

Tagging en folksonomies

Eric Sieverts

Inleiding

Onder de noemer *web 2.0* zijn op internet talloze diensten en technieken ontwikkeld, die als belangrijk gemeenschappelijk kenmerk hebben dat gebruikers het op allerlei terreinen zelf voor het zeggen hebben. Inhoudelijke ontsluiting is één van die terreinen. Veel van die nieuwe diensten bieden gebruikers namelijk de mogelijkheid *tags* toe te voegen. Dat kan bij eigen materiaal dat ze op internet aanbieden, zoals afleveringen van de eigen weblog, foto's die op Flickr zijn neergezet of videomateriaal dat bij YouTube staat. Toevoegen van tags kan ook bij al bestaand materiaal, bijvoorbeeld bij bookmarks die op social bookmarking sites als Delicious, Diigo, Bibsonomy, CiteULike of Connotea worden bewaard, bij webpagina's die op StumbleUpon worden aangemeld, bij muziek op Last.fm of bij boekbeschrijvingen in LibraryThing (Hammond, 2005; Lund, 2005). Hierop geïnspireerd laten ook sommige bibliotheken hun gebruikers tags aan objecten uit de catalogus toekennen. De ontsluitingssystemen die zo ontstaan worden ook wel met de term *folksonomies* aangeduid. Dit artikel gaat nader in op de eigenschappen van folksonomies, op voor- en nadelen die aan dit soort ontsluiting zijn verbonden, op een aantal toepassingsmogelijkheden en op daaromheen ontwikkelde diensten en technieken.

Het indexeringsproces

Informatiespecialisten en bibliothecarissen zijn gewend om materiaal - van welke aard dan ook - op inhoudelijke en formele kenmerken te ontsluiten. Als het goed is, ligt een systematische en planmatige aanpak ten grondslag aan dit proces van indexeren of toekennen van metadata. Vooraf zijn daarover vaak beslissingen en afspraken gemaakt. Die kunnen onder meer betrekking hebben op

- de keuze voor een bepaald gestandaardiseerd model voor metadatabeschrijving,
- de keuze tussen pre- en postcoördinatieve methoden,
- de keuze of thesaurustermen worden toegekend of dat materiaal op basis van een taxonomie of classificatie wordt ingedeeld,
- of globale of diepte-indexering wordt toegepast,
- een richtlijn welke gemiddelde indexeerdiepte - het aantal aan een object toe te kennen termen - wordt nagestreefd.

De keuze ten behoeve van ontsluiting een bepaalde bestaande informatietaal te gebruiken, impliceert meestal al een aantal van deze aspecten. Aan de ontsluiting zelf ligt vervolgens een zorgvuldige inhoudsanalyse van het te karakteriseren object ten grondslag (Becker, 2010 - hoofdstuk 8).

