen of een site zo'n faciliteit biedt is Google: `site:... subscribe`.

7.2. Tracking services

De 'Page Alerts' van Profusion zijn een voorbeeld van wat ook wel een 'tracking service' of 'Web page monitoring service' wordt genoemd, een dienst die veranderingen op webpagina's volgt en daarover per e-mail (of op een andere manier, bijvoorbeeld via ICQ) rapporteert. (ICQ staat voor 'I Seek You' en is een programma dat men kan downloaden om contact met anderen te houden als die online zijn.) Een ander goed voorbeeld van zo'n 'tracking service' is WatchThatPage (<http://www.watchthatpage.com>), dat op ongeveer dezelfde manier werkt. Er kan een onbeperkt aantal pagina's bij deze gratis dienst worden opgegeven, en ook hier is het mogelijk de wijzigingen zelf naar u te laten e-mailen of alleen de URLs van de pagina's waarop wijzigingen hebben plaatsgevonden. Net als bij Profusion kunt u zoektermen opgeven waarop die wijzigingen betrekking moeten hebben, en een verdere verfijning van dat filter is dat u die zoektermen per alinea kunt laten zoeken; WatchThatPage stuurt u dan niet alle veranderingen die het – op basis van de door u opgegeven zoektermen – in een veranderde pagina vindt, maar alleen de paragrafen waarin die zoektermen voorkomen. Een groot gemak is verder dat WatchThatPage u een knop in uw bookmarkbalk (zie blz. 264) laat zetten waarmee u een door u bezochte pagina direct als een te volgen pagina bij de dienst kunt opgeven.

De enige belangrijke beperking is dat de zoektermen die u als filter opgeeft voor alle pagina's gelden die u wilt laten volgen. Als dat – bijvoorbeeld wegens uiteenlopende belangstellingen – een bezwaar is, kunt u beter de Nederlandse dienst WebAgent (<http://www.webagent.nl/>) kiezen, die u per pagina laat instellen in welke gedeelten van de pagina veranderingen *niet* gevolgd moeten worden.

Een bijzonder handig overzicht van dit soort diensten, en nog talrijke andere, verschaft de toepasselijk genaamde pagina Tools (<http://www.geocities.com/hno1/Tools.html>) in het ook verder zeer nuttige Henks Portaal.

Trefwoorden om dit soort diensten met zoekmachines te vinden: "tracking services", "Web page monitoring services"

7.3. Discussiefora: nieuwsgroepen, message boards en mailing lists

In hoofdstuk 2.6. van dit boek zijn discussiefora ter sprake gekomen als plaatsen waar op sommige vragen antwoorden kunnen worden gevonden. Maar natuurlijk zijn discussiefora daar niet in de eerste plaats voor bedoeld. Net zo goed, zelfs beter, kunnen ze worden gebruikt om bij te blijven.

Op de sites die ik in hoofdstuk 2.6. heb genoemd is voldoende informatie te vinden over wat ervoor nodig is om u op een nieuwsgroep, forum, mailing list of nieuwsbrief te abonneren of u te laten registreren, reden waarom ik hier aan dat onderwerp geen aandacht besteed. Wel geef ik u enkele voorbeelden van mogelijkheden die hier liggen. Het eerste heeft betrekking op Google Groups (<http://groups-beta.google.com/>). Nadat u zich bij Google geregistreerd hebt (ook als dat bijvoorbeeld als gebruiker van Gmail is gebeurd), kunt u zich op een of meer van de groepen abonneren waar onderwerpen worden besproken die u interesseren. U vindt zulke groepen door met de zoekmachine naar een onderwerp te zoeken en te kijken in welke nieuwsgroepen dat wordt besproken (zie blz. 29, 114) of door vanaf de homepage van Google Groups de 'directory' van nieuwsgroepen te doorlopen.

Hebt u eenmaal een groep gevonden, dan dient u zich daarbij aan te sluiten om per e-mail van de berichten die daar worden gewisseld op de hoogte te worden gehouden. Kijk voordat u daaraan begint wel eerst goed naar wat er aan berichten circuleert en hoeveel berichten u kunt verwachten te krijgen. Klik dan de knop 'Join this group' die u op de berichtenpagina's van iedere openbare groep vindt en kies op het daaropvolgende scherm hoe u de berichtgeving wilt volgen. Ten slotte klikt u op 'Subscribe to this group', waarna deze groep voortaan op uw My Groups-pagina bij Google Groups verschijnt. De volgende keer dat u bij Google Groups inlogt zult u de inhoud van die map zien verschijnen, en daarbij de mogelijkheid uw abonnementen te wijzigen of te beëindigen.

Google Groups biedt nog een andere mogelijkheid om geïnformeerd te blijven over onderwerpen die u interesseren. Daarvoor hoeft u niet op een groep geabonneerd te zijn. In nieuwsgroepen worden onderwerpen ('topics') ter sprake gebracht en doorgaans in 'threads' bediscussieerd. Komt u bij een groep waarop u niet geabonneerd bent en

The screenshot shows the Google Groups interface for the group 'soc.history.war.misc'. At the top right, there are links for 'rvveen@gmail.com', 'My Groups', 'Help', and 'Sign out'. The group name 'soc.history.war.misc - Join' is displayed in a yellow banner. Below this, there are three sections: 'How do you want to read this group?' with radio buttons for 'No Email', 'Abridged Email (No more than 1 email per day)', and 'Digest Email (Approximately 4 emails per day)'; 'About this group' with details like 'Public - Usenet' and a description; and 'What nickname do you want people in this group to see?' with a text input field. A 'Subscribe to this group' button is located at the bottom of the form. At the very bottom of the page, there are links for 'Google Home', 'Terms of Use', 'Privacy Policy', 'Jobs, Press, & Help', and a copyright notice '©2005 Google'.

Het scherm waar u zich bij Google Groups op een groep abonneert.

waarop u zich ook niet wilt abonneren, een onderwerp tegen waarover u de discussie wilt volgen, klik dan, terwijl u bij Google Groups ingelogd bent, op de ster vóór de kop van het bericht dat het onderwerp inleidt. Het onderwerp komt dan als een 'My starred topic' op uw 'My Groups'-pagina te staan.

Anders dan nieuwsgroepen zijn forums, rondzendlijsten en nieuwsbrieven niet rondom één enkele zoekmachine georganiseerd. Abonneren en registreren zult u dus bij elk forum, elke lijst en elke nieuwsbrief apart moeten doen. Een uitstekend hulpmiddel om via trefwoorden te achterhalen wat de interessante forums etc. zijn is Lycos Discussions Web Search (<http://discussion.lycos.com/>). Zoek daar met specifieke termen. Als u bijvoorbeeld op zoek bent naar forums etc. waar zoekmachines ter sprake komen, zoek dan niet naar de combinatie google yahoo, maar naar [exalead gigablast](#) Over Google en Yahoo! heeft iedereen het, over Exalead en Gigablast alleen de pro's. Voor andere zoekmachines van forums, zie: Data Mining: Message Board Search (http://datamining.typepad.com/data_mining/2006/01/message_board_s.html).

7.4 Weblogs en RSS

Een kanaal waarlangs nieuws zich de laatste jaren razendsnel verspreidt, zijn de (volgens een recente berekening, zie: <http://www.blogherald.com/2005/10/10/the-blog-herald-blog-count-october-2005/>; zie ook:

<http://www.technorati.com/weblog/2006/02/81.html>) meer dan 100 miljoen weblogs of blogs die op het Web zijn ontstaan. Weblogs zijn een soort openbare dagboeken die door bepaalde bedrijven en organisaties worden gehost en waarvoor die de standaard software verschaffen die het bijhouden van zo'n dagboek tot een heel eenvoudige zaak maken.

Weblogs kunnen over van alles en nog wat gaan en kunnen heel verschillende functies hebben. Voor sommige bloggers vormen ze enkel een middel tot zelfexpressie, anderen zoeken een publiek, voor weer anderen vormen hun blogs een manier om met andere mensen in contact te komen en te communiceren. Blogs kunnen het werk van individuele personen zijn, van collectieven, en zelfs van bedrijven of instellingen. Voor ons als zoekers zijn ze in het algemeen gesproken om twee redenen van belang: ze verspreiden nieuws en geven uiting aan meningen of leveren commentaar of uitleg. In beide opzichten moet hun invloed niet worden onderschat: er zijn er die meer links trekken dan de complete sites van alle Nederlandse universiteiten bij elkaar.

Blogs bevatten typisch het soort informatie die in dit hoofdstuk past: je gebruikt ze eerst en vooral om op de hoogte te blijven. Hoewel ze ook opduiken in zoekresultaten, brengt het actuele karakter van blogs met zich mee dat de informatie die ze verschaffen de kans loopt verouderd te zijn; anders dan goede webpagina's worden de afzonderlijke berichten immers niet geupdated. De betrouwbaarheid van de informatie die ze verschaffen, loopt vanzelfsprekend ook uiteen. Een zekere controle in dit opzicht kan het feit bieden dat veel blogs lezers de gelegenheid bieden commentaar op de posts te leveren; in het beste geval kunt u als lezer getuige zijn van een discussie tussen experts. Om blogs te kunnen gebruiken zullen we ze eerst moeten vinden – althans, die blogs vinden die voor ons interessante informatie bevatten. Daarbij doet zich echter een probleem voor: algemene zoekmachines zoals Google zijn niet erg geschikt om blogs te vinden, eenvoudigweg omdat die als categorie niet zo eenvoudig van andere webdocumenten te onderscheiden zijn. Hoewel de zoekmachines ze wel fulltext indexerend, blijven de blogs meestal onzichtbaar te midden van de overvloed van andere zoekresultaten. Aan directories zoals die van Google en Yahoo! hebben we in dit opzicht ook niet veel, want in de eerste plaats bevatten die maar een klein deel van het totale aantal blogs die er zijn, en in de tweede plaats worden de meeste ervan als 'Personal' aangemerkt en zijn niet verder op onderwerp ingedeeld.

Speciale directories en zoekmachines van blogs helpen ons in dit opzicht ook niet veel

verder: de directories delen de blogs op één niveau in hele ruime categorieën in, en de zoekmachines indexeren alleen de, zoals gezegd, meestal nietszeggende namen en op z'n best een korte beschrijving (in de METAtags) die de weblogauteur zelf van zijn blog heeft gegeven. Nee, de beste manier om blogs te vinden is via een omweg, door naar interessante 'posts' te zoeken en zodoende de blogs waarin die posts voorkomen, te verzamelen. We gebruiken dan dus geen zoekmachines die *naar* blogs zoeken, maar zoekmachines die *in* blogs zoeken. Van zulke zoekmachines zijn er tientallen – voor twee overzichten, zie Peter Scotts voortreffelijke Weblogs Compendium (<http://www.lights.com/weblogs/>) en de lijst van blogs in de Internet Public Library (<http://www.ipl.org/div/blogs/>) – , maar zoals dat ook met gewone zoekmachines en directories het geval is, varieert de omvang daarvan aanzienlijk: van enkele tienduizenden blogs die worden gedekt, tot meer dan zeventien miljoen. Ik behandel hier daarom alleen de grotere zoekmachines, met de aantekening dat in het bovengenoemde Compendium ook speciale directories te vinden zijn van blogs voor speciale doelgroepen of uit bepaalde landen. De top vijf onder de zoekmachines zijn BlogPulse, Technorati, IceRocket, Feedster en, sinds kort, een bètaversie van een speciale blogzoekmachine van Google. Van deze vijf zijn Google Blog Search en Feedster eigenlijk zoekmachines die niet naar blogs als zodanig zoeken, maar naar 'feeds', waarover we het zo dadelijk nog zullen hebben. Verder hebben Technorati en Feedster een 'blog', resp. een 'feed finder', die het mogelijk moeten maken behalve in, ook naar blogs te zoeken. Met Google Blog Search kan dat ook, omdat deze zoekmachine het prefix `inblogtitle:` kent. Gezien wat hierboven is gezegd over de naamgeving van blogs is de waarde hiervan echter twijfelachtig.

The screenshot shows the homepage of 'nieuws.geenstijl.nl'. The navigation bar includes links for .home, GS, nieuwssubmit, archief, nieuwsbrief, rss, contact, goodies, and registreer. The main header features the word '.nieuws' in a blue box over a background image of satellite dishes. Below the header is a category menu with links for Laatste nieuws, Algemeen, Buitenland, Lifestyle, Provincie, Sport, and Tech&Internet. The main content area is divided into two columns. The left column contains a list of news items with timestamps, such as '20:44 Nawijn moet niet zeuren' and '18:30 Foto van leeuwin-palende tijger'. The right column features a featured article titled 'Blogzoeker van Google waardeloos' with a sub-headline 'Google heeft vandaag een nieuw type zoekmachine geïntroduceerd op blogsearch.google.com'. The article text discusses the introduction of a new search engine and the issue of spam blogs. At the bottom of the article, there is a footer with the text '.redactie | 14-09-05 | 16:13 | [link](#) | 18 reacties'.

Geenstijl, het bekendste van de naar schatting meer dan een half miljoen Nederlandse weblogs.

7.4.1. BlogPulse (<http://www.blogpulse.com>)

Ik begin met de BlogPulse, omdat die van de echte blogzoekmachines de meest geavanceerde zoekmogelijkheden biedt. Waar de meeste blog search engines weinig meer mogelijkheden bieden dan zoekacties met termen in een impliciete EN-relatie of met phrases, biedt het geavanceerde zoekscherm van BlogPulse ons de mogelijkheid om volledige Booleaanse vragen te stellen. Andere mogelijkheden zijn:

- beperking van de zoekactie tot een daterange of tot 7/30/60/90/**180** dagen (**NB:** BlogPulse gaat niet verder terug dan die 180 dagen, dus oudere berichten vindt u met deze zoekmachine niet!)
- zoeken naar blogs die naar een URL linken
- aantal zoekresultaten per pagina instellen op **10/25/50** per pagina
- de sortering kan worden ingesteld op **date**/relevance

Toegegeven, vergeleken bij de mogelijkheden van Google en Yahoo! is het niet indrukwekkend, maar het is genoeg om ons een eind op weg te helpen.

1. De eerste mogelijkheid om blogs te vinden die ons interesseren is via zoektermen. Stel nu dat we blogs willen vinden die over zoeken op het Web of over zoekmachines gaan, dan ligt het voor de hand dat we met één algemene term niet ver zullen komen. De term Google bijvoorbeeld levert meer dan 300.000 resultaten op, en zelfs met 'relevance' als sorteering verschijnen er bovenin de lijst geen posts van blogs die specifiek over zoekmachines gaan. Een stuk beter is al [Google Yahoo Ask A9](#), wat 96 treffers oplevert, met onder meer enkele posts van het blog van de Britse informatiespecialist Phil Bradley. Een aardige benadering is ook om een heel specifieke zoekterm te nemen, waarvan we mogen aannemen dat die alleen bij de experts bekend is. Een voorbeeld bij zoekmachines is Yahoo!'s prefix `originurlextension:`. Deze term leverde in juni 2005 bij BlogPulse 8 treffers op, waaronder posts uit [The Unofficial Google Weblog](#), [Search Engine Journal](#), [Internet News/Websearchguide](#) en [Researchbuzz](#).
2. Hebben we eenmaal één of twee interessante blogs gevonden, dan is het vinden van meer van hetzelfde een fluitje van een cent. Daartoe gebruiken we de mogelijkheid die BlogPulse biedt om blogs te vinden die naar een al bekende URL linken. Zo levert een zoekactie naar blogs die naar Search Engine Journal (<http://www.searchenginejournal.com>) linken een blog op dat Planet Search heet, met als ondertitel 'Aggregation of search related weblogs'.
3. Dit is niet alleen het mooiste resultaat dat we ons kunnen voorstellen, het wijst ons ook nog eens de weg naar een derde zoekmogelijkheid. Op de twee beweringen die in deze zin vervat liggen, ga ik hier nader in.

Hoezo een mooi resultaat? Omdat Planet Search (<http://www.johnhesch.com/planet/se/>) de posts van andere blogs die over zoeken gaan 'aggregeert', d.w.z. verzamelt. En het doet dat, zoals in de 'Subscriptions' blijkt, via 'feeds', een term waarop we zo dadelijk nog zullen ingaan. In elk geval bundelt Planet Search de posts van 21 andere voor ons interessante blogs, wat ons de moeite bespaart die een voor een te bekijken.

