

De kloof tussen zoeken en vinden

Orde scheppen in de informatiechaos is voor elke organisatie van levensbelang. Dat het ontsluiten van informatie geen sinecure is, ondervonden ook het Nederlandse Instituut voor Zorg en Welzijn (NIZW) en de gemeente Waalwijk. Achter de schijnbare eenvoud van zoeken en vinden schuilt vernuftige functionaliteit voor het classificeren en indexeren.

Rob Aaldijk

Het NIZW en de gemeente Waalwijk onderkennen de problematiek en weten inmiddels hoe de bezoeker van hun websites via eenvoudige ingangen – trechtermodellen – naar de gewenste informatie te (ge)leiden. Daartoe maken zij gebruik van thesauri en taxonomieën.

Maar waar heeft men het eigenlijk over als men spreekt over de inzet van taxonomieën in het informatiemanagement? En hoe levert de inzet hiervan een wezenlijke bijdrage aan het oplossen van de problematiek rond toegankelijkheid van informatie?

We raken steeds meer vertrouwd met termen als informatiemanagement, contentmanagement en kennismanagement. Echter, daar waar de massale belangstelling van het bedrijfsleven voor kennismanagement (KM) op zich aanzienlijk heeft ingeboet, lijkt de term contentmanagement inmiddels steeds vaker in één adem te worden genoemd met KM.

Hierbij wordt in de vakpers de laatste tijd met name de term ‘taxonomie’ gebezigd als nieuwste wondermiddel in het bestrijden van *information overload*, een fenomeen dat het behalen van de bedrijfsdoelstellingen bedreigt. In een onderzoeksartikel staat: ‘Taxonomies are *chic*’.

Onderstaand doe ik een poging om de strekking van de hulpmiddelen op dit terrein nader in kaart te brengen en de mogelijkheden en beperkingen op een rij te zetten.

Begripsverwarring

Over de term taxonomie bestaat in de advieswereld de grootst mogelijke verwarring. Deze verwarring, die ook bestaat rond het begrip thesaurus, is enigszins te vergelijken met de – nog steeds niet uitgewoede – discussie over het definiëren van het begrip ‘kennismanagement’.

Ook de meer recent verschenen artikelen over dit onderwerp hebben de doelgroep – de ICT-professionals – niet de gewen-

ste helderheid gebracht. Dat is misschien niet zo heel moeilijk te verklaren: de termen vinden niet hun oorsprong in de computerbranche. De gemiddelde ICT-er, die gewend is aan de definitie van datastructuren en kenmerken voor het indelen van informatie, kent doorgaans niet de implicaties van het opzetten van een thesaurus.

Thesaurus en taxonomie kennen hun toepassing in de werelden van respectievelijk de documentaire informatievoorziening en de biologie. Terwijl toepassing van thesauri en taxonomieën in informatiemanagement van relatief recente datum zijn. Dat geldt, met uitzondering van de professionele uitgeverij en het wetenschappelijk onderzoeksinstituut, in ieder geval voor het merendeel van de grote organisaties. Daarbij kunnen we stellen dat het hier in alle gevallen gaat om de toepassing van informatiemodellen in het kader van informatiemanagement.



Het in Utrecht gevestigde Nederlands Instituut voor Zorg en Welzijn (NIZW) is een onafhankelijke organisatie die zich inzet voor de kwaliteit en vernieuwing van de sector zorg en welzijn. In deze sectoren werken ruim 750.000 beroepskrachten en vele vrijwilligers. NIZW creëerde een uitgebreide thesaurus (zie www.nizw.nl/Thesaurus/index_tt.htm) met meer dan 13.000 trefwoorden die naast termen op het gebied van zorg en welzijn, ook termen ontsluit uit naastgelegen terreinen, zoals bijvoorbeeld veelgebruikte ICT-termen.

De in Smartsite gebouwde en door de thesaurusmedewerkers onderhouden en fors aangepaste applicatie kreeg de naam SmartIRIS.

