

tune  
101  
S  
Radio

# INFORMATIE PROFESSIONAL

01

## Vooruitblik

# NVB Congres

02

## Het vak, de professie en de bibliotheek

# KID- management

03

## Funfactor van apps

# SmartErfgoed

04

## Weekboek

# Het advocatenkantoor

# Semantisch web

## Zoek-machines van de toekomst

# Zoekmachines van de toekomst

Voor de manier waarop het web zich ontwikkelt, is de metafoor van het toekennen van versienummers algemeen in zwang geraakt. Na het statische en het interactieve web dat men met web 1.0 en 2.0 is gaan aanduiden, is dan nu de vraag welke ontwikkeling de meeste kans maakt met web 3.0 te worden aangeduid.

Edgar Meij zal op het NVB Jaarcongres op 17 november aanstaande betogen dat dit om het semantische web zal draaien, waarbij hij vooral ingaat op wat dat voor het zoeken betekent. Hier geeft hij vast een voorproefje.

Edgar Meij

**'De zoekmachine van de toekomst ziet er nu al slimmer, interactiever en socialer uit'**

**NVB  
Jaarcongres  
2011**

Edgar Meij spreekt tijdens de track  
**'Web 3.0'**

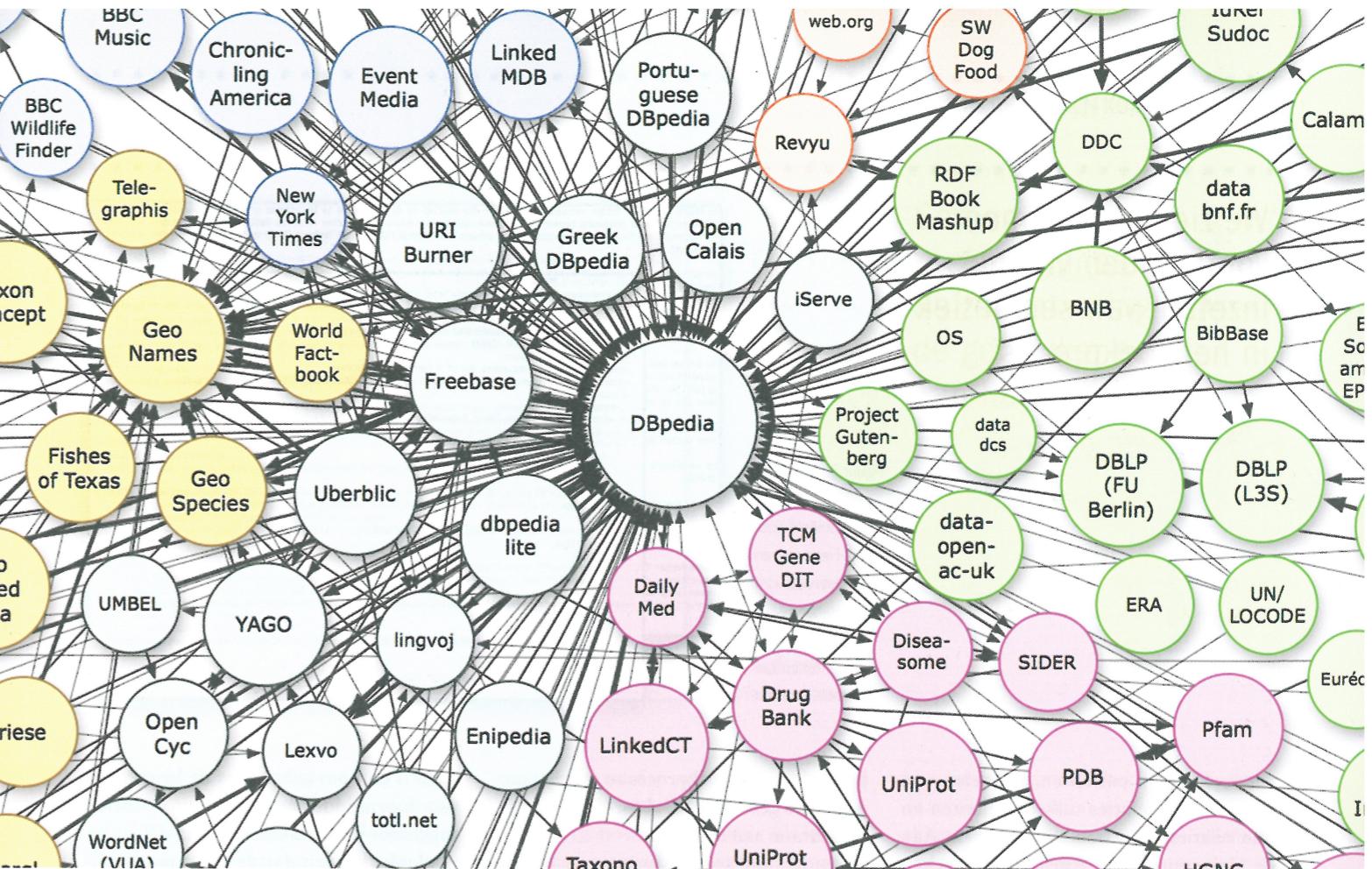


**Wie is...**

Edgar Meij is onderzoeker aan het Instituut voor Informatica van de Universiteit van Amsterdam. Hij doet onderzoek naar het intelligenter maken van zoekmachines door het toepassen van informatie uit externe kennisbronnen en richt zich op statistische methoden voor intelligente informatieontsluiting.

Er bestaat enige discussie over wat de logische opvolger zal zijn van web 2.0, waarin user-generated content, het delen van informatie en interoperabiliteit centraal stonden. Hoewel meer ideeën de ronde doen, is er veel steun voor het idee web 3.0 gelijk te stellen aan het semantische web. Het sturende idee achter het semantische web is dat alle dingen op internet worden beschreven in speciale talen, zodat computers de informatie die ze aantreffen kunnen 'begrijpen'. Een object kan bijvoorbeeld gekenmerkt worden als onderdeel van een voertuig, als persoon of als een fysieke plek op aarde. Zodra objecten op een dergelijke manier worden gekarakteriseerd, ontstaat een gigantisch netwerk van 'linked data' en kunnen computers verrassende en nuttige links ontdekken, die wellicht nooit door mensen ontdekt zouden zijn. Dit heeft de potentie een compleet nieuwe vorm van kunstmatige intelligentie