Planet Search doet bovendien wat veel andere blogs, die al dan niet aggregeren, ook doen: het noemt, onder Subscriptions, de blogs die het als bron heeft. Dit is karakteristiek voor het publicatiemedium weblog: weblogs zijn een kanaal waarlangs nieuws zich razendsnel verspreidt, ze zijn de rooksignalen, de grote trom van het Web. En zo'n overzicht van bronnen waaruit een weblog put, doorgeeft of citeert heeft ook een naam, ook al wordt die niet consequent door alle bloggers gebruikt: het is de 'blogroll'. Dat is dus een van de zoektermen (andere zijn 'subscriptions', 'links') waarmee we blogs kunnen vinden die hun bronnen noemen – en dus andere blogs die vermoedelijk geheel of ten dele over hetzelfde onderwerp gaan. Daarvoor moeten we dan wel een gewone zoekmachine zoals Google gebruiken, en niet een blogzoekmachine als BlogPulse, want die indexeert alleen de posts, niet de hele blogs (inclusief de sidebar met de blogroll).

Voor meer blogs en andere nieuwsbronnen over zoeken op het Web, zie Search Engine Detective (<http://www.pandia.com/searchworld/detective.html>). Een beroepsgroep die professioneel in zoeken geïnteresseerd is, zijn bibliothecarissen en informatiespecialisten. Voor de koppen van 264 bibliotheekblogs, zie [LISFeeds.com](#).

Planet Search

Aggregation of search related weblogs

June 01, 2005

Search Engine News

[Yet MORE Cool Tools that are also Free!](#)

(Direct Link) By Robin Nobles - Search Engine Guide - Here's Part 2 of our Cool Tools that are also Free. We all enjoy discovering free resources that genuinely merit some value, don't we?...

June 01, 2005 11:00 PM under Marketing

[Easy Tips for Adding some Zest to your Click Through Ratios](#)

(Direct Link) By John Alexander - Search Engine Guide - Of all of the other choices a potential visitor may have to choose from, within the top search results, you want to make yours just a little more unique, a little more...

June 01, 2005 11:00 PM under Marketing

[Are Google and Yahoo! Fund Suckers?](#)

(Direct Link) By Jennifer Laycock - Search Engine Guide - The phenomenal success of search engine leaders Google and Yahoo! is creating a sucking void in the bank accounts of other leading information companies....

June 01, 2005 11:00 PM under Industry

Subscriptions

[Google \(feed\)](#)
[Google Blogoscoped \(feed\)](#)
[Inside Google \(feed\)](#)
[Internet Search Engine \(feed\)](#)
[JenSense \(feed\)](#)
[John Battelle's Searchblog \(feed\)](#)
[MSN Search \(feed\)](#)
[Pandia Search World \(feed\)](#)
[Research Buzz \(feed\)](#)
[ResourceShelf \(feed\)](#)
[Search Engine Guide \(feed\)](#)
[Search Engine Journal \(feed\)](#)
[Search Engine Lowdown \(feed\)](#)
[Search Engine News \(feed\)](#)
[Search Engine Roundtable \(feed\)](#)
[Search Engine Usability \(feed\)](#)
[Search Engine Watch \(feed\)](#)
[Tim Bray On Search \(feed\)](#)
[Traffick \(feed\)](#)
[Unofficial Google Weblog \(feed\)](#)
[Yahoo Search \(feed\)](#)

About Planet Search

Planet Search is an aggregation of public weblogs on the topic of search engines. The opinions expressed in those weblogs and hence this aggregation are those of the original authors.

Planet Search. Rechts, onder Subscriptions, de 'blogroll'.

BlogPulse kan (vooraf, vanaf het Advanced Search-scherm) op actualiteit of op relevantie; standaard is actualiteit. In de resultatschermen toont het de treffers met een titel en de eerste regel(s) van de post, plus de URL van de post. Verder zijn bij elke treffer twee links beschikbaar, resp. met de tekst 'track conversation' en 'view blog profile'. De eerste biedt een interessant inkijkje in de structuur van de blogosphere. Uitgaande van de oorspronkelijke post, waarvoor de 'conversation' (bij posts in nieuwsgroepen zouden we van 'threads' spreken) wordt opgevraagd toont BlogPulse de posts van andere blogs die die post citeren, plus de blogs waardoor die posts zelf weer geciteerd worden. Daarbij geldt een standaard breedte van 25, d.w.z. per niveau worden niet meer dan 25 citerende blogs getoond, en een diepte van 3, d.w.z. verder dan het derde niveau van citeren gaat BlogPulse niet. Een voorbeeld biedt de '[conversation](#)' naar aanleiding van een post in The Blog Herald waarin de nieuwe blogzoekmachine Google Blog Search werd besproken.

Tools Overview

Trend Search

Featured Trends

Conversation Tracker

BlogPulse Profiles

General Overview

Focus on Technology

Home > Tools

Conversation Seed:

+<http://www.blogherald.com/2005/09/14/google-blog-search-reviewed/>

Want more?

- 09/14/2005 [Google Blog Search - First Look Review](#) (from: Blog Tips - ProBlogger)
- 09/14/2005 [Everybody's Talking About The New Google Blog Search](#) (from: Blog Marketing, Blog Promotion for Newbies)
- 09/14/2005 [Google Does Blog Search!](#) (from: Chris Pirillo)
 - 09/15/2005 [Google Blog Search Engine](#) (from: daveynin's thing)
- 09/14/2005 [Google Blog Search Launches](#) (from: Blog Blog)
 - 09/15/2005 [Google launches Blogsearch](#) (from: nithinkamath.info)
- 09/14/2005 [Google Launches A blog Search Engine](#) (from: Me, My Coke, And I)
- 09/14/2005 [Google Blog Search beta](#) (from: VoIP Blog - VoIP News, Gadgets)
- 09/14/2005 [Tivo / Dannon Yogurt / Claudia Deheza / Sigur Ros](#) (from: Prefixblog.com)
- 09/14/2005 [Google Blog Search](#) (from: SA Blog)
- 09/15/2005 [Google Blog Search](#) (from: Addicting Games Funny Junk Gag Video Clip - Kahsoon.com)
- 09/14/2005 [Google To Launch Blog Search](#) (from: Blog Tips - ProBlogger)
 - 09/14/2005 [Google adds blogsearch](#) (from: Sailing Close to the Wind)

Intelliseek's **BrandPulse™** tool analyzes all forms of consumer-generated media (CGM), including newsgroups, message boards and discussion forums. [Request a demo or more information](#), or call us toll free at 1-877-999-SEEK (7335)

Voorbeeld van een 'conversation'. De 'Conversation seed' is de post waarbij 'Track conversation' was aangeklikt. In deze gedeeltelijke weergave van de conversation ziet u 2 van de standaard 25 citerende posts op het eerste niveau. Daarvan heeft de eerste 8 citerende posts (tweede niveau), waarvan de tweede (van Chris Pirillo) er een op het [derde niveau](#) heeft.

De standaard breedte van 25 en de startdatum voor de citaties kunnen onderin het hierboven getoonde scherm worden gewijzigd.

Onder 'view blog profile' (zie volgende bladzijde) toont BlogPulse een aantal statistische gegevens over het blog in kwestie. Daartoe behoren onder meer de plaats die een blog op basis van aantallencitaties van andere blogs in de rangorde van door BlogPulse gedekte blogs inneemt, het gemiddelde aantal posts per week en de ontwikkeling van het aantal posts en het aantal citaties over de periode van een maand. Interessant zijn ook de 'Sources', die een overzicht geven van de bronnen die door het betreffende blog zelf geciteerd worden, en de 'Neighborhood', een tiental verwante blogs, waarbij de verwantschap berust op de mate waarin deze blogs eendere links en teksten citeren.

Dankzij deze uitgebreide linkanalyses heeft BlogPulse, dat volgens eigen opgave meer dan 16 miljoen blogs volgt, nog meer te bieden. Een van de doeleinden waarvoor blogs in het algemeen kunnen worden gebruikt is trend watching. Bloggers vormen (tot op zekere hoogte) een technologische voorhoede die 'het gesprek van de dag' weergeven. In dat gesprek van de dag vallen sommige termen vaker dan andere, en de verhouding tussen de frequentie van termen die tot hetzelfde interessegebied behoren, kan in grafieken worden weergegeven. Zo kan BlogPulse over een periode van zes maanden laten zien welke episodes van Star Wars het meest besproken werden, of hoe de termen iPod, MP3 en 'satellite radio' zich qua frequentie ten opzichte van elkaar verhielden. Een heel interessante om in 2005 te volgen is die van Nuclear threats: Iran vs. North Korea. In een Showcase laat BlogPulse zien hoe citatieonderzoek van blogs sociologisch en politicologisch kan worden uitgebuit. Ook wordt hier de mogelijkheid geboden om met de Blogpulse Trend Tool een eigen 'trend search' op te zetten.

De belangrijkste functies van BlogPulse staan vermeld in een horizontale kolom bovenin het venster. Een die hier nog niet is genoemd is 'Analyse', die populairste links opsomt (met de 'citations') die op een bepaalde datum in de door BlogPulse gedekte blogs zijn opgenomen.

BlogPulse Profiles Beta

The fun stuff available with BlogPulse Technology

Profile results for: [The Blog Herald: more blog news more often](#)New Profile Search:

OVERVIEW

POSTS

CITATIONS

TRENDS

SOURCES

NEIGHBORHOOD

General Information

Blog Title	The Blog Herald: more blog news more often
Blog URL	http://www.blogherald.com
Rank (by citation count)	48 (505 citations from 406 source(s) in past 30 days)
Post Frequency	29 per week

Rank Trend

A blog's rank is determined by the number of times it's cited by other bloggers over time. Rank is a moving average over 30 days.

Generated by BlogPulse Copyright 2005 Intelliseek, Inc.



Het 'profiel' van een blog bij BlogPulse.

7.4.2. Technorati (<http://www.technorati.com>)

Technorati is de grootste en de oudste van de hier behandelde blogzoekmachines. De kracht van deze zoekmachine ligt minder in de zoek- dan in de attenderingsmogelijkheden. Technorati 'kent' bijna 33 miljoen blogs en laat u die, net als BlogPulse, met zoektermen of een URL doorzoeken. Met de zoektermen vindt u 'posts' waarin deze voorkomen, met de URL van een site 'posts' waarin die site (of een pagina van die site) genoemd wordt. (De URL die u opgeeft hoeft zelf niet van een weblog te zijn. Zo vond ik in maart 2005 negen 'posts' in acht blogs bij de URL <http://www.hva.nl> van de Hogeschool van Amsterdam.) De term 'originurlextension' levert hier niets nuttigs op, maar de combinatie van 'exalead' en 'A9' gaf medio september 2005 20 'posts', afkomstig van onder meer [Phil Bradley's Blog](#), [Karen Blakemans Blog](#) en [The Search Lounge](#) van Chris Fillius, met besprekingen van zoekmachines. Technorati vermeldt bij zo'n zoekactie de duur van de periode waarover posts zijn gevonden. In dit geval 318 dagen. Het gaat dus aanzienlijk verder terug dan BlogPulse. Het oudste bericht dat ik vond was van 23 januari 2001, 1694 dagen vóór 14 september 2005, de datum waarop ik hiernaar zocht.

Net zomin als BlogPulse kan Technorati naar auteurs van blogs zoeken, maar evenals met BlogPulse is het mogelijk blogs te vinden naar de posts waarvan door anderen gelinkt wordt. Andere overeenkomsten met BlogPulse zijn dat Technorati posts kan

vinden die naar een site of pagina linken en dat het ook kan laten zien welke onderwerpen in de bloggerswereld de meest belangstelling trekken. Net als bij BlogPulse kun je via de posts de interessante blogs proberen op te sporen, maar een andere mogelijkheid is om dat via 'tags' te doen, trefwoorden die door de bloggers zelf aan hun blogs worden toegekend door een door Technorati beschikbaar gesteld stukje code aan hun blogs of posts toe te voegen. Gebruikers van Technorati kunnen zowel posts als hele blogs vinden waaraan een of meer trefwoorden zijn toegekend.

The screenshot shows the Technorati search interface. At the top, there are navigation tabs: Search, Tags, Blogs (selected), Popular, About, and Help. Below the tabs is a search bar containing the text 'searching' and a 'Search' button. The results section is titled '7 blogs about: searching'. There are three sorting options: 'Most Authority' (selected), 'Recently Updated', and 'Alphabetical'. The results are listed as follows:

- 1. A Consuming Experience**
By **Improbulus Improbulus**. Last updated **1 day ago**.
Tagged **A Consuming Experience**, **Blog Tools**, **Blogger**, **More >>**
140 links from 93 sites
- 2. Science Library Pad**
Thoughts on the use of technology and other issues for science libraries and science publishers.
By **Richard Akerman**. Last updated **15 days ago**.
Tagged **Conference**, **Digital Library**, **E-science**, **More >>**
46 links from 35 sites

On the right side of the results, there is a section titled 'Add Your Blog!' with the text: 'Do you blog about **searching**? There are two ways to add your blog to this page: **Become a member** and claim your blog to add your name, photo, and description. Or put **this code** on your site and we'll add you (no photo or description, though).'

Het resultaat van een zoekactie naar blogs met de tag: searching in Technorati.

Sorteren van posts die als treffers van een zoekactie worden gevonden, gebeurt met Technorati maar op één manier, op datum. Alleen bij het zoeken naar blogs met tags kunnen de resultaten op drie manieren worden gesorteerd, op 'Most Authority' (op basis van link popularity), actualiteit ('Recently Updated') en in alfabetische volgorde. Anders dan BlogPulse laat Technorati u van de zoekresultaten die u met de URLs van sites of met zoektermen vindt, zogeheten 'watchlists' maken. Hier manifesteert zich het attenderingsaspect van Technorati, want zo'n 'watchlist' attendeert u op nieuwe posts volgens het door u opgegeven vraagprofiel in de vorm van een zogeheten 'RSS feed'. Op wat dat is komen we zo dadelijk nog terug.

7.4.3. IceRocket (<http://www.icerocket.com>)

De derde grote blogzoekmachine is IceRocket, die in vergelijking met BlogPulse en Technorati weinig zoekmogelijkheden biedt. Wel biedt de Advanced Search de mogelijkheid om posts op auteursnaam te zoeken (= prefix `author:` in de Basic Search); daarnaast kan in de Advanced Search met 'all of these words', 'any of these words', 'the exact phrase' en 'none of these words' worden gezocht en kan daarbij op diverse manieren een 'Date Range' worden opgegeven; beperken tot een periode kan overigens ook achteraf via 'refine search to:'. Ook werkt IceRocket net als Technorati

met tags, die het in de resultatenlijsten laat zien (als 'topics', direct onder de titel van een treffer) en waarop je kunt klikken om alle posts met de betreffende tag te vinden. Daarbij blijkt dat het prefix `tag`: ook in de zoekregel kan worden gebruikt.

7.4.4. Feedster (<http://www.feedster.com>)

Te midden van andere blogzoekmachines is Feedster een geval apart. Feedster is namelijk niet zozeer een blogzoekmachine als wel een 'feedzoekmachine'. En om meer over Feedster te kunnen vertellen, moet ik eerst iets over 'Web feeds' of 'RSS feeds' uitleggen. Ik doe dat niet graag, want zoals iemand eens heeft opgemerkt: 'Explaining RSS is like explaining sex. You just don't get it until you do it.'

Om te beginnen dit: RSS staat voor 'Really Simple Syndication' (oorspronkelijk: 'Rich Site Summary') en is een standaard voor het publiceren van dynamische Web content (in het algemeen: nieuws op het Web).

