

Contentintegratie geeft overzicht en samenhang

Veel informatiespecialisten overwegen om contentintegratie binnen hun organisatie te introduceren. Speciaal voor medewerkers uit de advocatuur organiseerde Stichting Het Overleg¹ daarom samen met GO opleidingen een studiedag over dit onderwerp. Martijn van der Kaaij en Stef Verf gaan in dit artikel in op de algemene theoretische achtergronden van contentintegratie en de specifieke mogelijkheden voor advocaten.

Martijn van der Kaaij en Stef Verf

Het valt niet te ontkennen dat content-integratie in de mode is. Maar wat houdt het begrip precies in? Een eerste constatering is dat het hier om een ‘gekaapte’ term gaat, want aanvankelijk ging contentintegratie over marketing op websites (zie kader). Dat is dus níet de betekenis waarin het begrip tegenwoordig voornamelijk wordt gebruikt.

Hoewel er voor de huidige betekenis geen algemeen geaccepteerde definitie beschikbaar is, worden de volgende elementen vaak genoemd:

- > contentintegratie betreft het op inter- of intranet toegankelijk maken van informatie uit meerdere bronnen via één gebruikersinterface;
- > contentintegratie moet aansluiten op

het (vermeende) gedrag van de gebruiker en moet daarom altijd worden uitgevoerd in de context van een duidelijk omschreven doelgroep;

- > omgevingen met contentintegratie moeten de gebruiker de mogelijkheid bieden om de omgeving te personaliseren.

Waarom zou een organisatie in content-integratie moeten investeren? Efficiency is het meest gegeven antwoord op deze vraag. Medewerkers hoeven geen losse websites en databases te raadplegen maar hebben via één interface toegang tot informatie die voor hen is geselecteerd en op kwaliteit is gecontroleerd. Dit scheelt tijd en dus ook geld.

Een tweede argument is meer intern van aard: het centraal aanbieden van geïntegreerde content garandeert een minimum aan informatie voor alle medewerkers: wat wordt aangeboden, mag bekend worden verondersteld.

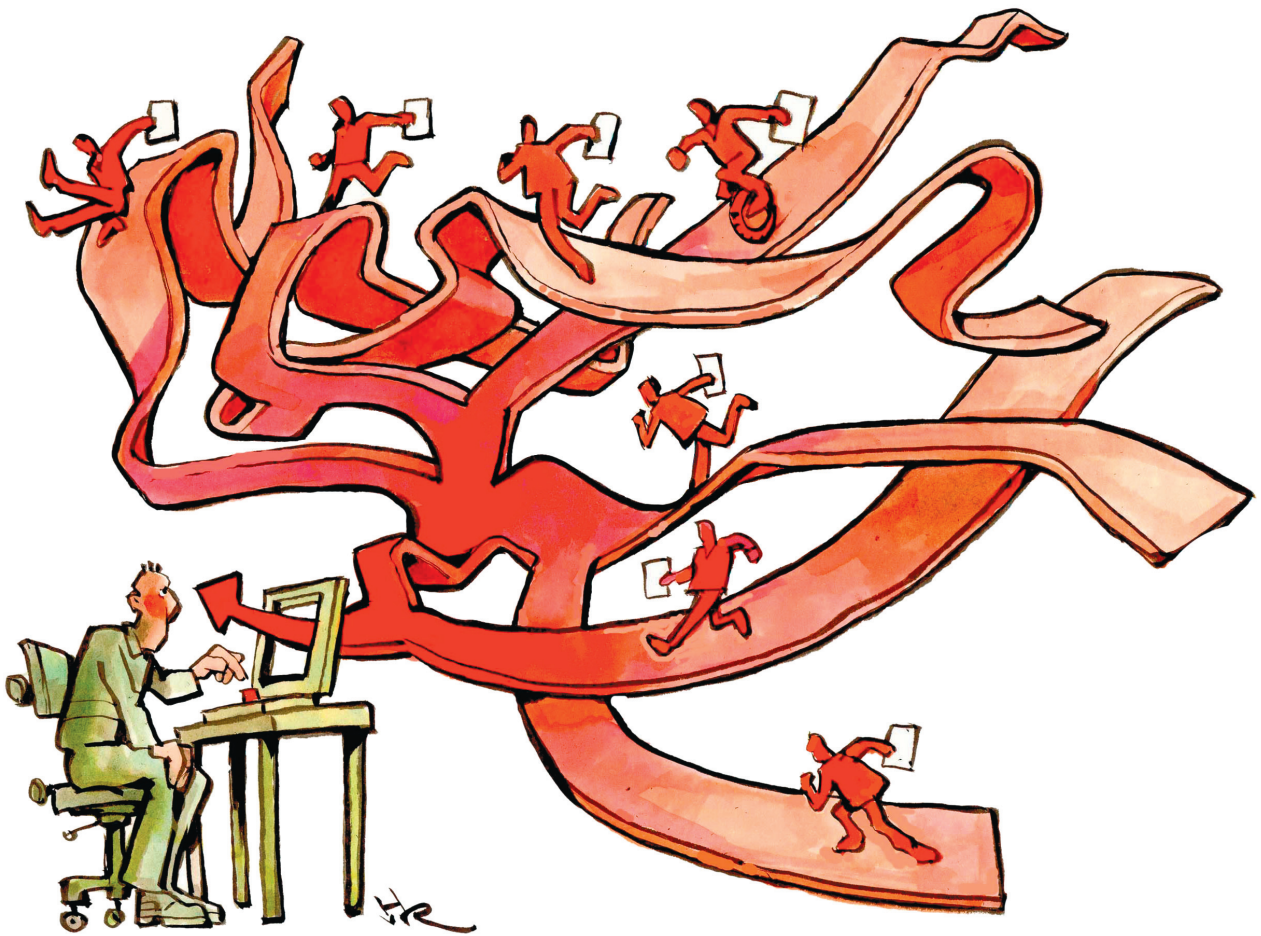
Tot slot kan goede contentintegratie ook een middel zijn om medewerkers te binden, aangezien het dagelijkse werk er door veraangenaamd wordt.

