

## Vermijd complexe ECM-systemen door goede toegangsstructuren en ontsluitingsmiddelen

# Toegang tot het ECM-systeem

*Eén van de grootste angsten van managers is dat hun net geïmplementeerde enterprise contentmanagementsysteem veel te complex blijkt te zijn. Met name de gebrekkige manier waarop content-items zijn terug te vinden en het beheer ervan, bezorgen gebruikers én management hoofdbrekens en verlagen de slaagingskans van de invoering van een systeem. Zaak is dus een goede toegangsstructuur te vinden op basis van de juiste ontsluitingsmiddelen. Leo Meerman geeft in dit artikel aan hoe dat kan. Stel de juiste vragen, zo betoogt hij, en u bent al een stuk verder in het maken van de juiste keuze.*

De complexiteit van een ECM-implementatie heeft een functionele en een technische component. In dit artikel behandelen we de functionele component. Voor wat betreft deze functionele component is de complexiteit van een ECM-implementatie voor een belangrijk deel op te lossen door een goede structuur te ontwerpen voor het toegankelijk maken van de content-items: de 'content access'. Het ontwerpen van goede toegangsstructuren met bijbehorende ontsluitingsmiddelen vereist speciale vaardigheden in het analyseren en synthetiseren (clusteren) van de content-items. Conceptueel denken en de concepten in hun context plaatsen (klassenstructuur) vergen domeinkennis en inventiviteit. Zonder op z'n minst enige domeinkennis is het daarom moeilijk om goede structuren te ontwerpen en te beheren. In eerdere artikelen in Business Process Magazine is aandacht besteed

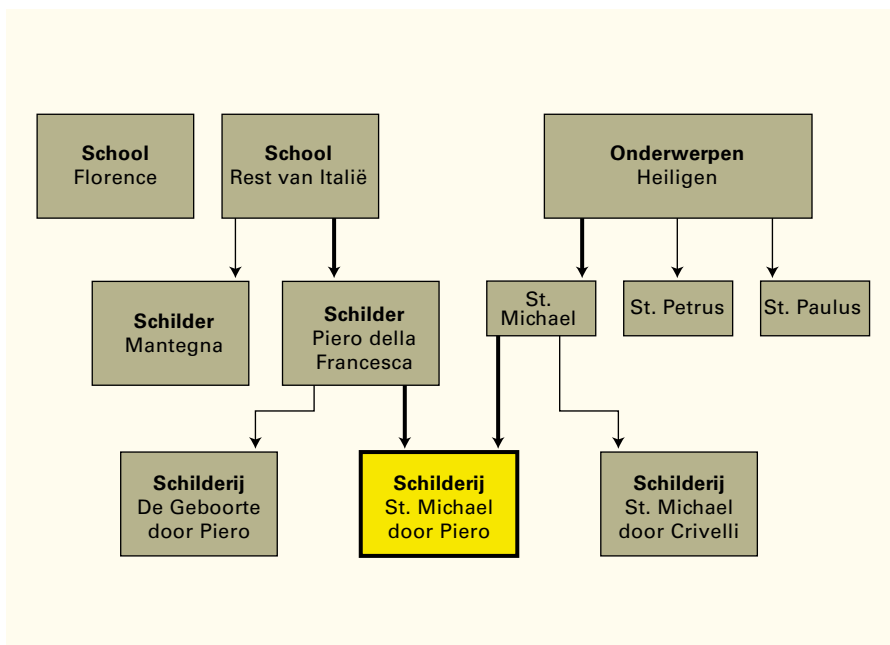
aan de ontwikkelingen rondom information retrieval met beschrijvingen van de ontsluitingsmiddelen voor content, zoals het gebruik van trefwoordenlijsten, thesauri, classificatieschema's en taxonomieën<sup>1</sup> en het belang van met name taxonomieën als middel om de content-items verantwoord op te slaan en terugvindbaar te maken<sup>2</sup>. Het doel van deze taxonomieën is de gebruiker zo efficiënt en effectief mogelijk naar de gewenste, gezochte of benodigde content te leiden. De taxonomie heeft hierbij de functie van zoekleiding. Er is een globaal stappenplan beschreven om te komen tot een opzet voor een taxonomie.

