

INFORMATIE ZOEKEN OP HET WEB

Directories, zoekmachines, folksonomieën... en ontologieën

Céline VAN DAMME

Assistent/Onderzoeker Vakgroep MOSI, Vrije Universiteit Brussel

Het artikel is opgesteld naar aanleiding van een conferentie gegeven door de auteur in het kader van *Inforum 2007, De 2.0 I&D professional: een alternatieve look op de nieuwe technologieën*, georganiseerd door de Belgische Vereniging voor Documentatie op 26 april 2007 te Brussel.

Article rédigé suite à la conférence donnée par l'auteur dans le cadre de l'*Inforum 2007, Le documentaliste 2.0 : un second regard sur les nouvelles technologies*, organisé par l'Association Belge de Documentation, le 26 avril 2007, à Bruxelles.

▪ Tegenwoordig worden we op het World Wide Web geconfronteerd met een overload aan informatie. De moeilijkheid bestaat erin om de gewenste informatie op een efficiënte manier terug te vinden. Momenteel beschikken we op het web over drie mogelijke zoektechnieken: directories, zoekmachines en folksonomieën. In de nabije toekomst zullen zoekmachines worden ondersteund door ontologieën. Ontologieën zijn één van de belangrijkste technologieën die worden gebruikt voor de ontwikkeling van het semantische web. Het semantische web, dat een uitbreiding vormt van het bestaande web, heeft tot doel het web interpreteerbaar te maken voor zoekmachines. In dit artikel zullen naast de informatie-explosie op het World Wide Web, de huidige zoektechnieken worden toegelicht en het begrip ontologieën worden besproken.

▪ À l'heure actuelle, nous sommes confrontés sur le World Wide Web à une surcharge d'information. La difficulté réside dans la manière efficace de retrouver l'information souhaitée. Pour l'instant, nous disposons sur le Web de trois techniques de recherche possibles : les répertoires, les moteurs de recherche et les folksonomies. Dans un proche avenir, les moteurs de recherche se verront renforcés par des ontologies. Les ontologies sont une des technologies les plus importantes utilisées dans le développement du Web sémantique. Le Web sémantique, qui constitue une extension du Web existant, a pour but de rendre le Web interprétable pour les moteurs de recherche. Dans cet article, l'auteur introduira, outre l'explosion informationnelle sur le World Wide Web, les techniques de recherche actuelles et abordera le concept des ontologies.

De informatie-explosie op het World Wide Web

De laatste jaren is er een ware informatie-explosie gaande op het World Wide Web. Volgens een studie¹ bedroeg het aantal indexeerbare webpagina's eind januari 2005: 11.5 miljard, terwijl in 2002 dit er nog maar 1,4 miljard² waren. Onder indexeerbare webpagina's worden webpagina's verstaan die kunnen worden geïndexeerd door een zoekmachine. Dit is mogelijk omwille van hun statische karakter: de webpagina is ten allen tijde beschikbaar op de webserver. Dit is echter niet het geval met dynamische webpagina's: deze worden gecreëerd op het ogenblik dat ze worden opgevraagd door de gebruiker, bijvoorbeeld het consulteren van een online woordenboek. Een door de gebruiker opgegeven zoekterm wordt opgezocht in een databank en vervolgens worden de zoekresultaten weergegeven in een nieuwe webpagina. Wanneer de gebruiker naar een andere webpagina verder surft of zijn browser sluit, verdwijnt deze webpagina. Hierdoor kan deze pagina niet worden geïndexeerd door een klassieke zoekmachine. De omvang van deze dynamische webpagina's, die samen het dark web vormen, wordt geschat 500³

keer groter te zijn dan het visible web, het web bestaande uit de verzameling indexeerbare webpagina's.

Deze informatie-explosie is gedeeltelijk toe te schrijven aan het feit dat de technische drempel voor het creëren van webpagina's grotendeels is verdwenen. Iedereen kan *content* publiceren op het web. Dit is mogelijk gemaakt door de intrede van nieuwe tools, de zogenaamde social software tools die web 2.0 vormen. Omwille van de eenvoudige en gebruiksvriendelijke interface is het informatiecreatieproces sterk vereenvoudigd. Kennis van een markup taal zoals HTML, is niet nodig voor het gebruiken van deze tools. Enkele voorbeelden zijn wikis en weblogs. Een weblog, of kortweg een blog, kan het best worden omschreven als een persoonlijke website onder de vorm van een dagboek. Elke post in het dagboek is een webpagina met een unieke URL. Volgens de statistieken van Technorati, een zoekmachine voor weblogs, waren er in april 2007 meer dan 70 miljoen blogs geregistreerd bij Technorati⁴. Webpagina's kunnen door een groep gebruikers op een eenvoudige en snelle manier gecreëerd en geëditeerd worden via een wiki. Wikipedia is momenteel het meest gekende voorbeeld van een dergelijke tool. Het bevat meer dan

1.700.000 artikelen in de Engelstalige Wikipedia⁵.