Volgens één van de negen parttime thesaurusmedewerkers van het NIZW, Richard Thiel, worden er bij elke term verschillende – bijvoorbeeld hiërarchische – relaties aangegeven. Zo kent *vervoermiddelen* als lager gelegen termen *fiets*, *auto*, *tram*, *trein*, enz. Behalve hiërarchisch (verticaal) kunnen termen ook horizontaal gerelateerd zijn: de termen *fietsen* en *fietspaden* bijvoorbeeld, maken geen deel uit van elkaar maar hebben wel met elkaar te maken. Daarnaast onderscheidt de thesaurus

synonieme relaties (rijwiel is synoniem aan fiets) en worden van veel termen korte begripsomschrijvingen – definities – gegeven. Thesaurusmedewerker Thiel is overtuigd van het nut: 'Zonder deze thesaurus zouden veel zaken en begrippen gewoon onvindbaar zijn. En vergelijk het eens met zoekmachine Google, als je daar *fiets* intikt krijg je wel enorme aantallen 'hits' waar het woord *fiets* in voorkomt, maar niet de hits met het woord *rijwiel*'. Of zoeken op bijvoorbeeld het woord *longziekten*, dat levert – ook zonder kennis van allerlei aan longziekten gerelateerde termen – talloze hits op. 'Met onze thesaurus worden dit soort synonieme begrippen wel meegenomen', zonder tussenkomst van andere personen.

Intern gebruikt het NIZW de thesaurus om aan de in databases opgeslagen begrippen trefwoorden toe te kennen die ook gebruikt worden voor NIZW-producten, zoals boeken, cd-rom's en aan derden geleverde databases. Dit biedt ook afnemers de mogelijkheid om er zelf trefwoorden aan toe te kennen en er bijvoorbeeld informatie uit hun eigen bibliotheek mee te ontsluiten.

Thesauri

Voor thesauri bestaat een gestandaardiseerde definitie: ISO standaard 2788. Volgens deze standaard is een thesaurus: "[...] a controlled vocabulary arranged in a known order and structured so that equivalence, homographic, hierarchical, and associative relationships among terms are displayed clearly and identified by standardized relationship indicators that are employed reciprocally". Een thesaurus is dus niet zomaar een lijst met termen, maar laat ruimte voor het leggen van verbanden tussen termen en hun betekenis vanuit meerdere invalshoeken.

Taxonomie

De definitie van een taxonomie is minder eenduidig vastgelegd. In de biologie wordt eronder verstaan de organisatie van een afgebakende set informatie voor een bepaald doel, doorgaans het vastleggen van de unieke locatie van een bepaalde soort in een complexe maar strikt toegepaste (afstammings) hiërarchie. In afwijking hiervan kunnen bij de toepassing van taxonomieën op andere kennisgebieden bepaalde items in meerdere taxonomische categorieën tegelijkertijd voorkomen. Juist deze laatste interpretatie wordt tegenwoordig steeds vaker toegepast. Als meest bekende voorbeeld van een breed toegankelijke en gebruikte taxonomie wordt aangehaald de structuur van de internetzoek- en vindsite Yahoo. Overigens, om de verwarring compleet te maken: de Yahoo website wordt ook wel omschreven als directory. Deze term is afkomstig uit de wereld van de openbare informatiegidsen (telefoonboek, Gouden Gids) en wordt – voor het gemak – gelijkgesteld aan de zojuist gegeven brede definitie van taxonomie. Contact- en andere gegevens van organisaties en personen worden hierbij in 'families' van producten en diensten ondergebracht.