In deze bijdrage wordt dieper ingegaan op de overwegingen die een rol spelen bij het kiezen van zo bruikbaar mogelijke structuren om toegang te krijgen tot de content (content access), voor niet alleen gebruikers, maar zeker ook voor de beheerders van content

in een ECM-systeem. Access heeft hierbij de betekenis van de mogelijkheid om kennis uit opgeslagen informatie te verwerven<sup>3</sup>. Er is een nauwe relatie tussen de hier bedoelde structuren en de in de vorige alinea genoemde ontsluitingsmiddelen. Aan bod komen de eisen die we moeten stellen aan de toegangsstructuren, de soorten die er zijn en wat de samenhang is met ontsluitingsmiddelen. Dit resulteert in een tweetal voorbeelden die als illustratie dienen voor de strategie waarmee verantwoorde keuzes voor de bestpassende toegangsstructuur annex ontsluitingsmiddel gemaakt kunnen worden.

### Eisen aan toegangsstructuren

De toegangsstructuren die we willen toepassen in ECM-systemen moeten in ieder geval aan een viertal eisen voldoen. Ze moeten allereerst in staat zijn om de grenzen van het contentdomein aan te geven. Bij iedere toegepaste structuur moet het mogelijk zijn om te zien over welk onderwerp informatie te vinden is. Met andere woorden: de 'kopjes' moeten de onderliggende content duidelijk weer geven. Ze moeten ook relaties aan kunnen geven tussen de content-items onderling. Een content-item kan over meerdere onderwerpen of concepten handelen en elk van de concepten zal niet zelden in andere content-items voorkomen. Ten derde moeten de toegangsstructuren de zogenaamde zoekleiding kunnen bieden. Zoekleiding is in deze context hulp die een gebruiker krijgt om te navigeren door het contentdomein. Die hulp kan be-



Afbeelding 1. Voorbeeld van een polyhiërarchische relatie, ontleend aan de museumwereld.

staan uit een trefwoordenlijst, een inhoudsopgave (bijvoorbeeld een site-map), een taxonomie en dergelijke. Als laatste moeten toegangsstructuren de mogelijkheid bieden om de content-items te ontsluiten. Wezenlijk is de mogelijkheid om metadata toe te kennen aan elk individueel content-item. Maar ook het toevoegen van een samenvatting aan een content-item en het definiëren van verwijzingen naar andere items is wenselijk, zo niet noodzakelijk<sup>4</sup>.

De belangrijkste toegangsstructuren hebben in het algemeen één van de volgende vier vormen. Er kan sprake zijn van hiërarchieën, kruisverwijzingen, indexen en/of volgordelijke structuren.

### Hiërarchieën

De hiërarchiestructuren zijn te verdelen in generieke en partiële vormen. Bij de generieke structuur is er sprake van een 'generiek-specifiek'-verhouding tussen de niveaus in de structuur (genus-species). In algemene bewoordingen kunnen we hiervoor ook de begrippen klassen en subklassen gebruiken. Een ICT-voorbeeld is:

- > generiek: software;
  - > specifiek: financiële software, logistieke software, retrieval software.
- De partiële structuur kunnen we beschrijven in de vorm van het relatiebegrip `Is_Part_Of`. Partiële hiërarchische relaties hebben betrekking op delen van een groter geheel. Stuklijsten van producten en machines zijn hiervan herkenbare voorbeelden.

Een complicerende hiërarchische relatie is de polyhiërarchische relatie. Bij een polyhiërarchische relatie heeft een klasse twee bovenschikkende ouders. In schema ziet dit fenomeen eruit als aangegeven in afbeelding 1, een voorbeeld uit de museumwereld. Bij dit type relatie is de zuiverheid van de hiërarchie in het geding. Gelukkig bieden de huidige softwaretechnieken een oplossing voor deze onvolkomenheid. De standaardoplossing voor dit probleem is te kiezen voor plaatsing op één positie in de zuivere hiërarchie en de andere relatie aan te brengen met behulp van een hyperlink. De keuze van de (generieke) hiërarchie (er zijn er immers bijna altijd meerdere) waarin de term wordt opgenomen, is een kwestie

van de inschatting via welke hiërarchie de gebruikers het meest zoeken<sup>5</sup>.