Uit bovenstaande cijfers kan worden geconcludeerd dat er heel wat informatie beschikbaar is op het web. Hierdoor rijst de vraag of het nog mogelijk is om gewenste informatie op een efficiënte manier terug te vinden.

Huidige zoektechnieken op het World Wide Web

Volgens ons inzien zijn er momenteel drie zoektechnieken beschikbaar op het web: directories, zoekmachines en folksonomieën. In onderstaande paragrafen zal respectievelijk dieper worden ingegaan op deze technieken.

Directories

Directories of folders zijn een techniek voor het opdelen of classificeren van informatie volgens hiërarchische categorieën: bijv. folders in een e-mailbox, documenten op een pc. Na de commerciële lancering van het World Wide Web heeft een professionele gebruikerscommissie van Yahoo zich ontfemd over het indelen van de websites in een directory of taxonomie: de Yahoo directories. Deze taxonomie wordt nog altijd geüpdate door deze commissie⁶. Door te navigeren in de directory kan de surfer zijn gewenste webpagina terugvinden. Ondertussen bestaan er heel wat andere projecten die hetzelfde doel beogen zoals het Open Directory Project⁷. Gezien de enorme toename in het aantal webpagina's, stellen we ons echter de vraag of dergelijke directories in staat zijn om alle webpagina's te classificeren.

Zoekmachines

Zoekmachines zijn een andere techniek voor het zoeken naar een specifieke internetpagina of informatie op het web. Zoekmachines bezitten kleine robots, ook wel webcrawlers of spiders genoemd, die constant alle internetpagina's kopiëren om ze dan vervolgens te indexeren en op te slaan in een databank. Telkens dat ze een hyperlink tegenkomen verwijzend naar een ander webpagina wordt de hyperlink toegevoegd aan het lijstje van te doorzoeken webpagina's. Dit is de reden waarom het soms een poosje duurt vooraleer een nieuwe webpagina wordt opgenomen in de resultatenlijst van een zoekmachine. Afhankelijk van de soort zoekmachine wordt de inhoud volledig ofwel gedeeltelijk geïndexeerd. De meeste zoekmachines doorbladeren vaak enkel het visible of surface web. Het visible web wordt gevormd door de hierboven reeds besproken indexeerbare webpagina's. Er zijn echter een aantal zoekmachines die louter het dark web indexeren zoals BrightPlanet⁸. Google indexeert hoofdzakelijk de statische webpagina's, hoewel

er ook een aantal dynamische webpagina's worden in opgenomen, bijvoorbeeld de dynamische webpagina's van het online Merriam Webster woordenboek. Bij het ingeven van de zoektermen "Merriam Webster tree" in Google, wordt de dynamische webpagina in de resultatenlijst afgebeeld. Wanneer dezelfde test wordt gedaan met "van Dale boom", dan komt de dynamische webpagina van het van Dale woordenboek helemaal niet voor in de resultatenlijst. Dit betekent dat de dynamische webpagina's van het van Dale woordenboek niet worden geïndexeerd door Google. Met een zoekmachine wordt dus niet alle op het web voorhanden zijnde informatie doorzocht.

Daarenboven doet er zich een ander probleem voor: semantiek. Welke betekenis hangt een gebruiker aan de woorden van de zoekopdracht en welke relaties bestaan er tussen de elementen? Dit is een probleem dat zich niet direct stelt op het niveau van de directories. Door een "drill-down" te doen kan direct worden afgeleid welke subfolders onderdeel uitmaken van een folder en kan de surfer zijn keuze direct corrigeren door een andere folder te selecteren. Bij een zoekmachine is dit niet echt mogelijk. Stel de gebruiker wil informatie terugvinden over de "openingsuren bibliotheek Brugge niet filialen". Bij het ingeven van een dergelijke zoekopdracht wordt helemaal niet de gewenste informatie als resultaat weergegeven. De zoekmachine is niet in staat te achterhalen welke betekenis de gebruiker hangt aan een bepaald woord en welke de relatie is tussen de ingegeven zoekwoorden. Het probleem inzake semantiek situeert zich niet alleen op het niveau van de zoekopdrachten van de gebruikers, maar ook bij het indexeren van de webpagina's. De web crawler indexeert de web pagina's als *plain* of *platte* tekst zonder de context van de woorden in acht te nemen. Een machine kan immers de context waarin de woorden worden gebruikt niet interpreteren zoals mensen dit doen. Wanneer een zoekmachine een zoekopdracht ontvangt van de gebruiker, wordt nagegaan met welke webpagina's er een match bestaat. Hiervoor wordt een vergelijking gemaakt met de indexen opgeslagen in de databank. Hoe groter de match, hoe meer vooraan de webpagina zal verschijnen in de lijst. Google maakt in tegenstelling tot de andere zoekmachines eveneens gebruik van het page ranking algoritme: hoe meer webpagina's linken naar een specifieke website, hoe hoger de ranking van deze website⁹.