Topic map

Als jongste telg in deze familie van informatiehulpmiddelen mogen we de topic map zeker niet vergeten. Een topic map is een technologie waarmee – op basis van een nieuwe ISO standaard voor de beschrijving van kennisstructuren waarbij kenniselementen worden gekoppeld aan informatiebronnen – de relatie tussen kenniselementen dynamisch maar tegelijkertijd gestandaardiseerd kan worden vastgelegd. Een topic map kan worden gezien als een functioneel huwelijk tussen thesaurus en taxonomie, met een (nog) hoger informatiegehalte. Topic maps worden doorgaans uitgedrukt in de taal XML (dit dialect heet XTM).

Informatie structureren

Welk doel dient de inspanning om tot een thesaurus of taxonomie te komen in de strijd tegen *information overload*? Om de informatie, die eenmaal is geproduceerd of verwerkt binnen de organisatie, te structureren moet deze – zoals wenselijk vanuit het oogpunt van kennismanagement – zodanig gestructureerd zijn dat de informatie ook teruggevonden kan worden. Dit met het doel om de informatie ook in nieuwe situaties te kunnen toepassen. In een kennisorganisatie is het vermogen om dit te doen zelfs cruciaal.

Nog vaker is sprake van de noodzaak voor de directe toepassing van actuele informatie in operationele situaties, bijvoorbeeld bij de behandeling van een verzekeringskwestie. In dat geval wordt de kennis die nodig is om de kwestie af te kunnen handelen, door een medewerker verzameld in een dossier. Zo'n verzekeringsdossier is opgebouwd rondom de 'case' ofwel de situatie van de klant.

Toch is het niet altijd eenvoudig om hierbij alle informatie te vinden die nodig is voor een snelle en efficiënte behandeling van

Zoals veel andere gemeenten heeft ook Waalwijk een eigen site op internet.

Het contentmanagement systeem (CMS) dat de gemeente Waalwijk gebruikt is Smartsite. Aan dit systeem is de module GPDC (Gemeentelijke Producten en Diensten Catalogus) van Seneca toegevoegd. Hiermee kunnen gemeenten hun diensten op gestandaardiseerde wijze aanbieden (daar is Vindcatalogus er één van). Op de site kan de burger via drie ingangen – trefwoorden, thema's en levensgebeurtenissen – zoeken naar informatie. De site onderscheidt een zevental thema's: bouwen & wonen, leefomgeving & veiligheid, milieu & afval, sport-recreatie-cultuur & kunst, verkeer & vervoer, werk & inkomen, en zorg & welzijn. Wie voor een van deze thema's kiest wordt via een soort trechtermodel naar de achterliggende producten en diensten geleid. Klikken op *verbouwvergunning* bijvoorbeeld, ontsluit informatie over de kosten van een aanvraag, de voorwaarden waaraan een *gevelaanzicht*

moet voldoen, of hoe daartoe een *dwarsdoorsnede* aan te leveren. Via een levensgebeurtenis als *trouwen* of het trefwoord *verbouwen* is het ook mogelijk om – via andere trechters – uit te komen bij de informatie over (ver)bouwvergunningen.

De GPDC-thesaurus wordt door de betreffende afdelingen zelf up-to-date gehouden: minstens één keer per jaar worden alle producten en diensten kritisch tegen het licht gehouden. Immers, wet- en regelgeving, tarieven, enz. veranderen constant. De verantwoordelijkheid voor de inhoudelijke informatie ligt bij de betreffende (gemeentelijke) afdelingen, de webmaster vervult daarbij een regiefunctie. Volgens gemeentevoorlichter Arnout Geus is de gemeente Waalwijk momenteel druk doende met een restyling van haar website. Bezien wordt of de huidige GPDC nog voldoet, of dat op een andere indeling moet worden overgestapt.

de case. Hier geeft de werkervaring van de betrokken verzekeringsmedewerker of van diens directe chef veelal de doorslag.

In beide situaties is het zaak om ongestructureerde informatie op enigerlei wijze te structureren, dat wil zeggen in een bepaalde, voor het praktisch gebruik relevante context te plaatsen. Hier ligt ook de relatie met kennismanagement: door het plaatsen in een praktische context kan informatie als kennis worden toegepast. Classificatie en indexering zijn de twee processen die structuur aanbrengen in ongestructureerde informatie.