Het belangrijkste criterium om de hiërarchische structuren zuiver te houden kan worden verwoord met het begrip 'omvattend knooppunt'<sup>6</sup>. Zo'n knooppunt omvat precies alle ondergeschikte termen die op hun beurt elkaar uitsluiten. Houden we ons aan dit criterium dan zal op elk niveau van de hiërarchie er steeds slechts één goede keuze voor de gebruiker zijn. De andere zijn immers uitgesloten! Dit uitsluitingsprincipe geldt op alle niveaus van de hiërarchie. Een prima test om na te gaan of aan dit principe is voldaan is het gebruiken van de 'alle/sommige'-toets: alle koeien zijn runderen, sommige runderen zijn koeien. Als we ons bij elke 'opsplitsing' naar subklassen deze vraag stellen zal de hiërarchie zuiver blijven. Hiërarchische structuren zijn te realiseren met classificatieschema's, taxonomieën en thesauri.

### Kruisverwijzingen

Kruisverwijzingen worden toegepast als we constateren dat er tussen content-items gelijkwaardigheidsrelaties of associatieve relaties bestaan. Bij gelijkwaardigheidsrelaties gaat het over termen (woorden of woordgroepen) met gelijke betekenis, synoniemen dus. Het vaststellen van een relatie die tussen termen bestaat is van belang, omdat in een latere fase van het ontwerp, één van de begrippen als voorkeursterm aangemerkt moet gaan worden. De andere term(en) worden dan in een synoniemenlijst opgenomen en met 'zie:'- en 'zie ook:'-verwijzingen aan elkaar gerelateerd. Een associatieve relatie is een relatie tussen content-items die inhoudelijke verwantschap hebben. Het probleem doet zich voor dat ook hiërarchische en gelijkwaardigheidsrelaties natuurlijke vormen van verwantschap hebben. In de praktijk wordt

dan ook bij voorkeur eerst nagegaan of er sprake is van hiërarchische of gelijkwaardigheidsrelaties. Is dat niet het geval dan wordt een eventuele associatieve relatie benoemd.

Echte associatieve relaties zijn (ontleend aan Woordsystemen?):

- Seismologie en aardbevingen (wetenschapsgebied en object van studie);
- Onderwijs en leraren (beroep en beoefenaar);
- Uitgeven en boeken (proces en resultaat);
- Erosie en woestijnvorming (causaliteit);
- Hout en planken (materiaal en product).

Ook een echte associatieve relatie bestaat er tussen oude en nieuwe termen voor een bepaald onderwerp, inclusief afkortingen. Naamswijzigingen van organisaties zijn het meest typerende voorbeeld hiervan. Uit de hiervoor gegeven voorbeelden blijkt dat associatieve relaties hiërarchie overstijgend zijn en daarom zullen ze met hyperlinks moeten worden geconstrueerd (vergelijk de polyhiërarchische relatie).

## Indexen

Indexen worden in principe opgebouwd uit de betekenisvolle woorden of termen die in content-items voorkomen. Een index is een - meestal op alfabet - geordende lijst van termen, die verwijzen naar relevante content-items. Behalve de index zoals we die kennen uit een boek, zijn ook trefwoordenlijsten verschijningsvormen van indexen. De relaties die in een trefwoordenlijst worden gedefinieerd, zijn beperkt tot de 'zie:'- en 'zie ook:'-verwijzingen. Behalve op trefwoord worden indexen ook gesorteerd op auteursnaam, datum, en andere door de organisatie zelf te kiezen metadata. Belangrijk in de ontwerpfase van het ECM-systeem is daarom te inventariseren op welke

kenmerken (metadata) straks moet worden gezocht. Voor elk kenmerk is immers een index nodig.

Uitgebreidere semantische relaties tussen termen zijn vast te leggen in thesauri. Hiervoor worden, behalve 'zie:'- en 'zie ook:'-verwijzingen, begrippen als broader term (BT) en narrower term (NT) gebruikt om boven- en onderliggende semantische relaties aan te geven. Thesauri worden onderscheiden in macro- en microthesauri. De macrothesaurus is de meest gangbare en voor het merendeel van de organisaties in een bepaald kennisdomein goed bruikbaar. Voor organisaties binnen het 'micro'-domein, zoals zeer gespecialiseerde onderzoeksinstellingen, kan het nodig blijken om in veel meer detail de terminologie te beschrijven. Uitgaande van de macrothesaurus uit het kennisdomein wordt dan een microthesaurus gemaakt.