mogelijk te maken [1]. Maar zover zijn we nog (lang) niet. Er zijn een hoop redenen aan te dragen waarom het semantische web uit de visie van Berners-Lee het niet zal reden. Een van de belangrijkste redenen is het feit dat veelal een enkele centrale ontologie ('model van de wereld') wordt verondersteld. Ook is nog niet onomstotelijk vastgesteld dat dergelijke technieken werken in de open omgeving van het web. In specifieke omgevingen, toepassingen en taken worden successen behaald, maar deze vertalen zich nog niet naar breed toepasbare initiatieven. Desondanks bestaat er inmiddels een groeiende 'wolk' van datasets met machineleesbare gegevens, die onderling gekoppeld worden ('linking open data' of LOD, zie figuur 1). Die wolk kan ingezet worden als semantische ruggengraat om allerhande taken aan te pakken.

Een andere reden is het feit dat bij alle-dagse gebruikers van het web weinig animo bestaat om deel te nemen aan het toekennen van metadata aan objecten. We zijn dusdanig gewend geraakt aan het gebruik van zoekmachines om een



Figuur 1. Linking Open Data cloud diagram; datasets met gestructureerde computerleesbare data en daartussen aangebrachte links (richard.cyganiak.de/2007/10/lod/)

weg door het web te vinden, dat geen dringende reden bestaat om bij te dragen aan de ontwikkeling van het semantische web – los van informele, 'lichtgewicht' semantiek in de vorm van persoonlijke bookmarks/tags<sup>1</sup> of simpele annotaties op webpagina's.

## Onder de motorkap

Waar we inmiddels wel steeds meer voorbeelden van zien is het inzetten van semantiek in het zoekproces, bijvoorbeeld verkregen uit LOD.<sup>2</sup> Wat houdt dat dan in? Recent ontwikkelingen maken het mogelijk dat computers begrijpen welke concepten gebruikt worden in stukken tekst, zoals webpagina's, zoekmachine queries, tweets, et cetera. Dit begrijpen houdt niets meer in dan het linken van (delen van) tekst aan meer gestructureerde informatie – veelal in de vorm van zogenaamde entiteiten – zoals je die bijvoorbeeld vindt in LOD.

Een veelvoorkomende definitie van een 'entiteit' omvat personen, bedrijven, producten, locaties, et cetera. Oftewel, alles wat een eigen Wikipedia-pagina kan hebben. Exaleads zoekmachine voor Wikipedia geeft een goed voorbeeld van entiteiten.