Het probleem inzake semantiek of betekenis kan worden opgelost door metadata of data over data mee te geven in de webpagina. Wij omschrijven dit als *interne metadata*. XML, een mark up taal, zoals HTML kan hiervoor worden gebruikt. XML is een uitbreiding op HTML die toelaat door middel van tags informatie mee te geven. De informatie

wordt geschreven tussen een begin en eind tag. Hieronder volgt een voorbeeldje.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<playlist name="mijn_lijst">
<song>
<title>Sad Day</title>
<artist>Rolling Stones </artist>
</song>
<song>
<title>Let it be</title>
<artist>The Beatles</artist>
</song>
</playlist>
```

Rolling Stones en The Beatles behoren beiden tot de categorie <artist> en Sad Day en Let it be vallen onder <title>. Het enige probleem dat zich hierbij stelt is dat er geen betekenis hangt aan deze tags. Zo heeft de tag <title> hier de betekenis van titel van een liedje, maar op een andere webpagina kan dit gaan om bijvoorbeeld de titel van een boek of film. Tags zijn met andere woorden context gevoelig.

Folksonomieën

De laatste jaren vindt een nieuwe vorm van categorisatie meer en meer opgang op het web: tagging en de resulterende folksonomie. Internetgebruikers kunnen alle informatie of objecten die identificeerbaar zijn met een unieke URL categoriseren met hun eigen sleutelwoorden of tags. Hierdoor vinden de gebruikers hun informatie veel gemakkelijker terug. De cognitieve overhead is hierbij zeer laag aangezien ze alle woorden mogen gebruiken die in hun hoofd opkomen. Het aggregeren of samennemen van al deze tags leidt tot een vlakke taxonomie of folksonomie. Het was Thomas vander wal die dit begrip introduceerde door de woorden folk en taxonomie samen te smelten¹⁰. Een taxonomie bestaat uit een verzameling woorden, een controlled vocabulary, waartussen hiërarchische relaties zijn gedefinieerd zoals een directory op een intranet, de Yahoo directories of Dewey Decimal Classification. Deze woorden zijn gekozen door een groep experts. Gebruikers die informatie moeten classificeren of opzoeken zijn gebonden aan deze taxonomie. Dit contrasteert natuurlijk met een folksonomie waar de gebruikers zelf de woorden kiezen. Een folksonomie is met andere woorden een taxonomie zonder hiërarchische relaties gecreëerd door de gebruikers. Er zijn ondertussen verschillende websites die een folksonomie-tagging functionaliteit hebben geïmplementeerd: Deli-

cious¹¹, Flickr¹², Technorati¹³ die respectievelijk het categoriseren van bookmarks, foto's en blogs gratis aanbieden.

Tags hebben niet alleen een meerwaarde voor de persoon in kwestie, maar ook voor alle andere internetgebruikers. De 3 partijen betrokken tijdens het tag-proces: tags, objecten en actoren zijn allen gelinkt aan elkaar zoals gevisualiseerd in onderstaande figuur en bijna altijd publiek beschikbaar.

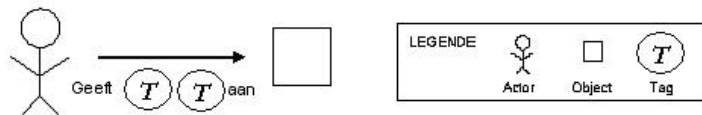


Fig. 1: Actor tagt object

Hierdoor krijgt dit proces een sociaal aspect. Internetgebruikers met dezelfde interesses kunnen elkaar gemakkelijk terugvinden op basis van de tags en objecten. Tags en objecten zijn immers een reflectie van de interesses en kennis van een actor. Vandaar dat een gemeenschappelijke tag of object een signaal kan zijn van overlappende interessevelden. Dit geeft aanleiding tot een nieuwe manier van zoeken naar informatie: sociale navigatie¹⁴.

