

ONTOLOGIEËN VOOR VOEDING

door Jeen Broekstra, Nicole Koenderink, Jan Top,
jeen.broekstra@wur.nl

Kwesties rond voeding vragen wereldwijd dagelijks onze aandacht. In grote delen van de wereld ontstaan voedseltekorten en in andere delen is juist sprake van overconsumptie. Producenten van voeding en overheden in de westerse wereld maken zich zorgen over voedselkwaliteit en het voorkomen van voedselgerelateerde gezondheidsproblemen.

Inleiding

Optimale inzet van informatie en kennis is tegenwoordig cruciaal bij de productie, verwerking en distributie van voeding. Informatie is nodig voor afstemming in steeds meer wereldomvattende productieketens, en kennis is nodig voor innovatie van producten en processen. Het ontsluiten van informatiebronnen, in het bijzonder via het web, en het bouwen van kennisintensieve webapplicaties wordt steeds belangrijker in deze sector.

De groep *Information Management* van Wageningen UR (Universiteit en Researchcentrum) past *Semantisch-Webtechnologie* toe in deze bedrijfstak. Enkele voorbeelden zijn:

- *Emerging Risks Detection System*: voortijdige detectie van mogelijke voedselveiligheidsproblemen in de toekomst. Het systeem signaleert mogelijk relevante ontwikkelingen op het web en vertaalt deze via kennisregels over voedselveiligheid naar risico's. Belanghebbenden zijn bijvoorbeeld de Voedsel- en Warenautoriteit (VWA) binnen Nederland en European Food Safety Association (EFSA) op Europees niveau.
- *Competence Guide* voor *Food & Nutrition Delta*: een digitale gids voor het vinden van experts op het gebied van innovatie in voedselverwerking.

Twee andere applicaties (het Tiffany systeem en de database gezonde inhoudsstoffen) worden verderop toegelicht.

Softwaretoepassingen voor de voedingswereld, al dan niet gebaseerd op het web, vragen steeds meer om standaardisatie, flexibiliteit en expressiviteit. Representatietalen als RDF en OWL komen tegemoet aan die vraag, vooral nu er steeds meer gebruiksvriendelijke gereedschappen beschikbaar komen. Voor het uitwisselen van data is het een hele vooruitgang dat naast losse concepten (met unieke identificatie) ook relaties tussen concepten in de vorm van *triples* uitgewisseld kunnen worden, en dat deze dankzij de inspanningen van W3C ook wereldwijd begrepen worden. De *Competence Guide* gebruikt triples om de gebruiker te helpen met het associatief formuleren van zoekvragen. Daarnaast zijn repositories gebaseerd op triples flexibeler en daardoor rijker aan informatie. Voor bovengenoemde applicaties gebruiken wij typisch het Sesame-raamwerk (zie <http://www.openrdf.org/>) om RDF-triples te ontsluiten.

Tenslotte is de sterkere expressiviteit van RDF en OWL boven XML

belangrijk om kennisregels uit te drukken, zoals in het geval van het genoemde *Emerging Risks Detection System*. Blijkbaar kunnen nieuwe representatieformaten zoals RDF en OWL veel betekenen voor de wereld van voeding. *Food Informatics*, onderdeel van het grootschalige onderzoeksprogramma *Virtual Lab for e-Science (vl-e)*, richt zich op het construeren en toepassen van (min of meer) formele vocabulaires bij voedingsonderzoek. De ontwikkelde technieken zijn echter breder toepasbaar. Een voorbeeld van een dergelijke techniek is ROC, een methode voor *Rapid Ontology Construction*. Hiermee kunnen leken op het gebied van kennismodellering zelf efficiënt vocabulaires samenstellen voor specifieke applicaties.