Zodra je daar een query intypt haalt de zoekmachine entiteiten op die gerelateerd zijn aan de query, die je vervolgens op een interactieve manier kan gebruiken om de resultaten aan te passen aan datgene wat je voor ogen had. In zekere zin is dit een moderne variant van faceted zoeken, waarbij metadata behorende bij of geëxtraheerd uit zoekresultaten, worden geanalyseerd en weergegeven, teneinde de zoekresultaten interactief te verbeteren. Een van de meest bekende voorbeelden hiervan is te vinden bij allerhande online winkels, waar je één of meer categorieën kan selecteren om de gevonden producten in te perken. Wikipedia is uitgegroeid tot een aanzienlijke kennisbron en het aantal Wikipedia-pagina's groeit gestaag. Een van de inte-

ressante eigenschappen van Wikipedia in de context van semantische zoekmachines is dat iedere entiteit geassocieerd is met een door mensen bewerkte beschrijving, zodat ze makkelijk gevonden en beschreven kunnen worden. Ook bevat Wikipedia structurele informatie in de vorm van categorieën, hyperlinks tussen artikelen en infoboxes en sjablonen. Infoboxen zijn de elementen die bij sommige Wikipedia-artikelen aan de rechterhand worden weergegeven en vaak semi-gestructureerde informatie bevatten. Hetzelfde 'soort' Wikipedia-artikelen heeft veelal hetzelfde soort infoboxes, zoals steden, zangers, et cetera. Sjablonen lijken hier enigszins op en bevatten ook gestructureerde informatie die specifiek is voor een bepaald soort onderwerp, zoals medicijnen of vliegvelden. DBpedia is een initiatief dat als doel heeft deze informatie automatisch uit Wikipedia te extraheeren en beschikbaar te stellen in LOD. Door zijn algemene aard speelt DBpedia hier een centrale rol, zoals in figuur 1 ook te zien is.

Methoden die tekst kunnen linken aan entiteiten gebruiken in belangrijke mate taaltechnologie in combinatie met machineleren, en zijn erop gericht om entiteiten,

**'Het semantische web uit de visie van Berners-Lee zal het niet reden'**

## 'We zien steeds meer voorbeelden van het inzetten van semantiek in het zoekproces'

Figuur 2. Voorbeeld van annotaties gemaakt door Fietstas, een webservice voor tekstanalyse die onder meer entiteiten kan herkennen in tekst

```

<NDMEGEN - B1> een Britse vrouwelijke militair die meeliep in de Vierdaagse is Mexicaanse groep vertegenwoordigde, gemeente Nijmegen donderdag bevestigd. De vrouw meldde zich voorspoedig ziek en wordt immiddels behandeld in virusremmer Tamiflu. Ze is opgenomen in het Universitair Medisch Centrum St Radboud (UMC) in Nijmegen en wordt verpleegd.
```

```

De patiënt maakt gebruik van XML version="1.0" type="text/xsl" href="http://fietstas.science.uva.nl/xslt/cloud.xsl">
<request xmlns="http://fietstas.science.uva.nl/cloud" xmlns:tgw="http://fietstas.science.uva.nl/tg_wiki">
  <xsl:stylesheet type="text/xsl" href="http://fietstas.science.uva.nl/xslt/cloud.xsl" />
  <date>2011-10-09 11:45:20.700039</date>
  <status>completed</status>
  <message>Processed 1 documents out of 1 (errors: 0)</message>
  <permalink>http://fietstas.science.uva.nl/cloud/2e28cff9dc27cb36a7c3ea03b867</permalink>
</request>
<items count="5" score="30.0" type="MISC" wikipedia:link="http://nl.wikipedia.org/wiki/Vierdaagse">
  <item count="5" score="30.0" type="LOC" wikipedia:link="http://nl.wikipedia.org/wiki/Nijmegen">
    Nijmegen
  </item>
  <item count="4" score="26.9497685552" type="LOC" wikipedia:link="http://nl.wikipedia.org/wiki/UMC">
    UMC
  </item>
  <item count="3" score="23.0173362787" type="MISC">
    De Britse
  </item>
  <item count="3" score="23.0173362787" type="LOC" wikipedia:link="http://nl.wikipedia.org/wiki/Mexico">
    Mexico
  </item>
  <item count="2" score="17.4748842776" type="LOC" wikipedia:link="http://nl.wikipedia.org/wiki/Hoitsna">
    Hoitsna
  </item>
</items>