Varianten

Er zijn in principe twee manieren om content te integreren: metasearch en indexering.

Oorspronkelijke definitie van contentintegratie

‘De mogelijkheid voor een adverteerder om zijn product te promoten op een pagina met gemengde inhoud. Voorbeeld: redactionele pagina over autorijden in de winter, aangevuld met betaalde (advertentie)tips van een bandenfabrikant en de mogelijkheid om tevens online een racespelletje te spelen.’²



Illustratie: Henk van Ruitenbeek

Metasearch

Bij de metasearchmethode wordt een zoekvraag doorgegeven aan de eigen zoekinterfaces van onderliggende bronnen. Waar mogelijk wordt ervoor gezorgd dat bij elke bron in het juiste veld wordt gekeken. Immers: wat bij bron 1 'trefwoord' heet, heet bij bron 2 nu eenmaal 'subject' of is wellicht zelfs opgesplitst in verschillende velden.

De metasearch heeft een belangrijke beperking: de zoektermen die de gebruiker opgeeft worden bij een metasearch in principe direct doorgegeven. Dit betekent dat men, eventueel binnen een beperkt aantal velden, in feite een pure textsearch doet. Als de gebruiker om het trefwoord 'fiets' vraagt, worden documenten met het trefwoord 'rijwiel' niet gevonden.

De resultaten van een metasearch kunnen ook niet eenduidig geïntegreerd en naar context worden gerangschikt. Na het ingeven van de zoekvraag is er geen gelegenheid om de antwoorden te classificeren op basis van de taxonomie en thesaurus. Een thesaurus kan natuurlijk wel aan de 'voorzijde' van een metasearch worden ingezet, zodat wanneer de gebruiker om 'fiets' vraagt, op basis van een thesaurus 'of rijwiel' aan de zoekvraag kan worden toegevoegd. Ook het zoekresultaat is

meestal niet voorzien van links of koppelingen naar gerelateerde termen.

Naast genoemde beperkingen biedt de metasearch ook grote voordelen:

- > een metasearch is 'real-time': er zit geen mogelijk verouderde index tussen de zoeker en de bronnen. Uiteraard moeten de bronnen zelf wel up-to-date zijn;
- > een eenmaal goed ingerichte metasearch vraagt om weinig onderhoudsinspanning en vereist nauwelijks opslagcapaciteit van de contentintegrator.