Objecten kunnen worden teruggevonden door een "tag zoekmachine" te gebruiken of een tag aan te klikken in de tag cloud. Deze tag cloud is een soort wolk waarbij de tekstgrootte van de tags een indicatie is van hun populariteit¹⁵. In onderstaande figuur behoren de tags: wedding, party, friends, family en Japan tot de meest populaire tags in de tag cloud.

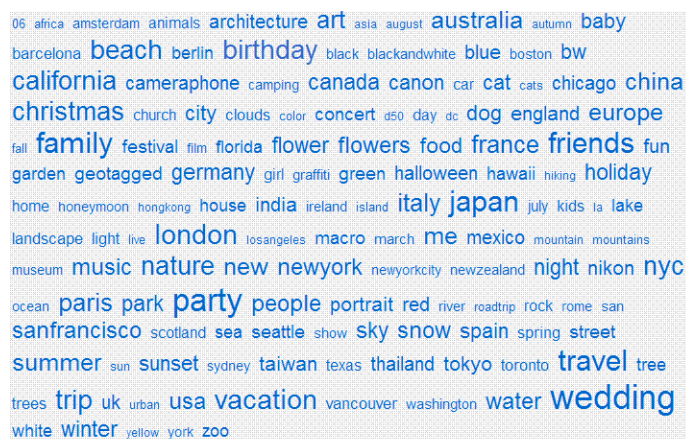


Fig. 2: Tag Cloud Flickr website (op 5 April 2007)

Bij het klikken op een tag wordt een overzicht gegeven van alle objecten die door deze tags zijn omschreven. Op sommige websites is het mogelijk deze zoekfunctie te verfijnen door op meerde-

re tags te gaan zoeken. Bij elk object staan meestal de actoren vermeld waardoor sociale navigatie mogelijk wordt. In het geval van een *broad* folksonomie, zullen er meerdere namen staan. Een *broad* folksonomie¹⁶ wordt namelijk gegeneerd in het geval de objecten niet gecreëerd zijn door de actoren vb. bookmarks en boeken. Deze objecten zullen door meerdere personen worden getagd. De tegenhanger van een *broad* folksonomie is een *narrow* folksonomie¹⁶. In dit geval is er vaak slechts één actor, namelijk de creator van het object of de content. Het zijn objecten die vaak enkel een meerwaarde hebben voor de actor zelf zoals bijvoorbeeld foto's.

Daarnaast bieden sommige van deze websites sociale feedback¹⁵ aan. Deze feedback kan zich in vele vormen uiten. In het geval van een *broad* folksonomy kan aan de actor een overzicht worden gegeven van de meest populaire tags voor dit object op voorwaarde dat reeds andere actoren dit hebben getagd. De feedback kan ook worden gegeven op basis van tags: vb. tag x wordt vaak gebruikt in combinatie met tag y.

Er zijn echter een aantal nadelen verbonden aan de tags in een folksonomie zoals: tik- en schrijffouten, meervouden, synoniemen, homoniemen, idiosyncratisch of egocentrisch taggen¹⁷. Een aantal van deze problemen kunnen volgens ons inzien op een eenvoudige manier worden opgelost door stemming algoritmes, die een woord herleiden naar zijn stam, en clustering. Problemen zoals homoniemen, een zelfde woord met verschillende betekenissen, wordt hiermee echter niet opgelost. Het zoeken naar informatie door middel van tags kan hierdoor leiden tot verkeerde informatie.

Delicious voorbeeld

Het meest gekende voorbeeld is ongetwijfeld Delicious, de social bookmark manager. Deze bookmark manager laat toe bookmarks of favoriete webpagina's te verzamelen op een persoonlijke account. Deze account wordt bewaard op de server van Delicious. Dit biedt het voordeel dat de bookmarks op elke pc ter wereld met een internetconnectie kunnen worden geconsulteerd.

Bookmarks worden toegevoegd door te surfen naar de delicious website en vervolgens in te loggen. Dit proces kan echter worden vereenvoudigd door het installeren van een kleine applicatie in



Fig.3: Deli.cio.us knoppen in browser

de browser. Bovenaan de browser zullen onderstaande knoppen zichtbaar worden.