```

netwerken van entiteiten, profielen van entiteiten of relaties tussen entiteiten en hun relaties in teksten te herkennen. Aan de Universiteit van Amsterdam werken we sinds 2008 aan een gedistribueerde omgeving genaamd Fietstas, die de vereiste functionaliteit als webservice aanbiedt. Naast verschillende vormen van tekstnormalisatie biedt Fietstas ook semantische functionaliteiten zoals het herkennen van entiteiten en relaties, het normaliseren van entiteiten en het genereren van 'profielen' van entiteiten. Dit maakt het bijvoorbeeld mogelijk om een willekeurige tekst aan te bieden en een lijst met gevonden entiteiten terug te krijgen. Zie figuur 2 voor een illustratie van het soort van documentannotaties dat daarbij door Fietstas gegenereerd wordt.

Wat kunnen we met de ontdekte links tussen tekst en Wikipedia? Vanwege de directe koppeling tussen Wikipedia en DBpedia kunnen de tools voor semantisch linken (zoals Fietstas of WikipediaMiner)<sup>3</sup> direct gebruikt worden om LOD-entiteiten te herkennen in willekeurige teksten, gebuikmakend van aanvullende informatie uit Wikipedia [5]. Zodra entiteiten zijn herkend, kunnen we deze inzetten tijdens het zoekproces.

### Semantisch zoeken en vinden

Semantische zoekmachines stellen ons in staat om relevante entiteiten en hun relaties te identificeren in grote hoeveelheden tekst en vervolgens in te zetten, bijvoor-

beeld door het zoekproces te sturen, suggesties te genereren of door de resultaatpresentatie aan te passen. De mogelijke toepassingen van het inzetten van semantiek tijdens het zoeken zijn legio en variëren van het simpelweg inzetten van alternatieve schrijfwijzen of synonymen (zoals de tilde operator in Google),<sup>4</sup> via specifieke interpretaties zoals WolframAlpha<sup>5</sup> die verzorgt, tot het volledig taalkundig analyseren van een query (zoals Powerset<sup>6</sup> en Ask.com dat beogen).

Laten we verder ingaan op het scenario van een zoekmachine voor het web. Het automatisch 'expanderen' van een query met alternatieve schrijfwijzen lijkt een veelbelovende strategie om op een automatische manier meer relevante webpagina's te identificeren. Waarom past een bedrijf als Google dit dan niet op grote schaal toe? Een belangrijk tegenargument is ambiguïteit; de gemiddelde query bestaat gemiddeld uit ongeveer 2,4 termen (vaak zelfs minder, zie [3,6]) en biedt vaak niet genoeg houvast om zeker te weten wat een gebruiker bedoelt. Een dergelijke strategie kan daarom al snel de verkeerde kant opgaan en termen introduceren die niet relevant zijn. Een ander tegenargument is het feit dat het zoekproces minder transparant wordt. Zodra er automatisch termen worden toegevoegd aan de query, bestaat de kans dat er documenten worden teruggegeven die niet relevant zijn, waardoor de kans groot is dat de gebruiker gefrustreerd raakt en uitwijkt naar een andere zoekmachine. Een betere stra-