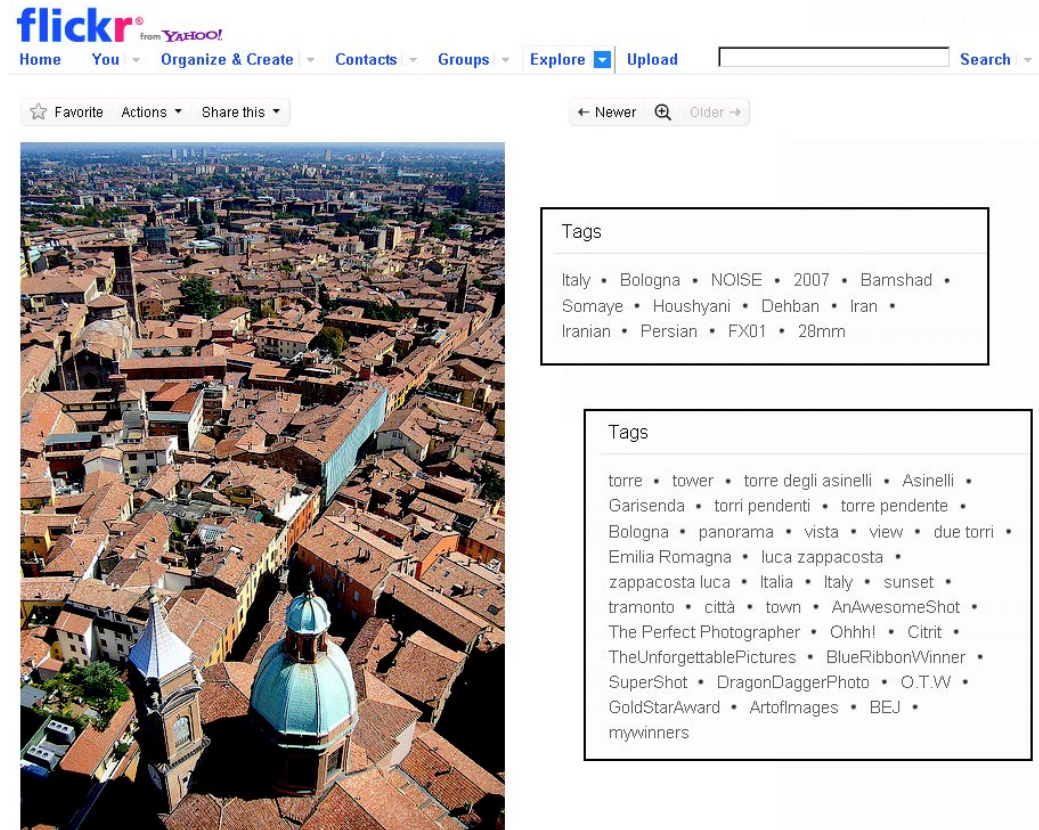
Bij het toekennen van tags zullen de meeste gewone gebruikers (uiteraard) niet op een dergelijke planmatige, systematische wijze te werk gaan. Toch zal iemand ook hierbij, veelal onbewust, een inhoudsanalyse maken van het object dat hij met tags wil karakteriseren. Het persoonlijke, veelal abstracte beeld dat hij zich van dat object heeft gevormd, zal hij vervolgens ook vertalen naar een aantal toe te kennen tags, of je dat nu een informatietaal wilt noemen of niet. Natuurlijk bestaat geen eenduidig beeld of iedere gebruiker zich bij deze stap volledig bewust zal zijn van de zoekster die dit object later moet kunnen terugvinden - of hij nu zelf die zoekster zal zijn, of dat juist anderen de doelgroep vormen. Daarbij zal men vrijwel altijd afzonderlijke tags toekennen voor de verschillende elementen en invalshoeken van de beschreven objecten, dus postcoördinatief te werk gaan. Daarnaast spelen begrippen als indexeerdiepte en globale of diepte-indexering bij tagging dus ook wel degelijk een rol.



Afbeelding 1 - Begin van de tagcloud van de tags die zijn toegekend aan de 60 miljoen in Librarything beschreven boekexemplaren (alfabetisch geordend). (<http://www.librarything.com/tagcloud.php>)

Een duidelijk verschil met pure inhoudelijke ontsluiting is dat tagging zich niet beperkt tot inhoudelijke elementen, maar dat formele kenmerken van objecten ook vaak in tags tot uiting worden gebracht. De tagcloud in afbeelding 1 laat een sterke vermenging zien van formele en inhoudelijke kenmerken, zoals die in LibraryThing voorkomt. Tevens zien we daaraan dat toegekende inhoudelijke tags voornamelijk de globale onderwerpen van beschreven boeken karakteriseren. Afbeelding 2 illustreert hoe verschillende gebruikers voor een zelfde foto op Flickr sterk verschillende indexeerdiepte kunnen toepassen en ook heel verschillend tags toekennen. Een analyse van de aan 50 miljoen foto's op Flickr toegekende tags wees

uit dat aan 30% slechts één tag was toegekend, terwijl maar 15% meer dan zes tags had meegekregen. Aan zes foto's waren zelfs meer dan 100 tags toegekend (Sigurbjörnsson, 2008). De meest voorkomende soorten facetten waartoe tags gerekend kunnen worden zijn: plaats, voorwerp, persoon, gebeurtenis, datum en fototechnisch.



Abbeelding 2: De tags die bij twee vrijwel identieke foto's van Bologna in Flickr door verschillende fotografen (een Italiaan en een Iraniër) zijn toegekend. Zowel de aard van de termen (inhoudelijk/formeel) als de indexeerdiepte verschillen aanzienlijk.

