

Ervaringen met geautomatiseerd literatuurzoeken

1. Inleiding

Aan de Technische Hogeschool in Delft wordt op de afdeling der Civiele Techniek, onder andere bij de Vakgroep Gezondheidstechniek, veel onderzoek verricht naar aan water verwante onderwerpen. Om publicaties te verkrijgen over het onderwerp van die onderzoeken, wordt vaak gebruik gemaakt van de diensten van de Centrale Bibliotheek van de TH Delft, waar men over vrijwel alle middelen beschikt om deze publicaties op te sporen en te verkrijgen. In H₂O verscheen al een artikel



IR. H. H. SIEBERS
Vakgroep Gezondheids-
techniek TH Delft



MEVR. ING. M. W. DE
JONG-HOFMAN
Centrale Bibliotheek
TH Delft

over het opsporen van 'waterliteratuur' met behulp van een computer in Aqualine [Newman, 1981], een bibliografisch bestand op het gebied van o.a. drinkwater-voorziening en afvalwaterbehandeling. Aqualine is een uitstekende bron van informatie; er zijn echter veel meer en soms ook goedkopere manieren om publicaties te verzamelen, afhankelijk van het gestelde doel en de financiële middelen die de onderzoeker ter beschikking staan. Hiernaast speelt de verkrijgbaarheid van documenten een grote rol. Wat in Nederland verkregen kan worden, is beslist goedkoper dan wat in het buitenland moet worden besteld. Aan de hand van twee, aan de realiteit ontleende voorbeelden, zullen we in dit artikel proberen een overzicht te geven van de diverse mogelijkheden, haken en ogen bij het literatuurzoeken.

2. Het opsporen van literatuur

Er zijn twee belangrijke wegen om publicaties op te sporen: met de hand en met behulp van een computer.

2.1. Opsporen van literatuur m.b.v. bibliografieën (manueel zoeken)

Omdat misschien niet aan iedereen bekend is wat een bibliografie is, volgt hier de definitie: 'Een bibliografie is een overzicht van publicaties'. Dat overzicht betreft gewoonlijk een bepaald vakgebied en het

wordt toegankelijk gemaakt door auteurs- en onderwerpregisters. Bibliografieën worden samengesteld door 'data base producers'. *Het is goed om te bedenken dat een data base producer een pakket abonnementen heeft op tijdschriften, congressverslagen etc. Dit pakket bevat beslist nooit alle literatuur die over een bepaald onderwerp is verschenen. Bovendien wordt de informatie binnen het gekozen pakket meestal niet volledig in de bibliografie opgenomen.* Iedere data base producer heeft 'reviewers' in dienst die beoordelen of de aangeboden literatuur wel aan bepaalde, gestelde eisen voldoet.

Het is daarom een illusie om te menen dat in één bibliografie alle publicaties over een bepaald onderwerp terug te vinden zijn. Van iedere op te nemen publicatie wordt een 'referaat' of 'abstract' gemaakt. Dit bestaat uit een titel, auteur, bronvermelding en een samenvatting van de inhoud van de publicatie. Aan iedere referentie/abstract worden door een indexeerder trefwoorden toegekend. Een deel van deze trefwoorden wordt gebruikt om het onderwerpregister samen te stellen. Er bestaan veel uitstekende bibliografieën op de vakgebieden van drink-, afval- en oppervlaktewater. Ze staan gerangschikt in tabel I en zijn voor eenieder gratis te raadplegen in de bibliografieënzaal van de bibliotheek van de Technische Hogeschool Delft. Literatuur zoeken in bibliografieën kost echter nogal wat tijd. Werkt men met enige nauwkeurigheid, dan is men al gauw enkele dagen bezig. Dit manueel zoeken biedt specifieke voordelen, zoals we verder in dit artikel zullen zien.

2.2. Opsporen van literatuur m.b.v. bibliografische bestanden (online zoeken)

2.2.1. De opbouw van bibliografische bestanden

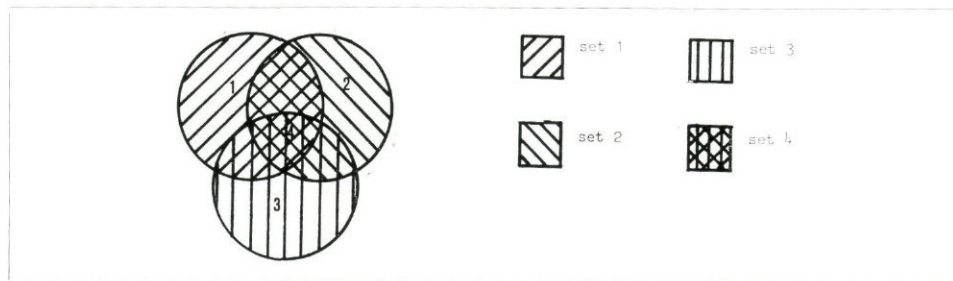
In de zestiger jaren begon men drukwerk

te vervaardigen met magneetband gestuurde zetapparatuur. Een plezierige bijkomstigheid van deze magneetbanden is dat de tekst van het drukwerk nu ook in computerleesbare vorm aanwezig is en gebruikt kan worden als basismateriaal voor geautomatiseerd literatuurzoeken.