Indexering

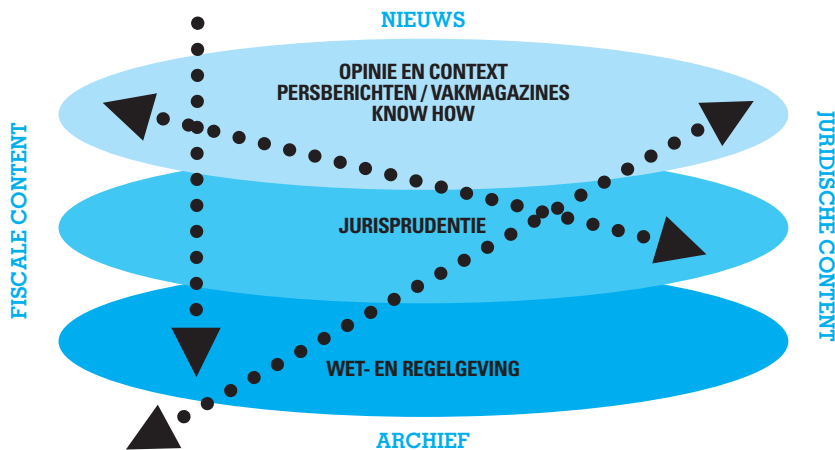
De indexeringsmethode is de tweede manier om content te integreren. Bij deze methode wordt een index gemaakt op de onderliggende bronnen. Voordat deze

aan de gebruiker wordt aangeboden, kunnen termen in de juiste context worden geplaatst en waar nodig eventueel worden gestandaardiseerd of vertaald. Zo biedt een index zowel een uniforme toegang als een toegang in context tot de onderliggende informatiebronnen.

Een index kan zeer uiteenlopende vormen hebben. Een index kan gecreëerd worden door voor elke bron een zeer gestructureerd en gestandaardiseerd metadata-record aan te maken waarin de bron wordt beschreven. Een index kan ook simpelweg bestaan uit een lijst van woorden die in de bronnen voorkomen. Indexeren kan handmatig, bijvoorbeeld door bronnen in een catalogus te beschrijven, maar vindt meer en meer

Federated search

Voor metasearch komt men ook wel de term 'federated search' tegen. Feitelijk gaat het om hetzelfde, maar bij de gangbare definitie van federated search worden ontubbeling van resultaten, uniforme presentatie van de resultaten en de mogelijkheid om de resultaten op bepaalde manieren te sorteren, meegenomen.



Figuur 1. Soorten content in de advocatuur

geautomatiseerd plaats. Bij automatische indexering wordt over het algemeen een ‘crawler’ ingezet. Dit is een applicatie die door de inhoud van de bronnen heen gaat. Op basis van de door de crawler verzamelde informatie wordt vervolgens een index aangemaakt. De meeste zoekmachinesoftware werkt op deze wijze.

De optimale wijze van indexering voor contentintegratie is een ‘broker-to-broker’-oplossing. Deze optie biedt de mogelijkheid om de contentintegratieoplossing rechtstreeks te laten zoeken in het redactiesysteem van de uitgeverij. Alle ‘intelligentie’ (bijvoorbeeld thesauri) en metadata worden meegenomen in de indexering, waarbij de betreffende uitgeverij garanties geeft op het gebied van kwaliteit van de zoekresultaten. Hiervoor is standaardisatie van classificatie wel noodzakelijk, hetgeen verregaande samenwerking en commitment vereist van de uitgeverijen en de contentintegrator.

Op basis van indexering kan een zogenaamde faceted search mogelijk worden gemaakt. Daarbij worden de zoekresultaten getoond met verwante en/of meer specifieke zoekmogelijkheden. Bij een via een woord in de titel gevonden muziekstuk kan bijvoorbeeld worden aangegeven, dat dit muziekstuk een genre heeft en dat men daarop eventueel verder kan zoeken. De full-text search wordt soms als een derde methode gezien. Dit is echter geen aparte mogelijkheid: vaak is een zogenaamde full-text search in werkelijkheid een search in een index. Daar waar werkelijk ‘live’ full-text wordt gezocht in verschillende documenten, is de full-text search in feite een metasearch zonder

‘Goede content-integratie kan ook een middel zijn om medewerkers te binden’

beperkingen omtrent de velden waarin gezocht moet worden.