Attenderingsdiensten zoals ze in het begin van dit hoofdstuk zijn besproken, gebruiken e-mail als kanaal om u over nieuwe webpagina's of veranderingen op bestaande webpagina's te informeren, en mailing lists, nieuwsbrieven en zelfs nieuwsgroepen doen of kunnen dat ook. Maar er is nog een ander kanaal waarlangs zulke informatie kan worden verspreid, en dat is via 'Web feeds' (of 'RSS feeds'). Dit zijn documenten met twee kenmerkende eigenschappen: in de eerste plaats zijn ze opgemaakt in een 'markup language' (vgl. HTML = 'hypertext markup language') die het mogelijk maakt ze te structureren alsof ze bestaan uit records in een database; deze taal heet XML, voor 'Extensible Markup Language', en RSS is dus een toepassing van die taal. Elk document bestaat uit een aantal van die, identiek gestructureerde, records, en als we ons voorstellen hoe een blog eruit ziet, wordt meteen duidelijk dat de 'posts' voor zo'n blog overeenkomen met de records in een 'Web feed'. Sterker nog, de oorsprong van Web feeds ligt in weblogs, waarin voor het eerst op grote schaal van XML gebruik werd gemaakt om dit bijzondere type documenten te structureren.

De andere kenmerkende eigenschap van 'Web feeds' is dat ze (althans de koppen ervan) per 'record' kunnen worden verspreid. Omdat dit in XML is, en niet in HTML, is er speciale software voor nodig om die records te kunnen lezen, en zoals vaker op het Web zijn daarvoor in principe weer twee oplossingen: u kunt die software – een zogeheten 'RSS reader' of 'feed reader' – op uw eigen computer hebben, of op een server op het Web. Een voorbeeld van het laatste zijn we al bij My Yahoo! (blz. 173 e.v.) tegengekomen; als u Web feeds op uw eigen computer wilt ontvangen en opslaan, hebt u een 'RSS reader' op uw eigen computer nodig. Zo'n programma, op een webserver of op uw eigen computer, leest de 'Web feeds' niet alleen, maar is ook geprogrammeerd om blogs – en andere webpagina's die daarvoor geschikt zijn gemaakt – te checken op nieuwe posts, en om vervolgens de koppen daarvan, met een link naar de post zelf, binnen te halen. Ze worden niet alleen 'readers' maar ook wel 'aggregators' genoemd, omdat ze feeds uit verschillende bronnen verzamelen en bijeenbrengen. Voor een overzicht van de talrijke RSS readers, op het Web en op uw eigen computer (Pocket PC, mobiele telefoon) zie het RSS Compendium(<http://allrss.com/rssreaders.html>).

De browsers Opera en Firefox kunnen inmiddels ook RSS-feeds lezen, en voor Internet Explorer zijn er plug-ins zoals Pluck (<http://www.pluck.com/>) die dat ook doen, maar de grotere functionaliteit van de meeste 'RSS readers' maakt het gebruik daarvan nog altijd te prefereren.

Hoewel RSS-feeds gebruik maken van XML, is het formaat van RSS feeds niet uniform maar is op verschillende plaatsen in diverse smaken ontwikkeld, zodat er verschillende versies van zijn. Het is niet zozeer een formaat als wel een familie van formaten. Voor ons als gebruikers van de informatie is dat niet zo belangrijk, want in de praktijk zijn

RSS readers erop berekend de diverse formaten te kunnen lezen en weergeven. Dat geldt ook voor een standaard die is ontworpen om de verschillen tussen de diverse formaten te overbruggen, Atom genaamd.

Hoewel de oorsprong van Web feeds dus in weblogs ligt, zijn ze allang niet meer alleen daartoe beperkt. Nieuwsbrieven, de resultatenpagina's van zoekmachines (MSN Search en andere), nieuwe items op webpagina's en zelfs nieuwe posts in nieuwsgroepen (Google Groups) kunnen in de vorm van Web feeds worden verspreid. In dit opzicht concurreren Web feeds dus met e-mail als verspreidingskanaal, een concurrentie die trouwens in een nog directere vorm twee kanten op werkt: er zijn diensten en programma's die e-mailberichten naar feeds vertalen (bijvoorbeeld [Mailbucket](#)), en andersom ([Bot a Blog](#)). Als u Opera als webbrowser gebruikt, zult u daarin een mogelijkheid aantreffen feeds als e-mailberichten te ontvangen (Opera 'leest' trouwens ook feeds); Firefox biedt u de mogelijkheid RSS feeds tot 'live bookmarks' te maken (via een icoontje in de statusbalk).

Tallose websites en webpagina's – en natuurlijk het overgrote merendeel van de blogs – bieden tegenwoordig feeds aan. Ze zijn te herkennen aan een link, een icoontje of een knop met de tekst XML, RSS, Atom of Syndication, en afhankelijk van welke oplossing u voor het lezen van de feeds kiest, moet met die link of de link onder het icoontje of de knop iets worden gedaan. Als u een 'feed reader' op uw computer hebt, is dat iets het toevoegen van de URL van de feed aan de feeds die u al volgt.

Feeds kunt u op verschillende manieren vinden, waarvan ik er hier drie noem.

- Met de prefixen [feed:](#) en [hasfeed:](#) biedt MSN Search de mogelijkheid feeds, resp. pagina's met een feed te vinden. Denk erom dat achter de dubbele punt van het prefix een spatie moet komen.

- Met behulp van de prefixen `filetype:` (Google en MSN Search) en `originurlextension:` (Yahoo!) kunt u via Google, MSN Search of Yahoo! Ook zelf naar feeds op zoek gaan. De extensies voor RSS feeds zijn `.xml`, `.rss` en `.rdf`. Wilt u dus van een bepaalde site weten of die ook feeds aanbiedt en zo ja welke, dan kan dat met de opdracht aan Google: `site:... filetype:xml OR filetype:rdf OR filetype:rss` (of een overeenkomstige opdracht in Yahoo!). Indien nodig kan, om de vraag toe te spitsen, als extra zoekterm nog `rss` worden toegevoegd. Voorbeelden:

Google: [site:nytimes.com rss filetype:xml OR filetype:rdf OR filetype:rss](#): 123 treffers ([vgl. Yahoo!](#))

Yahoo!: [site:bbc.co.uk AND rss AND \(originurlextension:xml OR originurlextension:rss OR originurlextension:rdf\)](#): 6290 treffers ([vgl. Google](#))

Op deze manier vindt u de feeds zelf, daarnaast bieden Yahoo! (`linkextension:`) en MSN Search (`contains:`) ook de mogelijkheid pagina's op te sporen waarop feeds worden aangeboden:

Yahoo!: [site:bbc.co.uk AND \(linkextension:xml OR linkextension:rdf OR linkextension:rss\)](#): 177.000 treffers ([vgl. MSN Search](#))

MSN Search: [site:nytimes.com AND \(contains:xml OR contains:rdf OR contains:rss\)](#) ([vgl. Yahoo!](#)).

- De derde mogelijkheid is via Feedster. Zoals eerder vermeld is Feedster niet zozeer een blogzoekmachine als wel een 'feedzoekmachine'. En net zoals BlogPulse en Technorati zoekmachines zijn die *in* en niet *naar* blogs zoeken, is Feedster in de eerste plaats bedoeld om *in* en niet *naar* feeds te zoeken. Dat zijn er meer dan 14 miljoen, en het gaat daarbij niet alleen om blogs, maar ook om feeds van andere nieuwsbronnen. Feedster levert met de zoektermen `originurlextension` en

Feedfinder :: Find the feeds you need

Use Feedfinder to search by keyword ([football](#), [microsoft security](#), [recipes](#), etc.) or by URL ([cnet.com](#), [itu.int](#), [microsoft.com](#), etc.) or a combination of the two ([amazon.com books](#)) in order to find feeds on that topic or by organization.

Search For:

[Add](#) - [About](#) - [Advertise](#) - [Blog](#) - [XML](#) - [Contact](#) - [Log In](#) - [Log Out](#) - [Privacy](#) - [Sign Up](#) - [Terms of Service](#) - [Help](#)

Copyright © 2003-2005 Feedster, Inc. All rights reserved. [Problems](#) or [suggestions](#)? We'd love to hear from you!

Feedfinder, de feeds zoekmachine van Feedster.

exalead A9 goede resultaten: voor [originurlextension](#) 21, voor [de combinatie van exalead en A9](#) niet minder dan 29. (Omdat Feedster maar een beperkte periode bestrijkt, variëren de resultaten natuurlijk in de tijd!) Die treffers kunnen vooraf op datum of relevantie worden gesorteerd. Bovendien biedt Feedster in een geavanceerd zoekscherm ook nog de mogelijkheid het zoekresultaat te beperken door in 'blogs only' te zoeken, of door de URL van een feed op te geven waarbinnen moet worden gezocht of die juist moet worden uitgesloten. Een voorbeeld is zoeken naar de term `exalead` met als filter: Limit to ONLY this RSS URL: <http://searchengineshowdown.com/blog/index.rdf>.

Feedster biedt nog meer aantrekkelijke mogelijkheden om blogs te vinden. In de eerste plaats heeft het ook een zoekfaciliteit *naar* blogs, Feedfinder genoemd. U kunt hier op twee manieren zoeken, met een of meer zoektermen of met een site (eventueel gecombineerd met een zoekterm). Het interessantst is zoeken met een site:

`nytimes.com` levert 87 treffers op, ongeveer hetzelfde aantal dus als we met Google: `site:nytimes.com rss filetype:xml OR filetype:rdf OR filetype:rss` vonden. Maar `bbc.co.uk` levert er 3338, dat is bijna zes maal zoveel als Google er vond.

De interactietaal waarvan u in de Feedfinder, maar ook in Feedster zelf, gebruik kunt maken is buitengewoon rijk. Ze omvat zoeken met

- de operatoren (impliciet) **AND**, **OR** en **NOT** (of -)
- de nabijheidsoperatoren **NEAR** (zoektermen in willekeurige volgorde, maximale afstand 10 woorden), **NEAR[#]** (zoektermen in willekeurige volgorde, maximale afstand # woorden), **NEAR[+#]** (zoektermen in de opgegeven volgorde, maximale afstand # woorden) en **{NEAR}** (aan het eind van een zoekvraag om alle voorgaande termmen in een NEAR-relatie met elkaar te verbinden)
- [phrases](#)
- [nesting](#)
- de uitdrukking **{RELAXEDBOOLEAN}** om partial matches in de zoekactie mee te nemen als een van de opgegeven zoektermen geen treffers oplevert
- de uitdrukking **{KEEPCASE}** om Feedster case-sensitive te laten zoeken

- de wildcards en truncatietekens `*` (truncatie midden en rechts, `?` (0 of 1 teken), `@` (0 of 1 letter), `%` (0 of 1 cijfer)
- prefixen als `title:`, `tld:`, `host`

Dit is maar een deel van het arsenaal dat Feedster de zoeker ter beschikking stelt; voor alle mogelijkheden (die naar mijn indruk niet allemaal werken), zie:

<http://www.feedster.com/help/search.php>.

Andere mogelijkheden van Feedster laat ik hier onbesproken, om in plaats daarvan nog even melding te maken van een mogelijkheid om blogs te vinden die ik zelf heel aardig vind: Andrew Grumets 'Find That Feed' (<http://grumet.net/syo/findThatFeed>). Op basis van gebruikersstatistieken laat deze site u eerst naar een feed zoeken met een naam die klinkt als ..., en laat u vervolgens zien op welke andere blogs de lezers van een bepaald blog geabonneerd zijn. Zo vindt u met de term 'battle' John Battelle's Searchblog, en vervolgens met `<sim>` andere blogs over zoeken. Hetzelfde principe dus dat Amazon op zijn site demonstreert met 'Customers who bought this book also bought...'

7.4.5. Google Blog Search Beta (<http://blogsearch.google.com/>, <http://search.blogger.com/>)

Van de zoekmachine die Google begin september 2005 in bèta geïntroduceerd heeft, bestaan twee versies, een met de bekende Google-interface en een met de interface van Google's eigen blogdienst Blogger. Functioneel is tussen beide geen verschil, zij het dat het zoeken naar posts die naar een adres (URL) linken, in de Advanced Search van de Blogger-interface expliciet wordt aangeboden en in die van de Google-interface niet. Hoe dan ook, de opdracht `link:www.hva.nl` levert alle posts in Google's blogdatabase op die naar een of meer pagina's van de Hogeschool van Amsterdam verwijzen.

Anders dan Technorati, BlogPulse en Ice Rocket indexeert Google Blog Search eigenlijk geen blogs maar `feeds`. Dat maakt de zoekmachine vergelijkbaar met Feedster, waarmee Google Blog Search ook gemeen heeft dat het niet alleen feeds van blogs indexeert, maar ook '`official news sources`'.

De zoekmachine is nog nieuw, en in de Advanced Search kan niet verder terug gezocht worden dan tot 1 maart 2005. Dat wil niet zeggen dat Google Blog Search geen oudere posts bevat: als die nog `op eenzelfde pagina` staan als die welke Google na 1 maart 2005 heeft gevonden, vindt Google ze ook.

Het aantal blogs dat door Google Blog Search wordt gedekt is op het moment dat ik dit schrijf, bedraagt volgens [The Blog Herald](#) bijna 9 miljoen, aanzienlijk minder dan door Technorati, BlogPulse en Feedster wordt opgegeven. Een van de oorzaken hiervan zou de nieuwheid van de zoekmachine kunnen zijn.

Van de zoekmogelijkheden noem ik in de eerste plaats de prefixen `allinblogtitle:` en `inblogtitle:`, dat zoeken *naar* blogs mogelijk maakt, zij het dat het resultaat van een zoekactie met dit prefix niet een lijst van blogs is met de zoekterm(en) in de titel, maar van de posts die tot zulke blogs behoren. Al te veel moet men zich de mogelijkheden om met titelwoorden naar blogs te zoeken niet voorstellen, de titels van blogs (Slashdot, Geenstijl enz.) zeggen vaak niet veel over de inhoud ervan. Daarbij komt dat Google de 'descriptions' van blogs, waarin vaak meer informatie over die inhoud wordt gegeven, niet mee indexeert.

Andere prefixen die in Google Blog Search kunnen worden gebruikt zijn `inpostauthor:` (werkt `soms`, maar `niet altijd`, vgl. [hier](#)), `allinposttitle:` en `inposttitle:` (overeenkomend met `allintitle:`, resp. `intitle:` in de webzoekmachine), `blogurl:` (= `site:` in de webzoekmachine, maar bij `blogurl:` kan ook een deel van een URL, bijvoorbeeld 'geenstijl' in plaats van 'geenstijl.nl' worden opgegeven). Mogelijkheden die alleen in de Advanced Search beschikbaar zijn, zijn het aantal treffers per pagina: 10/20/30/50/100, de actualiteit van de te vinden posts:

anytime/last hour/last 6 hours/last 12 hours/lastDay/last Week/last Month of een zelf op te geven periode, en resultaat wel of niet laten filteren door SafeSearch.

De operatoren, impliciet [AND], OR en – werken net als in de zoekmachine, evenals de [joker *](#); maar `numrange:` werkt [niet](#).

Google Blog Search sorteert zijn resultaten standaard op relevantie, maar biedt op het resultatenschem ook sorteren op datum aan. Verder bevat het resultatenschem in sommige gevallen een verwijzing naar 'Related Blogs'. Dit zijn blogs met de zoektermen in de titel.

Ten slotte biedt Google Blog Search onderaan het resultatenschem de mogelijkheid 10 dan wel 100 treffers als Atom- of RSS-feed te downloaden en in een RSS-reader te plakken.

Helaas is er bij de treffers geen cacheversie.

7.4.6. *Tot slot*

RSS feeds hebben diverse voordelen boven e-mail. Een daarvan is dat ze nieuws in afgemete hoeveelheden als koppen brengen en dus een snelle beslissing wat u wel of niet de moeite waard vindt om te bekijken, mogelijk maken. Een ander voordeel is dat feeds als kanaal lang niet zo door spam vervuild worden als uw postbus. En ten slotte is het heel gemakkelijk weer van een feed af te komen: u hoeft hem alleen maar uit uw reader te verwijderen. Dat u van feeds af zult willen, is een ding dat zeker is. Over welk onderwerp u ook geïnformeerd wilt worden, er zijn meer feeds over beschikbaar dan u zult kunnen behappen.

Het zoekproces

In het laatste hoofdstuk van dit boek gaan we het over u hebben: de man of vrouw aan de knoppen van al die zoekinstrumenten die in de vorige hoofdstukken de revue gepasseerd zijn.