Hierdoor kan op een gebruiksvriendelijke manier elke interessante webpagina worden toegevoegd aan de bookmark collectie. Door een druk op de "tag" knop bovenaan de browser, kan de webpagina die op dat ogenblik in de browser staat, worden toegevoegd aan de collectie. Onderstaand scherm verschijnt waarbij een aantal velden reeds automatisch zijn ingevuld zoals de URL en titel van de webpagina.

Het rest de gebruiker om zijn tags in te geven en

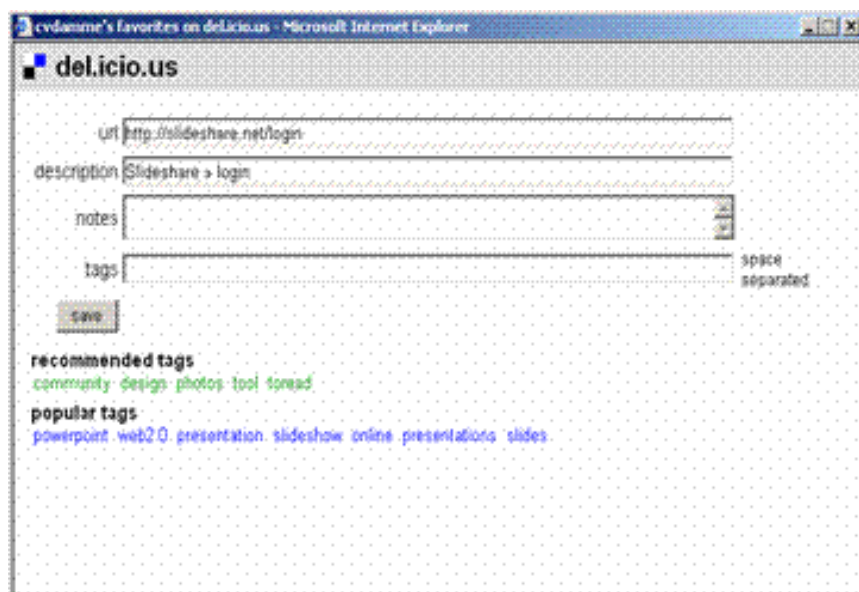


Fig. 4: Toevoegen van bookmark via "Tag" knop in browser

indien hij dat wenst kan hij commentaar toevoegen. Het taggen van een bookmark wordt ondersteund door sociale feedback. Er wordt namelijk nagegaan of deze bookmark reeds is toegevoegd door andere gebruikers van de Delicious website. Indien dit het geval is, worden de meest populaire tags voor dit object gegeven. In bovenstaande figuur zijn dit: powerpoint, web2.0, presentation, slideshow, online, presentations, slides. Daarnaast worden tags aanbevolen op basis van de doorsnede van de tags van de gebruiker en populaire tags gebruikt voor dit object: community, design, photos, tool en toread. Na het klikken op de bewaarknop verdwijnt dit scherm en kan de gebruiker opnieuw verder surfen.

Ten allen tijde kan de overzichtspagina van de bookmarks worden bekeken. Bij elke bookmark staat vermeld hoeveel andere Delicious leden deze in hun collectie bezitten. Verder kan de De-

licious user beslissen om zijn tags zelf te gaan clusteren in zogenaamde bundels. Hierdoor kan het terugvinden van informatie worden bevorderd en problemen zoals enkelvoud, meervoud, synoniemen enz. worden opgelost.

Door te klikken op "other people" bij een bookmark wordt een overzicht genereerd van actoren die dezelfde bookmark in hun collectie hebben.

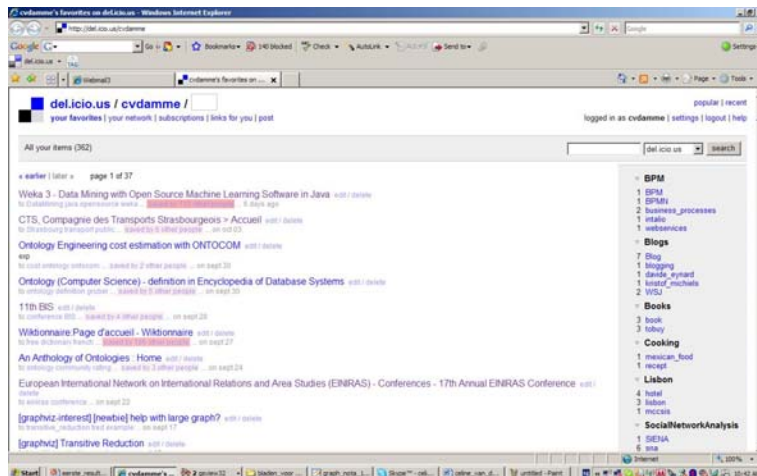


Fig. 5: Overzicht van de bookmarks van de Del.icio.us account
<http://del.icio.us/cvdammeH>

Vervolgens kan de account van de persoon in kwestie worden geconsulteerd en worden overgegaan tot sociale navigatie.

Ook bestaat de mogelijkheid om een actor toe te voegen tot een persoonlijk netwerk. Op deze manier wordt op de hoogte gebleven van de nieuwe bookmarks van deze personen.

Folksonomie versus Taxonomie

Na het toelichten van een folksonomie en het uitdiepen van een voorbeeld rijst de vraag wat de concrete verschillen zijn tussen een taxonomie en een folksonomie. In onderstaande tabel bespreken we heel kort de grote verschillen die volgens ons tussen beide technieken bestaan.

Uit deze tabel kunnen we concluderen dat zowel taxonomieën als folksonomieën sterktes en zwaktes hebben. Het combineren van beide technieken zou wel eens waardevolle synergieën kunnen genereren.

Ontologieën

De laatste jaren wordt er volop gewerkt aan een uitbreiding van het bestaande web, het semantische web. De bedoeling van dit nieuwe web is om alle informatie interpreteerbaar te maken voor machines¹⁸. Op die manier zullen de semantische problemen bij zoekmachines worden opgelost. Om te komen tot een semantisch web is een architectuur uitgedacht, waarvan ontologieën het belangrijkste deel van uitmaken¹⁹.

Het woord ontologie vindt zijn oorsprong in de filosofie. De term is afgeleid van het Grieks en betekent de facto de zijnsleer. In de jaren 90 is het begrip geïntroduceerd binnen het domein van de computerwetenschappen en is er een nieuwe betekenis toegekend aan dit concept¹⁹. Het wordt door Gruber omschreven als "an explicit specification of a conceptualization"²⁰. Een ontologie tracht