Beide manieren van contentintegratie worden ook vaak samen gebruikt. Om bijvoorbeeld een index te kunnen maken die ook inhoud van het diepe web bevat, kan indexeersoftware een metasearch op een aantal databases uitvoeren. De resultaten van deze search worden vervolgens in de index opgenomen.

De contentintegratieoplossingen die op dit moment in de advocatuur worden gebruikt, werken over het algemeen met een index en enkele zelfs met een taxonomie en thesaurus van het Nederlands (fiscaal) recht. De meeste van deze oplossingen bieden echter ook de mogelijkheid om een deel van de bronnen met een metasearch te doorzoeken.

Middelen voor contentintegratie

Bij het zoeken naar manieren om contentintegratie tot stand te brengen, is het van belang dat men zich steeds realiseert dat contentintegratie bestaat bij de gratie van

een aantal standaarden:

- > toegang via inter- of intranet: het http-protocol;
- > standaard vraagtaal voor databases: in de bibliotheekwereld zijn nu vooral Z39.50³ en SRU⁴/SRW⁵ (aanpassing van Z39.50 aan het webtijdperk) in gebruik; verder de meer generieke standaarden SQL⁶ voor relationele databases en XPath⁷ voor xml-databases;
- > standaard syntax voor metadata: bijvoorbeeld RDF⁸ en Dublin Core⁹;
- > standaardwaarden voor metadata-elementen: thesauri en taxonomieën.

Al deze zaken betreffen de systemen die geïntegreerd moeten worden opgenomen (de bronnen dus) en niet de ‘schil’ waar de leverancier deze systemen in aanbiedt. Ook dit is een reden om niet te beginnen bij de leverancier.

De eerste stappen op weg naar contentintegratie zijn:

- > definitie van de informatiebehoefte van de doelgroep (vaststellen zoekgedrag en classificatiewensen);
- > selectie van informatiebronnen die qua inhoud bij deze doelgroep passen;
- > verdere selectie op basis van toegankelijkheid.

Pas als deze stappen doorlopen zijn, mag de software voor de ‘schil’ aan de orde komen.

Contentintegratie in de advocatuur

Voor de advocatuur is contentintegratie met name van belang voor het ontsluiten van wet- en regelgeving, jurisprudentie, opinie en vakliteratuur (figuur 1). Dit houdt in dat de geïntegreerde content vooral gebruikt zal worden in de fase van informeren binnen het primaire proces (figuur 2) van een advocaat.

Het delen en hergebruiken van kennis en ervaringen binnen de organisatie kan door contentintegratie zeer worden gestimuleerd, als tenminste de mogelijkheid wordt geboden om nieuwe bronnen ‘aan te melden’. Tijdens de studiedag van Stichting Het Overleg en GO opleidingen bleek echter dat dit delen en samenwerken in de advocatuur niet vanzelf gaat. Informatie is kapitaal en men is niet per definitie bereid om dat kapitaal met de collega-concurrenten te delen. Ook het argument van tijdsbesparing zou in de advocatuur wellicht niet zo zwaar wegen:

Houthoff Buruma: 'Strategische beslissing'

Mieke Steinberg, manager Informatiedienst bij Houthoff Buruma, voerde eind 2003 al de eerste gesprekken om te komen tot integratie van content van verschillende uitgevers in één portaal. De beslissing om tot aanschaf van een pakket voor contentintegratie over te gaan was een strategische. Steinberg: 'Ik zat al direct op één lijn met het bestuur. De wijze van zoeken en benaderen via één zoekingang en met behulp van websearching was een innovatie. Zij zagen direct in dat dit ten goede zou komen aan de kwaliteit van onze dienstverlening en dat het daarnaast ook veel gebruiksgemak met zich mee zou brengen.'

