



INFO MANAGEMENT

Dox*i***s**
V*i***NDT**



Jaargang 17, Nr. 1, april 2002

Eén van de lastigste dingen in de digitale informatiemassa die op ons af komt, is het vinden van de juiste informatie. Op allerlei manieren wordt getracht daar oplossingen voor aan te dragen. Vooral op het web speelt dat probleem. Daar lijken alle schotten tussen verschillende informatiebronnen immers verdwenen. In een vorige aflevering heb ik daar al het nodige over gezegd. In deze aflevering wil ik in dit verband nader ingaan op het zogeheten Semantic Web, een poging om meer betekenis toe te voegen aan de massa informatie om zo beter en gericht zoeken mogelijk te maken met behulp van slimme programmaatjes, zogeheten agents. Tevens nog enkele woorden over bewaring, omdat op dat terrein veel initiatieven spelen die vermeldenswaardig zijn.

Het Semantic Web

De nieuwste ontwikkeling in relatie tot het zoeken op internet is die van het zgn. Semantic Web. Er bestaat geen eenduidige definitie van, maar de bedoeling is om de inhoud die op het web wordt gepubliceerd meer betekenis te geven, zodat zogeheten (intelligent) agents het web kunnen afzoeken en gemakkelijker in staat zijn informatie te identificeren die voldoet aan de vraag (profiel) van een gebruiker.¹ Met andere woorden: de informatiemassa op het web wordt meer geschikt gemaakt om door 'machines' of programma's verwerkt en gebruikt te worden. Men hoopt dat te bereiken door gebruik te maken van eXtensible Markup Language (XML) en Resource Description Framework (RDF). Het laatste is een nadere toepassing van XML. Met XML kan de structuur van documenten, in dit geval webpagina's of delen ervan, beschreven worden door de onderdelen

van documenten van zogeheten tags (etiketten, labels) te voorzien. XML zegt op zich echter niets over de betekenis van de tags, maar dat kan wel met behulp van RDF. In RDF kan worden aangegeven dat onderwerpen (personen of dingen) bepaalde eigenschappen (bijv. een relatie, 'is auteur van' en dergelijke) bezitten die een bepaalde waarde (bijvoorbeeld een andere persoon of ding) kunnen hebben. Voorbeeld: De heer Jansen woont in de Voorstraat te Juinen. Elk van die elementen kan met behulp van een Uniform Resource Identifier (URI; zoals bijvoorbeeld een URL) beschreven worden. Aanvullend is het noodzakelijk om de gehanteerde begrippen in een bepaald domein te definiëren en de onderlinge relaties ervan vast te stellen, het ontwerpen van een zogeheten ontologie. Dat is een soort classificatiesysteem, waarmee objecten kunnen worden ingedeeld en onderverdeeld en met elkaar in verband gebracht. Er ontstaan zo klassen, subklassen enzovoorts die

bepaalde, nader gedefinieerde relaties met elkaar hebben.² Een bepaalde klasse kan eigenschappen erven van hoger geordende klassen en zelf eigenschappen doorgeven aan lagere klassen. Simpel voorbeeld: indien een bepaalde straat zich in een bepaalde stad bevindt en die stad ligt in Nederland, dan ligt die straat ook in Nederland. Zo wordt betekenis toegevoegd aan de informatie en kunnen op basis hiervan afleidingsregels (logische redeneringen) worden toegepast door de genoemde agents, bij het zoeken naar informatie die aan bepaalde criteria voldoet. De ontologie maakt het voor de agents mogelijk om te 'beredeneren' of bepaalde informatie relevant is voor de vraag die ze bij zich hebben of niet.

De stand van zaken rond het Semantic Web op dit moment is dat het allemaal nog erg in een beginstadium is. Een werkgroep van het bekende World Wide Web consortium (W3C) is er druk mee bezig, maar er moet nog zeer veel gedaan worden om de verwachte voordelen inderdaad te realiseren. Er zijn echter al wel beperkte voorbeelden van Semantic Web applicaties (zie verwijzingen).

Wat betekent dit nu voor toegankelijkheid? De gedachte achter het Semantic Web is dat het onmogelijk is een wereldwijde, gezamenlijke, gestandaardiseerde taal te ontwikkelen die voor het gehele web geldt. Een dergelijke gemeenschappelijke taal kenmerkt zich bovendien in de regel door beperking en starheid. In een dergelijke benadering worden de mogelijke vragen beperkt tot die vragen die het systeem kan beantwoorden. In plaats daarvan wil men iets ontwikkelen dat uiteindelijk wel hetzelfde doel zou moeten dienen, maar opener, flexibeler en gemakkelijker te beheren is. En daarmee is deze benadering meer in lijn met het hele karakter van het World Wide Web: geen centrale benadering, maar ieder voor zich, mits men zich aan een beperkt aantal regels houdt (o.a. gebruik van RDF, XML,

zelfde beschrijvingstaal, ontologie). Het betekent ook dat in deze omgeving niet op alle vragen een antwoord mogelijk zal zijn, maar dat is op zich niet erg.

Kijkend naar de archiefwereld zou men een inventaris of zelfs een heel systeem van inventarissen kunnen zien als een ontologie die de relaties tussen documenten beschrijft (op verschillende niveaus). Indien men een beschrijving geeft in bijvoorbeeld RDF, wordt dergelijke informatie ook voor het Semantic Web bruikbaar. Dat kan gedaan worden op het niveau van inventarissen, maar uiteraard ook op dat van organisaties of de documenten die ze produceren bij de uitvoering van taken.