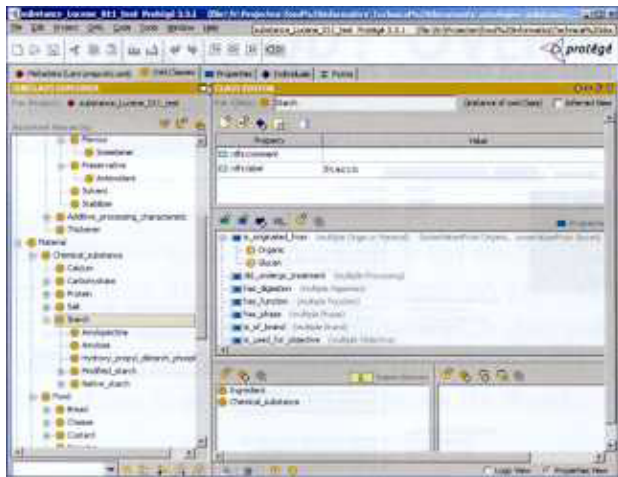


Research Management in voedingsonderzoek

Een van de grote uitdagingen waar onderzoekers en producenten van voedsel en voeding zich mee geconfronteerd weten, is om de 'gezonde keuze' ook de 'makkelijke keuze' te maken. Om deze uitdaging aan te gaan is het van belang dat industrie en onderzoeksorganisaties kennis en capaciteit voor multidisciplinair onderzoek delen. Het *TI (Top Institute) Food and Nutrition* is een samenwerkingsverband tussen publieke en private partners dat zich bezighoudt met wetenschappelijke doorbraken in voeding en gezondheid. Dit heeft geresulteerd in innovatieve producten en technologieën die voldoen aan consumenteneisen voor veilige, gezonde en smakelijke voeding.

Onderzoeksprojecten bij *TI Food and Nutrition* worden uitgevoerd, in de laboratoria van de onderzoeksorganisatie, door multidisciplinaire teams. Het onderzoek is georganiseerd in drie hoofdthema's waaraan – gezamenlijk door onderzoeks- en industriepartners – onderzoeksprojecten worden uitgevoerd door multidisciplinaire teams in de laboratoria van de onderzoeksorganisatie.

Een belangrijke missie van *TI Food and Nutrition* is kennisoverdracht tussen onderzoekers en tussen de verschillende partners van het instituut. De uitdagingen in het faciliteren hiervan zijn legio: aangezien onderzoekers veelal op verschillende locaties werken



Figuur 1: Tiffany systeem

is kennisoverdracht via seminars of mondeling overleg niet een structurele oplossing. Kennisoverdracht en kennisborging (dat wil zeggen het vastleggen van kennis op een manier die terugvinden en overdracht in de toekomst mogelijk maakt) zullen gefaciliteerd moeten worden via systemen die voor onderzoekers op locatie toegankelijk zijn.

Het Tiffany systeem

Het Tiffany systeem is een *research management systeem*, ontwikkeld in opdracht van *TI Food and Nutrition*. Het systeem stelt gebruikers (onderzoekers en partners van *TI Food and Nutrition*) in staat gedetailleerde informatie over onderzoek op te slaan, terug te vinden en met elkaar uit te wisselen. Tevens biedt het systeem actieve ondersteuning voor *workflows*, dat wil zeggen de onderzoeker wordt begeleid in elke stap van het onderzoeksproces de juiste informatie te formuleren en de juiste volgende stap te selecteren.

Het onderzoeksproces binnen *TI Food and Nutrition* is opgedeeld in vijf hoofdactiviteiten:

- (1) het formuleren van de onderzoeksvraag,
- (2) voorbereidende activiteiten,
- (3) het uitvoeren van een experiment,
- (4) het analyseren van de resultaten en
- (5) het rapporteren van de bevindingen.

Om deze activiteiten (en de producten die uit elke activiteit volgen) te formaliseren is een centraal element van Tiffany een in OWL geformuleerde *procesontologie*. Tiffany bevat ook een thesaurus met daarin domeinspecifieke terminologie rondom voedsel en voedselonderzoek. Hierin liggen domeinspecifieke termen als “cheese”, “dairy” enz., alsmede de relaties tussen deze termen, vastgelegd. Ook deze thesaurus is gemodelleerd in OWL.

De rol van Semantisch-Webtechnologie

Zoals aangegeven maakt Tiffany gebruik van Semantisch-Webtechnologie. De kern van het systeem zijn twee in OWL gemodelleerde ontologieën, de procesontologie en de domeinontologie. Zowel deze ontologieën als de onderzoeksinformatie die de gebruikers

toevoegen worden opgeslagen via het Sesame RDF framework en ontsloten via het gebruik van SPARQL queries gegenereerd vanuit een .Net webapplicatie.

