

VU Research Portal

Social Networks and the Semantic Web

Mika, P.

2007

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Mika, P. (2007). *Social Networks and the Semantic Web*.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Samenvatting

Zelfs terugkijkend is het moeilijk te zeggen of wij het Web hebben veranderd of dat het Web ons heeft veranderd.

Maar hoewel het proces een vraag is, de veranderingen zijn een feit. Een recent grootschalig onderzoek naar het Internetgebruik door Amerikanen heeft de enorme verandering laten zien in de manier waarop men de online wereld benadert [Boase et al., 2006]. Wanneer we het Web beschouwen als een enorm mededelingenbord, dan plaatste in het begin soms iemand eens een krabbel, maar men liep er vooral langs heen, turend van de ene naar de andere mededeling. Tegenwoordig ‘surfen’ we niet alleen meer, maar is het Web een verlengstuk van onszelf geworden dat ons helpt om anderen te bereiken. We hebben geleerd om het reclamebord te gebruiken om actief anderen te zoeken en hebben het zo een ontmoetingsplaats gemaakt. Rondom het bord discussiëren we, vragen en geven we advies en delen we onze smaken, ervaringen en ideeën met vrienden en onbekenden, en doen op deze manier nieuwe contacten op.

Het resultaat blijkt duidelijk uit het onderzoek: de contacten die we opdoen gebruiken we zelfs in situaties waarin we vroeger alleen een beroep deden op nauwe verwanten, goede vrienden en familie. Het Internet heeft de grootte en samenstelling van onze sociale verbanden veranderd: onze verbanden zijn uitgebreid met een heel veel ‘vage kennissen’, d.w.z. al die bekende gezichten in onze *blog rolls*, *buddy lists*, *chat groups*, *web fora*, mailinglijsten en de vele andere fora waarin we met anderen communiceren.

Het moge duidelijk zijn dat het mededelingenbord mee moest veranderen om zich aan zijn nieuwe functie aan te passen. De term *Web 2.0* wordt tegenwoordig gebruikt als overkoepelende term voor al deze geleidelijke veranderingen. Ten eerste is het Web eindelijk het lees en schrijf Web geworden zoals zijn uitvinder het oorspronkelijk bedoeld had: de nu populaire plaatsen zijn door en voor mensen gemaakt. De kennis van de massa wordt gebruikt om grote kennisbronnen te bouwen en onderhouden, zoals de online encyclopedie Wikipedia. Maar die ‘massa’ schrijft niet alleen mee aan een encyclopedie, zij deelt ook foto’s en muziek, jaagt op nieuws en boeken, ontwerpt software, schrijft verhalen, en nog veel meer. Het Web heeft zich technisch gezien ook moeten aanpassen. Door gebruik te maken van AJAX technologie voelt de interactie met web-sites veel natuurlijker aan, terwijl *RSS feeds* en andere technieken zorgen voor een betere verbinding tussen de inhoud van het Web en de gebruikers. De enorme populariteit van *scripting* programmeertalen suggereert dat ook het programmeren is gedemocratiseerd en nu geworden is tot een kunst in plaats van een ambacht. Tenslotte, het gevoel van (gemeenschappelijk) bezit blijft mensen stimuleren om creatief te experimenteren met

de inhoud van het web in de vorm van *mash-ups*, web toepassingen die door gebruikers gecreëerde informatie uit verschillende bronnen combineert.

In tegenstelling tot het Web 2.0 is het Semantisch Web een meer bewuste inspanning onder leiding van het *World Wide Web Consortium* (de standaardisatie organisatie van het Web) om het Web toegankelijker te maken voor machines. Op dit moment is de meeste informatie op het Web alleen toegankelijk voor mensen. Het Semantische Web voegt een aantal extra lagen toe aan de architectuur van het Web die het mogelijk maken de inhoud van het Web te beschrijven met behulp van gemeenschappelijk vocabulaires, ook wel ontologieën genoemd. Dit zou computers in staat moeten stellen te redeneren over de kennis die in de ontologieën is uitgedrukt, bijvoorbeeld door relevante informatie uit verschillende bronnen te combineren en daaruit conclusies te trekken op een manier die lijkt op hoe mensen zouden redeneren. Hoewel het Semantisch Web een infrastructuur voor machines is, zal de kennis waaruit het is opgebouwd en de afleidingsregels waarmee geredeneerd kan worden uiteindelijk door mensen moeten worden geleverd. Kortgezegd is er geen semantiek zonder mensen, waardoor het Semantische Web net zo goed een sociaal als een technologisch systeem is.

Deze ontwikkelingen zijn zowel interessant voor onderzoekers in de sociale wetenschappen en de informatiekunde, als voor ontwikkelaars van sociaal semantische software voor het Web. Aan de ene kant biedt het zich ontwikkelende *Sociale Web* onverwachte mogelijkheden om sociaal gedrag te observeren door de interactie op het Web te volgen. Aan de andere kant vraagt de metadata die door gebruikers is aangeleverd om een andere behandeling dan de andere informatie. De door gebruikers aangeleverd informatie kan namelijk niet los gezien worden van de sociale context waarin het is gegenereerd. Met deze kennis, met name de sociale netwerken van gebruikers, kunnen de machines ook weer redeneren. Dit biedt ongezien mogelijkheden om informatie systemen te bouwen die rekening houden met de sociale context. Voor het Semantische Web betekent dit in het bijzonder dat intelligente toepassingen kunnen worden gebouwd die rekening houden met het sociale karakter van betekenis.