Een van de uitgangspunten van Houthoff Buruma was het werken met een onafhankelijke leverancier. Legal Intelligence was destijds de enige aanbieder in de markt. Het duurde tot eind 2006 voordat kantoorbrede implementatie volgde met een totaal vernieuwde gebruikers-

interface van het portaal en een nog betere zoekfunctionaliteit. 'We hebben veel vooronderzoek verricht, allereerst met de kenniswerkers, maar daarna ook met een uitgebreide gebruikersgroep. Dat gaf ons veel inzichten over de manier van zoeken, hetgeen leidde tot een groot aantal verbeterpunten. Legal Intelligence heeft deze in relatief korte tijd in de functionaliteit verwerkt.' Met de uitgevers hebben we goed samengewerkt zodat we in het voortraject over de meeste bronnen konden beschikken. Tegelijk is het voor de uitgevers lastig gebleken om hun businessmodel hierop aan te passen. Het leveren van informatiebronnen aan een tussenpersoon, de wijze van levering en de prijsbepaling zijn al enige tijd onderwerp van gesprek. 'We denken actief mee over de uitgangspunten hiervoor en hebben binnen PAUC, het Platform Advocatuur en Uitgevers Contacten, zelfs een aparte werkgroep ingesteld die alle betrokkenen hierover zal adviseren.'



Mieke Steinberg

Mieke Steinberg heeft nog een paar praktische tips voor kantoren die met een project contentintegratie willen beginnen: 'Bedenk eerst goed wat contentintegratie voor je organisatie betekent. Onderzoek de huidige werkwijze van de eindgebruikers als het gaat om het gebruik van elektronische bronnen en bedenk dat de investering niet alleen gaat om geld, maar vooral ook om kwaliteitsverbetering en tijdsbesparing.'

uren worden immers doorberekend aan de klant.

De voordelen voor de advocatuur moeten vooral gezocht worden in de kwaliteit van het zoekresultaat. Die kwaliteit zit in de combinatie van het eerder genoemde selectieaspect, de zoekinterface en de onderliggende zoektechnologie. Daarbij is vooral het kunnen 'zoeken in context' van belang. Het woord contract bijvoorbeeld kan in een arbeidsrechtelijke context iets geheel anders betekenen dan binnen een belastingrechtelijke context. Alleen een contentintegratieoplossing die dit 'begrijpt', biedt werkelijk waar voor zijn geld.

Valkuilen

Het ligt voor de hand om snel van start te gaan met contentintegratie. Het lijkt er immers op dat het invoeren van contentintegratie alleen maar voordelen heeft. Het succes valt of staat echter met voldoende interne capaciteit voor onderhoud aan de voorziening. Wordt dit onderhoud niet vanaf het begin geborgd, dan zal deze doodbloeden.

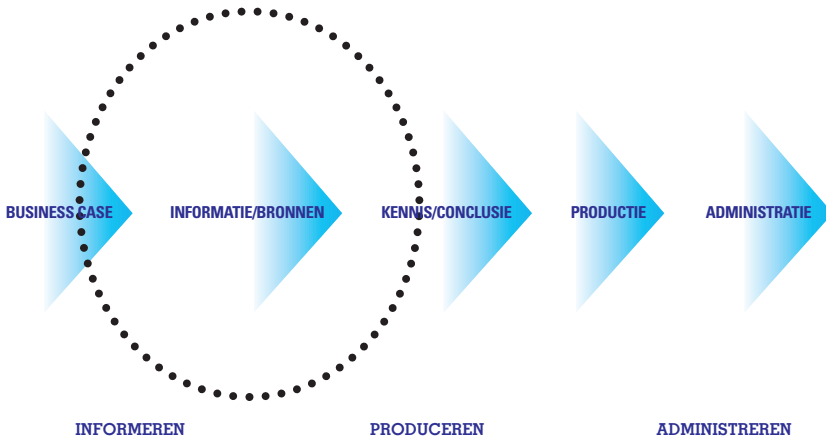
'Het delen en hergebruik van kennis kan door contentintegratie worden gestimuleerd'

Invoer van iets nieuws omdat het er nu eenmaal is, mag nooit de aanleiding zijn. Hoe mooi de verkooppraatjes van leveranciers ook zijn, het levert de organisatie zelden iets op. Bovendien zijn er in het geval van contentintegratie vele wegen die naar Rome leiden (zie hierboven). De introductie van contentintegratie moet een weloverwogen beslissing zijn en kan pas gestart worden indien men er werkelijk aan toe is.