Classificeren en indexeren

Classificeren is het structureren van bestaande informatie volgens een conceptueel model. Elke computergebruiker classificeert al: bijvoorbeeld door informatie in bestandsvorm op te slaan in een structuur van mappen. Deze structuur wordt niet voor niets een directory genoemd. We geven namen aan deze mappen die passen bij de begrippen uit onze belevingswereld. De structuur voor deze mappen is (de representatie van) ons persoonlijk informatiemodel.

Zoals uit het genoemde verzekeringsvoorbeeld blijkt, is het nodig dezelfde informatie vanuit verschillende invalshoeken te classificeren, zodanig dat er later gemakkelijker dwarsverbanden te maken zijn. Niet zelden hebben we – ook in ons eigen computergebruik – de behoefte om die informatie op een andere manier te structureren. Dat is lastig, aangezien het opslaan van een kopie van het bestand in een andere map, het beheer van dat bestand niet ten goede komt. We moeten dan terugvallen op het toekennen van extra kenmerken aan deze bestanden.

Moeten we die informatie ook nog delen, dan is het van even groot belang dat de toegepaste classificatie aansluit bij de beleving van de verschillende doelgroepen.

Het is cruciaal voor een goede classificatie, want voordat ermee kan worden begonnen moeten de potentiële gebruikers overeenstemming hebben bereikt over het toe te passen

conceptuele model. Verder behoeft het geen betoog dat deze inspanningen (tot consensus) vaak evenredig groot zijn als het aantal en de omvang van de gebruikersgroepen die daarover met elkaar overeenstemming moeten bereiken. En wordt het model daarmee erg groot, dan bestaat het gevaar dat het informatiemodel al bij oplevering geen actualiteitswaarde meer heeft.

Stelt u zich in het ergste geval eens voor dat u opdracht krijgt om alle woorden uit de Dikke van Dale aan elkaar te relateren in 'de ultieme' thesaurus voor het Nederlands taalgebied. Dan nog zou u daarmee het gros van de in Nederlandse organisaties gebruikte vaktermen missen!

Classificeren wordt dan toegepast bij het aanmaken, bewerken of produceren van informatie, indexeren lijkt hierop. Echter, indexeren wordt toegepast op een verzameling reeds bestaande – gestructureerde dan wel ongestructureerde – informatie. Indexeren betreft het achteraf reconstrueren van een conceptueel model, met of zonder relaties. Met als resultaat een index, die kan fungeren als hulpmiddel bij het vinden van informatie.

De kloof overbruggen

In hoeverre ondersteunen thesauri, taxonomieën en andere informatiemodellen het classificeren en indexeren?

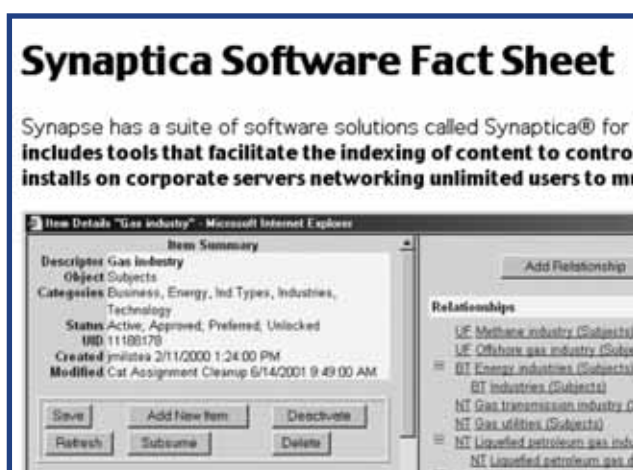
Toepassing van een thesaurus of taxonomie bij classificatie heeft als voordeel dat de wijze van opslag van informatie niet beperkend is voor gebruik door een grotere doelgroep.