In dit boek beschrijven we twee grote *case studies* waarin beide mogelijkheden gedemonstreerd worden. De eerste studie laat zien hoe een onderzoeksgemeenschap via de interactie op het Web in kaart kan worden gebracht en geanalyseerd, waarbij ook gebruik wordt gemaakt van andere databronnen, zoals e-mails en publicaties (hoofdstuk 9). Het distilleren van een sociaal netwerk uit de informatie op het Web speelt in deze studie een grote rol. De grootschalige en dynamische gegevens over het netwerk die deze methode levert zouden niet met enquête methoden verzameld kunnen worden. Andersom is de semantische technologie essentieel voor het representeren en aggregeren van informatie vanuit verschillende bronnen. Methoden uit de sociale netwerkanalyse worden gebruikt om voorspellingen over de prestatie van onderzoekers te genereren.

Omdat de methoden die we gebruiken breder toepasbaar zijn dan in ons onderzoek over een wetenschapsgebied besteden we meer aandacht aan het beschrijven van de methoden dan aan het bespreken van de resultaten. In hoofdstuk 4 vatten we de mogelijkheden van het (her)gebruik van elektronische gegevens voor netwerkanalyse samen. Hoofdstuk 8 bevat een gerelateerd onderzoek waarin twee methoden voor het verkrijgen van sociale netwerk-gegevens vanuit het Web worden geëvalueerd. De semantische technologie voor het aggregeren van sociale netwerkgegevens wordt beschreven in de hoofdstukken 5

en 6. Tenslotte beschrijven we in hoofdstuk 7 hoe onze methoden zijn geïmplementeerd in het onderscheiden Flink systeem. Deze beschrijvingen maken het niet alleen mogelijk om onze resultaten te reproduceren, maar ook om onze methoden in verschillende soorten situaties toe te passen. Hiervoor zullen de methoden aan de andere situatie en informatiebronnen moeten worden aangepast, maar blijven de voordelen van het volledig geautomatiseerde proces gebaseerd op elektronische data behouden.

Ons tweede onderzoek belicht de rol van de sociale context in door gebruikers gecreëerde classificaties van inhoud, in het bijzonder in zogenaamde *tagging*-systemen, ook wel *folksonomies* genoemd (hoofdstuk 10). *Tagging* wordt vaak gebruikt om de inhoud van veel Web 2.0 diensten te ordenen, bijvoorbeeld in del.icio.us, het systeem voor gemeenschappelijke bladwijzers op het internet, en Flickr, de website die met mogelijk maakt om foto's te delen. Wij beschouwen *folksonomies* als lichtgewicht semantische structuren, waarbij de semantiek van de *tags* (labels) na verloop van tijd boven komt drijven door de manier waarop de labels worden gebruikt. Voor het bestuderen van de *tagging*-systemen gebruiken we de concepten en methoden van de netwerkanalyse. We laten zien dat *folksonomies* inderdaad een veel rijkere semantiek hebben dan op het eerste gezicht lijkt en dat deze afhankelijk is van de sociale context van de applicatie. Deze resultaten zijn in het bijzonder nuttig voor het ontwikkelen van het Semantische Web via een *bottom-up* aanpak, waarin de samenwerking tussen mensen een belangrijke rol speelt. Het plaatsen van de beschikbare kennis in een sociale context maakt ook de weg vrij naar gepersonaliseerde toepassingen, zoals sociaal zoeken.

Zoals duidelijk wordt uit bovenstaande beschrijving worden beide studies gekenmerkt door de interdisciplinaire aanpak, waarbij concepten en methoden vanuit de Kunstmatige Intelligentie worden gecombineerd met die vanuit de Sociale Netwerk Analyse. Het is echter niet noodzakelijk om voorkennis over deze gebieden te hebben, de hoofdstukken 2 en 3 bevatten de benodigde inleidingen. Hierdoor zou ons werk zowel toegankelijk moeten zijn voor sociale wetenschappers met interesse in elektronische data als informatiekundigen met interesse in sociaal-semantische toepassingen.

Ons belangrijkste doel is niet om beide gebieden in detail te behandelen, maar om zowel sociale wetenschappers als informatiekundigen inzicht te geven in de concepten en methoden die buiten hun eigen gebied vallen. We laten iets zien van de voordelen die dit begrip kunnen bieden voor het oplossen van de complexe problemen die een inherent interdisciplinair karakter hebben. Onze hoop is dat we mensen inspireren om verdere creatieve experimenten uit te voeren die ons een beter begrip zullen geven over zowel de sociale interactie die online plaatsvindt als over de eigenschappen van menselijke kennis. Dit begrip is onmisbaar in een wereld waarin de afstand tussen beide disciplines naar verwachting steeds kleiner zal worden door de online omgevingen die een steeds grotere rol in ons sociale leven gaan spelen, zoals de virtuele werelden van *Second Life*. Alleen met dit goede begrip zullen we in staat zijn om systemen te ontwerpen die zowel in hun redenen als in hun sociale mogelijkheden echte intelligentie laten zien en ons zo kunnen leiden door een steeds complexere online wereld.

De auteur bedankt de Vrije Universiteit Research School for Business Information Sciences (VUBIS) voor de steun bij het uitvoeren van het onderzoek in dit boek.