

EXPERTS AAN HET STUUR – MODELLEREN MET ROC

Nicole Koenderink, Lars Hulzebos, Mark van Assem, Jeen Broekstra, Remko van Brakel en Jan Top

Wageningen UR – Information Management
<http://www.afsg.nl/InformationManagement>

De agrifood sector heeft te maken met een sterk concurrerende wereldmarkt. Innovatie is dan ook van levensbelang. Optimaal gebruik maken van beschikbare kennis is een basisvoorwaarde voor innovatie. Kennis over de eigen business en kennis uit externe bronnen moeten in kaart worden gebracht en worden ontsloten. Binnen het Food Informatics project hebben we onderzocht hoe Semantic Web technologie hiervoor gebruikt kan worden. Een belangrijke conclusie is dat relevante **kennismodellen** onmisbaar zijn om de aanwezige informatie te ontsluiten. Om het kostenefficiënt bouwen van kennismodellen mogelijk te maken hebben we de **ROC-methode** ontwikkeld.

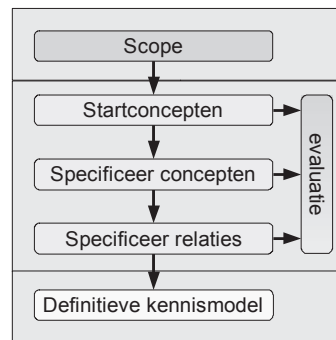
Kennismanagement in de praktijk

Iedereen kent de situatie dat iemand afscheid neemt en een kast met projectmappen en een computer met relevante files overdraagt aan een collega. In de praktijk blijkt de overgedragen informatie nooit meer gebruikt te worden, omdat het lastig is om snel de juiste documenten te vinden. Of een situatie waarin je weet dat je ooit iets hebt geschreven over een bepaald onderwerp, maar je geen idee meer hebt in welke context dat was. Het terugvinden van een dergelijk document is dan bijna onmogelijk zonder gebruik van slimme zoekmethodes. Semantic Web technologie biedt de mogelijkheid om bestaande kennis beter te ontsluiten.

De laatste jaren vindt een opmars plaats van semantic web technologie op internet. Tags – labels die beschrijven waar een item overgaat – zijn op websites als del.icio.us (www.delicious.com) en Flickr (www.flickr.com) al volledig ingeburgerd. In zoekfuncties zoals die van de openbare bibliotheek van Amsterdam (<http://zoeken.oba.nl/>) worden thesauri gebruikt om zoektermen (grafisch) in hun context te plaatsen. Hierdoor kunnen gebruikers hun zoekvraag aanscherpen en gemakkelijker de gewenste documenten vinden. Dit zijn slechts enkele voorbeelden van informatiemangement in de praktijk. De ontwikkelingen staan niet stil en de toepassingen van kennisintensieve applicaties zijn eindeloos.

De ROC-methode

Bij de ontwikkeling van kennisintensieve applicaties is vrijwel altijd een kennismodel van het domein van de applicatie nodig. Een dergelijk model legt de belangrijke begrippen in de gewenste applicatie en hun onderlinge relaties vast. De applicatie krijgt als het ware inzicht in de betekenis van de informatie, waarmee taken zoals het vinden van informatie,