Een groot voordeel van deze aanpak is de *flexibiliteit* van het informatiemodel: semantiek en structuur zijn losgekoppeld en expliciet gemaakt, waardoor het formuleren van queries puur in termen van de betekenis van informatie gebeurt. Dit maakt het mogelijk dat zoeken naar informatie op vele aspecten van die informatie (niet alleen keywords, maar ook type, eigenaar, datum, project, enz. enz.) kan gebeuren. Deze functionaliteit is door de keuze voor Semantisch-Webtechnologie zeer snel te ontwikkelen en aan te passen. Bovendien laat de flexibiliteit van het systeem ingrijpende veranderingen in de modellering toe, zonder dat deze hoeven te leiden tot drastisch herontwerp van de applicatie zelf.

Een tweede voordeel is *hergebruik* van informatie. De eerder genoemde domeinontologie zoals gebruikt in Tiffany is een thesaurus van termen en relaties uit het voedseldomein. Deze thesaurus is niet specifiek voor Tiffany ontwikkeld maar in een ander voedselgerelateerde project. Door de keuze voor RDF/OWL is hergebruik van deze structuur triviaal.

Database Gezonde Inhoudsstoffen

Een tweede voorbeeld van het gebruik van ontologieën in het agrifood domein is de database gezonde inhoudsstoffen.



Het *Innovatiecentrum Gezonde Voeding (ICGV)* is enige tijd geleden opgericht als een unieke publiek-private samenwerking van bedrijfsleven, kennisinstellingen en overheden. Wageningen UR en Universiteit Maastricht zijn samen met VGZ, Nunhems Vegetable Seeds, LLTB, en de gemeentes Venlo en Horst aan de Maas de partners in deze samenwerking. Het ICGV richt zich op productinnovatie bij midden- en kleinbedrijven in de agro-foodketen. Het brengt daartoe kennisinstellingen en bedrijven uit de agro-foodketen bij elkaar in zogenaamde kennisplatforms.

Een van de speerpunten van het ICGV is informatie bieden over gezondheid van voeding. Er is wereldwijd redelijke consensus onder voedingskundigen dat groenten en fruit belangrijke bronnen zijn van gezondheidsbevorderende inhoudsstoffen en dat zij bescherming bieden tegen diverse ziekten. Als voorbeeld noemen we de stof ‘lycopeen’ die aanwezig is in tomaten en bescherming biedt tegen prostaatkanker. In welke rassen tomaten lycopene in de hoogste mate aanwezig is, in hoeverre het bewerken van tomaten tot tomatensoep of spaghettisaus de hoeveelheid lycopene beïnvloedt, en hoe robuust het effect van lycopene is aangetoond, zijn vragen die lastig te beantwoorden zijn. De wetenschappelijke bewijzen voor dergelijke gezondheidsclaims zijn niet altijd aanwezig, laat staan dat ze via één plek toegankelijk zijn.

Om een begin te maken met het samenbrengen van dergelijke informatie heeft AFSG op verzoek van het ICGV en in samenwerking met domeinexperts binnen Wageningen UR en Universiteit Maastricht, een webapplicatie ontwikkeld, waarin van verschillende gewassen (blauwe bes, paprika, koolsoorten) dergelijke gegevens zijn opgeslagen. Deze informatie is met name interessant voor groentetelers en de verwerkende industrie, maar ook voor diëtisten, huisartsen en voor cheffoks en instellingenkokks.

Systeembeschrijving

De bovengenoemde webapplicatie bestaat uit vier componenten: een RDF-datamodel, een bijbehorende Sesame repository, een .NET user interface, en een tussenlaag om de Sesame-data aan de user interface te koppelen. In het RDF-datamodel zijn concepten als gewas, inhoudsstof, bewerking, gezondheidseffect, etc. via relaties aan elkaar verbonden. De domeinexperts is vervolgens gevraagd om gegevens in te voeren over paprika, blauwe bes en verschillende koolsoorten. Hierbij gaat het om algemene gegevens over bijvoorbeeld de Latijnse en Engelse naam, een overzicht van de soorten inhoudsstoffen in het gewas en de bijbehorende gezondheidseffecten.