U arriveert per trein op het station van een grote stad, waar u iets of iemand moet bezoeken. Zo kunnen we ons het begin van een zoektocht op het Web voorstellen. Er zijn nu verschillende situaties denkbaar die het vervolg van uw zoektocht kunnen bepalen:

8.1 Known-item search

U weet precies waar u zijn moet. In de stad betekent dat: u kent de naam en adres van de persoon of de instelling die u wilt bezoeken, en/of u weet hoe u daar moet komen: met bus 16 tot de halte Parklaan, oversteken en dan de derde straat links. Op het Web betekent het: u kent de URL van de pagina die u wilt bezoeken, of de URL van een pagina die u daar direct of via een bekende route heen voert. In het eerste geval typt u de URL in de adresregel van uw browser en komt direct op de plaats waar u moet zijn, in het tweede typt u de URL van een verwijzende pagina in en komt indirect op uw bestemming aan.

Als ik iemands telefoonnummer moet opzoeken, ga ik altijd eerst naar Startpagina. De link naar twee telefoongidsen staat daar ergens bovenaan in de middelste kolom, dus in twee stappen ben ik waar ik zijn moet.

Waarom onthoud ik het adres van de telefoongids zelf niet, in plaats van dat van Startpagina. Wel, in de eerste plaats omdat er meer telefoongidsen zijn, met verschillende mogelijkheden en resultaten, maar ook omdat ik Startpagina ook gebruik om de pagina van het busvervoer te vinden (www.9191ov.nl, of was het www.2121ov.nl?), eerste kolom, ergens in het midden, en zo nu en dan een kaartje onder Routeplanners (eerste kolom, meen ik, iets verder naar beneden). Als ik in een grote stad drie adressen (antiquariaten, tweedehands CD-winkeltjes) in één buurt moet bezoeken, is het simpeler te onthouden dat ik bus 16 moet nemen en op halte Parklaan moet uitstappen, dan voor elk adres afzonderlijk een aparte route uit te stippelen. In de bibliotheekwereld spreken we in zo'n geval van een 'known-item'-search. Of de zoeker loopt direct naar de studiezaal en kast waar het gezochte boek staat, ofwel hij/zij gebruikt een catalogus om de plaats van het boek, waarvan hij/zij de titel kent, op te sporen.

In een situatie als deze hebt u op het Web geen zoekmachine, directory of ander zoekinstrument nodig. (Ja, ik weet het, Startpagina is ook een soort zoekinstrument, maar is hier maar een toevallig voorbeeld; ik kan ook de homepage van de KB gebruiken om bij de database van BL-Online te komen.)

Er zijn nog andere mogelijkheden om zonder gebruik van zoekinstrumenten known-item searches te doen. Typ in de lege adresregel van uw Firefox-browser in: openbaar vervoer, geef Enter, en u komt direct op de pagina van 9292ov.nl | OV-reisinformatie; typ: scheepvaartmuseum, en u komt op de homepage van het Nederlands Scheepvaartmuseum Amsterdam; typ: Nederlandse Spoorwegen, en u krijgt meteen een treinplanner op uw scherm te zien. Voor simpele vragen die in trefwoorden te vangen zijn is dit een snelle weg, en zelfs telefoonnummer Smit Leiden levert meteen het gewenste resultaat (maar mij vindt u zo niet). Firefox maakt hiervoor gebruik van Google.

Dat brengt ons op zoekmachines. Known-item searches zijn ook zoekacties in die gevallen waarin u precies weet wat u te weten wilt komen en hoe het moet worden benoemd. Een zoekmachine is in zo'n situatie het voor de hand liggende instrument, omdat hij documenten fulltext indexeert. Dat betekent dat de kans om het gezochte

document te vinden mits u het jargon dat bij de vraag hoort maar kent, behoorlijk groot is. Zelfs al zou u bij het eerste schot te veel, te weinig of de verkeerde zoektermen gebruiken, bij een tweede of een derde keer lukt het wel.

Als u een grote, mondiale zoekmachine gebruikt, zult u hoogst waarschijnlijk kunnen profiteren van het geavanceerde rankingmechanisme waarover zo'n zoekmachine beschikt. Een aantal van de factoren die zoekmachines daarbij mee laten wegen, zijn op blz. 54 opgesomd. Niettemin is het in veel gevallen verstandig gebruik te maken van de niet minder geavanceerde mogelijkheden die de zoekmachine u biedt om uw vraag – en daarmee het antwoord van de zoekmachine – fijn te regelen. Hoe geavanceerd de rankingalgoritmen van de grote mondiale zoekmachines ook zijn, en hoe succesvol ze vaak ook zijn in het gissen naar onze bedoelingen, des te meer zinvolle informatie we ze in de zoekvraag verschaffen, des te beter de zoekresultaten zullen zijn en des te minder tijd we zullen hoeven besteden aan het vinden van het juiste antwoord op onze vraag.

We hebben gezien dat MSN Search ons als gebruiker de mogelijkheid biedt zoekresultaten op die verschillende manieren te sorteren: op de overeenkomst tussen zoektermen en documenten die die termen bevatten (exactness), op (link) populariteit en op actualiteit. Als we in het bewustzijn van de mogelijkheid van die verschillende sorteringen naar de zoekresultaten van zoekmachines als Google en Yahoo! kijken, die hun gebruikers niet in die vorm bieden, gaan we beseffen dat ergens in hun rankingalgoritme een compromis van 'van alles wat' verscholen zit: bij, zeg, de eerste treffers zitten er waarbij exactheid, populariteit of actualiteit net iets meer gewicht in de schaal legt dan de beide andere factoren. Nu, term weighting, hoewel niet vaak gebruikt, kan ook bij Google en Yahoo! een middel zijn om de ranking van de gevonden documenten op het punt van exactness te beïnvloeden, en beperking van het zoekresultaat tot die documenten die de laatste week of maand zijn verschenen of herzien geeft wat meer gewicht aan de factor actualiteit.

We helpen een zoekmachine ook door de ruimte te beperken waarin hij moet zoeken. De gTLD's die op blz. 7 zijn genoemd, kunnen aspecten van onze vraag tot uitdrukking brengen die op een andere manier nauwelijks in de vraag onder te brengen zijn: willen we informatie over een geneesmiddel die van een bedrijf afkomstig is, van een overheidsinstelling of van een wetenschappelijke instelling? Net zo kunnen vraagbeperkingen met ccTLD's (blz. 7) helpen: moet de Nederlandstalige informatie waarnaar we op zoek zijn uit Nederland of uit Vlaanderen afkomstig zijn, of misschien juist uit geen van beide?

De grote mondiale zoekmachines zijn allemaal dol op frases. Dat komt doordat phrases meestal op een eenduidiger manier begrippen aanduiden dan woorden dat doen. Ook al zijn de verschillen in het gebruik van phrases in de verschillende talen (bijvoorbeeld Engels en Nederlands) aanzienlijk, desondanks verdient het *altijd* aanbeveling te overwegen of het document dat u zoekt niet een of twee phrases zou moeten bevatten. Behalve begrippen omschrijven phrases ook antwoorden, en documenten op het Web staan daar vol van; dus als u op zoek bent naar het antwoord op een vraag, ligt het voor de hand dat dat antwoord te vinden is in een phrase die het document bevat: "Beethoven wrote his violin concerto in".

Net als een boek of een krant hebben documenten op het Web een structuur: sommige delen in het document zeggen meer over de globale inhoud of het onderwerp van het document dan andere. Webdocumenten hebben een adres, de URL, en een titel, en de rankingalgoritmes van zoekmachines maken daar mede gebruik van om de door hen gevonden documenten te sorteren. Maar omdat het voorkomen van een zoekterm in die titel of URL maar één van de vele factoren is die bij die ranking een rol spelen, kunt u het gewicht van die factor verhogen door aan te geven waar in het document de zoekterm(en) wel of niet moet(en) voorkomen. Met name de URL biedt interessante

mogelijkheden: denk eens aan de bedrijven die gerichte mailings doen op basis van postcodes; net zoals bewoners van bepaalde buurten aan bepaalde profielen voldoen, kunnen webpagina's met bepaalde adreselementen dat ook doen. Maar ook de titel van een document is een belangrijk aanknopingspunt voor wie informatie over een onderwerp zoekt. Een interessante aanvulling op de mogelijkheden die de grote mondiale zoekmachines bieden om een zoekactie tot de titel van documenten te beperken, is het prefix `inanchor:`, waarmee documenten kunnen worden gevonden onder titels (hyperlinks) die anderen dan de auteur zelf eraan toe hebben gekend. Webdocumenten bevatten niet alleen letters en woorden, maar ook cijfers en getallen. Die vormen een taal die veel preciezer is dan alleen maar tekst, en van dat feit kunnen we gebruik maken om sneller bij de gewenste informatie te komen dan mogelijk zou zijn als we alleen met sleutelwoorden zouden moeten zoeken. Als er een kans bestaat dat het door u gezochte document een getal bevat dat u kent of waarnaar u bij benadering kunt gissen, gebruik dat getal dan als een van uw zoektermen of gebruik Google's range-operator `~`.

Documenten op het Web staan vrijwel nooit op zichzelf: het Web zou geen web zijn als het alleen maar uit losse punten bestond. Links leggen verbanden tussen (delen van) documenten die over hetzelfde gaan, en voor wie daar niet genoeg aan heeft kunnen de grote mondiale zoekmachines op basis van nog andere criteria gerelateerde documenten opzoeken. Zoeken op deze manier representeert een van de natuurlijkste en meest toegepaste manieren van kennisverwerving: zoeken naar wat je niet weet door aan te haken bij wat je wel weet. In die zin is zoeken met `link:` en `related:` ook een soort `known-item` zoeken. Een interessante aanvulling op deze manieren van zoeken is het gebruik van een directory om documenten te vinden die door mensen (redacteuren) zijn beoordeeld als te behoren in dezelfde rubriek of categorie als het al bekende of gevonden document. De zoekmachine Exalead biedt daartoe consequent een directe mogelijkheid, andere zoekmachines doen dat bij gelegenheid – als een treffer ook in hun directory voorkomt – ook, maar ook als dat niet het geval is kunt u natuurlijk altijd kijken of een via een zoekmachine gevonden treffer ook in een directory – niet noodzakelijk de directory van de gebruikte zoekmachine voorkomt.

Zoeken naar gelinkt of gerelateerd is zoeken naar meer van hetzelfde. Net zoals het via categorieën of rubrieken mogelijk is meer items te vinden die tot die rubrieken of categorieën behoren of gerekend worden, zo is het door het zoeken naar items met dezelfde hoedanigheden mogelijk zowel een omvattende categorie of rubriek te vinden, als andere items die tot diezelfde categorie of rubriek behoren. [Google Sets](#) vormt een illustratie van dat principe, maar het principe zelf is universeel toepasbaar.

Het grote merendeel van de documenten op het Web bestaat uit bestanden in HTML, maar daarnaast zijn er nog allerlei andere bestandsformaten met een vaak heel specifiek karakter. Maak daar gebruik van. Een PowerPoint-presentatie is – ook inhoudelijk – iets anders dan een PDF- of een PostScript-bestand, een audiobestand of een bestand waarin op de een of andere manier beelden een rol spelen. Met `filetype:` (Google, MSN Search) en `originurlextension:` (Yahoo!) kunt u die speciale bestandsformaten opsporen, en indirect ook via `linkextension:`, `outgoingurltype:`, `feature:` (alledrie Yahoo!) en `contains:` (MSN Search).

Bij `known-item` zoeken hoeft het antwoord op een vraag niet onmiddellijk te worden gevonden in een van de treffers die een zoekmachine biedt. Soms kan het nodig zijn een van die treffers, waarin het antwoord niet staat, wat nader te bekijken en dan een link te volgen naar een niet door de zoekmachine gepresenteerd (of lager gerankt) document waarin het antwoord wel staat.

8.2. Re-finding

Een iets andere situatie is die waarin u een plaats moet zien terug te vinden waar u

eerder geweest bent. Het adres (de URL) herinnert u zich niet meer, en ook de route daarheen, inclusief het gebruikte vervoermiddel (bus, zoekmachine), staat u niet meer helder voor de geest; het enige wat u nog weet is dat u er een maand of twee geleden geweest bent, dat u langs een brede weg met aan weerszijden bomen (de tweede resultatenpagina in de gebruikte zoekmachine) bent gekomen en dat het gezochte gebouw (URL) een balkon (domain name tv) had. Ook herinnert u zich dat het u indertijd nogal wat moeite heeft gekost het adres te vinden. Tja, daar zijn bookmarks voor...

In zo'n geval kunt u weer gebruik maken van de mogelijkheden die uw browser biedt. Deze houdt namelijk een lijst bij van de door u bezochte adressen en toont die als u een URL intypt of op het pijltje achterin de adresregel klikt.

Maar uw browser biedt veel uitgebreidere mogelijkheden om eerder bezochte URLs terug te vinden. Zowel in Internet Explorer als in Firefox roept de toetsencombinatie Ctrl-H de history-zijbalk op, waarin de door u bezochte adressen standaard op datum gesorteerd worden getoond; andere sorteringen zijn ook mogelijk, en beide browsers bieden de mogelijkheid de history-database van uw browser (want dat is het) te doorzoeken. Daarbij zoekt het programma naar strings, niet naar woorden, zowel in de titels als in de URLs van de bezochte pagina's, en is het bijvoorbeeld mogelijk met .tv naar pagina's met die ccTLD te zoeken.

Zowel Internet Explorer als Firefox heeft een standaard aantal dagen dat de geschiedenis van bezochte adressen bewaard wordt. In beide is dat aantal aan te passen, maar bij Internet Explorer (Extra – Internet Opties – Geschiedenis) is het maximum 99 dagen; in Firefox (Tools – Options – Privacy – History) kan de standaard bewaartijd van 9 dagen in een veel groter getal (bijvoorbeeld 1000) worden gewijzigd, maar of dit er inderdaad toe leidt dat de URLs van bezochte pagina's jarenlang bewaard blijven heb ik niet met zekerheid kunnen vaststellen. Mijn indruk, gebaseerd op enig geëxperimenteer met de datum van mijn computerklok, is van wel, maar dan wordt alles wat ouder is dan zes dagen in één grote folder ondergebracht.

De 'browsing history' van uw browser laat u naar strings in de titels en de URLs van door u bezochte pagina's zoeken, maar er zijn gevallen waarin dat niet voldoende is; wat u eigenlijk zou willen is de vragen terugzien en terugvinden die u aan een zoekmachine gesteld hebt. Dat kan, op een beperkte schaal, als u uw browser uitrust met een of meer toolbars van zoekmachines. In de zoekregel daarvan zult u dan enkele tientallen vragen terugvinden die u eerder gesteld hebt. In de Groowe Search Toolbar (zie blz. 268), die tevens als springplankzoekmachine dienst doet, geldt dat vragen die u in diverse zoekmachines hebt gesteld. Maar de mogelijkheden hiervan zullen meestal te beperkt zijn. Voor wie meer wil bieden de diverse grote mondiale zoekmachines ook gepersonaliseerde zoekdiensten aan, waarvan ik Google Personalized Search Beta (<http://www.google.com/psearch>) als voorbeeld neem. Deze dienst, beschikbaar voor iedere geregistreerde gebruiker van Google (bijvoorbeeld gebruikers van Gmail), bewaart uw hele zoekgeschiedenis (voor zover gedaan met de webzoekmachine, Google Images, Google News of Froogle, in de vorm van de zoekopdrachten met daarbij treffers die u vervolgens bekeken hebt. Dit is gecombineerd met een bookmark manager, die u in staat stelt uw bookmarks bij Google op te slaan.

Google Personalized Search laat u niet alleen door uw zoekvragen en de daarbij aangeklikte resultaten browsen, maar binnen de ruimte van die zoekresultaten ook zoeken. Het gaat daarbij niet enkel om de zoekvragen en de titels van de aangeklikte treffers, maar fulltext binnen de ruimte van alle pagina's die u bekeken hebt.