De eenvoud van het vermengen van inhoudelijke en formele kenmerken, is meteen ook een zwakke kant van deze methodiek. Die kan namelijk aanleiding geven tot zeer slechte precisie van zoekresultaten. Wie in Flickr naar foto's zoekt met de zoekvraag *Haarlem AND Eerste wereldoorlog*, vindt op het moment van schrijven grote aantallen foto's uit de collectie van het Nationaal Archief, die zijn gemaakt rond de slagvelden in Noord-Frankrijk en Vlaanderen. Dat die foto's niettemin de tag 'Haarlem' hebben meegekregen, komt door het feit dat ze afkomstig zijn uit het omvangrijke fotoarchief van de vroegere tijdschriftuitgever Spaarnestad die inderdaad in Haarlem was gevestigd (zie de betreffende termen in de tagcloud van de foto's van het Nationaal Archief in afbeelding 3). Die specifieke context van 'Haarlem' gaat verloren in een simpel systeem van tagging, waarbij geen nadere betekenis - naar aard en inhoud - aan tags kan worden meegegeven.

1928 1930 1938 1941 19de eeuw 19th century 441048 accidents airplane amerika amsterdam
 animal animals apron archief armisticeday artis auto autumnfashion aviation
 beach beate bicycle boys burgerluchtvaart car child childlabor children christmas
 christmas tree civilaviation classroom cycloracing cyclists dieren dierentuin dutch eerste
 eerstewereldoorlog emigranten england factory fashion feminism fiets
 firstworldwar france frankrijk gewonden girls haarlem hat herfstmode holland
 housekeeping huishouden ice industry internationalevrouwendag inventions jacdenijs japan jvaneijk
 kaart kerst kimbei kinderen klm koningin juliana koninklijk leerling leger luchtvaart machine
 man manhattan men militairen mode nationaal nationaalarchief
 nederland nederlandsindisch netherlands newyork newyorkcity nijmegen nyc
 oldfashionedprofessions olympicgames olympische ongelukken ongevallen
 oudeberoepen outdoors portrait prinses juliana protest pupil queen juliana refugees
 repatrianten retouche rondevanfrankrijk rusland russia school schooltijd soldiers sovietunion
 spaarnestadphoto spelen sport sports standing strand street thegreatwar
 thenetherlands tourdefrance tweedewereldoorlog uitvindingen uniforms unitedstates
 usa verenigdestaten vonstillfried vrouw vrouwen vs wereldoorlog wielrennen wielrenners winter
 winterfashion wintermode wintersport wintersports WOI woman women womensday
 work workers worldwari wounded WWI wwii zoo

Afbeelding 3 - Tagcloud voor de foto's die het Nationaal Archief aanbiedt op Flickr
 (<http://www.flickr.com/photos/nationaalarchief/>)

De gebruiker centraal

Zoals uit het laatste voorbeeld al bleek, zijn tags in feite gewoon ongecontroleerde trefwoorden. Er zijn vrijwel nooit afspraken over gebruik van enkel- of meervoud, over gebruik van zelfstandig naamwoorden of werkwoorden of over synoniemen. Ze hebben dan ook last van alle nadelen die aan ongecontroleerde ontsluitingsmethoden kleven (Macgregor, 2006). Doordat er meestal geen consistentie is in de door een gebruiker toegekende tags, en zeker niet tussen de tags die verschillende gebruikers voor gelijksoortige onderwerpen toepassen, zal iemand die een zoekactie uitvoert op basis van tags, vaak veel relevante informatie missen. De vangst is dus over het algemeen slecht. Anderzijds kan de precisie van zoekacties - hoeveel van het gevonden materiaal relevant is voor de gestelde zoekvraag - ook vaak slecht zijn, doordat veelvuldig tags met een te ruime of een dubbelzinnige betekenis zijn toegekend.

Tags hebben echter het grote voordeel dat ze berusten op het woordgebruik van de informatiegebruikers zelf. Dit soort ontsluitingsystemen wordt daarom ook wel aangeduid als *folksonomies*. Deze term is in 2004 door Thomas Vander Wal bedacht, als samentrekking van de woorden *folks* (mensen) en *taxonomy*. Het zijn dus door *folks* zelf opgezette taxonomieën (Vander Wal, 2004). In het Engelse taalgebied spreekt men in dit verband ook wel van *grass-roots* taxonomieën. Hierbij wordt de gebruiker geacht zelf het best te weten waarop hij voor zijn onderwerp moet zoeken of onder welke term hij zijn eigen materiaal verwacht terug te kunnen vinden. Bovendien ziet men tagging als hulpmiddel bij het samenwerken van

mensen. Ook dat is een belangrijk kenmerk van web-2.0-toepassingen. In dit verband spreekt men daarom ook wel van *social tagging* of *collaborative tagging*. Het taggen van materiaal kan dan helpen virtuele gemeenschappen te laten ontstaan van mensen met gelijke interesse en daardoor hopelijk gelijksoortig woordgebruik.