Door middel van computersystemen, in dit geval door tussenkomst van een zogenaamde hostorganisatie of kortweg host, worden deze magneetbanden met bibliografische informatie zó bewerkt, dat ze gebruikt kunnen worden voor het geautomatiseerde literatuurzoeken. Grote hostorganisaties zijn ESA in Italië, DIS en SDC in de USA, INKA in Duitsland en Telesystèmes in Frankrijk (zie voor de gebruikte afkortingen paragraaf 7). De hostorganisatie bevindt zich vaak op grote afstand, maar met een terminal en een, al dan niet speciaal daarvoor aangesloten telefoonverbinding, is het mogelijk de gegevens die op afstand zijn opgeslagen, te raadplegen; dat gebeurt dus 'on line'. De informatie die in het computergeheugen wordt opgeslagen, is in principe dezelfde als die, welke in gedrukte vorm verschijnt, namelijk een overzicht van publicaties. Dat overzicht heet nu echter niet meer 'bibliografie' of 'referaatijdschrift', maar 'bibliografische bestand' of kortweg 'data base'. Om in die data bases ook nog iets terug te kunnen vinden, worden er indexen of registers op gemaakt. De belangrijkste daarvan zijn het auteurs- en het trefwoordenregister, maar afhankelijk van de data base en de hostorganisatie die de data base afzoekbaar maakt, kunnen er andere registers zijn: op taal, op publicatiedatum, op documenttype, etc. Een goed literatuuroverzicht over een bepaald onderwerp kan slechts verkregen worden wanneer een behoorlijke kennis van de inhoud van de beschikbare data bases en de wijze waarop ze worden

Afb. 1 - Voorbeeld van een online referentie.

geheugenadres	81-03630 AQUALINE 00018636
titel	Simulation of the deep-shaft process.
auteur(s)	VELDHAMP R., J. SIEBERS M. W.
bronvermelding	WHD 1980, 13, no. 15, 425-431 (in Dutch, English summary p. 401). Document Origin: NETHERLANDS *RC INFORMATION Vol 07
samenvatting	Purification of sewage by the deep-shaft process is described. Treatment is performed in a vertical shaft (100-300 m, below groundlevel) in which circulates a mixed liquor. A model is described whereby a small volume of mixed liquor is subjected to recirculation, alternating pressure and air injection in the treatment of industrial waste water. Theoretical considerations are given with regard to the air injection, the flow of liquid in the shaft and the oxygen transfer. The mean TOD, COD and BOD concentrations of the waste water were 10,300, 7544 and 4968 g per m ³ ; overall BOD, COD and TOD reductions were 76.0, 66.5 and 64.7 per cent, respectively; mean daily sludge loading varied between 1.5 and 13.5. The sludge had a high chemical oxygen demand and high sludge index between days 9 and 12. Oxygen requirements decreased with increasing sludge loading.
classificatiecodes	Classification Codes: 6, Sewage / 7, Industrial effluents
geografische bepaling	Geo-Location: NETHERLANDS
toegekende trefwoorden	Controlled Terms: OXYGEN / PRESSURE / PURIFICATION / SEWAGE / SHAFTS / SLUDGE / TREATMENT / CONCENTRATION / RECYCLING / REDUCTION / REQUIREMENTS / SHAFT / THEORY / INDUSTRIAL WASTE WATER / VARIATION / WASTEWATER / MEAN / ALTERNATING / CIRCULATION / DAILY / DESCRIPTION / LIQUIDS / OXYGEN TRANSFER / VERTICALLY / DAY / GROUND / OVERALL / OXYGEN DEMAND (BIOCHEMICAL) / DEEP SHAFT PROCESS / HIGH / OXYGEN DEMAND (TOTAL) / MODELLING (=GENERAL-) / VOLUME / MIXED LIQUOR / INCREASING / ACTIVATED SLUDGE / AERATION (SEE ALSO OXYGENATION) / RE-OXYGENATION / OXYGEN DEMAND (CHEMICAL) / FLOW



Afb. 2 - Schematische voorstelling van de met de computer gevonden sets.

samengesteld, aanwezig is. En om hierbij de kosten niet de pan uit te laten rijzen, is ook een perfecte kennis vereist van de commandotalen die nodig zijn om de, bij de diverse hostorganisaties, opgeslagen informatie te kunnen raadplegen. Hoe gaat dat in zijn werk?

Het trefwoordenregister van het online bestand is veel uitgebreider dan dat van het gedrukte equivalent van de data base. Dat is nodig om de informatie snel en efficiënt terug te kunnen vinden, wat mogelijk gemaakt wordt door het feit dat de opgeslagen bestanden via een computer snel te inverteren zijn. Deze inversie vindt dus plaats van elementen (onderdelen) van de referentie naar het geheugenadres: bijv. van auteurs die worden opgenomen in het auteursregister met verwijzing naar het geheugenadres waar de betreffende auteur te vinden is. Afb. 1 geeft een voorbeeld van een referentie.

Het trefwoordenregister is gewoonlijk samengesteld uit:

— *titelwoorden*: woorden uit de titel van de publicatie;

— *trefwoorden*: deze worden door een bij de data base producer werkzame indexerder aan de betreffende publicatie toegekend. Deze kunnen tweeledig zijn:

• trefwoorden, door de indexerder gekozen uit een woordenlijst (thesaurus) die speciaal samengesteld is voor een bepaalde data base; dit zijn de 'controlled terms' of 'descriptors'. Dit zijn tegelijkertijd ook de trefwoorden van het register van de gedrukte editie,

• trefwoorden, door de indexerder gekozen uit de titel of (korte) inhoud van de publicatie; dit zijn de 'uncontrolled terms', ook wel 'identifiers' genoemd;

— *abstractwoorden*: soms zijn ook alle woorden van de korte inhoud van een publicatie in het trefwoordenregister opgenomen (met behoud van enkele stopwoorden).