tegie is dan om gebruikers de keus te geven door middel van zijdelings vermelde suggesties in de resultaatpagina's.<sup>7</sup> Ondanks het feit dat de genoemde technieken niet automatisch worden toegepast, werken de drie grote zoekmachines voor het web (Google, Bing en Yahoo!) aan verbeteringen van algoritmes en methoden die met hoge precisie kunnen aangeven welke entiteit(en) worden bedoeld in een query. Zij passen dergelijke technieken trouwens al enige tijd in beperkte mate toe. In het bijzonder bij het zogenaamde 'vertical search', waarin een incrementeel lerend algoritme beslist wat de intentie van de gebruiker is. Aan de hand van de voorgestelde intentie ('ik wil een boek kopen', 'ik wil meer informatie', et cetera) wordt vervolgens besloten of er, naast de normale webresultaten, ook resultaten worden weergegeven uit 'verticale' sub-zoekmachines die gespecialiseerd zijn in een bepaald type informatie. Denk aan Google Maps, YouTube of plaatjes, maar ook aan resultaten uit online winkels, blogs, het nieuws, et cetera. In het meest voorkomende geval worden de resultaten uit deze 'verticals' verweven met de normale zoekresultaten. De Yahoo!-zoekmachine gaat hiermee zelfs nog een stap verder en biedt specifieke overzichtspagina's aan voor sommige queries [2].<sup>8</sup> Ook het simpelweg herkennen van bepaalde patronen kan worden beschouwd als een vorm van intelligent zoeken. Probeer maar eens de Google queries 'time in Amsterdam', '1+1' of '1 dollar in euro'. Andere manieren om concepten en entiteiten in te zetten tijdens het zoeken vinden we in zogenaamde Microformats en in RDFa. Dit zijn semantische talen die als doel hebben om semantiek in te bedden in XHTML-opmaakcode. Waar Microformats een specifiek vocabulaire behelst, kan RDFa gebruikt worden om ieder willekeurig kennismodel in te zetten in de annotatie van (gedeeltes van) webpagina's. Het doel hiervan is om op een backwards-compatible manier semantische webfunctionaliteit aan webpagina's te kunnen toevoegen. Dit wordt veelal op automatische wijze toegepast (zie figuur 3). Zoekmachines kunnen hier vervolgens gebruik van maken, bijvoorbeeld door zogenaamde 'rich snippets', waarin dergelijke gestructureerde informatie wordt getoond in het scherm met zoekresultaten.

Denk aan resultaten van LinkedIn, waarbij de huidige baan en locatie van een persoon worden getoond, aan recensiescores voor een bepaald product of recept, maar denk ook aan het weergeven van de tijden en locaties waarop een bepaalde film wordt vertoond.<sup>9</sup>

Zoekgedrag van de gebruikers is een van de ontwikkelingen op weg naar web 3.0. Waar we in eerste instantie gewend waren aanzielijke aantallen webpagina's uit de resultaatpagina's van zoekmachines te bekijken, wordt tegenwoordig doorgaans slechts de eerste pagina bekeken. Tien jaar geleden keek minimaal 58 procent van de gebruikers alleen naar de eerste resultaatpagina; in 2005 was dit percentage zelfs opgelopen tot meer dan 80 procent [3,4,6,7]. Ook zien we in het zoekgedrag