Figuur 1: De ROC-workflow

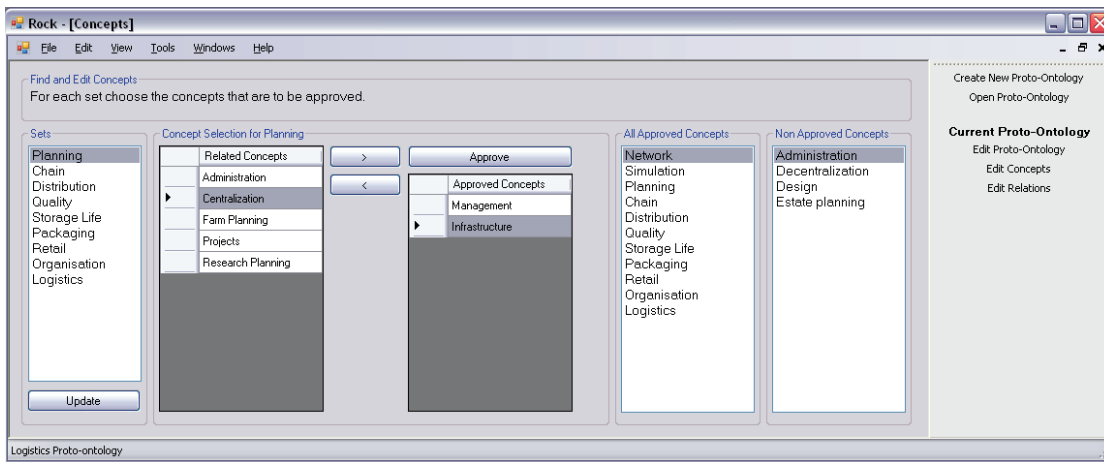
het annoteren van bronnen en het automatisch redeneren met bestaande kennis mogelijk worden. Deze activiteiten zijn belangrijk voor innovaties in een organisatie.

Het maken van een dergelijk kennismodel – een ontologie – is vaak een duur en intensief proces. Traditioneel vinden interviews plaats door een informatiekundige met een domeinexpert. De domeinexpert draagt zijn kennis over aan de informatiekundige, die er vervolgens een model over opstelt. De domeinexpert moet in een dergelijk traject alle basiskennis in detail uitleggen; de informatiekundige wordt opgeleid in het vakgebied van de domeinexpert. Dit proces is voor beide deelnemers niet efficiënt.

In het Food Informatics project hebben we de ROC-methode ontwikkeld. ROC staat voor Rapid Ontology Construction en is een methode die speciaal ontworpen is om het kennismodelleringsproces efficiënter te maken. Met ROC brengt de domeinexpert zelfstandig de relevante kennis in. De informatiekundige voert vervolgens de modelleerstep uit. Er zijn drie belangrijke spelers in de methode. De **problemeigenaar** overziet het doel van het kennismodel en is verantwoordelijk voor het bepalen van de scope en het evalueren van het gemaakte model. De **domeinexpert** is inhoudelijk deskundig op zijn eigen vakgebied, is verantwoordelijk voor het vastleggen van kennis in het model en voor evaluatie van het model samen met de probleemeigenaar. De **informatiekundige**, tot slot, is verantwoordelijk voor het formaliseren van het kennismodel

De belangrijkste kenmerken van ROC zijn:

- Het **scheiden van de rollen** van domeinexpert, probleemeigenaar en informatiekundige, zodat iedereen die taken uitvoert waar hij goed in is.



Figuur 2: Een screenshot van de ROC-tool. De domein expert wordt gevraagd om de conceptselectie uit te voeren.

- De focus van het modelleringsproces op de uiteindelijke taak; de proto-ontologie is pas nuttig als hij is afgestemd op de applicatie die hij moet ondersteunen. De scope, gespecificeerd door de probleemeigenaar is essentieel voor de bruikbaarheid van de proto-ontologie.
- Het hergebruik van semi-gestructureerde bronnen verkregen via internet. Dit zijn thesauri, verklarende woordenlijsten, tabellen met informatie etcetera. Deze bronnen worden voorafgaand aan het ROC-proces gescand en korte kennisfragmenten worden eruit geëxtraheerd. De gevonden kennisfragmenten worden opgeslagen als RDF-triples¹ en in het ROC-proces aan de domeinexpert aangeboden als suggesties voor interessante toevoegingen aan het model. Door hergebruik van bestaande kennis wordt een efficiëntieslag gemaakt.

Het ROC-proces wordt ondersteund met de ROC-tool. Bovenstaande spelers voeren m.b.v. deze tool taken in een workflow uit die goed bij hun expertise passen. De workflow is afgebeeld in figuur 1. De eerste stap in het proces is het vastleggen van de scope van de gewenste applicatie. De probleemeigenaar kan een beschrijving van de gewenste applicatie geven, een indicatie van het beoogde domein, het type gebruikers van de applicatie, etcetera. De scope van de applicatie kan gedurende het ontwikkelingsproces van het kennismodel indien nodig worden bijgesteld.