Controlled terms/descriptors zijn aan een vastgelegde spelling gebonden en terug te vinden in een gedrukte woordenlijst of

thesaurus). Alle andere trefwoorden zijn afkomstig van de auteur(s) en dus afhankelijk van zijn/haar taalgebruik. De ervaring leert dat dat heel veelzijdig kan zijn. Ze worden opgenomen zoals ze er staan en kunnen dus voorkomen:

— in enkel- en meervoud;

— met een verschillende spelling, bijv. Amerikaans of Engels: aluminum - aluminium; organisation - organization etc.;

— los of aan elkaar geschreven: off-peak - off peak.

2.2.2. Het zoekprofiel

Iedere geautomatiseerde zoekgang bestaat in principe uit de volgende handelingen:

1. beschrijf het onderwerp in enkele woorden zo nauwkeurig mogelijk; bijv. onderzoek naar de nitrificatie bij de zuivering van afvalwater;

2. verdeel het onderwerp in aspecten; in hetzelfde voorbeeld:

— aspect a: zuivering,

— aspect b: afvalwater,

— aspect c: nitrificatie;

3. zoek trefwoorden bij deze aspecten in de taal of talen, waarin de gewenste data base(s) afzoekbaar is (zijn) gemaakt;

4. maak een zogenaamd zoekprofiel: dat wil zeggen: maak een opsomming van de trefwoorden die het aspect beschrijven en

combineer deze trefwoorden op één of meerdere manieren.

Het zoekprofiel in het voorbeeld is in zijn eenvoudigste vorm als volgt:

— aspect a: het trefwoord *treatment* levert bij zoeken set 1 op,

— aspect b: het trefwoord *wastewater* levert set 2 op,

— aspect c: het trefwoord *nitrificatie* levert set 3 op,

— een combinatie van set 1, 2 en 3 levert set 4 op. Met deze set wordt een doorsnede van set 1, 2 en 3 verkregen;

5. kies een hostorganisatie en een gewenste data base;

6. voer de search uit: de computer legt nu de sets referenties vast volgens de hierboven genoemde voorwaarden:

— set 1 bestaat uit alle referenties in de gekozen data base die het trefwoord *treatment* bevatten,

— set 2 idem m.b.t. het trefwoord *wastewater*,

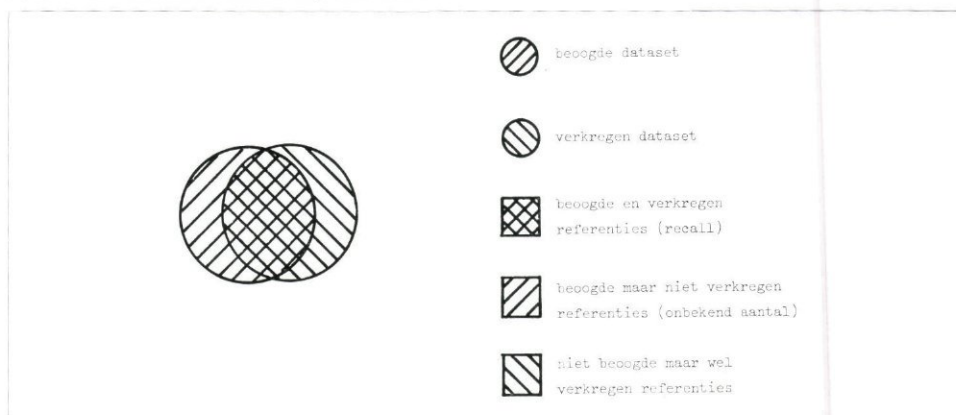
— set 3 idem m.b.t. het trefwoord *nitrification*,

— set 4 ten slotte bestaat uit alle referenties die zowel het trefwoord *treatment*, het trefwoord *wastewater* als het trefwoord *nitrification* bevatten (wiskundig een 'en' voorwaarde, waarmee een deelverzameling van de sets 1, 2 en 3 verkregen wordt). Dit wordt voorgesteld in afb. 2.

Nogmaals en met nadruk: bovengenoemde gang van zaken is slechts een simpele voorstelling van de werkelijkheid.

Als voorbeeld moge het volgende dienen: vóór het uitvoeren van de search moeten per aspect ook synoniemen in het zoekprofiel opgenomen worden: in het gegeven zoekprofiel *wastewater*, *waste water*, *sewage*, etc. (wiskundig een 'of' voorwaarde, waarbij de verzamelingen per trefwoord — set 2a, 2b en 2c — gesommeerd worden).

Afb. 3 - Schematische voorstelling van datasets.



Volgens een gekozen trefwoordencombinatie wordt online een dataset verkregen (bijv. set 4). Deze dataset bevat relevante en niet relevante referenties. Men vindt met de gekozen trefwoordencombinatie namelijk niet alleen dië literatuur waar men op uit is. Schematisch wordt dit verschijnsel voorgesteld in afb. 3. Een ideale search is dan uitgevoerd wanneer de cirkels (die deelverzamelingen uit de data base representeren) elkaar overlappen. Dat zal in het algemeen niet de realiteit zijn. Bij een strenge eis wordt in het algemeen een hoge precisie bereikt: enkele zeer relevante referenties. Het risico bestaat echter dat de eis te streng is en dat helemaal geen referenties gevonden worden. Bij een minder strenge eis wordt een hogere recall bereikt: een hogere opbrengst, met acceptatie van minder relevante referenties. Als voorbeeld kan genoemd worden de eis dat de gezochte literatuur de trefwoorden 'treatment' en 'wastewater' moet bevatten.