Figuur 3. Voorbeeld RDFa-code met Dublin Core-metadata

```

<div about="myWebApp">
  <!--1. The Essentials about myWebApp-->
  <div about="http://example.com/mywebapp" typeof="am:WebApplication">
    <p>myWebApp, created by <span rel="dctypes:creator" resource="http://example.com/AOELtd">AOE Ltd</span>, is a [...]<br/>The application is provided to you and maintained by <span rel="dctypes:publisher" resource="http://example.com/AOELtd">AOE Ltd</span>.</p>
    <p>The myWebApp is powered by the<br/><a rel="am:uses" resource="http://src.example.com/mywebapp/software" href="#myWebAppSoftware">myWebApp software</a> described below, last updated on <span property="dctypes:modified" datatype="xsd:date">2010-01-01</span>.</p>
    <!--2. Components-->
    <p>The main components of myWebApp are:</p>
    <ul style="list-style-type: none;">
      <li>1.1 Components</li>
      <li>1.2 Dependencies</li>
      <li>In myWebApp, the following software and services are used:</li>
      <ul style="list-style-type: none;">
        <li>2.1 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/Apache" href="#Apache">Apache</span> typeof="doap:Project">
              <span property="doap:name" href="http://httpd.apache.org/">Apache</span>
              <span rel="doap:homepage" href="http://httpd.apache.org/">http://httpd.apache.org/</span>
            </li>
          </ul>
        <li>2.2 am:depends</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/RedlandRaptor" href="#RedlandRaptor">Redland's Raptor</span> typeof="doap:Project">
              <span property="doap:name" href="http://librdf.org/raptor/">Raptor</span>
              <span rel="doap:homepage" href="http://librdf.org/raptor/">http://librdf.org/raptor/</span>, release version <span property="doap:release">1.4.20</span> <span property="doap:Version">1.4.20</span>
              <span property="doap:revision" href="http://librdf.org/raptor/">1.4.20</span> <span property="doap:created" href="2009-11-28">2009-11-28</span>
            </li>
          </ul>
        <li>2.3 am:requires</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/Java" href="#Java">Java</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/">Java</span>
              <span rel="doap:homepage" href="http://java.sun.com/">http://java.sun.com/</span>
            </li>
          </ul>
        <li>2.4 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaScript" href="#JavaScript">JavaScript</span> typeof="doap:Project">
              <span property="doap:name" href="http://www.mozilla.org/projects/javascript/">JavaScript</span>
              <span rel="doap:homepage" href="http://www.mozilla.org/projects/javascript/">http://www.mozilla.org/projects/javascript/</span>
            </li>
          </ul>
        <li>2.5 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/MySQL" href="#MySQL">MySQL</span> typeof="doap:Project">
              <span property="doap:name" href="http://www.mysql.com/">MySQL</span>
              <span rel="doap:homepage" href="http://www.mysql.com/">http://www.mysql.com/</span>
            </li>
          </ul>
        <li>2.6 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/PHP" href="#PHP">PHP</span> typeof="doap:Project">
              <span property="doap:name" href="http://php.net/">PHP</span>
              <span rel="doap:homepage" href="http://php.net/">http://php.net/</span>
            </li>
          </ul>
        <li>2.7 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/Python" href="#Python">Python</span> typeof="doap:Project">
              <span property="doap:name" href="http://www.python.org/">Python</span>
              <span rel="doap:homepage" href="http://www.python.org/">http://www.python.org/</span>
            </li>
          </ul>
        <li>2.8 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/SQLite" href="#SQLite">SQLite</span> typeof="doap:Project">
              <span property="doap:name" href="http://www.sqlite.org/">SQLite</span>
              <span rel="doap:homepage" href="http://www.sqlite.org/">http://www.sqlite.org/</span>
            </li>
          </ul>
        <li>2.9 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/Tomcat" href="#Tomcat">Tomcat</span> typeof="doap:Project">
              <span property="doap:name" href="http://tomcat.apache.org/">Tomcat</span>
              <span rel="doap:homepage" href="http://tomcat.apache.org/">http://tomcat.apache.org/</span>
            </li>
          </ul>
        <li>2.10 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServerFaces" href="#JavaServerFaces">JavaServer Faces</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServerFaces/">JavaServer Faces</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServerFaces/">http://java.sun.com/javaServerFaces/</span>
            </li>
          </ul>
        <li>2.11 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JBoss Seam" href="#JBoss Seam">JBoss Seam</span> typeof="doap:Project">
              <span property="doap:name" href="http://jboss Seam.org/">JBoss Seam</span>
              <span rel="doap:homepage" href="http://jboss Seam.org/">http://jboss Seam.org/</span>
            </li>
          </ul>
        <li>2.12 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Pages" href="#JavaServer Pages">JavaServer Pages</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Pages/">JavaServer Pages</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Pages/">http://java.sun.com/javaServer Pages/</span>
            </li>
          </ul>
        <li>2.13 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Faces" href="#JavaServer Faces">JavaServer Faces</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Faces/">JavaServer Faces</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Faces/">http://java.sun.com/javaServer Faces/</span>
            </li>
          </ul>
        <li>2.14 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Pages" href="#JavaServer Pages">JavaServer Pages</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Pages/">JavaServer Pages</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Pages/">http://java.sun.com/javaServer Pages/</span>
            </li>
          </ul>
        <li>2.15 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Faces" href="#JavaServer Faces">JavaServer Faces</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Faces/">JavaServer Faces</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Faces/">http://java.sun.com/javaServer Faces/</span>
            </li>
          </ul>
        <li>2.16 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Pages" href="#JavaServer Pages">JavaServer Pages</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Pages/">JavaServer Pages</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Pages/">http://java.sun.com/javaServer Pages/</span>
            </li>
          </ul>
        <li>2.17 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Faces" href="#JavaServer Faces">JavaServer Faces</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Faces/">JavaServer Faces</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Faces/">http://java.sun.com/javaServer Faces/</span>
            </li>
          </ul>
        <li>2.18 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Pages" href="#JavaServer Pages">JavaServer Pages</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Pages/">JavaServer Pages</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Pages/">http://java.sun.com/javaServer Pages/</span>
            </li>
          </ul>
        <li>2.19 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Faces" href="#JavaServer Faces">JavaServer Faces</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Faces/">JavaServer Faces</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Faces/">http://java.sun.com/javaServer Faces/</span>
            </li>
          </ul>
        <li>2.20 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Pages" href="#JavaServer Pages">JavaServer Pages</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Pages/">JavaServer Pages</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Pages/">http://java.sun.com/javaServer Pages/</span>
            </li>
          </ul>
        <li>2.21 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Faces" href="#JavaServer Faces">JavaServer Faces</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Faces/">JavaServer Faces</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Faces/">http://java.sun.com/javaServer Faces/</span>
            </li>
          </ul>
        <li>2.22 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Pages" href="#JavaServer Pages">JavaServer Pages</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Pages/">JavaServer Pages</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Pages/">http://java.sun.com/javaServer Pages/</span>
            </li>
          </ul>
        <li>2.23 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Faces" href="#JavaServer Faces">JavaServer Faces</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Faces/">JavaServer Faces</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Faces/">http://java.sun.com/javaServer Faces/</span>
            </li>
          </ul>
        <li>2.24 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Pages" href="#JavaServer Pages">JavaServer Pages</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Pages/">JavaServer Pages</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Pages/">http://java.sun.com/javaServer Pages/</span>
            </li>
          </ul>
        <li>2.25 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Faces" href="#JavaServer Faces">JavaServer Faces</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Faces/">JavaServer Faces</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Faces/">http://java.sun.com/javaServer Faces/</span>
            </li>
          </ul>
        <li>2.26 am:dependsOn</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Pages" href="#JavaServer Pages">JavaServer Pages</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Pages/">JavaServer Pages</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Pages/">http://java.sun.com/javaServer Pages/</span>
            </li>
          </ul>
        <li>2.27 am:uses</li>
          <ul style="list-style-type: none;">
            <li><span about="http://src.example.com/mywebapp/dependencies/JavaServer Faces" href="#JavaServer Faces">JavaServer Faces</span> typeof="doap:Project">
              <span property="doap:name" href="http://java.sun.com/javaServer Faces/">JavaServer Faces</span>
              <span rel="doap:homepage" href="http://java.sun.com/javaServer Faces/">http://java.sun.com/javaServer Faces/</span>
            </li>
          </ul>
        <li>2.28 am:dependsOn</li>
          <ul
```