Ook is er gevraagd om gedetailleerdere informatie over de specifieke gehaltes van de inhoudsstoffen in verschillende rassen en over de invloed van bewerkingen (bewaren, snijden, blancheren, magnetron,...). Deze informatie is opgeslagen in een Sesame repository. Via een aantal modules in de *Data Acquisition Layer* (DAL) van de ontwikkelde .Net-applicatie worden de gegevens in de Sesame repository opgehaald en in de user interface weergegeven. In de user interface kan door inzoomen meer rasspecifieke of bewerking-specifieke informatie worden opgevraagd over een bepaalde inhoudsstof in het gewas en het bijbehorende gezondheidseffect. De huidige versie van de webapplicatie is te vinden op <http://www.afsg.nl/icgv>.

De keuze voor een ontologie

Voor het realiseren van deze applicatie hebben we gekozen voor een RDF-datamodel. De reden hiervoor is de flexibiliteit die een dergelijk model biedt met betrekking tot toekomstige uitbreidingen. In eerste instantie is het datamodel puur gericht op gewassen – rassen – inhoudsstoffen – gezondheidseffecten. In de voorbereidende fase werd echter al aangegeven dat men allerlei uitbreidingsmogelijkheden zag. Men zou meer informatie over de invloeden van teeltcondities, procescondities en logistieke condities kunnen opnemen met betrekking tot de inhoudsstofgehaltes. Ook was er uitbreiding mogelijk richting de gezondheidsinformatie toegespitst op specifieke bevolkingsgroepen zoals senioren, kinderen, etc. Door het datamodel in RDF op te zetten kunnen concepten en daaruit voortvloeiende veranderende relaties gemakkelijk worden toegevoegd of gewijzigd, zonder dat de data opnieuw moet worden ingevoerd. Ook kunnen



Figuur 2: Database Gezonde Inhoudsstoffen

gebruikers hun interface zelf instellen, los van de inhoud waar ze naar kijken.

Conclusie

Modelvorming is altijd een van de belangrijkste aspecten van ICT-toepassingen geweest, en zal dat ook wel blijven. De wijze waarop een organisatie, een proces, een taak etc. digitaal wordt gepresenteerd heeft zich echter in de loop van de tijd ontwikkeld richting betekenisvolle communicatie. RDF en OWL passen in die ontwikkeling. Ook in de wereld van voedselproductie en -distributie is er steeds meer behoefte aan methoden en systemen om informatie en kennis te delen op conceptueel niveau. Ook bijvoorbeeld een organisatie als FAO (*Food and Agriculture Organisation*), onderdeel van de Verenigde Naties, is erg actief op dit gebied.

In de toekomst zal ook het verwerken van gegevens steeds belangrijker worden, bijvoorbeeld voor het afleiden van nieuwe feiten of het controleren van de kwaliteit van de informatie. Daarin krijgen standaarden als RDF en OWL, of mogelijk nieuw te ontwikkelen 'semantische' standaarden ongetwijfeld een rol. Als het om specifieke niches gaat is dit al het geval, zoals we hier hebben aangegeven. In mainstream-toepassingen, zoals logistieke orderverwerking en financiële transacties, zullen de huidige XML-gebaseerde oplossingen voorlopig de overhand hebben. Het is echter zinvol om te onderzoeken of ook in die gebieden de meer geavanceerde standaarden voor kennisrepresentatie toch een meerwaarde bieden. Omdat hierbinnen veel modelleringskeuzen al gemaakt zijn hoeft dit geen onoverkomelijke stap te zijn. Bovendien komen er steeds meer algemene ontologieën beschikbaar op het gebied van voeding waaruit geput kan worden.

◆ ■ ■ ■ Jeen Broekstra, Nicole Koenderink, en Jan Top zijn allen werkzaam bij de groep Information Management van de Agrotechnology and Food Sciences aan de Wageningen UR (Universiteit en Research Centre). Voor meer informatie over deze groep, zie <http://www.afsg.nl/InformationManagement/> ■ ■ ■ ◆