In dit kader kunnen we ook onderscheid maken tussen persoonlijk georiënteerde en sociale tags. De eerste soort gaat alleen uit van het persoonlijke gezichtspunt ('Oom Jan') of het persoonlijke oordeel ('interessant', 'nog lezen'). Daardoor hebben die tags eigenlijk alleen betekenis voor degene die de tag heeft toegekend, of dat nu ten behoeve van het zoeken is, voor het filteren van resultaten of voor een snelle beoordeling van gevonden materiaal. Met de tweede soort, hopen gebruikers juist anderen te bereiken en te interesseren voor het eigen (of andermans) product, of dat nu een blogpost, een gevonden wetenschappelijk artikel, een favoriet boek, een collectie op Flickr geplaatste foto's of een zelf naar YouTube geüploade video is. Juist bij foto- en videomateriaal zijn dergelijke tags vaak ook de enige tekstuele informatie waarop dat materiaal kan worden teruggevonden.

Zoeken of laten aanraden

Het voorgaande betekent in feite dat tagging vaak voor een ander soort toegankelijkheid is bedoeld dan klassieke ontsluiting. Dat het bij zoekacties tot slechte vangst en precisie kan leiden, wordt daarbij veelal minder als een bezwaar gezien dan bij zoeksystemen die gebruik maken van "echte" ontsluiting of de digitale tekst van documenten zelf doorzoeken. Tags zijn meer bedoeld als ondersteuning bij het browsen. Gebruikers neigen ook meer af te gaan op suggesties die door collega's en vrienden door middel van tags zijn gedaan, dan dat ze op eigen initiatief diepgravend naar informatie zoeken. *Tagclouds* (woordenwolken) laten bij dergelijke systemen ook grafisch zien welke tags het meest frequent en/of het meest recent zijn gebruikt, hoe groter de letters hoe populairder de woorden (zie afbeeldingen 1 en 3).

Zo vormen tags ook een hulpmiddel voor recommandatie. Analyse van toegekende tags kan daarbij helpen. Andere gebruikers die veelvuldig dezelfde tags toekennen als jijzelf worden in dat verband wel aangeduid als *tag buddies* (Peters, 2009). Het is waarschijnlijk dat ze dezelfde interesse en dezelfde smaak hebben. De door hen getagde items kunnen dus ook voor jou interessant zijn. Je kunt zelfs zeggen dat ze met hun tags indirect hun eigen interesse inhoudelijk toegankelijk maken. Deze manier van analyseren van toegekende tags lijkt in feite sterk op wat onder meer Amazon doet op basis van het koopgedrag van zijn klanten. Op social bookmarking site Delicious kun je dergelijke *tag buddies* ook in je *network* opnemen, waardoor je al hun bookmarks in één overzicht gepresenteerd krijgt. Onderzoek heeft aangetoond dat semantische gelijkenis tussen toegekende tags op muzieksite Last.fm zelfs voorspellende waarde heeft voor het bestaan van vriendschappen in het echte leven (Schifanella, 2010).

Waarom gebruikers tags toevoegen kan heel verschillende redenen hebben. In de voorgaande paragrafen gingen we er vooral vanuit dat eigenbelang een belangrijke drijfveer is, omdat het helpt om makkelijk materiaal terug te kunnen vinden dat anders onvindbaar dreigt te worden. Dit wordt wel aangeduid als de behoefte te categoriseren. Onderzoek wijst echter uit dat er ook andere motieven kunnen zijn,

zoals de behoefte om te beschrijven, zonder daarbij direct de terugvindbaarheid voor zichzelf in gedachten te hebben. Welke motivatie mensen hebben blijkt ook binnen eenzelfde systeem sterk te kunnen variëren (Strohmaier, 2010). Of gebruikers gemotiveerd blijven om tags toe te voegen wordt wel eens betwijfeld. Zolang het een nieuw en daardoor spannend fenomeen was, deed iedereen daar graag aan mee. Nu het zo algemeen geworden is, komt de vraag op of gebruikers het zullen blijven opbrengen hier tijd en aandacht aan te besteden (Sieverts, 2009).