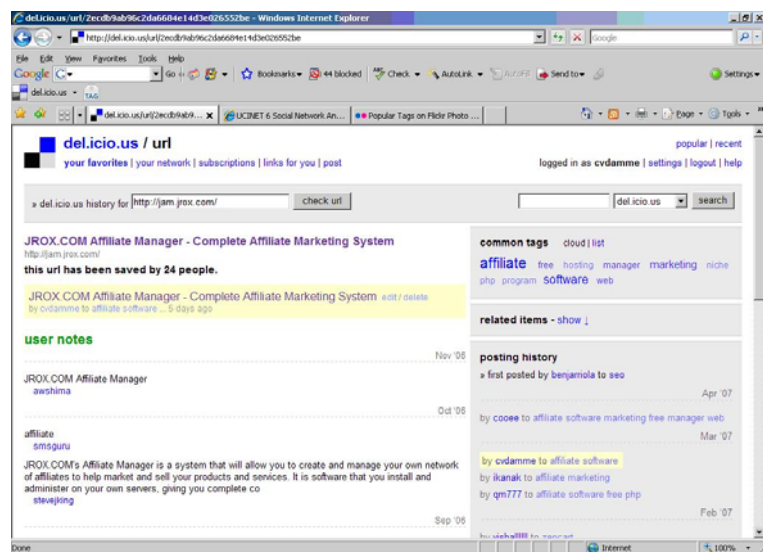


Fig. 6: Sociale Navigatie Del.icio.us

met andere woorden de natuurlijke taal van een domein (concepten, instanties en relaties) te expliciteren.

Ontologieën kunnen worden opgedeeld in twee categorieën: de lightweight ontologieën en de heavyweight ontologieën. Eerstgenoemde categorie vormt een uitbreiding op de taxonomieën waarin er meer relaties worden geëxpliciteerd. Heavyweight ontologieën zijn lightweight ontologieën aangevuld met axioma's en restricties. In een pizza ontologie kan bijvoorbeeld de volgende regel of restrictie worden ingevoerd: een margha-

rita pizza heeft tomaten en kaas als topic, maar geen vlees²¹.

verschillende ontologietalen die toelaten een ontologie formeel uit te drukken zoals OWL, RDF, shoe, oil, daml + oil representation, ontolingua

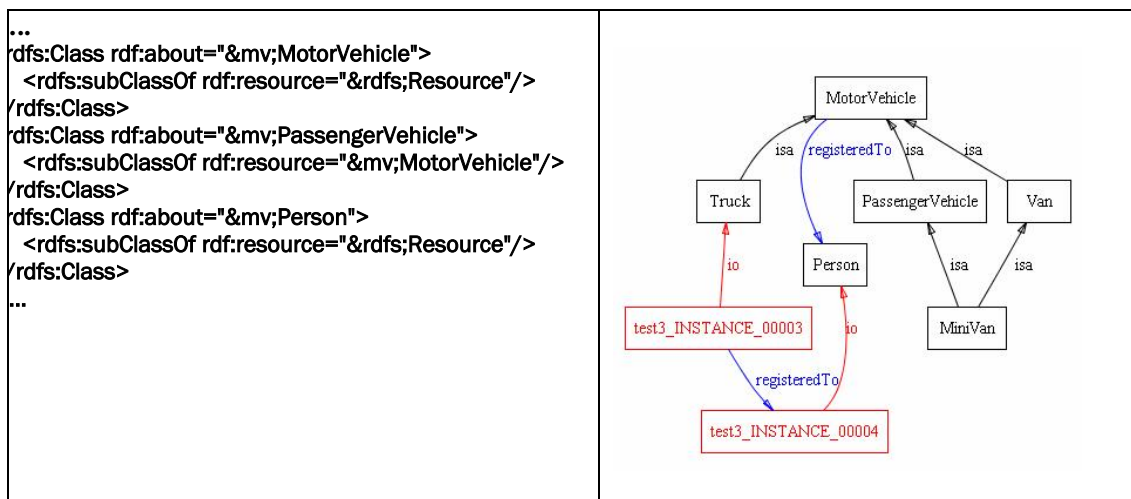
Tab. 1: Verschil tussen folksonomieën en taxonomieën

Folksonomieën	Taxonomieën
<ul style="list-style-type: none"> • Zelfgekozen sleutelwoorden: iedereen beschikt over de vrijheid om eender welk woord in te voeren, dit hoeft door niemand te worden goedgekeurd. • Lage kost: het zijn de gebruikers die deze externe vorm van metadata genereren • Lage cognitieve overhead: elk woord die in de hoofden van de gebruikers opkomt kan worden gebruikt, • Geen relaties: alles is vlak • Sociale navigatie en feedback • Een object is niet specifiek gelocaliseerd op één plaats 	<ul style="list-style-type: none"> • Controlled vocabulary: vaste woordenlijst opgesteld en onderhouden door een groep experts • Hoge kost: het opstellen en onderhouden is zeer tijdsintensief en bijgevolg zeer duur • Hoge cognitieve overhead: gebruikers moeten de woorden kiezen die deel uitmaken van de controlled vocabulary, ze kunnen met andere woorden niet hun eigen terminologielijst hanteren • Hiërarchische relaties • Hiërarchische navigatie en geen sociale feedback • Een object kan slechts op één plaats voorkomen in taxonomie, tenzij kopieën worden gemaakt van het object. Dit is mogelijk bij een directory op een pc

Volgens Uschold en Gruninger²² kunnen we drie doelen detecteren waarvoor ontologieën worden gebruikt: (1) communicatie tussen personen, (2) interoperabiliteit tussen machines en (3) systeemontwikkeling. Het doel van de ontologie zal bepalen hoe formeel een ontologie wordt uitgedrukt. Zo zal een ontologie voor de communicatie tussen mensen eerder informeel worden weergegeven, de zogenaamde vocabularia. Terwijl de interoperabiliteit tussen machines veel meer nood heeft aan een formelere vorm²². Er bestaan

etc. OWL en RDF zijn door W3C, het world wide web consortium, voorgesteld als standaard ontologietalen²³. Deze ontologietalen zijn meestal zeer complex en moeilijk om te coderen vandaar dat er vaak wordt gebruik gemaakt van tools, die een eenvoudige interface hebben en zorgen voor een vertaling naar een ontologietaal. Zo laat de open source ontologie tool Protégé²⁴, ontwikkeld aan de Stanford Universiteit, toe een ontologie te vertalen naar RDF of OWL. De elementen (concepten, instanties, relaties, restricties) van een