## 'Alle Facebook-acties, waaronder de bekende Like-knop, zijn geschreven in RDFa'

van intelligente manieren van interactie of de presentatie van zoekresultaten, alle voorbeelden tot dusver laten zien dat semantische zoekmachines de zoekervaring van de gebruiker nu al aan het veranderen zijn.

### Recente ontwikkelingen

Eerder kwamen Microformats en RDFa al aan de orde en in het bijzonder hoe deze toegepast worden om de resultaatpagina's van zoekmachines 'intelligenter' te maken. De wildgroei aan verschillende dialecten en de verschillende mate van ondersteuning bij verschillende zoekmachines deden Bing, Google en Yahoo! recentelijk besluiten een gezamenlijk voorstel voor een vocabulaire om entiteiten te beschrijven naar buiten te brengen, waarmee gestructureerde informatie opgenomen kan worden op webpagina's. Op de bijbehorende website, schema.org<sup>10</sup> genaamd, staan alle categorieën die ondersteund worden, alsmede documentatie voor ontwikkelaars en webmasters. Onder de categorieën bevinden zich entiteitstypen als personen, films, producten, locaties en meer. Ook wordt ondersteuning gegeven voor semantische webtalen als OWL en RDF, met als doel schema.org-annotaties deel uit te laten maken van LOD.

Een andere recente ontwikkeling is Open Graph van Facebook. We kennen allemaal de Facebook 'Like'-knoppen op websites en dit is de volgende versie daarvan. Facebook erkende een veelgehoorde klacht van gebruikers dat een simpele 'Like' in veel gevallen misstond. Gebruikers wilden bijvoorbeeld alleen maar aangeven dat ze iets gezien of gehoord hadden, zonder een impliciet positief waardoordeel te geven. En, hoewel het in eerste instantie al mogelijk was om

een zogenaamd objecttype (zoals 'acteur', 'stad', et cetera) mee te geven aan datgene dat 'geliked' kon worden, werd er maar sporadisch gebruik van gemaakt, voornamelijk vanwege de erg beperkte lijst met ondersteunde types. Open Graph heeft als doel beide bezwaren aan te pakken. Allereerst kan men, naast de 'Like'-actie, eigen acties definiëren, zoals 'Listened', 'Watched', et cetera. Ten tweede kunnen nu ook entiteiten aangeduid worden (zogenaamde 'objects'), waarop de acties betrekking hebben. Maar wat heeft dit te maken met het semantische web? En met zoeken? Het aanduiden van objecten en mogelijke acties die men daarop kan ondernemen, heeft wel degelijk het karakter van het semantische web. Daarnaast zijn alle Facebook-acties, waaronder de bekende 'Like'-knop, geschreven in RDFa.