Search History (Beta) for MRJHofstede@gmail.com

Search History

- Web
- Images
- News
- Froogle

[Select all](#)[Pause](#)[Remove items](#)[Trends](#)

Bookmarks

[Add bookmark](#)

Mar 31, 2006

"My Google" - [Related history](#)

- [Google Launches "My Google" Style Feature](#) - 12:56am
blog.searchenginewatch.com/blog/050519-141359

Mar 30, 2006

Firefox "browsing history" Or "visited pages" 99

- [Information Search and Re-access Strategies of Experienced Web Users](#) - 8:57pm
www.2005.org/cdrom/docs/p583.pdf

Firefox "browsing history" Or "visited pages"

- <http://librenix.com/?page=Firefox> - 8:31pm

antiquariaat leiden

- [Antiquariaat van Paddenburgh - Photographic Literature](#) - 1:17pm
www.antiqua.nl/paddenburgh/



Google Personalized Search Beta. Rechts een kalender waarin u toegang hebt tot de zoekvragen en -resultaten van eerdere dagen en maanden. De link Trends in de linker kolom geeft toegang tot enkele statistieken van uw zoekacties.

Mits u ervoor zorgt bij Google ingelogd te zijn als u met die zoekmachine zoekacties uitvoert, scheidt u zo een persoonlijk archief van zoekacties en -resultaten waarin dingen die u eerder gezien hebt maar niet gebookmarkt of opgeslagen, toch redelijk gemakkelijk terug te vinden zijn.

Diensten als Google's Personalized Search, en natuurlijk ook Gmail, brengen mee dat allerlei privacy-gevoelige zaken op de servers van een bedrijf komen te staan waar personen en instanties die daar niets mee te maken hebben inzage kunnen krijgen of vorderen. In feite is dat al zo als u een zoekmachine gebruikt om iets op te zoeken, of als u zomaar een site bezoekt. Zoals op blz. 3 is uiteengezet, heeft de computer waarop u werkt, op het moment dat u zich op het Web begeeft een IP-adres, en dat is – als uw provider meewerkt of wordt gedwongen dat te doen – vrijwel altijd te achterhalen. Hoe dan ook, als u uw zoekgeschiedenis niet aan Google of een andere zoekmachine wilt toevertrouwen maar een kopie van de door u bezochte pagina's op uw eigen computer wilt opslaan, kunt u daarvoor, mits u gebruikt maakt van Internet Explorer, de Seruku Toolbar (zie: <http://www.seruku.com/Toolbar/HowTheToolbarWorks.html>) gebruiken.

8.3. Exploratory search

Bij dit type vragen arriveert u op het station van een stad waarin u de weg niet kent. U weet wel waar u naartoe wilt, maar om te beginnen al niet of u de voor- of de achterzijde van het station uit moet, of u de bus, de metro, de taxi of de benen moet nemen, kortom, voordat u zich op weg kunt begeven zult u zich eerst moeten oriënteren.

Vertaald naar het Web gaat het hier om vragen waarbij u wel weet wat u te weten wilt komen, maar niet goed hoe u dat onder woorden moet brengen en waar of waarmee u moet zoeken. Een voorbeeld vormen vragen met een statistisch aspect, waarbij de vraag 'hoeveel...!', 'hoe groot...!', 'hoelang...' vaak moet worden vertaald naar het antwoord 'statistiek van...!'

Op het Web kunt u bij dit soort vragen het beste bij een directory als die van Google of Yahoo!, een directory plus als About.com of een startpagina zoals die van Startnederland.nl beginnen. Die bieden de mogelijkheid vragen vanuit een meer algemeen perspectief te benaderen. Op een station zou u eerst eens naar een stadskaart

op zoek gaan, in een bibliotheek misschien eerst maar eens naar een encyclopedie. Een artikel daarin behandelt in het kort de voornaamste aspecten van een onderwerp en laat u kennismaken met het 'jargon'. Net zo kan op het Web gebruik worden gemaakt van Wikipedia (http://en.wikipedia.org/wiki/Main_Page), of, nog beter, van Answers.com (http://www.answers.com/main/top_five_reasons.jsp). Voor meer gespecialiseerde vragen komen gespecialiseerde portals of directories in aanmerking – te vinden als Web directories via de directory van Yahoo! – of gespecialiseerde online encyclopedieën zoals de Stanford Encyclopedia of Philosophy, The Canadian Encyclopedia of de Encyclopaedia Mythica. Bij mijn weten bestaat er geen omvattend overzicht van alle belangrijke encyclopedieën op het Web, maar via de directories van Google en Yahoo! moeten de meeste ervan wel te vinden zijn. In het algemeen gesproken kunt u bij dit soort vragen het beste beginnen in de afdeling Naslagwerken, waartoe niet alleen encyclopedieën, woordenboeken, adresboeken etc. behoren, maar ook sites zoals HowStuffWorks (<http://www.howstuffworks.com/>) en How Things Work (<http://rabi.phys.virginia.edu/HTW/>), die antwoorden geven op een hele categorie van vragen die met een gewone directory of zoekmachine moeilijk te beantwoorden zijn. Overzichten van algemene en speciale naslagwerken bieden het artikel Nachschlagewerke im Internet van de Duitstalige Wikipedia (http://de.wikipedia.org/wiki/Wikipedia:Nachschlagewerke_im_Internet), Martindale's 'The Reference Desk' (<http://www.martindalecenter.com/>) en de pagina Enzyklopädien, Lexika, Wörterbücher (<http://www.fh-augsburg.de/informatik/projekte/mebib/fidb/lexika.html>) van de Fachhochschule Augsburg. De Librarians' Internet Index heeft een selecte lijst van Ready Reference & Quick Facts (<http://www.lii.org/pub/topic/reference>).

Als u een grote mondiale zoekmachine gebruikt om zoekacties van dit type te ondernemen, kunt u gebruik maken van diverse hulpmiddelen om de benodigde zoektermen op het spoor te komen.

Een grote hulp bij het zoeken in catalogi en databases kunnen alfabetische indexen vormen, omdat ze de alfabetische context van een zoekterm laten zien, en uit die context suggesties voor andere zoektermen kunnen worden gehaald. [Google Suggest](#) biedt hier de illustratie van het principe, maar bij zoekmachines (Exalead) die truncatie van zoektermen ondersteunen kan daar op een meer algemene manier gebruik van worden gemaakt.

Trunceren is een van de manieren waarop OF-relaties tussen zoektermen kunnen worden gelegd. Zoeken met OR is dé manier om uw netten verder uit te werpen, en de OF-relatie hoeft natuurlijk niet alleen maar gelegd te worden tussen zoektermen die alfabetisch bij elkaar in de buurt liggen. Juist bij dit soort zoekacties, waarbij de precieze te gebruiken termen niet bekend zijn, kan het gebruik van semantisch verwante zoektermen in een OR-relatie nodig zijn om het antwoord op uw vraag, of althans de juiste zoektermen, te kunnen vinden. Daartoe biedt Google de zogeheten 'synoniem-operator' ~, die u met de opgegeven zoekterm en een aantal daaraan verwante termen (breder, synoniem, nauwer, gerelateerd) laat zoeken. In plaats daarvan kan ook gebruik worden gemaakt van een thesaurus, een woordenboek of woordenlijst waarin begrips- of betekenisclusters bij elkaar zijn gebracht. Diverse van de grote mondiale zoekmachines bieden daartoe een directe toegang nadat u eenmaal met een of meer zoektermen een search hebt gedaan. Zo heeft Google in de regel waarin het zijn aantal resultaten meldt een [definition]-link, die naar het hierboven genoemde Answers.com met onder meer de zoekterm als thesaurusterm voert.

Suggesties voor andere zoektermen die aan een zoekvraag kunnen worden toegevoegd om die te preciseren, worden gedaan door zoekmachines en metazoekmachines die categorisatie ondersteunen.

8.4. Don't know what you need to know

In deze situatie arriveert de treinreiziger op het station van een grote stad zonder een vast omlind idee van wat hij of zij daar wil. Een voorbeeld op het Web kan zijn dat iemand de homepage van een organisatie of instelling bezoekt om meer over producten, organisatie, openingstijden en voorzieningen te weten te komen. Of hij/zij kan er een navigatiesysteem, een site search engine of een sitemap hopen te vinden die hem/haar op een simpeler manier bij een resultaat brengt dan een grote nationale of mondiale zoekmachine doet. Op de homepages van veel instellingen en organisaties verschijnt het laatste nieuws, bijvoorbeeld over aankopen, tentoonstellingen, speciale sluitingstijden, dat nog niet door zoekmachines geïndexeerd is. Het gaat hierbij vaak om informatie die in een verband moet worden gezien, bij voorbeeld bij de vraag: Welke cursussen in de filosofie die mij vijf studiepunten opleveren worden er aan de Leidse universiteit in het collegejaar 2005-2006 gegeven die in de periode augustus-december niet op een donderdag vallen? Zulke vragen ontstaan meestal gaandeweg, en het is ook niet zinvol de antwoorden daarop op een andere manier dan via homepages te benaderen.

Homepages – ik bedoel daarmee niet 'persoonlijke homepages'-- zijn op diverse manieren te vinden. Yahoo! heeft daarvoor twee prefixen: `depth=0` en `feature:index`, die aan een zoekvraag kunnen worden toegevoegd, maar de meeste sites hebben vanaf vrijwel elke pagina wel een link naar hun eigen homepage. Wie nog sneller de was wil doen bij dit type zoekacties kan ook direct via een zoekmachine naar de sitemap of de A-Z index van een site op zoek.

8.5. Visible en Invisible Web

In hoofdstuk 5 van dit boek is het Invisible Web ter sprake gekomen, de wereld van dynamische en anderszins niet door zoekmachines geïndexeerde pagina's die daardoor buiten het bereik van de standaard zoekinstrumenten op het Web blijft liggen. Een van de lastigste vragen die zich bij het zoeken op en via het Web voordoen, is wanneer je moet besluiten dat het antwoord op een vraag niet op het Web maar in het Invisible Web moet worden gezocht – als het al te vinden is!

Een oplossing voor dat probleem valt niet te geven, simpelweg omdat visible en invisible web niet duidelijk van elkaar zijn afgebakend. Welke pagina's door zoekmachines worden geïndexeerd, en welke niet valt in zijn algemeenheid niet te zeggen, enkel dat pagina's met veel inkomende en uitgaande links of waar veel bezoekers langs komen op weg naar een andere pagina, een grotere kans maken door de spider van een zoekmachine te worden bezocht dan pagina's waarvoor dat allemaal niet geldt. Wat dynamische pagina's betreft, resultaat van een zoekactie in een database, valt ook nauwelijks te voorspellen of ze in de resultaten van een zoekmachine zullen opduiken of niet.

De enige raad die ik u in dit opzicht kan geven is om u vertrouwd te maken met de diversiteit van het Invisible Web en het soort vragen dat daar gesteld kan worden, zodat als het moment komt waarop het antwoord op een vraag zich met alle gebruikelijke zoekinstrumenten op het zichtbare Web niet laat vinden, het idee bij u opkomt op zoek te gaan naar die speciale database waarin het misschien wel te vinden zal zijn.

Zie: Donna Maurer – Four Modes of Seeking Information and How to Design for Them - Boxes and Arrows

(http://www.bboxesandarrows.com/view/four_modes_of_seeking_information_and_how_to_design_for_them)

Aanbevolen: Search Strategies: Search with Peripheral Vision

(<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html>)

The Telson Spur – Internet Search (<http://www.snark.org/search.htm>)

Fravia's web-searching lore (<http://www.searchlores.org/>).

Appendix A De mogelijkheden van drie grote mondiale zoekmachines

<i>Functie</i>	<i><u>Google</u></i>	<i><u>Yahoo!</u></i>	<i><u>MSN Search</u></i>	<i>Opmerking</i>
Algemeen				<p>Op deze pagina vindt u de zoekmogelijkheden van de drie grootste mondiale zoekmachines – Google, Yahoo! en MSN Search – bijeengebracht, met hier en daar een verwijzing naar speciale mogelijkheden van drie andere grote mondiale zoekmachines: Ask, Gigablast en Exalead. Grote mondiale zoekmachines bieden veel mogelijkheden om informatie op het web te vinden, maar die zijn algemeen van aard. Om u een idee te geven van de talloze speciale zoekmogelijkheden die het web daarnaast biedt, verwijs ik hier naar een paar sites die zich richten op van alles wat hier buiten beschouwing blijft:</p> <ul style="list-style-type: none"> • <u>Choose the Best Search for Your Information Needs</u> • <u>Phil Bradley: Finding what you need with the best search engines</u> • <u>NoodleQuest - The Search Strategy Wizard</u> • <u>Guides To Specialized Search Engines (with Descriptions)</u>
Help/Info	<u>Advanced Google Search Operators</u> (zie ook: <u>Google Guide: Making Searching Even Easier</u>)	<u>Inktomi's Search Syntax</u> ; zie ook: <u>Yahoo! Search Engine Review on Search Engine Showdown</u>	<u>Web Search Help: Search Builder and advanced search options</u> (zie ook <u>MSN Search's WebLog : New Operators Explained</u>)	
Blogs	<u>Google Weblog</u> (Aaron Swartz); zie ook: <u>Google Blog Inside Google - part of the Blog</u>	<u>Yahoo! Search blog</u> ; zie ook: <u>The Unofficial Yahoo! Weblog</u>	<u>MSN Search's Weblog</u>	

	News Channel New Google Blog - News and tips about Google The Unofficial Google Weblog Google Blogscoped Google Weblog jeroen.com/weblog/ Nuzee: Google Weblog Steve Johnston, the Google Blog of a Google Consultant Google, Inc. News (Topix.Net) Google Interviews GoogleGuy Says GoogleGuy Says - Google Ranking Info Zorgloob - Tout savoir sur Google Google Blog referencement et moteurs.			
Cijfers				
Aantal geïndexeerde pagina's	8,8 miljard	8 miljard	7,1 miljard	Volgens een studie van Gulli en Signorini, The Indexable Web is more than 11.5 billion pages.
Maximale indexeeromvang	htm, 101K of 513K txt, ca. 513K pdf, ca. 500K	htm, ca. 524K txt, ca. 215K pdf, 33K	htm, ca. 1Mb txt, ? pdf, 765K	De enige zoekmachine met op dit punt eenduidige resultaten is Google: voor html-bestanden vind je hier twee limieten, 101K en 513K; vermoedelijk is Google bezig naar de hoogste lengte over te stappen. In het txt-bestand (8pilg10.txt in de directory http://www.gutenberg.org/dirs/etext04/ van het Gutenberg-project), Love's Pilgrimage van Upton Sinclair, dat volgens de summary 1Mb groot is, vond Google als laatste tekst nog de phrase: "sleep fled from him like a mistress spurned" , de laatste zin van de cacheversie, die overeenkomstig Google's

				<p>opgave ca. 513K lang is. Wat het pdf-bestand betreft vond Google nog de phrase "since the agent does not know exactly what observations" op blz. 98 van de 148, wat op een indexeerlengte van ca. 500K (in dit geval 153K platte tekst) wijst. Yahoo! vindt in Love's Pilgrimage van Upton Sinclair als laatste tekst nog de phrase "made me think of you as some old Viking's", wat op een indexeerlengte van ca. 215K wijst en niet van 525K, zoals Yahoo! zelf opgeeft; ca. 215K is ook de werkelijke omvang van de cacheversie. Wat het pdf-bestand betreft, de laatste phrase die Yahoo! daarin vindt, duidt op een indexeerlengte van ca. 33K platte tekst. MSN Search indexeert htm-documenten tot ca. 1 Mb, zoals uit de cacheversie van het ca. 2,6 Mb grote document blijkt. Wat het document in txt-formaat betreft, MSN-search heeft het wel in zijn database, maar er is geen tekst in te vinden. Het pdf-bestand, dat een omvang van ca. 765K heeft, heeft op de laatste van de 148 bladzijden nog de phrase "intensity increasing at each time", die door MSN gevonden wordt.</p> <p>NB: Het kan zijn dat de hier opgegeven getallen niet voor alle documenten gelden die de zoekmachines in hun database hebben.</p>
Maximale omvang zoekargument in het zoekvenster	32 zoektermen, placeholders (*) niet meegerekend; NB: deze beperking kan niet worden omzeild door de URL te manipuleren	100 tekens; NB: deze beperking kan worden omzeild door de URL te manipuleren of door een korte zoekvraag vanuit het resultatscherm uit te breiden.	onbeperkt, maar bij invoer van meer dan 19 termen in een EN-relatie (in phrases maximaal 24) vindt MSN het document niet meer	URL-manipulatie: het veranderen van parameters in de URL van een resultatenpagina.
Toont:	nooit meer dan 1000 treffers	nooit meer dan 1000 treffers	maximaal ca. 250 treffers	
Prefixen				Prefixen zijn wat je bij

				databases veldlabels of voorloopteksten zou noemen. Functioneel zou je ze beter limiters of qualifiers kunnen noemen, want hun functie is de algehele verzameling van documenten op het Web tot een deelverzameling te beperken. Je zou daarbij onderscheid kunnen maken tussen "documentbegrenzers" en "tekstbegrenzers". De eerste hebben betrekking op eigenschappen (adres, taal, bestandstype etc.) van het document als zodanig, de tweede op de interne indeling van de documenten. Zo is site: een documentbegrenzer, en intitle: een tekstbegrenzer. Een aantal prefixen – zoals link: – valt echter buiten deze boot.
<i>Documentbegrenzers</i>				
Sites die worden gehost door een opgegeven Internet Provider (IP-adres)	n/a	n/a	ip:	Kan met een zoekterm, al dan niet met een prefix, worden gecombineerd. Gebruik DNS Stuff City from IP of IP Routing Lookup om land en plaats bij een IP-adres te vinden; zie ook Fagan Finders URLinfo beta onder General. Zoeken op een IP-adres kan ook bij Gigablast . Daar kan ook een getrunceerd adres worden opgegeven, wat bij MSN search niet kan.
De domain of host van het gezochte document (generieke en 'country code' TLD's , plus second level, third level etc.)	site:	site: of domain:	site:	Met site: beperkt u de vraag tot sites met een opgegeven – generiek of "country code" – TLD (Top Level Domain) of een (country code) domain op een lager (second level, third level etc.) niveau; gebruik de in de volgende rij genoemde mogelijkheden om niet alleen sites met een opgegeven ccTLD te vinden, maar ook die met een generiek TLD die met een bepaald land geassocieerd zijn.
Het land waar het document op een server staat	[Language Tools]	[Advanced Search]	loc: of location: (vgl.)	Zie boven bij site:; loc: omvat ook de generieke TLD's die in een opgegeven land in gebruik zijn.