Wanneer de scope bepaald is, kan de domeinexpert aan de slag. Zijn eerste taak is om een lijst te produceren van domeinconcepten die belangrijk zijn voor de taak, de zogenaamde *startconcepten*. De ROC-methode zal vervolgens suggesties doen voor concepten die mogelijk ook relevant zijn voor de scope. Deze suggesties worden geput uit bestaande semi-gestructureerde bronnen. Wanneer alle concepten verzameld zijn, worden de relaties tussen de concepten vastgelegd door de domeinexpert. Het resultaat van deze exercitie is een proto-ontologie – een eerste aanzet voor het kennismodel. Deze proto-ontologie kan worden opgeslagen als SKOS-model² en is in die vorm in veel gevallen al bruikbaar.

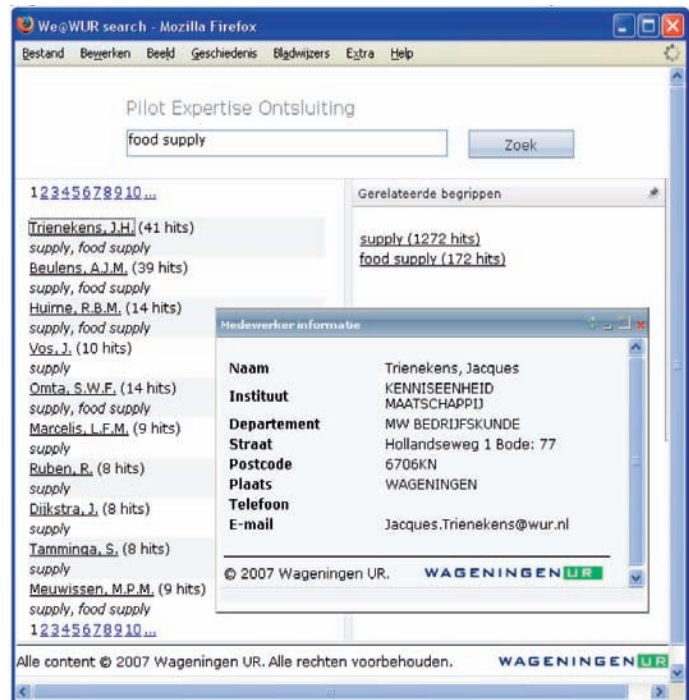
¹ RDF = Resource Description Framework. Zie <http://www.w3.org/RDF/> voor gedetailleerde informatie.
² SKOS = Simple Knowledge Organisation System. Zie <http://www.w3.org/2004/02/skos/> voor meer informatie.

Een ROC-gebaseerde pilot-applicatie

We hebben met ROC een pilot-applicatie ontwikkeld om experts binnen WUR te vinden (taak) op het gebied van *supply chain management* (domein). De applicatie moest geschikt zijn voor *journalisten* (gebruikers). De taak, het domein en de gebruikers leggen samen de scope van de applicatie vast.

We hebben een domeinexpert gevraagd om belangrijke begrippen binnen het logistieke domein aan te geven. De 49 startconcepten die de domeinexpert aandroeg, zijn vervolgens één voor één opgezocht in een database met vooraf verzamelde semi-gestructureerde bronnen (de Agrovoc thesaurus, Wikipedia's pagina over logistiek, etc.). Deze zoekactie leverde een rijtje mogelijk interessante concepten op die door de expert werden beoordeeld (zie figuur 2). Door het opzoeken van goedgekeurde termen steeds te herhalen, is een kennismodel gemaakt met 236 relevante concepten.

Dit model is gebruikt om in de applicatie zoektermen en gerelateerde begrippen via publicaties aan experts binnen



Figuur 3: Een screenshot van de expert finder.

| Proto-ontologie | # interviews | # concepten | Tijd domein expert | Tijd informatiekundige | Totale tijd |
|-----------------|-----------------|-------------|--------------------|------------------------|-------------|
| Planten | 13 (13 experts) | 37 | 58 uur | 88.5 uur | 146.5 uur |
| Ingrediënten | 4 (4 experts) | 120 | 10 uur | 30 uur | 40 uur |
| Supply chains | 2 (1 expert) | 236 | 5 uur | 16.5 uur | 21 uur |
| Geometrie | 12 (1 expert) | 208 | 20 uur | 25 uur | 45 uur |
| Aardappels | 8 (1 expert) | 279 | 8 uur | 32 uur | 40 uur |