Zolang er maar:

- één of meerdere classificatietermen als kenmerk worden meegegeven;
- er een hulpmiddel beschikbaar is waarmee uit eenmaal geclassificeerde informatie geput kan worden;
- en daar vervolgens een index mee kan worden geproduceerd.

Bekende classificatie- en indexeringstools zijn onder meer de producten van Autonomy en Verity.

Ook zijn er voorbeelden van toepassing van topic maps, waarbij informatie die met een term geassocieerd is in een logische structuur, via diezelfde structuur kan worden teruggevonden. Zo wordt als het ware een soort routekaart gecreëerd van met elkaar verbonden begrippen. Deze informatie moet dan wel eerst weer zijn geïndexeerd. Hiertoe worden gespecialiseerde producten aangeboden (zie bijvoorbeeld www.ontopia.net).



Zoals eerder gesteld zit de inspanning hem juist in het komen tot en het actueel houden van het informatiemodel zelf. Is er eenmaal een taxonomie of thesaurus, dan moet er – op basis van mogelijk nieuwe informatie – op worden toegezien dat de informatie compleet blijft en dus actueel. Verder is het ook van belang te constateren dat juist indexering hierbij van dienst kan zijn. Door de opgebouwde index te vergelijken met de bestaande thesaurus, kan worden vastgesteld of het conceptuele model na verloop van tijd moet worden aangepast, bijvoorbeeld aan de hand van een lijst van nieuw gevonden – nog niet in de thesaurus voorkomende – termen.

Conclusie

Een taxonomie of ander gestandaardiseerd informatiemodel is een in potentie krachtig hulpmiddel bij het classificeren en indexeren van informatie. Echter een strategie die berust op het in korte tijd bedrijfsbreed inzetten van een allesomvattend model, is gedoemd te mislukken.

Daar waar het ondoenlijk blijkt om via een centraal initiatief tot een breed toepasbare thesaurus of taxonomie te komen, is delegeren soms een haalbare optie. Spraakmakend voorbeeld is het Open Directory Project waarbij specialisten op een bepaald kennisniveau zich kunnen aanmelden als beheerder van het betreffende onderdeel in de complete taxonomie. Het resultaat hiervan kan worden gezien als het equivalent van de Yahoo-taxonomie, maar dan met een meer democratisch beheerproces als grondslag – iedereen kan zich in principe aanmelden als (onbetaald) redacteur.

En zoals bij alle complexe trajecten met een behoorlijke impact op de organisatie, is het vaak raadzaam om wel gewoon te beginnen maar bij aanvang de verwachtingen niet al te hoog te stellen. Met duidelijke doelstellingen over een niet al te lange tijdsspanne, is het komen tot een meer gestructureerde vorm van informatiemanagement zeker een haalbare kaart. De technologie is daarbij een hulpmiddel, maar ook hier schuilt het werkelijke succes toch in mensenwerk. **IK**

drs. R.G.A. Aaldijk – rgaaldijk@seneca.nl – adviseur Seneca Information Consultancy en docent contentmanagement aan de Erasmus Universiteit Rotterdam.

Referenties

- Edols, L., *Taxonomies are what?*, free print issue #041001, 2003
- Roberts-Witt, S.L., *Practical taxonomies: hard-won wisdom from creating a workable knowledge classification system*, University of Torun, Poland, 2002
- ANSI/NISO, *Guidelines for the construction, format, and management of monolingual thesauri* (ISO standard 2788), Bethesda 1994, ISBN 1-880124-04-1
- Di Maio, P., *Online taxonomies: the next shift for information science*, 2001 (www.contentwire.com/Taxonomies/Index.cfm?ccs=82&cs=8)
- Oxbrow, N., and Abell, A, *What next? Life after Knowledge Management?*, 2002 (www.tfpl.com/assets/applets/What_next._Information_Outlook_180202.pdf)