De regio waar het document op een server staat	n/a	region:	n/a	Voor een overzicht van de beschikbare regio's en de landen die onder elke daarvan vallen, zie hier . Naast Yahoo! biedt ook Ask de mogelijkheid met het prefix geoloc: naar regio's te zoeken.
De computer waarop het opgegeven document staat	n/a	hostname:	n/a	hostname: bij Yahoo! doet niets wat site: niet ook kan.
De taal van het document	[Advanced Search, Preferences, Language Tools]	[Advanced Search]	language:	Klik hier voor een lijst van taalcodes bij MSN Search. Naast MSN Search hebben ook Ask (lang:) en Exalead een language:-prefix.
Het type document (op basis van de file-extensie)	filetype: of ext:	originurlextension:	filetype:	Voor meer informatie over filetypes, zie Greg Notess' Fiddling with File Types , beslist de moeite van het lezen waard!; hoewel de diverse zoekmachines opsommen welke meest gebruikte filetypes ze indexeren, blijken dat er in de praktijk altijd meer (voorbeeld: OpenOffice.org .sxw) te zijn, voor een uitgebreid overzicht van filetypes, zie FILExt - The File Extension Source . Fagan Finder heeft een handige filetype-kiezer, Search by File Format (beta).
xml-bestanden die als 'feed' dienen	n/a	n/a	feed:	Zoekacties met dit prefix vinden de feeds (xml-bestanden) zelf, niet de pagina's die zo'n feed bevatten; gebruik daarvoor hasfeed:
De opgegeven term in de URL	inurl:	inurl:	inurl:	inurl: kan onder meer worden gecombineerd met site:; als zoekterm achter inurl: kan bij alle drie de zoekmachines ook een host + pad worden opgegeven.
Alle opgegeven termen in de URL	allinurl:	n/a	n/a	allinurl: heeft betrekking op alle termen in de zoekvraag, vóór en na het prefix, zij het dat deze door OR met elkaar verbonden kunnen zijn; allinurl: duldt geen andere prefixen naast zich, met uitzondering van site:
De opgegeven term in het pad	n/a	path: of originurlpath:	n/a	Het path omvat alle subdirectories, maar niet de

				host en geen bestandsnamen met een extensie
Het (maximale) niveau waarop het gezochte document in de directorystructuur van een site te vinden moet zijn (mogelijke waarden 0-4)	n/a	depth:	n/a	De waarde 0 staat voor de eerste directory na de hostname; in het voorbeeld vindt Yahoo! pagina's waarin de gezochte term ofwel in de hostname voorkomt, ofwel in een bestandsnaam die er direct op volgt (inurl:, depth:0), maar niet in een path (-path:)
De pagina met het opgegeven adres, evt. met extra informatie	info: of id:	url:http:// of originurl:http://	url:	Google's info: of id: geeft toegang tot de cacheversie, eendere ("related") pagina's, pagina's die naar het opgegeven document linken, pagina's van de site waartoe de opgegeven pagina behoort, en pagina's waarin de opgegeven URL als link of als tekst voorkomt. Voor meer informatie over sites, zie Fagan Finders URLinfo beta
Alleen pagina's met een Last Modified-datum binnen de opgegeven periode	[Advanced Search]	[Advanced Search]	n/a	Tot voor kort ondersteunde Google naast Date in de Advanced Search ook daterange:[datum uit de Juliaanse kalender], waarmee pagina's konden worden gevonden met een Last Modified date binnen een vrij door de gebruiker op te geven range . Dit lijkt niet meer te werken (de gekozen daterange is hier de periode 1-10 september 2001). Gebruik Ask Advanced Search of Ask's prefixen afterdate:[yyyymmdd], beforedate:[yyyymmdd] en betweendate:[yyyymmdd.yyyymmdd] voor een zoekactie van dit type die verder gaat dan de mogelijkheden die Google en Yahoo! bieden; zie Ask's resultaat voor de mislukte zoekactie bij Google. Naast de hier genoemde mogelijkheden biedt Ask ook nog het prefix last: gevolgd door day, #days, week, #weeks of year, #years
<i>Tekstbegrenzers</i>				
De opgegeven term in het	intitle:	title: of intitle:	intitle:	Yahoo! rekt ook termen

titelveld				die in de META keywords en description voorkomen tot de titel!; NB: gebruik bij voorkeur title: en niet intitle:
Alle opgegeven termen in het titelveld	allintitle:	TITELWORDS (...)	n/a	TITELWORDS (...) is niet gedocumenteerd.
De opgegeven term in de (body van de) tekst	intext:	n/a	inbody:	De body van de tekst is het gedeelte van een HTML-bestand tussen de tags <body> en </body>.
Alle opgegeven termen in de (body van de) tekst	allintext:	n/a	n/a	allintext: bij Google heeft betrekking op alle termen in de zoekvraag, vóór en na het prefix, zij het dat deze door OR met elkaar verbonden kunnen zijn; allintext: duldt geen andere prefixen naast zich, met uitzondering van site:
Pagina's met een of meer getallen binnen het opgegeven bereik	numrange:[#-#] of: #.#	n/a	n/a	Gebruik numrange: in alle gevallen waarin u vermoedt dat het te vinden document een getal bevat en u kunt gissen wat dat getal ongeveer is; in het algemeen is een getal een betere zoekterm dan een woord.
Documenten waarnaar door andere documenten wordt verwezen onder een titel (het anchor) die de opgegeven term bevat	inanchor:	n/a	inanchor:	inanchor: betreft het document waarnaar verwezen wordt (en dat de opgegeven term dus zelf niet hoeft te bevatten); dat document wordt gepresenteerd, niet het document dat de verwijzende tekst bevat (hoewel het die natuurlijk wel kán bevatten); NB: als u inanchor: meermalen in één zoekvraag gebruikt, kunnen de zoektermen uit diverse anchors op diverse pagina's komen!
Documenten waarnaar door andere documenten wordt verwezen onder een titel (het anchor) die de opgegeven termen bevat	allinanchor:	n/a	n/a	allinanchor: bij Google heeft betrekking op alle termen in de zoekvraag, vóór en na het prefix, zij het dat deze door OR met elkaar verbonden kunnen zijn; allinanchor: duldt geen andere prefixen naast zich, met uitzondering van site:; NB: allinanchor: houdt niet in dat alle daarbij gebruikte zoektermen in één anchor of zelfs diverse anchors op één pagina

				moeten voorkomen!
Pagina's met een of meer verwijzingen (links) naar een opgegeven MIME-type	n/a	outgoingurltype:	n/a	Voor een overzicht van MIME-types, zie Mimetypes .
Pagina's met een of meer verwijzingen naar (een) bestand(en) met een bepaalde extensie	n/a	linkextension:	contains:	Voor een uitgebreid overzicht van file extensions, zie FILExt - The File Extension Source .
Pagina's met een opgegeven kenmerk (script, form, table etc.) of een link naar een opgegeven bestandstype of formaat	n/a	feature:	n/a	Yahoo! heeft dit prefix geërfd van de Inktomi-databases; er zijn tal van features, die bij elkaar een beeld geven van alles wat er op het web te koop is, voor zover HTML-pagina's daar toegang toe verschaffen. Op zoek naar plaatjes die u niet bij Google of Yahoo! Images kunt vinden? Combineer een zoekterm (evt. in een titel) met de opdracht feature:image en met wat geluk vindt u een foto bij de webpagina's die over het onderwerp gaan; op zoek naar databases van het onzichtbare web? Combineer een zoekterm met feature:form, wat wil zeggen dat u zoekt naar pagina's met een regel of een formulier waarin bezoekers iets kunnen invullen. Voor een uitgebreide lijst van de mogelijkheden van feature: zie Inktomi's search syntax .
Documenten die een definitie van de opgegeven term bevatten	define:	n/a	n/a	Hiermee is hier niet bedoeld de woordenboekfunctie die alle drie de grote zoekmachines als shortcut hebben, waarbij de opgegeven term in Answers.com (Google), Yahoo! Education (Yahoo!) of Encarta (MSN Search) wordt opgezocht, maar een zoekactie naar de opgegeven term, beperkt tot sites waar definities van termen worden gegeven.
Documenten die een RSS- of een Atom-feed bevatten			hasfeed:	Te gebruiken in combinatie met site: om pagina's op een site te vinden die feed bevatten.

<p>Operatoren</p>				<p>Hiertoe reken ik alle middelen die de relatie tussen zoektermen besturen: unitaire en Booleaanse operatoren, phrases, nabijheidsoperatoren, nesting, stemming en truncatie (deze twee omdat ze een patroon gebruiken om een impliciete OF-relatie tussen diverse termen leggen die aan dat patroon voldoen) en "term weighting", waarbij het belang van de ene zoekterm ten opzichte van de andere wordt gemanipuleerd.</p>
<p>Unitair verplicht aanwezig ("must contain")</p>	<p>\pm (vgl.); ander gebruik: \pm (vgl.)</p>	<p>+</p>	<p>+</p>	<p>Door het standaardgebruik van de impliciete EN-relatie tussen zoektermen heeft de unitaire operator + bij de drie hier vergeleken grote zoekmachines merendeels andere functies gekregen dan het verplicht stellen van de term waaraan hij voorafgaat. Bij Google dient de +-operator ter precisering: om een zoekactie met een geaccentueerde term tot de geaccentueerde vorm te beperken, en om automatic stemming uit te schakelen. + hoeft bij Google niet meer te worden gebruikt om stopwoorden in phrases in de zoekactie mee te nemen. Bij Yahoo! hoeft dat ook niet, en voor zover mij bekend heeft + bij Yahoo! geen nuttige functie; wel maakt het gebruik van + verschil voor de ranking als dat bij stopwoorden buiten phrases gebeurt: de ranking wordt dan anders, net zoals trouwens ook bij Google het geval is. Bij MSN Search speelt de + geen enkele rol.</p>
<p>Unitair verplicht afwezig ("must not contain")</p>	<p>=</p>	<p>=</p>	<p>=</p>	<p>De min-operator kan zowel met een term als met een phrase worden gecombineerd. Voeg elke term die u wilt uitsluiten in de impliciete EN-relatie toe met het min-teken ervoor. NB: De Search Builder van MSN bouwt een zoekvraag met meer minnen als volgt op: boerenkoolvoetbal - (doelpunt ploeg trainer). Dat</p>

				geeft een onjuist resultaat, vgl. boerenkoolvoetbal -doelpunt -ploeg -trainer .
Unitair facultatief aanwezig ("should contain")	n/a	n/a	prefer:	Behalve MSN Search ondersteunt ook Exalead deze "Best match" zoekmogelijkheid met de operator OPT . Ask Advanced Search wekt ten onrechte de indruk "best match" te ondersteunen.
Impliciet EN	[voorbeeld]	[voorbeeld]	[voorbeeld]	Dit is de standaard voor alle grote zoekmachines.
Booleaans EN	n/a	AND	AND (&)	Theoretisch zou er nog een verschil in ranking kunnen bestaan tussen impliciet en Booleaans EN. In de praktijk van de hier behandelde zoekmachines is dat niet zo. Dat Google geen AND ondersteunt, heeft alleen consequenties bij nesting. NB: Schrijf de Booleaanse operatoren altijd in hoofdletters. Ook al vereist niet iedere zoekmachine of database dat, het helpt u uw vraag te structureren.
Booleaans OF	OR	OR	OR ()	Schrijf de Booleaanse operatoren altijd in hoofdletters. Ook al vereist niet iedere zoekmachine of database dat, het helpt u uw vraag te structureren.
Booleaans NIET	n/a	NOT (AND NOT)	NOT	Als u meermalen NOT gebruikt, moet dat bij Yahoo! in de vorm: boerenkoolvoetbal NOT (doelpunt OR ploeg OR trainer) , niet in de vorm: boerenkoolvoetbal NOT doelpunt NOT ploeg NOT trainer ; bij MSN Search moet dat in de vorm: boerenkoolvoetbal NOT doelpunt NOT ploeg NOT trainer , en niet in de vorm boerenkoolvoetbal NOT (doelpunt OR ploeg OR trainer) NB: Schrijf de Booleaanse operatoren altijd in hoofdletters. Ook al vereist niet iedere zoekmachine of database dat, het helpt u uw vraag te structureren.
Nesting	n/a [voorbeeld1] ,	[voorbeeld1] ,	[voorbeeld1] ,	Nesting is het gebruik van

[voorbeeld2],
[voorbeeld3],
[voorbeeld4]

[voorbeeld2],
[voorbeeld3],
[voorbeeld4] (vgl.)

[voorbeeld2],
[voorbeeld3],
[voorbeeld4]

haakjes in een zoekvraag met diverse Booleaanse operatoren waardoor het computerprogramma opdracht wordt gegeven bewerkingen uit te voeren in een andere dan de standaardvolgorde. Doorgaans is die standaardvolgorde NOT, AND, OR: in een Booleaanse vraag worden eerst de NIET-bewerkingen uitgevoerd, dan de EN-bewerkingen en ten slotte de OF-bewerkingen. Door het gebruik van haakjes kan die standaardvolgorde worden doorbroken, net zoals dat bij wiskundige bewerkingen kan waarin optellingen, aftrekkingen, vermenigvuldigingen etc. voorkomen. In de vraag: [mensen EN klonen OF kloneren](#) zou de standaardvolgorde van bewerkingen zijn: eerst: mensen EN klonen (=set 1), daarna OF (is set 1 uitbreiden met) kloneren (=set 2). Het eindresultaat is set 3, bestaande uit treffers over het klonen van mensen, plus treffers over kloneren in het algemeen. Dit is duidelijk niet de bedoeling: vandaar dat door middel van nesting wordt aangegeven dat eerst de OF-relatie moet worden uitgevoerd: mensen EN (klonen OF kloneren). Klonen OF kloneren wordt nu set 1, mensen set 2, en set 1 verkleind met set 2 (EN) wordt set 3. Google werkt anders: tussen termen bestaat een impliciete EN-relatie, die automatisch haar prioriteit verliest als twee termen (of phrases) in de zoekvraag met een OR verbonden worden; deze OR-relatie geldt dan alleen voor de twee zoektermen (of phrases) aan weerszijden van de OR-operator. Voor een simpele vraag als [mensen klonen OR kloneren](#) is nesting dus niet nodig, omdat de OF-bewerking hier automatisch eerst wordt uitgevoerd, en daarna de

impliciete EN-bewerking. Dat gaat ook op als er meer OF-bewerkingen in een zoekvraag zijn: [apen OR stieren klonen OR kloneren](#). Een probleem krijgt Google pas met een vraag waarbij een EN-bewerking prioriteit zou moeten krijgen boven een OF-bewerking: [salmons cloning OR genetic engineering](#). In de interpretatie van Google is set 1 hier cloning OR genetic, en wordt deze vervolgens verkleind met set 2, salmons, en set 3, engineering; het eindresultaat is, in volgorde van bewerkingen: cloning OR genetic AND salmons AND engineering. Doordat Google de combinatie genetic engineering bij het ranken als een phrase blijft behandelen, is bij de eerste treffers niet direct zichtbaar dat het resultaat niet het verwachte is; dat dat toch zo is, blijkt als we genetic engineering tot een phrase maken: [salmons cloning OR "genetic engineering"](#). Dat dit resultaat kleiner is dan bij de vraag zonder de phrase, komt doordat daar aan de treffers die de term cloning bevatten ook nog de extra eis wordt gesteld dat ze engineering moeten bevatten.