3. Beschikbare bibliografieën en bibliografische bestanden voor aan water verwante onderwerpen

Deze zijn samengevat in tabel I. Alle genoemde bibliografieën zijn beschikbaar op de Bibliografieënleeszaal van de Centrale Bibliotheek TH Delft, en zijn gratis te raadplegen, behalve uiteraard ** (zie tabel I). Alle bovengenoemde bibliografische bestanden kunnen daar, tegen betaling, geraadpleegd worden. Op de afdeling Literatuurinformatie van de Centrale Bibliotheek beschikt men hiertoe over twee terminals. Voor het geautomatiseerd zoeken kunt u een afspraak maken (tel. 015 - 78 56 54). Informatiespecialisten kunnen geïnteresseerden adviseren en assisteren. Het gebruik van de computer kost gemiddeld f 250,— tot f 300,— per uur. Indien er 'off-line'-verwijzingen worden afgedrukt, komt daar nog een bedrag per verwijzing bij, variërend van f 0,50 tot f 0,85 per stuk, prijzen per 1 januari 1982. Voor de TH-bevolking gelden gereduceerde tarieven.

4. Literatuurzoeken, voorbeeld 1

Welke voortreffelijke resultaten er kunnen worden bereikt met een computersearch, blijkt uit de resultaten van zo'n onderzoek over het onderwerp 'De praktijk van de flotatie', verricht door H. C. Rijpstra, destijds medewerker van de Vakgroep Gezondheidstechniek (Afdeling der Civiele Techniek) van de TH Delft.

Hij raadpleegde voor dit onderwerp de volgende bibliografische bestanden: Aqualine, CA Search, Compendex, Enviroline,

TABEL I - Bibliografieën en bibliografische bestanden op het gebied van drink-, afval- en grondwater.

Naam van de bibliografie	Naam van het bibliografisch bestand	Beschikbaar bij hostorganisatie ²
* Biological Abstracts	BIOSIS	ESA, DIS, SDC, e.a.
* Bulletin Signalétique	PASCAL	ESA, Telesystèmes
* Chemical Abstracts	CA Search	ESA, DIS, SDC, e.a.
Dokumentation Wasser	—	—
Ecological Abstracts	—	—
* Ecology Abstracts	IRL Life Sciences	DIS
* Engineering Index	Compendex	ESA, DIS, INKA, SDC
* Environment Abstracts	Enviroline	ESA, DIS, SDC
* Environmental Periodicals Bibliography	als bibliografie	DIS
* Excerpta Medica	als bibliografie	DIS
* Oceanic Abstracts	Oceanic	ESA, DIS
* Pollution Abstracts	Pollution	ESA, DIS
* Selected Water Resources Abstracts	als bibliografie	SDC
* US Government Reports	NTIS ¹	ESA, DIS, SDC
Wasser, Abwasser, Luft und Feste Abfallstoffe	—	—
* WRC Information	Aqualine ¹	ESA, DIS
**	Acompline	ESA

¹ de computerversie bevat meer referenties dan de gedrukte versie

² zie lijst van gebruikte afkortingen (§ 7)

* zowel als gedrukte versie (bibliografie) als data base (bibliografisch bestand) beschikbaar

** uitsluitend als data base beschikbaar

PASCAL en Pollution, alsmede de bibliografie Dokumentation Wasser.

Er werden online in totaal 362 relevante referenties gevonden waarvan 267 verschillende. In dit geval werd het grootste aandeel geleverd door PASCAL, gevolgd door respectievelijk Aqualine, Pollution,

Enviroline, CA Search en Compendex. Een overzicht van de opgespoorde referenties wordt weergegeven in tabel II.

Hieruit blijkt dat de precisie, het percentage relevante literatuur, zeer hoog is. Voor deze search werden ca. 75 computerminuten gebruikt. Door de hostorganisatie, in dit geval ESA, werd hiervoor (in kosten per 1-1-1982) ca. f 400,— berekend. De kosten zijn, uitgesplitst naar data base, weergegeven in tabel III.

Een met de hand uitgevoerd literatuuronderzoek in 'Dokumentation Wasser' leverde in totaal 70 relevante plaatsen op, waarvan 47 unieke niet op andere wijze gevonden referenties. Duidelijk is, dat wanneer men een uitgebreid literatuuroverzicht wil hebben, het beslist noodzakelijk is om meerdere bestanden te raadplegen en met de hand te zoeken. Voor verdere gegevens over deze search wordt verwezen naar Rijpstra [1980].

TABEL II - Aantallen gevonden en relevante referenties bij aparte raadpleging van de diverse bestanden.

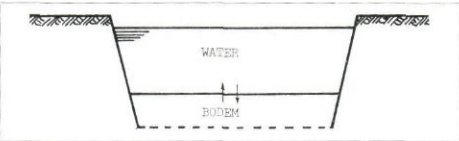
Geraadpleegd bibliografisch bestand	Totaal aantal gevonden referenties	Aantal relevante referenties	% relevant (precisie)
Aqualine	79	79	100
CA Search	34	33	97
Compendex	30	29	97
Enviroline	54	52	96
PASCAL	135	112	83
Pollution	59	57	97
Totaal	391	362	93
Overlap geëlimineerd		267	

TABEL III - Kosten van de search van Rijpstra.

Geraadpleegd bibliografisch bestand	Search-tijd	Kosten v/h gebr. bestand	Search-kosten	Print-kosten	Totale kosten	Aantal relevante referenties	Kosten per relevante referentie
	min.	f1/uur	f1	f1	f1		f1
Aqualine	14.58	146.65	35.64	39.50	75.14	79	0.95
CA Search	6.59	194.20	21.33	19.04	40.37	33	1.22
Compendex	7.19	200.20	23.99	22.20	46.19	29	1.59
Enviroline	9.22	205.90	31.64	29.16	60.80	52	1.17
PASCAL	28.30	135.35	63.84	55.50	119.34	112	1.06
Pollution	10.25	191.75	32.76	31.86	64.62	57	1.13
Totaal resultaat (overlap geëlimineerd)					406.46	267	1.52

Bovenstaande prijzen moeten worden verhoogd met:

- de kosten, die de PTT berekent voor het gebruik van Euronet; afhankelijk van het aantal overgevoerde karakters is dat f 15,— tot f 25,— per uur;
- de kosten van de telefoonverbinding van de searchplaats naar Amsterdam.