## Volgordelijke structuur

Dit zijn de structuren van content-items die we standaard in gedrukte media zoals boeken en tijdschriftartikelen zien. De volgorde van de hoofdstukken en de paragrafen ligt vast en daarmee zijn de content-items in deze vorm niet anders te benaderen dan sequentieel. Bij omvangrijkere werken is dit een groot nadeel. De aanwezigheid van inhoudsopgave en

auteurs- en trefwoordenindexen zijn dan ook van wezenlijk belang voor het ontsluiten van deze contentstructuren. Overigens is de volgordelijke structuur een belangrijke eis bij het formuleren van een goed betoog of het schrijven van een essay en is daardoor als verschijningsvorm niet weg te denken.

## Ontsluitingsmiddelen

In de vierdeling van toegangsstructuren duiken regelmatig de verscheidene ontsluitingsmiddelen op waarmee de structuren zijn te realiseren. De samenhang tussen beide is in tabel 1 weergegeven. De geschikt/ongeschikt-aanduiding is slechts indicatief bedoeld. Er zijn meer criteria om de keuze te maken voor welke toegangsstructuur en daarmee voor welk ontsluitingsmiddel gekozen moet gaan worden. Hierna wordt dit door middel van een voorbeeld verder uitgewerkt.

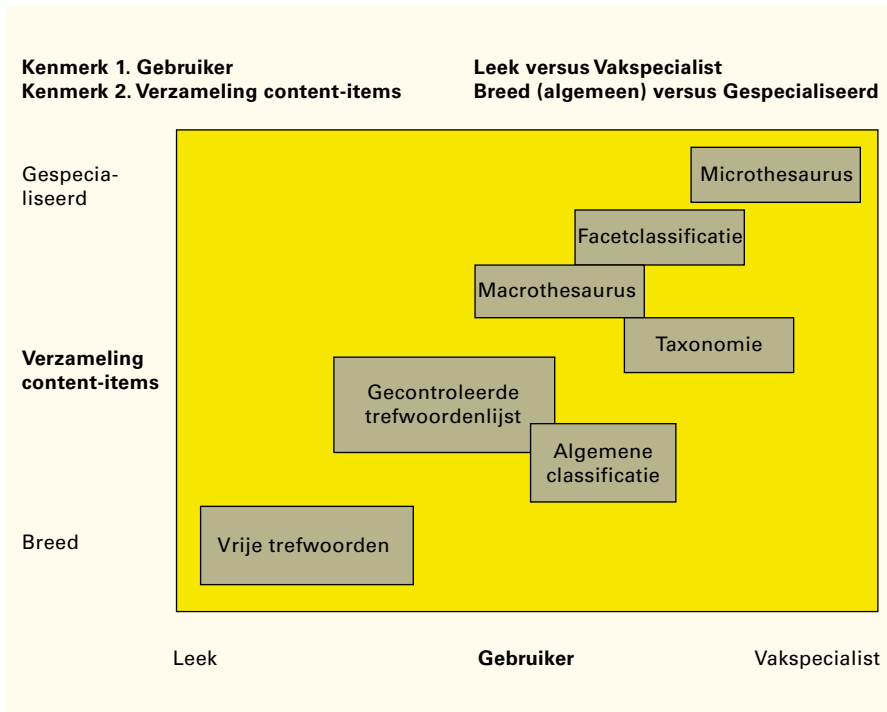
## Kiezen geschikte toegangsstructuur

De complexiteit van ECM-systemen wordt voor een belangrijk deel bepaald door de grote variatie in keuzemogelijkheden met betrekking tot de genoemde toegangsstructuren. Dé oplossing is niet te geven. Wel is het mogelijk om aan de hand van kenmerken van componenten van het ECM-systeem oplossingsrichtingen aan te geven. We gaan hierbij uit van

Toegangsstructuren	Hiërarchie	Kruisverwijzing	Index	Volgordelijk
Ontsluitingsmiddelen				
Trefwoordenlijst	-	-	+	-
Thesaurus	-	+	+	-
Classificatie	+	-	-	-
Taxonomie	+	+	+	-

+ geschikt  
 - wel mogelijk, niet leidend  
 - ongeschikt, niet relevant

Tabel 1. Samenhang tussen toegangsstructuren en ontsluitingsmiddelen.