Uit de bij Google gegeven voorbeelden mag blijken dat het gebruik van phrases en placeholders in phrases het gebrek aan Booleaanse nesting voor een deel kan compenseren; een phrase is immers een speciale manier om een EN-relatie te leggen die prioriteit heeft boven OF-bewerkingen. Zonder de phrase zou de hierboven vermelde vraag met Booleaanse nesting overigens luiden: [salmons AND \(cloning OR genetic AND engineering\)](#): ook binnen de haakjes blijft de computer de EN-relatie tussen genetic en engineering standaard prioriteit geven boven de OF-bewerking.

Voor een overzicht van

				<p>Google's combineermogelijkheden, zie Search Engine Showdown Analysis: Boolean Searching on Google.</p> <p>Voorbeeld 4 maakt duidelijk dat nesting ook prefixen kan omvatten. Wel dient u er bij Yahoo! op te letten dat als u nest met een OF-relatie tussen twee of meer zoektermen, de AND-operator expliciet moet worden gebruikt voor de EN-relaties!</p>
De opgegeven zoektermen als phrase	<u>"..."</u>	<u>"..."</u>	<u>"..."</u>	<p>Google accepteert phrases (en niet-phrases) van tot en met 32 termen (placeholders [*] niet meegeteld) en verwaarloost het meerdere; Yahoo! heeft geen of een heel hoge limiet; MSN Search accepteert in phrases tot en met 24 termen (buiten phrases maximaal 19) in een EN-relatie en geeft een nulresultaat als u er meer gebruikt.</p>
Een willekeurige term binnen een phrase ("placeholder")	<u>*</u>	n/a	n/a	<p>Google is de enige zoekmachine die placeholders ondersteunt; bij Yahoo! kon daar vroeger een stopwoord voor worden gebruikt, maar dat werkt niet meer. Google accepteert meer dan een placeholder (achter elkaar of op verschillende plaatsen) binnen een phrase; placeholders tellen niet mee voor de limiet van 32 termen in één zoekvraag die Google hanteert.</p>
Proximity operators	nabijheid van termen is een van de factoren die een rol spelen in het ranking algoritme	nabijheid van termen is een van de factoren die een rol spelen in het ranking algoritme	nabijheid van termen is een van de factoren die een rol spelen in het ranking algoritme	<p>Exalead heeft een NEAR-operator, die pagina's zoekt in een afstand van maximaal 16 woorden van elkaar, maar het gebruik van deze operator is niet zonder problemen.</p>
De opgegeven term, plus de termen waarin de opgegeven term een woordstam is ([automatic] stemming)	(in beperkte mate automatisch; gebruik +[term] om automatic stemming uit te schakelen; voorbeelden [let op de vet weergegeven	stem:	n/a	<p>Google past in sommige gevallen – en alleen bij meer dan één zoekterm – automatic stemming toe; net als Google ondersteunt de zoekmachine Exalead automatic stemming, maar op meer uitgebreide schaal en enkel wel/niet voor alle</p>

	termen]: 1 , 2 , 3 , 4)			zoektermen in de zoekvraag.
Truncatie	n/a	n/a	n/a	Niet een van de grote zoekmachines biedt de mogelijkheid tot trunceren. Gebruik daarvoor Exalead , dat zowel rechts trunceren met * mogelijk maakt, als zoeken met regular expressions
De opgegeven term, plus gerelateerde termen	≈	n/a	n/a	Ask Jeeves heeft suggesties voor gerelateerde termen onder Expand your search in het resultatenschermb van een zoekactie met één term. Zie ook onder Categorisatie.
Term weighting	Plaats term voorop (vgl.); herhaal term (vgl.)	Plaats term voorop (vgl.); 0.1-9.9*[zoekterm] (vgl.)	n/a	Door de hier vermelde mogelijkheden te gebruiken geeft u de zoekterm waarop de bewerking wordt toegepast meer (of in het geval van Yahoo! evt. ook minder) gewicht dan andere termen in de zoekvraag. De kleinere zoekmachine Gigablast gebruikt het teken om treffers die alle zoektermen bevatten op de term na te sorteren. Vgl. voeding kanker met kanker voeding .
Sortermiddelen				Behalve MSN Search biedt vrijwel geen enkele zoekmachine zijn gebruikers de mogelijkheid een andere dan de standaard sortering op het zoekresultaat toe te passen.
Sorteer op actualiteit (Last Modified)	n/a	n/a	{frsh=[0-100]}	Vindt veel blogs, nieuwsbrieven, etc.; naast MSN Search biedt ook Exalead een sort:new / sort:old mogelijkheid. Maar daarnaast kunt u ook overwegen speciale nieuws-/blog-zoekmachines, zoals Google News , Yahoo! News , RocketNews , Feedster , BlogPulse en Technorati te gebruiken.
Sorteer op populariteit	n/a	n/a	{popl=[0-100]}	Controleer het resultaat door met link: de pagina's te tellen die naar de volgens MSN Search populairste pagina's over de opgegeven zoekterm(en) linken. Vgl.

				Hubs en authorities hieronder.
Sorteer op relevantie	n/a	n/a	{mch=[0-100]}	Naar mijn mening heeft MSN Search hier Exact match en Approximate match met elkaar verwisseld: {mch=100} zou moeten betekenen dat de overeenkomst tussen zoekvraag en eerst gepresenteerde documenten maximaal is (Exact match).
Overige				
De documenten die verwijzen (linken) naar de opgegeven pagina	link:	link:http://	link: (of link:http://)	link: bij Google geeft maar een beperkt aantal resultaten en kan niet met andere zoektermen of prefixen (bij voorbeeld: site: link:) worden gecombineerd; bij Yahoo! en MSN Search kan dat wel; interessante toepassingen zijn onder meer: link:[URL 1] link:[URL 2] etc. om "hubs" (pagina's met veel outgoing links) te vinden; link:http://[URL 1] -site:[de site waartoe de pagina behoort] om alleen de links van buiten te vinden. Voor meer mogelijkheden om naar verwijzende pagina's te zoeken, zie Fagan Finders URLInfo beta onder Links.
De documenten die verwijzen (linken) naar alle pagina's binnen het opgegeven domain	n/a	linkdomain:	linkdomain:	linkdomain: kan zowel bij Yahoo! als bij MSN Search met andere termen en prefixen worden gecombineerd.
Pagina's die eender (similar) zijn aan de pagina met het opgegeven adres	related:	n/a	n/a	Google's related: kan niet met andere termen of prefixen worden gecombineerd. Voor meer mogelijkheden om naar eendere pagina's te zoeken, zie Fagan Finders URLInfo beta onder Similar. Exalead biedt de mogelijkheid via Related categories door te klikken naar andere sites die in eenzelfde rubriek van het ODP zijn ondergebracht.

Webpagina's vertalen	[link Translate this page in resultatenlijst] ; daarnaast via [Google Language Tools]	[link Translate this page in resultatenlijst]	n/a	Voor meer vertaalmogelijkheden, zie Fagan Finders Translation Wizard (beta) en Free Online Language Translation .
De cachekopie van het document met de opgegeven URL	cache:	[link Cached in resultatenlijst]	[link Cached page in resultatenlijst]	Voor een overzicht van mogelijkheden om oudere kopieën van webpagina's te vinden, zie Fagan Finders URLinfo beta , onder de tab Cache. Overigens is de ene cachekopie de andere niet. Vgl. de lengte van die van Google met die van Yahoo! .
Zoeken in natuurlijke taal	How high is the second-highest mountain in the world? Who is the current secretary-general of NATO? How many species of tarantula are there? Which part of Aristotle's Ars Poetica is lost? When was the library of Alexandria destroyed? What does Parkinson's second law say? What is the title of Dan Brown's first novel?	How high is the second-highest mountain in the world? Who is the current secretary-general of NATO? How many species of tarantula are there? Which part of Aristotle's Ars Poetica is lost? When was the library of Alexandria destroyed? What does Parkinson's second law say? What is the title of Dan Brown's first novel?	How high is the second-highest mountain in the world? who is the current secretary-general of NATO? How many species of tarantula are there? Which part of Aristotle's Ars Poetica is lost? When was the library of Alexandria destroyed? What does Parkinson's second law say? What is the title of Dan Brown's first novel?	Vergelijk deze resultaten met die van Ask : How high is the second-highest mountain in the world? Who is the current secretary-general of NATO? How many species of tarantula are there? Which part of Aristotle's Ars Poetica is lost? When was the library of Alexandria destroyed? What does Parkinson's second law say? What is the title of Dan Brown's first novel?
Hubs en authorities	n/a	n/a	Sorteer op popularity (authorities)	Hubs en authorities zijn kort gezegd sites met resp. veel uitgaande en veel inkomende links. Hoewel het rankingalgoritme van Google aan beide een zeker gewicht toekent, zijn er zoveel andere factoren die de ranking beïnvloeden , dat je met Google niet op een enigszins systematische manier hubs en authorities kunt identificeren.

				<p>Voor de autoriteiten zou dat in theorie wel met MSN Search moeten kunnen door in de Search Builder de slider voor popularity naar het hoogste punt te slepen en de andere twee sorteropties op nul in te stellen. Maar in de praktijk is onduidelijk waarop de populariteitsranking van MSN Search gebaseerd is. Met de vraag "forensic psychiatry" {popl=100} is het aantal inkomende links van de eerste twaalf treffers als volgt: 6, 65, 3103, 19, 1456, 12.206, 9, 4556, 4, 30, 2, 19.299. Vergelijk dit met de aantallen inkomende links van de treffers op de plaatsen 250-239: 4, 5, 2, 26, 11, 43, 88, 2, 32, 15, 2, 10. Het belang van zulke aantallen uitgaande links voor het opsporen van hubs wordt gerelativeerd doordat de echt grote hubs hun links in een database hebben.</p>
Spellingsuggesties	[voorbeeld1], [voorbeeld2], [voorbeeld3]	[voorbeeld1], [voorbeeld2], [voorbeeld3]	[voorbeeld1], [voorbeeld2], [voorbeeld3]	<p>Van de drie hier genoemde zoekmachines heeft Google zonder twijfel verreweg de beste spellingchecker. Dislektisi die aan Google niet genoeg hebben, kunnen de Advanced Search van Exalead proberen, met de daar onder Search method beschikbare Phonetic search en Approximate spelling.</p>
Categorisatie, suggesties voor meer / andere / specifiekere zoektermen	n/a	[voorbeeld]	n/a	<p>Ask (Narrow Your Search en Expand Your Search), Gigablast (Gigabits) en Exalead (Related terms en Related Categories) bieden diverse voorzieningen.</p>
Gerelateerde namen	n/a	n/a	n/a	<p>Ask laat bij het zoeken met de namen van bekende personen, functies, werken, gebeurtenissen, e.d. in de rechter marge van het resultatscherm "Related Names" zien.</p>
RSS feeds	n/a	via link "View as XML" bij treffers die een feed bieden	via RSS-knop onderaan resultatenpagina	<p>Van de drie grote mondiale zoekmachines is MSN Search de enige die zijn resultatenpagina's als RSS-feed verschaft. Dat betekent dat de feed van zo'n pagina in</p>

				<p>een RSS-reader kan worden geplakt, waarna nieuwe resultaten van de zoekactie met de door u opgegeven frequentie als koppen in dat programma verschijnen. Yahoo! doet het anders. Als tussen de resultaten sites zitten die feeds ondersteunen, wordt bij de treffer een link met de tekst "View as XML" aangeboden, waarna het adres van het XML-bestand in de RSS-reader kan worden geplakt. Nieuwe items voor die site worden dan als koppen in de RSS-reader getoond.</p>
Local search	via Google Local	via Yahoo! Local	<p>Geef onder Settings bij Location een plaatsnaam op waarvoor u lokale resultaten wilt zien, en zoek vervolgens met de knop Near me.</p>	<p>Gebruik voor Nederland: Wijsvinger.</p>

Appendix B

Lijst van termen

+: (unitaire) operator die de aanwezigheid van de ermee verbonden zoekterm in een document verplicht stelt (in menusystemen ook aangegeven door *All of the words of Must include, Must contain, Very important*).

-: (unitaire) operator die aangeeft dat de te vinden documenten de ermee verbonden zoekterm niet mogen bevatten (in menusystemen ook aangegeven door *Must not include, Must not contain, Exclude*).

accenten: zie diacrieten.

algorithmic results: treffers in de resultatenlijst van een zoekmachine die hun plaats in de ranking te danken hebben aan het algoritme dat de zoekmachine voor de ranking hanteert. Ook bekend als natural of organic results, in tegenstelling tot de zogeheten 'sponsored results'.

anchor: een verwijzing in de vorm van een tekst; document of plaats in een document waarnaar een link (uit een ander of hetzelfde document) verwijst.

AND: (Booleaanse) operator die de aanwezigheid van twee (of meer) erdoor verbonden zoektermen verplicht stelt. Komt ook voor als &.

bloggers: personen die een journaal (weblog) op het Web bijhouden en die zo'n weblog op het Web publiceren dan wel via nieuwsletters rondzenden.

blogroll: lijst van andere blogs waarnaar door een blogger verwezen wordt of die hij of zij volgt.

browse: grasduinen, neuzen. Het doorlopen van de hiërarchische structuur van een directory of het bekijken van een index; in tegenstelling tot zoeken, waarbij direct zoektermen aan een zoekmachine worden opgegeven.

browser: programma op de PC of een andere terminal dat HTML-pagina's opvraagt en leest (= toont).

bulletin board: zie forum.

cache: stukje intern geheugen op de PC of een terminalserver waarin een opgevraagde webpagina tijdens een sessie bewaard blijft. Bij de zoekmachine Google een kopie van een webpagina in Google's eigen database.

case-sensitivity: onderscheid tussen hoofd- en kleine letters dat zoekmachines wel of niet maken. In het algemeen vinden de vormen met kleine letters alle vormen, de vormen met hoofdletters (b.v. letterwoorden, WHO, World Health Organization, in plaats van who, wie) alleen de exact overeenkomende vormen.

categoriseren: bij zoekmachines de automatische indeling van zoekresultaten in categorieën die het mogelijk maken de vraag achteraf te specificeren.

classificatiesysteem: stelsel waarin beschrijvingen van zaken in een hiërarchisch verband, meestal op basis van decimale codes, zijn samengebracht en ingedeeld. Aan de gebruiker presenteert zo'n classificatiesysteem zich als een zoekboom waarin hij van algemeen naar bijzonder wordt geleid.

client: computer, terminal of programma waarmee informatie kan worden opgevraagd bij een server.

clustering: bij zoekmachines de presentatie van één of enkele treffers van dezelfde site als zijnde representatief voor alle resultaten van die site; wordt gedaan om gebruikers niet met ellenlange lijsten van sites te confronteren die niets relevant te bieden blijken te hebben (hoewel hun treffers wel aan de zoekcriteria voldoen)

cookie: bij het gebruik van zoekmachines: boodschap die een door u bezochte host op uw computer zet met informatie over door u opgegeven instellingen als u de Customize-functie gebruikt. Deze boodschap wordt door de host gevonden als u er terugkeert, waarna de host zich aan uw instellingen aanpast.

customize: het aan de persoonlijke voorkeuren aanpassen van een gebruikersinterface,

bij voorbeeld een zoekscherm. De gekozen instellingen kunnen meestal worden opgeslagen ('gesaved') waardoor ze in een bestandje (cookie) op uw computer komen te staan.