Tabel 1: Ontwikkeling van een aantal ontologieën in termen van inhoud en ontwikkeltijd. De supply chain en aardappelontologie zijn gearceerd omdat ze met elkaar vergeleken worden. De geometrie- en de aardappelontologie zijn ontwikkeld met de ROC methode. Het aantal interviews geeft aan hoeveel interviews en met hoeveel experts nodig waren om tot de proto-ontologie te komen. Het aantal concepten geeft aan hoeveel concepten de uiteindelijke proto-ontologie bevatte. De tijd van de domein expert resp. informatiekundige geeft weer hoeveel tijd de proto-ontologie ontwikkeling in totaal heeft gekost voor de domein experts gezamenlijk resp. de informatiekundige. Deze getallen zijn opgeteld weergegeven in de Totale tijd kolom.

Wageningen UR te linken. De gebruiker van de applicatie kan een zoekterm invoeren. Deze term wordt in het kennismodel opgezocht en alle gerelateerde begrippen worden teruggegeven. Aan de begrippen wordt een gewicht toegekend en in de database van de WUR-bibliotheek wordt opgezocht welke medewerker de meeste publicaties heeft met de betreffende begrippen. Deze medewerker eindigt hoog in de expertlijst (zie figuur 3).

Evaluatie ROC

Doordat de domein expert geen inhoudelijke kennisoverdracht aan een informatiekundige hoeft te doen, doordat hij kennis specifiek voor een taak verzameld, en doordat hij op weg geholpen wordt met de automatische associaties, wordt een efficiëntieslag behaald in het maken van de ontologie. Deze efficiëntie voor DE en KE zien we terug in onderstaande tabel. In deze tabel zijn vijf door A&F ontwikkelde kennismodellen weergegeven. De planten, ingrediënten en aardappelontologieën zijn via traditionele interviews ontwikkeld, de geometrie en supply chains ontologieën via ROC. Van deze vijf kennismodellen zijn de supply chains ontologie en de aardappelontologie het best met elkaar te vergelijken qua aantal concepten. Bovendien is voor deze ontologieën steeds één domein expert geïnterviewd. We zien dat voor het opstellen van de supply chain ontologie circa 50% minder tijd nodig was dan voor het opstellen van de aardappelontologie. De tijdswinst zit hem met name in de inzet van de informatiekundige, maar ook de domeinexpert heeft baat bij de ROC-aanpak.

De kwaliteit van de uiteindelijke ontologie kan gemeten wor-

den door te onderzoeken hoe goed de ontologie functioneert in een applicatie. De expertidentificatie-applicatie die ontwikkeld is, blijkt goed te functioneren: bij een tiental willekeurig gekozen zoekopdrachten die als test zijn uitgevoerd, blijken de verwachte experts ook daadwerkelijk bovenaan de lijst gevonden te worden.

Conclusie

De ROC-methode levert een belangrijke bijdrage aan kennisacquisitie. Tot nu toe waren er voor domein experts geen werkbare tools beschikbaar om op snelle en efficiënte wijze een ontologie te ontwikkelen die bruikbaar is in de praktijk. Door de rollen van de domeinexpert en de informatiekundige van elkaar te scheiden, is ROC een methode geworden die specifiek gericht is op mensen die geen achtergrond hebben in kennismanagement. De bijbehorende ROC-tool is gebruikersvriendelijk en helpt de domeinexpert met het snel in kaart brengen van zijn expertisegebied.

Met behulp van de ontologieën die ontwikkeld worden met ROC kunnen de gewenste informatie-intensieve applicaties worden gebouwd. Bestaande kennis kan worden ontsloten en betekenisvol worden gecombineerd. Beslissingsondersteuning, efficiënt zoeken, intelligente modelkoppeling, eBusiness applicaties en ketenbrede kennisuitwisseling zijn slechts enkele voorbeelden van toepassingen die mogelijk worden. Concluderend kunnen we zeggen dat met Semantic Web technologie een actieve impuls gegeven wordt aan innovatie binnen de agrifood sector en dat ROC een nuttige bouwsteen is in dit proces.