Afbeelding 2. Samenhang tussen gebruikerstype en inhoud van content-items.

twee van de belangrijkste componenten: de gebruikers en de inhoud van de content-items.

Kenmerken van de gebruiker:

- > Leek versus vakspecialist;
- > Gebruiker zoekt zelf versus literatuuronderzoeker/informatiespecialist als intermediair;
- > 'Information literate' versus 'information //literate'.

De inhoud van content-items is te karakteriseren als:

- > Breed versus gespecialiseerd;
- > Platte teksten versus blobs (Binary Large Objects, zoals video en audio-opnamen);
- > Eén taal versus meertalig;
- > Homogene verzameling versus heterogene verzameling (alle platte teksten tot en met blobs in één multimediale database).

Welke toegangsstructuren moeten we nu kiezen als we beide eerste kenmerken van de gebruikers en de inhoud van de content aan elkaar willen relateren? We zetten daartoe beide

uitersten af op de assen van een rechtehoek en positioneren de geschikte ontsluitingsmiddelen in de ontstane ruimte (zie afbeelding 2). Is het doel van het ECM-onderdeel om vakspecialisten gespecialiseerde content-items te leveren, dan moeten we een microthesaurus aanbieden, een facetclassificatie of een taxonomie. Daaruit volgt direct dat we (zie tabel 1) ons moeten concentreren op een hiërarchische structuur, aangevuld met kruisverwijzingen. Is het doel om voor de gemiddelde gebruikers content-items aan te bieden met een gemiddelde diepgang, dan komen een trefwoordenlijst of een algemene classificatie in aanmerking. We zouden hier mogelijk kunnen volstaan met een eenvoudige hiërarchische structuur, aangevuld met indexen. Deze zijn vanzelfsprekend sneller te realiseren dan de vereiste structuren uit het eerste voorbeeld.

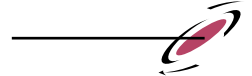
De werkvolgorde bij dit deel van de functionele analyse voor een ECM-systeem is:

1. kiezen welke componenten van het ECM-systeem als uitgangspunt dienen;

2. welke ontsluitingsmiddelen daarvoor geschikt zijn;
3. welke toegangsstructuren nodig zijn voor de realisatie.

### Haalbaarheid

Een goede toegangsstructuur op basis van de juiste ontsluitingsmiddelen is van wezenlijk belang voor het slagen van een ECM-implementatie, maar de realisatie ervan is lastig en gecompliceerd. De kosten zijn behoorlijk. Er is veel denkwerk nodig, er is veel tijd mee gemoeid en dus geld. Bovendien is het onderhoud van groot belang. Bij het niet onderhouden zal het nut snel en drastisch afnemen. De functies van taxonomist of van de recentelijk ontstane functie van ontologist zullen in belang toenemen. Management zal ook steeds meer een business-case verlangen, waarin ze de haalbaarheid aange-  
toond wil zien!



### Literatuur

1. L. Meerman, Ontwikkelingen in information retrieval, Business Process Magazine 4 2002, pag. 20-23.
2. L. Meerman, Taxonomieën cruciaal in enterprise content management, Business Process Magazine 4 2003, pag. 6-11.
3. Design criteria standard for electronic records management software applications, DoD 5015.2-STD.
4. Handboek Content Management, Bob Boiko, Den Haag, Academic Service, 2002.
5. Three steps to a corporate taxonomy, Content Management 2003 June, pag. 16-19.
6. M. Stefik, Introduction to Knowledge Systems, Morgan, 2000.
7. Woordsystemen: theorie en praktijk van thesauri en trefwoordensystemen / H. Magrijn, S.A.Th.M. Pontzen, G.J.A. Riesthuis [et al], Den Haag, Biblion, 2000.

### Leo Meerman

Leo Meerman is directeur van Celt Consultancy, een onafhankelijk adviesbureau op het gebied van kennis-, enterprise content- en documentmanagement (lmeerman@celt.nl).