Tab. 2: Een ontologie in Protégé over motorvoertuigen²⁴



ontologie worden ingegeven via verschillende invulschermen. Daarna kan de gebruiker vragen om deze ontologie te exporteren naar OWL of RDF. De ontologie kan eveneens gevisualiseerd worden via de plugin Ontoviz. In onderstaande tabel wordt er een voorbeeld gegeven van een ontologie gemaakt in Protégé. De rechterkolom bevat de visualisatie en de linkerkolom de formele kant van de ontologie uitgedrukt in RDF.

Ontologieën kunnen in het kader van "informatie retrieval" op 2 manieren worden ingezet: 1) zoekopdracht van de gebruiker kan worden uitgebreid met synoniemen, gerelateerde woorden of andere concepten waartussen een relatie bestaat en 2) de documenten kunnen worden geïndexeerd op basis van de ontologie²⁵.

De ontwikkeling van deze ontologieën is een arbeidsintensief proces en wordt uitgevoerd door experts. Gebruikers worden echter niet actief betrokken bij de ontwikkeling. Ze worden afgeschrikt door de technische en ongebruiksvriendelijke kant van de ontologietaal. Deze kunstmatige taal wijkt ver af van de natuurlijke taal. Daarenboven duurt het een hele poos vooraleer nieuwe woorden, relaties etc. worden opgenomen in de ontologie. Het zijn immers de experts die ervoor zorgen dat deze worden opgenomen in de ontologie²⁶. Binnen het onderzoeksdomein van het semantische web wordt er momenteel volop onderzoek verricht om deze problemen op te lossen. Het betrekken van folksonomieën bij de creatie en onderhoud van respectievelijk nieuwe en bestaande ontologieën, is volgens ons een oplossing voor dit probleem en dit wordt eveneens geargumenteed in een aantal papers^{27, 28}.

sing voor dit probleem en dit wordt eveneens geargumenteed in een aantal papers^{27, 28}.

Conclusie

In dit artikel hebben we aangetoond dat er een enorme hoeveelheid informatie beschikbaar is op het web. De huidige technieken (directories, zoekmachines en folksonomieën) laten ons toe informatie te zoeken op het World Wide Web. De gewenste informatie kan niet altijd worden teruggevonden met de huidige zoektechnieken. Het semantische web zal ervoor zorgen dat zoekmachines betere zoekresultaten kunnen genereren. Op dit web zal namelijk alle informatie interpreteerbaar zijn voor machines. Ontologieën zijn een van de belangrijkste technologieën die hiervoor worden gebruikt. De ontwikkeling van deze ontologieën kampt met een aantal nadelen. Deze nadelen zouden volgens ons kunnen worden gereduceerd door folksonomieën te betrekken bij de ontwikkeling en onderhoud ervan.

Céline Van Damme
Vrije Universiteit Brussel
Pleinlaan 2
1050 Brussel
celine.van.damme@vub.ac.be

Mei 2007

Nota's

- 1 Gulli, A.; Signorini, A. The indexable Web is more than 11.5 billion pages. *Poster. Proceedings of the 14th international conference on World Wide Web*, 2005, pp. 902-903.
- 2 O'Neill, E.T.; Lavoie, B.F.; Bennett, R. Trends in the evolution of the public Web. *D-Lib Magazine*, April 2003, 9(4), online beschikbaar
<<http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>> (bezoekt op 5 april 2007).
- 3 <<http://www.sims.berkeley.edu/how-much-info-2003>> (bezoekt op 5 april 2007).
- 4 <<http://www.sifry.com/stateoftheliveweb/>> (bezoekt op 29 april 2007).