'dead links': ook wel 'broken links' of 'dode links': bestands- of IP-adressen die door de bewoner verlaten zijn.

description: 1. beschrijving van de inhoud van een HTML-bestand in de META-tag van dat bestand. Wordt door de browser niet zichtbaar gemaakt, maar is te bekijken via de menu-opties View – (Page) Source. 2. Synoniem voor presentatieformaat van een link bij een zoekmachine.

diacrieten: in het algemeen bijzondere schrifttekens, meer speciaal geaccentueerde letters.

discussiefora: overkoepelende term voor nieuwsgroepen, message boards en mailing lists.

directories: 1. de mappen of folders waarin het bestandensysteem op een computer is ingedeeld. Verdeeld in een hoofddirectory of 'root' en subdirectories. Samen vormen ze een path naar een bestand. 2. zoekinstrumenten op het Web die een op onderwerp ingedeeld overzicht van het hele Web of een deel ervan geven. Staan ook bekend als Webgidsen, Webwijzers, en (minder juist) Webindexen.

discussielijst: groep van personen met dezelfde interesse op internet die in een openbare of besloten omgeving via e-mail met elkaar communiceren. Zie ook: rondzendlijst.

discussion board: zie forum.

domain: internet is ingedeeld in domeinen. Daarmee kunnen zijn aangeduid: landen of typen organisaties (.nl, .co, .edu), organisaties (leidenuniv.nl) en onderdelen van organisaties (let.leidenuniv.nl). Elke host (www.let.leidenuniv.nl) heeft een domain name (let.leidenuniv.nl), die deel uitmaakt van de unieke naam van de host.

downloaden: een bestand naar de eigen PC kopiëren.

'experts': technische term voor personen die op het Web bereid zijn vragen over een onderwerp te beantwoorden. Doorgaans georganiseerd in AskAnExpert-diensten.

expressie, Booleaanse (ook: Boolean phrase): zoekvraag aan een zoekmachine of database waarin zoektermen met elkaar verbonden zijn door Booleaanse operatoren.

extensie: in een bestandsnaam het gedeelte achter de punt dat een type bestand aanduidt.

feed: een bestand in een speciaal formaat (XML, RDF) dat de koppen en beschrijvingen van een aantal nieuws- of blogitems e.d. bevat; zie ook RSS

forum: elektronisch prikbord dat niet tot de officiële Usenet-hiërarchie behoort; ook wel aangeduid als prikbord, message board, bulletin board, discussion board, Web forum of Web conference.

frames: aparte HTML-bestanden die tot één webpagina kunnen worden gecombineerd waarin afzonderlijke delen onafhankelijk navigeerbaar zijn.

fuzzy searching: ook wel 'sound-alike' searching genoemd: het zoeken van termen waarin een of meer letters door jokers (zie daar) vervangen kunnen zijn. Hierbij geeft de gebruiker niet door middel van jokers aan welke letters dat zijn, maar vindt het zoekprogramma automatisch alle woordvormen waarin een of meer letters op een willekeurige plaats door andere letters vervangen zijn. Voorbeeld: stroop vindt ook strook en siroop.

hit popularity: rankingmechanisme van zoekmachines waarbij na een zoekactie die documenten hoog in de resultatenlijst worden geplaatst die aan de zoekcriteria voldoen en waarnaar veel bezoekers doorklikken om er daarna geruime tijd door te brengen.

host: een computer op internet waarop bestanden staan die door andere computers (terminals) kunnen worden opgevraagd. Ook de organisatie die eigenaar is van de betreffende computer.

HTML: Hypertext Markup Language. De opmaaktaal waarin webpagina's zijn opgemaakt en die het uiterlijk van de pagina's in een browser bepaalt.

HTTP: Hypertext Transfer Protocol. De taal waarin computers op het Web met elkaar communiceren.

hyperlink: verwijzing naar een (plaats in een) ander document of een andere plaats in hetzelfde document.

hypertext: tekst in computerbestanden die directe, aanklikbare verwijzingen bevat naar andere documenten (tekst, beeld of geluid).

internet: netwerk van computernetwerken.

Invisible Web: die gedeelten van het Web die niet niet door zoekmachines geïndexeerd en doorzoekbaar gemaakt worden

IP-adres: uniek adres van een computer op internet, bestaande uit vier door een punt gescheiden reeksen getallen tussen 1 en 255. Als u in de adresregel van uw PC een URL opgeeft en die opstuurt, wordt het domeinnaamgedeelte daarvan door een zogeheten Domain Name Server (DNS) in een IP-adres vertaald en wordt met de computer met dat adres een verbinding tot stand gebracht.

joker (wildcard): teken dat in een zoekterm kan worden gebruikt als vervanging voor een of meer willekeurige letters of cijfers.

keywords: trefwoorden die door de auteur van een webpagina in de METAtag kunnen worden toegevoegd.

Last Modified-datum: de datum waarop de laatste versie van een bestand op een server is opgeslagen.

Limited Area Search Engines (LASEs): zoekmachines die gespecialiseerde servers spideren en indexereren.

link: het adres (de URL) dat in een hyperlink naar een ander document of een andere plaats in hetzelfde document verwijst. Dit adres kan absoluut (d.w.z. volledig uitgeschreven) of relatief (ten opzichte van het adres van het verwijzende document) zijn. Er hoort altijd een anchor bij. Samen vormen ze een hyperlink.

link popularity: rankingmechanisme bij zoekmachines waarbij na een zoekactie die documenten boven in de resultatenlijst worden geplaatst die niet alleen aan de zoekcriteria voldoen maar ook veel links 'trekken' van webpagina's die zelf ook een hoge 'link popularity' genieten.

listing: lijst van treffers in een directory of zoekmachine

maskeren: het gebruik van een speciaal teken (bij voorbeeld %) dat een willekeurige letter binnen een woord vervangt. Voorbeeld: categori%ation voor categorisation of categorization. Vergelijk trunceren.

message board: zie forum

META: tag of veldaanduiding in een HTML-bestand waarmee informatie over het betreffende bestand wordt gegeven. Speciaal van belang voor het zoeken zijn de metatags met de inhoud 'keywords' en 'description'.

metazoekmachine: programma dat een zoekvraag doorgeeft aan verschillende zoekmachines en de resultaten van die opdrachten aan de gebruiker presenteert.

minteken: zie -.

nabijheidsoperator: AND-operator die ook de maximale afstand specificeert waarbinnen twee erdoor verbonden zoektermen mogen voorkomen.

natuurlijke taal: taal zoals mensen verondersteld worden die onder elkaar te gebruiken. Bij zoekmachines een (korte en duidelijke) vraag zoals die (zonder franje) aan een persoon zou kunnen worden gesteld.

natural results: zie algorithmic results

NEAR: nabijheidsoperator. Een speciale vorm van de (Booleaanse) AND-operator waarmee wordt aangegeven dat twee erdoor verbonden termen niet alleen samen in een document moeten voorkomen, maar dat ook nog eens moeten doen binnen een bepaalde

afstand van elkaar. Van de grote zoekmachines ondersteunt alleen AltaVista deze operator (= binnen een afstand van 10 woorden of minder).

nesting: het gebruik van (ronde) haken in een Booleaanse vraag met de operatoren AND en OR om ervoor te zorgen dat een bepaalde bewerking eerder wordt uitgevoerd dan een andere. Komt in de praktijk meestal neer op 'brood AND (honing OR jam)' om aan te geven dat men brood met (honing of jam) wil en niet (brood met honing) of jam.

netiquette: gedragsregels op internet. Wie zich daaraan niet houdt, kan een 'flame' (uitbrander) verwachten.

newsreader: softwareprogramma waarmee de berichten in nieuwsgroepen kunnen worden bekeken

nieuwsgroep: elektronisch prikbord waarop berichten kunnen worden gezet en vragen worden gesteld. Meer speciaal Usenet newsgroup.

NOT: (Booleaanse) operator waarmee wordt aangegeven dat de zoekterm die erop volgt niet in de te vinden documenten moet voorkomen. Wordt gebruikt om irrelevante zoekresultaten uit te sluiten. Komt ook voor als ANDNOT, AND NOT of !.

operatoren: woordjes die in een zoekopdracht aan de computer een logisch verband tussen zoektermen leggen (Booleaanse operatoren) of aan één term een logische functie meegeven (unitaire operatoren). Booleaans zijn AND, OR en NOT, unitair + en -.

OR: (Booleaanse) operator die aangeeft dat een van beide zoektermen die erdoor verbonden worden in een te vinden document aanwezig moet zijn; beide mag ook.

organic results: zie algorithmic results

path: De reeks van subdirectories (folders, mappen) in het bestandensysteem op een computer tussen de 'root' of hoofddirectory en een bestand.

PDF: bestandsformaat van het programma Acrobat van Adobe, waarmee op computers met verschillende besturingssystemen documenten in hun oorspronkelijke vorm kunnen worden getoond.

phrase: opeenvolging van woorden in een tekst, woordgroep. Voorbeelden van phrases zijn persoonsnamen ("Bill Clinton"), titels van boeken of films ("Silence of the Lambs") en vaste samenstellingen ("neusje van de zalm").

plusteken: zie +

popular sites: bij zoekmachines treffers die voorin het zoekresultaat worden gezet op basis van hun populariteit.

post: een bericht in een nieuwsgroep, forum, bulletin board, weblog etc.

protocol: taal waarmee computers onderling en computernetwerken onderling boodschappen en opdrachten uitwisselen.

ranking, relevancy ranking: de volgorde waarin zoekmachines documenten die aan de zoekcriteria van de gebruiker voldoen in de resultatenlijst presenteren. Zoekmachines hanteren daarbij verschillende methodes om de meest 'relevante' documenten het eerst te presenteren.

resources: bronnen, alles wat internet aan informatie te bieden heeft.

rondzendlijst (mailing list): lijst van personen met dezelfde belangstelling die via e-mail met elkaar communiceren (discussielijst) of berichten ontvangen (nieuwsbrieven en e-zines).

RSS: 'Really Simple Syndication' (oorspronkelijk: 'Rich Site Summary'), een standaard voor het publiceren van dynamische Web content (in het algemeen: nieuws op het Web)

server: een computer waarop server-software draait die kan samenwerken met clientsoftware (de browser) op de PC of een terminal.

shortcuts: prefixen bij zoekmachines waarmee snel allerlei praktische vragen kunnen worden opgelost, zoals het omrekenen van maten en gewichten, het opvragen van weersinformatie voor alle grotere steden van de wereld en het volgen van postpakketten.

site search engine: de eigen zoekmachine van een site, die enkel de documenten op die

site ontsluit.

sorteren: (treffers) in een andere dan de standaardvolgorde (laten) presenteren. Sorteercriteria kunnen zijn: datum, grootte bestand, plaats bestand, een zoekterm waarvan de frequentie, of de aan- dan wel afwezigheid bepaalt op welke plaats een treffer in de resultatenlijst terechtkomt.

spider: onderdeel van een zoekmachine, programma dat webpagina's opvraagt en verzamelt ('spidert') door hyperlinks te volgen.

sponsored (featured) links (sites, results, listings): links in de resultatenlijst van een zoekmachine of directory waarvoor de organisatie of het bedrijf achter de link aan de zoekmachine of directory geld betaalt.

springplankzoekmachine: vraagformulier dat zo is ingericht dat van daaruit diverse zoekmachines achtereenvolgens met dezelfde kunnen worden bevraagd.

startpagina: pagina waarop in een groot aantal populaire rubrieken links bij elkaar zijn geplaatst. Rubrieken kunnen verder zijn onderverdeeld en de overige onderwerpen, die op de startpagina zelf geen plaats hebben gevonden, kunnen in 'dochterpagina's', die meestal door verschillende personen worden onderhouden, aan bod komen.

stopwoorden: kleine, veel gebruikte woorden die door zoekmachines niet worden geïndexeerd.

summary: de beschrijving van een link die een zoekmachine aan een webpagina ontleent en die bij de link wordt gepresenteerd. Ook wel 'teaser' genoemd.

tags: veld aanduidingen waarmee secties of elementen in een document of databaserecord worden gemarkeerd. In HTML meestal bestaande uit een begin- en een eindtag. Tags zijn tegelijk codes die de browser vertellen hoe een element op het scherm van de gebruiker moet worden weergegeven.

TCP/IP: Transmission Control Protocol/Internet Protocol. De interactietaal tussen computers op internet die de overdracht van bestanden tussen computers op internet regelt.

terminalserver: server die een aantal terminals of werkstations bedient.

thesaurus: lijst van indextermen die begrippen term- en taalonafhankelijk beschrijven, in hun betekenisrelatie tot elkaar.

thread: een reeks van berichten over hetzelfde onderwerp in een nieuwsgroep, forum, bulletin board, weblog etc.

title: de titel van een HTML-pagina zoals die in de titelbalk van de browser getoond wordt; in het HTML-bestand het gedeelte van de tekst dat omsloten wordt door de tags <title> ... </title>.

treffers (hits): links en hun beschrijvingen (of in een database: databaserecords) die aan de zoekcriteria voldoen en door een zoekmachine gepresenteerd worden.

trunceren: de vervanging van 0-x of 1-x letters aan het begin van (links), aan het eind van (rechts) en/of binnen in (intern) een woord door een joker of wild card om in één keer termen met dezelfde stam te vinden. *kanker vindt dan ook longkanker, f*t, vindt foot en feet, femini* vindt feminine, feminist, feminism. Vergelijk maskeren.

URL: Uniform Resource Locator. Het volledige adres van een bestand (resource) op een computer in internet.

velden: gedeelten van een document of databaserecord die ten opzichte van andere delen van de tekst of het record gemarkeerd zijn en die afzonderlijke elementen beschrijven. In webpagina's alles wat door middel van tags van andere tekst onderscheiden is. Sommige velden worden ook aangeduid als 'filters'.

Web conference: zie forum.

Web forum: zie forum.

webdocument: ieder soort document – beeld, tekst, geluid – dat tot het Web behoort.

webgids: 1. Nederlandse vertaling voor directory 2. meer speciaal een directory of anders georganiseerd overzicht van resources op een bepaald terrein.

weblog (blog): een op het web bijgehouden journaal.

webpagina: document in HTML op het Web zoals het zich in de browser van een gebruiker als eenheid presenteert.

website: samenstel van webdocumenten (of één enkele webpagina) binnen een gemeenschappelijk domein die één samenhangend geheel vormen en door links met elkaar verbonden zijn. In Yahoo!: webpagina's die door Yahoo's interne zoekmachine binnen de directory worden gevonden.

word stemming: bij zoekmachines een vorm van trunceren die zich beperkt tot meervouden en soortgelijke veel voorkomende woorduitgangen. Gebeurt automatisch dan wel via een door de gebruiker in te stellen optie. Anders dan bij trunceren hoeft er geen joker voor te worden gebruikt. Maar de ene word stemming is de andere niet.

World Wide Web: dat deel van internet dat als protocol HTTP gebruikt.

zie-verwijzing: bij directories verwijzing van een plaats (rubriek) in een tak van de zoekboom naar een andere plaats waar dezelfde rubriek voorkomt.

zoekboom: bij een directory het systeem van rubrieken en subrubrieken via hetwelk een gebruiker van algemene naar bijzondere begrippen wordt geleid.

zoekcriteria: de opgegeven zoektermen en hun onderlinge relaties.

zoekinstrumenten: overkoepelende term voor directories (zie daar) en zoekmachines (zie daar)

zoekmachines: doorzoekbare databases van webpagina's. webpagina's worden door spiders (zie daar) opgevraagd en in een database opgeslagen. De database wordt geïndexeerd, waarna de gebruiker deze kan doorzoeken op termen die in de webpagina's voorkomen. Daarbij vindt hij dan het adres (de URL) met een kort uittreksel van de pagina's waarbij dat het geval is.

Zie ook: Search Engine terms - a glossary of internet search technology

(http://www.cadenza.org/search_engine_terms/) en de Nederlandse vertaling Internet zoek machine technology - een overzicht met search engine terms (<http://www.cetest.nl/nlgloss.htm>).

Glossary of Internet & Web Jargon

(<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html>)