Een andere valkuil is wachten tot er een breed draagvlak is. Draagvlak is het vijgenblad van de Nederlandse managementcultuur. Omdat er altijd wel ontevreden zijn, kan met 'te weinig draagvlak' elke verantwoordelijkheid voor het mislukken van een project op de medewerkers worden afgeschoven. Ook bij de invoer van contentintegratie is het voldoende als er een paar enthousiaste voorlopers zijn, mits het management voldoende duidelijk maakt dat contentintegratie – op termijn – van levensbelang is voor de organisatie.

De toekomst

Met de middelen die op dit moment beschikbaar zijn, kunnen al heel aardige contentintegratieoplossingen worden gebouwd. Maar het werk ligt wel voor een zeer groot deel bij de aanbieder van de 'schil' rond de content: hij moet indexen maken, veldnamen 'mappen', beheercapaciteit ter beschikking stellen of veel geld betalen aan een marktpartij die dat voor hem wil doen. Op termijn kan dat anders en beter.



Figuur 2. Primaire proces in de advocatuur

Standaardisatie

Onder de vleugels van het World Wide Web Consortium (W3C), de beheerders van het WWW, is er de afgelopen jaren gewerkt aan standaardisatie op het gebied van metadata. Termen die hierbij horen zijn rff (het resource description format, een samenhangende verzameling afspraken om de inhoud van het WWW

op redelijk uniforme wijze van metadata te voorzien), Dublin Core (vijftien ‘universele’ metadatavelden) en xml (de syntax om de metadata in op te slaan). Bij het selecteren van contentintegratieoplossingen is het van belang om de leverancier naar ondersteuning van dergelijke standaarden te vragen. Op het moment dat aanbieders van bron-

nen tot afspraken kunnen komen om de gestandaardiseerde metadatastructuren op uniforme wijze te vullen, bijvoorbeeld door gebruikmaking van dezelfde thesauri, breekt er een nieuwe fase in contentintegratie aan. Op basis van gestandaardiseerde metadata kunnen, uiteraard, betere indexen gebouwd worden, maar ook voor de metasearch bieden gestandaardiseerde metadata-records nieuwe mogelijkheden. Op het moment dat men in, bijvoorbeeld, de juridische sector uit kan gaan van de aanwezigheid van gestandaardiseerde metadata-records, wordt het mogelijk om in een gegeven domein (= in een aantal bronnen) real-time op zoek te gaan naar recent toegevoegde informatie. De tussenstap van een – mogelijk verouderde – index is dan niet meer nodig. De verantwoordelijkheid voor de kwaliteit van de integratie komt dan weer te liggen waar deze eigenlijk hoort: bij de eigenaar van de bron, en niet bij de indexeerder. En dat zal de kwaliteit van contentintegratieoplossingen alleen maar ten goede komen.

**Allen & Overy:
 'Het is nog maar net begonnen'**

Mart van de Kerkhof is kennismanager van Allen & Overy LLP in Amsterdam. Al in de vorige eeuw stelde hij zich achter het concept van één punt van toegang tot relevante juridische informatie voor alle juridische professionals. Toen het eind 2002 de Commissie Uitgevers Contacten (CUC) – waar Allen & Overy in deelneemt – niet lukte om de grote uitgevers te verleiden gezamenlijk één portaal voor juridische professionals op te zetten, besloot hij dat voor zijn kantoor dan maar zelf in te richten.

Met iLLUMiON, het bedrijf waarmee Allen & Overy een intranet had opgezet met CMS Smartsite, werden begin 2003 de eerste stappen gezet op weg naar een eigen portaal op basis van Verity (K2). De grootste uitdaging was in eerste instantie niet zozeer de techniek, als wel het verkrijgen van de medewerking en ondersteuning van de uitgevers bij het aankoppelen van de bronnen. Geen van de uitgevers had de passende licenties klaar voor de gewenste wijze van contentontsluiting. Hoe om te gaan met spide-

ren, het omzeilen van productinterfaces, het doorzoeken van indexen en deeplinken?