Afb. 4 - Schematische voorstelling van het onderzoekonderwerp van voorbeeld 2.

- Conclusies uit de genoemde search zijn:
1. PASCAL en Aqualine leveren de grootste bijdrage: resp. 42 en 30 % van het totaal aantal gevonden relevante referenties. Het percentage éénmalig gevonden literatuur bedroeg 80 %. Hieruit blijkt de noodzaak om meerdere bestanden te raadplegen.
 2. Het zoekprofiel is opgezet in Aqualine. Daarom is de searchtijd in Aqualine hoger dan in de andere bestanden. Uitzondering hierop vormt PASCAL want:
 3. PASCAL moet zowel met franstalige als engelstalige trefwoorden worden afgezoekt, waardoor de searchtijd ten minste 2 x zo hoog is als bij de andere bestanden.
 4. COMPENDEX is het duurste bestand geweest, Aqualine het goedkoopste.
 5. De precisie is zeer hoog (93 %), omdat het onderwerp zeer gemakkelijk in trefwoorden was om te zetten.

Dat het zoeken vaak niet zo eenvoudig is, moge blijken uit het in de volgende paragraaf te behandelen voorbeeld.

5. Literatuurzoeken, voorbeeld 2

5.1. Inleiding

Dit onderzoek wordt eveneens bij de Vakgroep Gezondheidstechniek van de Afdeling der Civiele Techniek van de TH Delft uitgevoerd. Het onderzoek heeft als doelstelling een bijdrage te leveren aan de kennis over de vraag wanneer, hoeveel, hoe snel en hoe lang fosfaten vanuit het water in de bodem (sediment) van meren terechtkomen en vanuit de bodem vrijkomen onder wisselende omstandigheden (afb. 4).

Om zoveel mogelijk kennis te vergaren over wat anderen reeds op dit onderzoeksterrein publiceerden

- werden vaktijdschriften, alsmede de attenderingsbladen sectie 61 van Chemical Abstracts en sectie 42 van Excerpta Medica bestudeerd en bijgehouden;
- werd contact onderhouden met vakgenoten.

Teneinde zo volledig mogelijk te zijn is besloten met behulp van een computer literatuur te zoeken en nauwkeurig te analyseren, hoe belangrijk deze informatie

is ten opzichte van op andere wijzen verkregen literatuur.

Het onderzochte onderwerp bevat 5 verschillende aspecten:

A interactie

B sediment

C water

D model

E stoffen

Het was de bedoeling om een zo volledig mogelijk literatuuroverzicht samen te stellen. Daarom is, naast bestanden waar op voorhand veel van werd verwacht, ook in minder bekende bestanden zoals CAB, Oceanic en EDF-DOC gezocht. Deze bleken te weinig op te leveren, zodat hiermee niet verder gewerkt is.

Bovendien werd een zeer uitgebreid zoekprofiel gemaakt, waarin voor alle aspecten zoveel mogelijk trefwoorden bij elkaar zijn gezocht; bijv. voor aspect C: lake, lakes, limnology, freshwater, clear water, water. Bovendien hebben we van de diverse aspecten verschillende combinaties gemaakt met een wiskundige 'en' relatie: ABCDE, ABCD, ABCE, ABE en CDE. Deze combinaties leveren deelverzamelingen op van de door de computer per aspect gecreëerde afzonderlijke sets (op dezelfde manier zoals

TABEL IV - Aantallen gevonden en relevante referenties bij aparte raadpleging van de diverse bestanden.

Geraadpleegd bibliografisch bestand	Totaal aantal gevonden referenties	Aantal relevante referenties	% relevante (precisie)
Aqualine	147	57	39
Biosis	75	35	47
CA Search	339	83	25
Compendex	27	13	48
Enviroline	55	8	15
NTIS	15	3	20
PASCAL	224	55	25
Pollution	22	4	18
Totaal	904	258	29
overlap geëlimineerd		193	

TABEL V - Searchkosten, uitgedrukt in prijzen per 1 jan. 1982, van de search van voorbeeld 2.

Geraadpleegd bibliografisch bestand	Searchkosten	Searchtijd	Searchkosten	Kosten offline prints	Kosten online prints	Totale kosten	Aantal relevante referenties	Kosten per relevante referentie
Aqualine ¹	146.65	33.0	80.66	50.00	9.40	140.06	57	2.46
BIOSIS ²	149.45	8.95	22.29	21.00	—	43.29	35	1.24
CA Search ²	194.20	42.23	136.68	162.96	13.44	313.08	83	3.77
Compendex ²	200.20	6.23	20.79	19.98	—	40.77	13	3.14
Enviroline ¹	205.90	10.44	35.83	19.44	—	55.27	8	6.91
NTIS ²	132.55	10.87	24.01	—	—	24.01	3	8.00
PASCAL ¹	135.35	68.0	153.40	85.00	8.50	246.90	55	4.49
Pollution ²	191.75	9.10	29.08	—	—	29.08	4	7.27
Totaal resultaat, overlap geëlimineerd						892.46	193	4.62

Bovenstaande prijzen moeten worden verhoogd met

a. de kosten, die de PTT berekent voor het gebruik van Euronet; afhankelijk van het aantal overgeseinde karakters is dat f 15,— tot f 25,— per uur;

b. de kosten van de telefoonverinding van de searchplaats naar Amsterdam.