- 5 <http://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia> (bezoekt op 1 mei 2007).
- 6 <<http://help.yahoo.com/help/us/dir/basics/basics-03.html>> (bezoekt 28 april 2007).
- 7 <<http://dmoz.org/>> (bezoekt op 29 april 2007).
- 8 Bergmann, M.K. Deep Web: surfacing hidden value. *The Journal of electronic publishing*. 2001, 7(1). Online beschikbaar <<http://www.press.umich.edu/jep/07-01/bergman.html>>
- 9 Brin, S.; Page, L. The anatomy of a large-scale hypertextual web search engine. *Proceedings Seventh International World Wide Web Conference*, 1998, pp.107-117.
<<http://vanderwal.net/folksonomy.html>> (bezoekt op 5 april 2007).
- 10 <<http://del.icio.us/>> (bezoekt op 15 april 2007).
- 11 <<http://www.flickr.com/>> (bezoekt op 15 april 2007).
- 12 <<http://technorati.com/>> (bezoekt op 15 april 2007).
- 13 Moreville, P. *Ambient Findability*. O'Reilly, 2005, 204p.
- 14 Sinclair, J.; Cardew Hall, M. The folksonomy tag cloud: when is it useful? *Journal of Information Science*, 31 May 2007, pp. 1-18.
<http://www.personalinfocloud.com/2005/02/explaining_and_.html> (bezoekt op 20 april 2007).
- 15 Quintarelli, E. Folksonomies: power to the people. Paper presented at the *ISKO Italy-UniMIB meeting*, June 2005. Online beschikbaar <<http://www.iskoi.org/doc/folksonomies.htm>>
- 16 Berners-Lee, T.; Hendler, J.; Lassila, O. The semantic Web. *Website ScientificAmerican.Com* 2001, online beschikbaar
<http://www.scientificamerican.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>

- 19 Fensel, D. *Ontologies: a silver bullet for knowledge management and electronic commerce*. Berlin: Springer, 2004, 162p.
- 20 Gruber, T. R A translation approach to portable ontology specifications, *Knowledge Acquisition*, 1993, Vol. 5(2), pp. 199-220.
- 21 Gomez-Perez, A. ; Fernandez-Lopez, M. ;Corcho, O. *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web*. London: Springer Verlag, 2004, 403p.
- 22 Uschold, M. Building ontologies: towards a unified methodology", *Proceedings of Expert Systems '96, the 16th Annual Conference of the British Computer Society Specialist Group on Expert Systems*, 1996, 18p.
<<http://www.w3.org/TR/2004/REC-owl-guide-20040210/#OwlVarieties>> (bezocht op 25 april 2007).
- 23 <<http://protege.stanford.edu/>> (bezocht op 1 mei 2007).
- 24 Gamper, J.; Nejdli, W.; Wolpers, M. Combining ontologies and terminologies. *Information Systems. Proc. of the 5th International Congress on Terminology and Knowledge Engineering*, 1999, Innsbruck, Austria.
- 25 Hepp, M.; Bachlechner, D.; Siorpaes, K. Harvesting wiki consensus using Wikipedia entries as ontology elements. *Proceedings of the 1st Workshop: SemWiki2006 - From Wiki to Semantics, co-located with the 3rd Annual European Semantic Web Conference (ESWC 2006)* . 2006: Budva, Montenegro.
- 26 Specia, L.; Motta, E. Integrating folksonomies with the semantic Web. in: E. Fraconi, M. Kifer, and W. May (Eds.): *Proceedings of the European Semantic Web Conference (ESWC 2007)*, Innsbruck, Austria. Springer 2007, pp.624-639.
- 27 Van Damme, C.; Hepp M.; Siorpaes, K. FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. *Proceedings of the ESWC 2007 Workshop "Bridging the Gap between Semantic Web and Web 2.0"*, 2007, Innsbruck, Austria.