Een belangrijke doorbraak was een technische oplossing en een passende service level agreement voor een directe koppeling van Allen & Overy's zoekmachine met de zoekindexen van de bronnen van uitgevers. Kenmerkend voor de gekozen wijze van contentintegratie is, dat naast de zoekfunctie van het portaal de interfaces naar de producten van de uitgevers (vaak bron-specifieke zoekmachines) vooralsnog beschikbaar blijven voor de gebruikers. Van de Kerkhof: 'De stap van vertrouwde interfaces als van de ADW of Opmaat naar één geheel nieuw interface acht ik te groot voor gebruikers, ook al omdat daarmee de aansluiting met wat er verder in het beroepsveld wordt gebruikt verdwijnt.' Van de Kerkhof is er beducht op dat de oplossing van Allen & Overy geen exclusieve wordt: 'De mogelijkheden van contentintegratie zijn nog lang niet uitgeput en die verzilver je alleen als je zoekt naar open oplossingen en open standaarden



Mart van de Kerkhof

toepast. Daar wordt iedereen beter van.' Allen & Overy heeft in kleine stappen samen met iLLUMiON en met de medewerking van uitgevers in een aantal jaren het zoekportaal ontwikkeld. Momenteel is het volledig operationeel en in gebruik bij de professional support lawyers en informatiespecialisten. In de loop van het jaar volgt de kantoorbrede uitrol. Van de Kerkhof: 'Het was een lang en taai project, maar ik ben erg blij dat er nu een resultaat ligt. Een resultaat waar we nog wel even op zullen voortbouwen, want contentintegratie is nog maar net begonnen.'

Integratie in enterpriseportals

Op termijn zullen contentintegratieoplossingen aantrekkelijker worden voor de gebruikers. Nu nog is het vaak een kwestie van browsen door lijstjes en kiezen uit categorieën. In de nabije toekomst zal het mogelijk worden om delen uit een contentintegratieoplossing op te nemen in zogenaamde 'enterpriseportals', die toegang tot bedrijfsapplicaties geven. Zo komt de geïntegreerde content steeds makkelijker ter beschikking op plaatsen waar zij gebruikt

moet worden. Bovengenoemde standaarden zullen het eenvoudig maken om het bedrijfsaanbod van informatiebronnen vervolgens te filteren of uit te breiden door middel van eigen, wellicht meer verfijnde, metasearches, om zo tot een gepersonaliseerd aanbod te komen. Zo wordt het organisatiebrede belang van een goede contentintegratie verzoend met de individuele (en individualistische) informatiebehoefte van de advocaat. Als dat de advocaat niet over de streep trekt, wat dan wel? <

Martijn van der Kaaij is docent informatie-management bij GO opleidingen en zelfstandig adviseur. Stef Verf is directeur van iLLUMiON, specialist op het gebied van intranet en kennismanagement.

Noten

- 1) Stichting Het Overleg is een platform voor kennismanagers, bibliothecarissen, informatiespecialisten, knowhowmedewerkers, documentaristen werkzaam voor advocaten, (kandidaat-)notarissen, fiscalisten en rechtsbijstandverzekeraars. Het Overleg biedt gelegenheid tot gedachtewisseling met vakgenoten, bijvoorbeeld in de vorm van workshops en ronde tafel bijeenkomsten, met onderwerpen als: beroepsprofiel, kennismanagement, kwaliteitsonderzoek, opleidingen en vakkenis. www.hetoverleg.org
- 2) www.breekpunt.nl/woordenboek.aspx?index=C&wbitem=71
- 3) www.loc.gov/z3950/agency/
- 4) www.loc.gov/standards/sru/
- 5) www.loc.gov/standards/sru/srw/
- 6) nl.wikipedia.org/wiki/SQL
- 7) www.w3.org/Style/XSL/
- 8) www.w3.org/RDF/
- 9) dublincore.org/

Verder lezen

Ontsluiting elektronische informatie in nieuw stadium : Hans, van der Klis. In: *Advocatenblad*; vol. 86 (2006), afl. 13, pag. 598-606 (9) / 2006. www.hetoverleg.org/files/Advocatenblad-13-Digitale_ontsluiting.pdf

Advocaat en Kennismanagement : M. Apistola. Amsterdam: Vrije Universiteit, 2007. Academisch proefschrift. dare.uvu.vu.nl/handle/1871/10608