¹ search uitgevoerd in november 1980; ² search uitgevoerd in maart 1981.

set 4 gecreëerd wordt in het voorbeeld van paragraaf 2 en afb. 2). ABCDE is daarbij de strengste eis: een referentie voldoet pas dan aan deze eis wanneer van elk aspect minstens één trefwoord in de aan de referentie toegekende trefwoorden voorkomt.

5.2. Resultaten en kosten van de online search

Het totaalresultaat van deze search – (relevante) referenties en kosten – is samengevat in de tabellen IV en V. De getallen gelden voor de gehele search.

De precisie ligt veel lager dan bij vb. 1: 29 %. CA Search geeft het beste resultaat, gevolgd door Aqualine en PASCAL. De kosten van één relevante referentie liggen hier hoger dan in de search van vb. 1. Ook de searchtijden zijn langer, enerzijds omdat het zoekprofiel meer elementen bevat, anderzijds omdat de grote bestanden niet in één zoekgang doorzocht konden worden. Ook hier wordt ca. 80 % slechts eenmalig gevonden. In het vervolg van dit artikel zijn alleen referenties uit de publicatiejaren 1978 en 1979 in beschouwing genomen. Deze publicatiejaren leveren 96 referenties op waarvan er 70 verschillend zijn.

5.3. Resultaten van manueel zoekwerk

Naast het online zoeken is gedurende het onderzoek ook manueel gezocht d.m.v. lezing van vaktijdschriften en attenderingsbladen. Er bestaan veel uitstekende tijdschriften waarin over het onderhavige onderwerp regelmatig gepubliceerd wordt. Enkele hiervan zijn H₂O, Journal of the Water Pollution Control Federation, Freshwater Biology, Water Research en Limnology and Oceanography. Dit is maar een kleine greep uit het grote aanbod. Attenderingsbladen zijn o.a. WRC-Information, sectie 60/61 van de Chemical Abstracts, Dokumentation Wasser etc. Ook

hier zijn voor dit rapport alleen publicaties uit de publicatiejaren 1978 en 1979 in beschouwing genomen. De resultaten van het met de hand zoeken blijken werkelijk niet gering. Er werden 104 stuks originele literatuur of literatuurverwijzingen gevonden. Dit kan voor een deel komen omdat zoveel goede tijdschriften en attenderingsbladen direct voorhanden zijn geweest. Het goed zoeken met de hand kost echter steeds bijzonder veel arbeidstijd. Daarbij spelen het toeval en de aandacht een grotere rol dan bij het online zoeken; het toeval omdat een computer nu eenmaal systematischer zoekt, de aandacht omdat deze verflauwt bij het enkele dagen achtereen in bibliografieën speuren.

5.4. *Vergelijking van de online search en manueel zoekwerk*

Een volledig overzicht wordt gegeven in De Jong-Hofman en Siebers [1982]. Bij vergelijking van de online search en het manueel zoekwerk worden de volgende totalen aan relevante referenties gevonden (searchperiode 1978 en 1979):

– uitsluitend online gevonden	:	29
– uitsluitend manueel gevonden	:	63
– zowel online als manueel gevonden:		41
totaal aantal gevonden referenties		: 133

63 referenties (47 %) van het totaal zijn dus niet met de computer teruggevonden.

Oorzaken hiervoor zijn o.a.:

— aan ongeveer de helft ontbraken enkele of alle essentiële trefwoorden;

— het niet opnemen van een deel van de publicaties in de data bases.

Deze conclusies komen volledig overeen met de resultaten van een eerder gedaan onderzoek [De Jong-Hofman, 1981]. Bij de niet opgenomen publicaties bevinden zich 10 nederlandstalige.

Het blijkt dus dat speciale problemen rijzen bij het opsporen van nederlandstalige literatuur. Dat is echter bij zowel het zoeken in bibliografieën als in bibliografische bestanden het geval.

5.5. *Verkrijgbaarheid en beoordeling van de oorspronkelijke publicaties*

In de publicatiejaren 1978 en 1979 zijn door gecombineerd manueel en online zoeken in totaal 133 literatuurplaatsen in eerste instantie als waarschijnlijk relevant beschouwd.

In het algemeen is pas na lezing van het originele document in voldoende mate te beoordelen of de gevonden referentie een voor het onderwerp bruikbare publicatie beschreef. Daarom is het belangrijk of de publicaties wel of niet in Nederland ver-

krijgbaar zijn en op welke termijn deze verkregen kunnen worden. De kosten spelen hierbij een belangrijke rol: een bij de TH gemaakte kopie van een in deze bibliotheek aanwezig document kost f 0,40 per bladzijde, terwijl een in het buitenland gemaakte kopie excl. verzendkosten al gauw f 1,25 kost. Een overzicht van deze kosten verscheen onlangs in Login [Van de Gein, 1981]. Bij de hier beschreven search is in eerste instantie getracht de originele documenten in het bestand van de (Centrale) Bibliotheek van de TH in Delft te verkrijgen. Wanneer het document hier niet beschikbaar bleek te zijn, is vervolgens uitgezocht of het document bij een andere bibliotheek in Nederland verkregen kon worden. In dit geval vnl.: de Koninklijke Nederlandse Academie van Wetenschappen, het Shell Expl. Laboratorium, het Rijksinstituut voor de Zuivering van Afvalwater, het Rijksinstituut voor Drinkwatervoorziening en het Rijksinstituut voor Visserij-onderzoek. Van de 133 oorspronkelijke documenten konden er 108 in Nederland verkregen worden (81 %), waarvan 85 (64 %) in de Centrale Bibliotheek van de TH Delft. Na lezing bleven van deze 108 documenten 88 relevante documenten over (66 % van het totaal van 133 referenties), 25 van de 133 literatuurplaatsen (19 %) waren niet in Nederland verkrijgbaar. Ten aanzien van deze literatuur zal men moeten beslissen of deze in het buitenland besteld moet worden (tegen vaak hoge kosten en lange wachttijden) of dat men wellicht gaat trachten op andere manieren aan deze literatuur te komen. Het materiaal dat zowel met de computer als met de hand is gevonden, blijkt hoog relevant (78 %) en goed verkrijgbaar (92,6 %). Het materiaal dat uitsluitend met de computer is gevonden is niet zo relevant (48,3 %) en ook moeilijk verkrijgbaar (69 %). Hieruit blijkt hoe belangrijk het met de hand zoeken kan zijn. Voor een goed en zo compleet mogelijk overzicht over een te onderzoeken onderwerp is het dus raadzaam om zowel manueel als online literatuur op te sporen. In een onderzocht geval zouden over de publicatiejaren 1978 en 1979 bij alleen online resp. manueel zoeken 70 (53 %) resp. 104 (78 %) van de in totaal 133 relevante literatuurplaatsen teruggevonden zijn. Hierbij moet worden aangetekend dat er vrijwel zeker meer dan 133 relevante literatuurplaatsen bestaan.