En, wellicht nog belangrijker, het ondersteunt het sociale karakter van internet en brengt dit naar het semantische web. Met dergelijke annotaties kunnen namelijk gepersonaliseerde suggesties worden gedaan, bijvoorbeeld aan de hand van acties van vrienden of van personen die erg lijken op een gebruiker.

Tot slot is het open source contentmanagementsysteem (CMS) Drupal een vermelding waard. Dit van oorsprong Belgische project wordt over de gehele wereld als back-end voor allerlei websites gebruikt. In de laatste versie (Drupal 7) wordt er out-of-the-box ondersteuning geboden voor RDFa.<sup>11</sup> Dit betekent dat alle content op een website met Drupal als CMS standaard ontsloten kan worden met RDFa. Denk aan personen op een forum, producten in een webshop, of bijvoorbeeld recensies. Door het open karakter van Drupal zijn objecten en types vrijelijk te definiëren en kunnen deze dus met minimale inspanning aangeboden worden als machineleesbare informatie en deel gaan uitmaken van het semantische web.

### Tot slot

Met het steeds breder beschikbaar komen van machineleesbare informatie, methoden om deze informatie te consumeren en produceren, en met het steeds slimmer worden van automatische methoden om entiteiten te herkennen in tekst, komt het semantische web steeds dichterbij. Met behulp van moderne, semantische zoek-

machinetechnologieën zijn inmiddels de eerste stappen – van hoge kwaliteit en op grote schaal – gezet om deze informatie in te zetten om gebruikers sneller te laten vinden wat ze zoeken. Of dit alles daadwerkelijk web 3.0 gaat voorstellen zullen we pas over enige tijd weten, maar voor nu ziet de zoekmachine van de toekomst er in ieder geval al slimmer, interactiever en sociaal uit.

### Noten

- [1] Op websites zoals delicious.com en www.diigo.com.
- [2] Overigens vindt het omgekeerde ook plaats, dat wil zeggen het inzetten van zoekalgoritmes zoals Google die gebruikt om wijs te worden uit de gigantische hoeveelheid aan data die op het semantische web in de vorm van gelinkte data te vinden is. Dat is echter niet het onderwerp van dit artikel.
- [3] wikipedia-miner.cms.waikato.ac.nz.
- [4] Zie www.google.com/support/websearch/bin/answer.py?answer=35890#synonym.
- [5] Zie bijvoorbeeld www.wolframalpha.com/input/?i=tortoise+vs+hare.
- [6] en.wikipedia.org/wiki/Powerset\_(company).
- [7] Zie bijvoorbeeld de 'related searches' aan de linkerhand van search.yahoo.com/search?p=picasso.
- [8] Zie bijvoorbeeld search.yahoo.com/search?p=amsterdam.
- [9] www.google.com/movies.
- [10] schema.org.
- [11] semantic-drupal.com.

### Referenties

- [1] Tim Berners-Lee, James Hendler, Ora Lassila: The Semantic Web. *Scientific American Magazine*, 2001.
- [2] Ricardo Baeza-Yates, Andrei Broder, Yoelle Maarek, Prabhakar Raghavan: The new frontiers of Web search: going beyond the 10 blue links. SIGIR 2010 Industry Track (2010).
- [3] Bernard J. Jansen, Amanda Spink, Tefko Saracevic: Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management* 36(2): 207-227 (2000).
- [4] Evangelos P. Markatos: On caching search engine query results. *Computer Communications* 24(2): 137-143 (2001).
- [5] Edgar Meij, Marc Bron, Laura Hollink, Bouke Huurnink, Maarten de Rijke: Mapping queries to the Linking Open Data cloud: A case study using DBpedia. *Web Semantics: Science, Services and Agents on the World Wide Web* (2011).
- [6] Craig Silverstein, Monika Rauch Henzinger, Hannes Marais, Michael Moricz: Analysis of a Very Large Web Search Engine Query Log. *SIGIR Forum* 33(1): 6-12 (1999).
- [7] Amanda Spink, Bernard J. Jansen, Chris Blakely, Sherry Kosman: A study of results overlap and uniqueness among major Web search engines. *Information Processing and Management* 42(5): 1379-1391 (2006).