Alles samenvattend is het hele idee van een Semantic Web een boeiende en veelbelovende ontwikkeling die voor iedere informatiespecialist van groot belang is. Ik heb hier slechts aan de oppervlakte van het onderwerp gekrabbd.

Duurzaamheid in de politiek en praktijk

Digital preservation Charter

Behalve dat er veel activiteit is om te komen tot oplossingen (al dan niet tijdelijk) op het terrein van digitale bewaring, wordt er ook op politiek vlak aandacht aan besteed. Het gaat in dit geval om pogingen te komen tot een zogeheten 'digital preservation charter'. Een dergelijk charter beoogt een soort verklaring te zijn, waarin aandacht voor digitale bewaring en het belang ervan voor cultureel erfgoed wordt gevraagd. Tevens wil het samenwerking bevorderen. Het lijkt ineens in de lucht te zitten. Zowel bij de UNESCO als de Europese Commissie, in het bijzonder via het Spaanse voorzitterschap van de EU en het ERPANET-project, wordt op dit moment aan een dergelijk charter gewerkt.

Digital Preservation Coalition

In dit verband is recentelijk in het Verenigd Koninkrijk een nieuw initiatief van de grond gekomen dat op 27 februari 2002 gelanceerd is. Het gaat om de zogeheten Digital Preservation Coalition (DPC; zie www.jisc.ac.uk/dner/preservation). Deze coalitie omvat zo'n 17 organisaties, waaronder het Public Record Office, de British Library, de National Archives of Scotland en het Joint Information Systems Committee of the Higher and Further Education Funding Councils (JISC), en beoogt samenwerking binnen het Verenigd Koninkrijk en internationalisering op dit terrein te bevorderen, om zo tot oplossingen te komen. Dit nieuwe initiatief richt zich vooral op cultureel erfgoed, zoals ook de deelnemerslijst laat zien. Hiermee is er weer een nieuwe loot aan de toch al volle stam van samenwerkingsverbanden op het terrein van digitale bewaring bij gekomen. Of dit ook echt bijdraagt tot oplossingen zal de tijd leren. Het toont wel hoe het onderwerp in de belangstelling staat.

Safekeeping project van NLA

Een interessante ontwikkeling is te vinden in het verre Australië. Daar is de National Library of Australia (NLA) het zogeheten safekeeping project gestart. Het is een gezamenlijke aanpak van de Council on Library and Information Resources (CLIR) en Preserving Access to Digital Information (PADI). Idee hierachter is dat het wel leuk is dat er veel websites zijn en informatie is over digitale bewaring, maar dat het ook goed zal zijn dat dergelijke bronnen voor de lange termijn bewaard worden. Het project beoogt een gedistribueerde 'bibliotheek' van belangrijke publicaties en bronnen over digitale bewaring op te zetten. Elke bron die op die manier bewaard wordt heeft het kenmerk 'safekept'. Elke organisatie kan deelnemen, mits men de zorg voor de bewaring van één of meer dergelijke digitale bronnen op zich neemt. Voor meer informatie zie de PADI-website (www.nla.gov.au/padi).

3e DLM-Forum in Barcelona

Tenslotte wordt dit jaar weer een DLM-Forum georganiseerd. Ditmaal nu eens niet in Brussel, maar in Barcelona. Op 7 en 8 mei 2002 wordt deze conferentie gehouden onder de titel '@cces and preservation of electronic information: Best practices and

solutions'. Dat deze dit keer in Spanje wordt gehouden, is verbonden aan het Spaanse voorzitterschap van de Europese Unie. De Catalaanse overheid is één van de sponsors, naast AIIM en DLM-Forum (EU). Zoals gewoonlijk is het programma weer ingedeeld volgens parallelsessies. De thema's zijn dit keer:

- The memory of the information age (preservation, migration and long-term availability)
- The use of public information (security, protection and control)
- Content management and delivery (establishing and maintaining archival systems, best practices and easy-to-use-access)
- Organising records and archives (metadata, standardisation and model requirements)
- Improving access to knowledge (training and education for information experts and users)
- Capturing information (high volume information transformation and automatic indexing)

Tevens is er een drietal workshops, respectievelijk over 'Metadata, standaards...', 'De implicaties van digitale handtekeningen' en 'De juridische bewijskracht van digitale opslag'. Tijdens de conferentie zal tevens een zestal 'white papers' over de ontwikkelingen in de informatietechnologie door de industrie worden gepresenteerd. Daarmee drukt ze een tamelijk zwaar stempel op de conferentie. Interessant is dat men tijdens deze conferentie tevens een zogeheten DLM-network of excellence on electronic archives wil lanceren. Men hoopt hiervoor subsidie van de Europese Commissie te verkrijgen. Hoewel in de kern niet veel anders dan het DLM-Forum, heeft het DLM-netwerk een meer projectmatig karakter en wordt het gesubsidieerd, dat wil zeggen indien het voorstel wordt geaccepteerd.

Zie voor meer informatie www.dlmforum.eu.org of http://europe.eu.int/historical_archives/dlm_forum.

1 Zie o.a. Tim Berners-Lee, James Hendler en Ora Lassila, The Semantic Web, in: Scientific American, (May 2001); zie www.scientificamerican.com/2001/0501issue/0501berniers-lee.html. Zie ook www.semanticweb.org www.w3.org/2001/sw/ en Sean B. Palmer, The Semantic Web: An Introduction, (September 2001), op <http://infomesh.net/2001/swintro>.

2 Zie o.a. Robbert Hoeffnagel, 'De toekomst van internet. Het semantische web', in: Tijdschrift voor multimedia, nr.6 (2001).