6. **Conclusies**

— Het uitvoeren van een online search is een gerieflijke en zeer tijdbesparende manier om literatuur op te sporen.

— Het is raadzaam om enige tijd te

investeren in het voorbereiden van de search, het vinden van trefwoorden, synoniemen etc., dus in het maken van een zoekprofiel en het uitstippelen van de zoekstrategie.

— Welk percentage van de gepubliceerde relevante literatuur met een online search opgespoord zal worden, is niet te voorspellen. Dat is afhankelijk van het onderwerp, het (de) gebruikte zoekprofiel(en) en de gekozen data base(s).

— Hoe verschillend de resultaten kunnen zijn, blijkt uit de twee in dit artikel behandelde voorbeelden. Het eerste onderwerp is zeer gemakkelijk om te zetten in trefwoorden, het andere juist heel moeilijk. Dit blijkt van grote invloed op de precisie en de kosten. Onder precisie wordt verstaan het percentage gevonden relevante referenties van het totaal aantal gevonden referenties. In de beide voorbeelden bedraagt deze precisie gemiddeld resp. 93 % en 29 %.

— Het zoekprofiel in voorbeeld 2 werd noodgedwongen erg lang en bovendien konden de grote bestanden CA Search en PASCAL niet in één search doorzocht worden, omdat dit een geheugenoverflow in de computer veroorzaakte. Dit heeft grote invloed op de kosten van een relevante referentie gehad: f 1,52 in vb. 1 en f 4,62 in vb. 2 per relevante referentie.

— In beide voorbeelden bedraagt het percentage éénmalig gevonden literatuurverwijzingen ca. 80 %, waaruit de conclusie mag worden getrokken dat voor het verkrijgen van een zo volledig mogelijk literatuuroverzicht altijd meerdere bestanden moeten worden geraadpleegd. De belangrijkste bibliografische bronnen voor het vinden van publicaties op de in dit rapport behandelde vakgebieden, zijn Aqualine, CA Search en PASCAL.

— Ook het manueel zoeken in bibliografieën en dagelijkse aandacht voor attenderingsbladen en vaktijdschriften blijkt onmisbaar. Men moet dan natuurlijk wel over zowel tijd als literatuur kunnen beschikken.

— Het manuele zoekwerk leverde een aantal publicaties op die niet met de online searches teruggevonden werden. De voornaamste redenen hiervoor zijn:

- er vindt een selectie door reviewers plaats, waardoor belangrijke publicaties niet worden opgenomen in de data bases;
- door onvoldoende indexering en mager taalgebruik van auteurs, ontbreken essentiële sleutelwoorden aan de referenties, zowel in de vorm van toegekende trefwoorden, als in de vorm van andere afzoekbare woorden zoals titel- en abstractwoorden.

— Nederlandstalige publicaties worden maar ten dele in de data bases opgenomen, derhalve zijn ze heel moeilijk terug te vinden. Wil men toch op de hoogte blijven van hetgeen in de Nederlandse taal over een onderwerp wordt gepubliceerd, dan moet men daarvoor de vakliteratuur lezen.

— Van de originele publicaties die relevant leken voor het onderzoek, kon 64 % door de Centrale Bibliotheek van de TH Delft geleverd worden; andere bibliotheken in Nederland leverden nog eens 17 %. 19 % was niet in Nederland verkrijgbaar.

— Van de originele publicaties bleek na lezing 81 % relevant voor het onderzoek te zijn (vb. 2).

— Uit een achteraf controle van de bronnen door middel waarvan deze originele relevante literatuur is opgespoord, blijkt dat het lezen van vaktijdschriften en attenderingsbladen onmisbaar was voor het verkrijgen van deze bundel publicaties, evenals ieder van de searches in elk van de data bases.

Het volledige rapport, waarvan dit artikel een samenvatting is [De Jong-Hofman en Siebers, 1982], is bij de auteurs te bestellen à f 9,50 (excl. verzendkosten).

7. Lijst van afkortingen en begrippen

ESA	European Space Agency
INKA	Informationszentrum Karlsruhe
IRL	Information Retrieval Limited
DIS	Dialog Information Systems
NTIS	National Technical Information Service
PASCAL	Programme Appliqué de la Sélection et à la Compilation Automatique de la Littérature
SDC	System Development Corporation
TELE	Telesystèmes-Questel
WRC	Water Research Centre

Literatuur

- Gein, H. A. G. M. van de. *Overzicht van prijzen voor documenten, besteld via ESA - IRS Primordial*. LOGIN, 1981, nr. 4, 13-15.
- Jong-Hofman, M. W. de. *Comparison of selecting, abstracting and indexing by Compendex, Inspec and PASCAL and the impact of this on manual and automated retrieval of information*. Online Review, vol. 5, 1981, nr. 1, 25-36.
- Jong-Hofman, M. W. de en Siebers, H. H. *Ervaringen met geautomatiseerd literatuurzoeken; het opsporen van literatuur op het vakgebied van drink-, afval- en oppervlaktewater: van referaat tot originele publicatie*. Geautomatiseerd en manueel zoeken vergeleken. Rapport TH Delft, februari 1982, 32 pp.
- Newman, L. et al. *Waterliteratuur toegankelijk per computer: Aqualine*. H₂O (14), 1981, 340-341.
- Rijpsma, H. C. *Een online literatuuropsporing*. Delft, oktober 1980, 12 pp.

Verslag van het Thematische gedeelte van de Jaarvergadering van de N.V.A. op 12 mei 1982 in het RAI-Congresgebouw te Amsterdam

De eerste voordracht met als titel „De bestrijding van de waterverontreiniging in Nederland en de E.E.G.” werd gehouden door ir. J. H. Jansen, hoofdingenieur-directeur van het R.I.Z.A. te Lelystad.

Deze duidelijke en overzichtelijke voordracht, waarin werden belicht:

- de rol van de E.E.G.
 - de wetgeving in Nederland
 - het Nederlandse lozingsbeleid
 - de stand van zaken van de afvalwaterzuivering in de E.E.G.
 - de slibproblematiek in Nederland
 - financiële- en economische aspecten kan als volgt worden samengevat:
 - a. Internationale bestrijding van de waterverontreiniging is voor Nederland erg belangrijk. De E.E.G. speelt hierbij een grote rol.
 - b. Zwarte-lijst stoffen dienen in E.G.-kader versneld te worden aangepakt.
 - c. Vergelijking van de bestrijding van de waterverontreiniging in de diverse E.G. lidstaten is moeilijk wegens gebrek aan betrouwbare statistische gegevens.
 - d. Wat betreft afvalwaterzuivering slaat Nederland in E.E.G.-verband een goed figuur.
 - e. In Nederland wordt relatief veel zuiveringsslib in de landbouw toegepast.
 - f. De eisen, die in Nederland worden gesteld t.a.v. toelaatbare concentraties aan zware metalen, zijn aanmerkelijk hoger dan die van de betreffende E.G.-Commissie.
 - g. T.a.v. het beginsel „de vervuiler betaalt” bestaan er binnen de E.E.G. grote verschillen.
 - h. Verdere harmonisatie binnen de E.E.G. betreffende de bestrijding van waterverontreiniging is van groot belang voor het milieu in Nederland.
- De tweede voordracht was voor rekening van dr. ing. E. H. E. Kuntze, Leiter der Stadtentwässerung Hamburg, voorzitter van de A.T.V. van de BRD en tevens voorzitter van de E.W.P.C.A.
- Hij gaf een overzicht van „Abwasserbeseitigung in der Bundesrepublik Deutschland 1982”.
- Diverse aspecten werden behandeld:
- Wetgeving en hun doel

- Uitvoering van en controle op wetten
- Technologie en praktijk van afvalwaterzuivering
- Slibverwerking
- Kosten en investeringen.

Sinds 1958 zijn door de Bondsregering veel raamwetten gemaakt voor het bestrijden van waterverontreiniging, bijv. de Wet op de Waterhuishouding en de Wet op de Afvoer van Afvalwater. Zij hadden veel invloed op het tot standkomen van een groot aantal industriële afvalwaterzuiveringsinstallaties, het wijzigen van produktieprocessen, waardoor minder afvalwater ontstond.

De heffingen op basis van bovengenoemde wetten moeten worden besteed voor het bouwen van afvalwaterzuiveringsinstallaties, het verwerken van zuiverings-slib, opleiding van personeel van zuiveringsinstallaties, verbetering van de wateren.

De toepassing van slib in de landbouw zal binnenkort door een wet worden geregeld. Evenals in Nederland speelt ook in de BRD het toelaatbare gehalte aan zware metalen in slib een belangrijke rol.

Naast industrie en grote gemeenten zijn er in de BRD zgn. „Abwasserzweckverbände”, meestal meerdere gemeenten, die gezamenlijk het transport en zuiveren van afvalwater verzorgen.

De bondsregering controleert de afvoer van zuiveringsinstallaties e.d. in rivieren en andere wateren.

In de BRD heeft de A.T.V. (Abwasser Technische Vereinigung) een belangrijke taak bij het tot standkomen van regelingen en richtlijnen voor bestrijding van waterverontreiniging.

Bij de afvalwaterzuivering wordt gestreefd naar een volledige biochemische zuivering, zodat in veel gevallen een 2e biochemische trap wordt toegepast.

Vermindering van afvoer van schadelijke stoffen zal in de komende jaren bij de bron moeten worden aangepakt.

Door de A.T.V. worden in samenwerking met de overheid veel cursussen georganiseerd voor opleiding van medewerkers. De diploma's worden binnenkort door de overheid erkend.

Ook de BRD kent het probleem van de afzetmogelijkheden van zuiveringsslib. Het storten in meren kan niet meer, bij verbranden kan luchtverontreiniging optreden, de as kan vaak alleen naar speciale stortplaatsen worden afgevoerd.

Er bestaat een tendens naar het bouwen van kleinere zuiveringsinstallaties, omdat meer verdeelde, kleinere toevoeren naar rivieren beter worden geacht dan één grote op één plaats.

De investeringen door gemeenten voor