Toch signaleert Shirky in zijn boek "Cognitive surplus" (2010) een sterke trend dat mensen steeds meer geneigd zijn om hun vrije tijd actief participierend, creatief en intellectueel in te vullen. Hij illustreert dit aan de Wikipedia, waarvan hij schat dat er belangeloos al meer dan 100 miljoen uur intellectuele arbeid in geïnvesteerd is. Vergeleken met de 200 miljard uur die in de VS jaarlijks wordt besteed aan (passief) televisie kijken is dat nog niet eens veel. Er zijn ook manieren om gebruikers te verleiden tot het taggen van al in systemen aanwezig materiaal, door daar een spel- of wedstrijdelement aan toe te voegen. Dat is bijvoorbeeld gebeurd bij het Waisda-project van Beeld en Geluid. Daarbij kregen deelnemers videofragmenten voorgeschoteld, waarbij zij punten konden verdienen als ze daaraan dezelfde tags toekenden als een andere gebruiker die dat fragment ook te zien kreeg (Belice Baltussen, 2010; Oomen, 2010).

Tagging in een professionele omgeving

Voor kritische toepassingen in een bedrijfsomgeving kan men zich niet veroorloven om met een zoekactie relevante documenten te missen. Daar zul je dus niet kunnen volstaan met tagging als enige hulpmiddel om die documenten vindbaar te maken (Alexander, 2009). Slechte vangst is daar meestal een te groot bezwaar. Ook ontbreekt in die omgeving meestal de tijd om de vele niet-relevante documenten te moeten doornemen die het resultaat zijn van de slechte precisie van een zoekresultaat. In gevallen waar geen menskracht beschikbaar is voor gecontroleerde ontsluiting door informatiespecialisten, kan tagging echter een zinnig alternatief zijn. Er wordt dan gebruik gemaakt van wat wel *crowd-sourcing* wordt genoemd, het uitbesteden van de activiteit aan "de massa", de *crowd*. Die *crowd* moet dan echter wel groot genoeg zijn in vergelijking tot de hoeveelheid te taggen materiaal. Zelfs bij een grote organisatie als een universiteit, kan niet verwacht worden dat bijvoorbeeld 20.000 studenten of 2000 wetenschappers een tastbare bijdrage leveren aan het inhoudelijk ontsluiten van een evenredig grote collectie van 500.000 boeken of tien miljoen wetenschappelijke artikelen.

Tagging kan vaak wel nuttig zijn voor het ontsluiten van niet-tekst materiaal, waar een gewone zoekmachine geen woorden aan de documenten zelf kan ontleen. Zo zijn er steeds meer archieven die gedigitaliseerde foto's op internet zetten, met het verzoek aan bezoekers van de site om onbekende details toe te voegen. Ditzelfde geldt ook voor gescande tekstdocumenten waarvan de kwaliteit te slecht is om ze met optische karakterherkenning enigszins betrouwbaar computerleesbaar te maken, zoals bij handschrift en oud drukwerk. Een overzicht van meer toepassingen in het erfgoed-domein - ook bij museumobjecten en videomateriaal - is te vinden bij Oomen (2011). In dat artikel komen overigens meer soorten van crowdsourcing aan de orde dan alleen het "classificeren" waartoe "social tagging" gerekend wordt.

Standaardisatie van tags

Aanbieders van web 2.0 diensten blijken de nadelen van ongecontroleerde ontsluiting wel te onderkennen. Daarom ondersteunen veel systemen het tag-proces, door lijstjes te presenteren van eerder al toegekende tags - hetzij door dezelfde gebruiker voor ander materiaal, hetzij door andere gebruikers voor hetzelfde materiaal. Zo hoopt men gebruikers tot meer consistent tag-gedrag te verleiden.

Daarnaast worden bij specifieke gebeurtenissen - of dat nu een specialistisch congres met slechts honderd deelnemers is, of een wereldwijd nieuwsfeit - wel vrijwillige afspraken gemaakt om een standaard *hashtag* te gebruiken. Dat zijn tags die beginnen met een #hash-teken. Daarmee kunnen deelnemers en belangstellenden dan alle materiaal taggen, dat op die gebeurtenis betrekking heeft, op welke site dat zich ook bevindt. Zo werd alles rond de tumultueuze verkiezingen in Iran in 2009 getagd met *#iranelection*, of dat nu op Twitter, op Flickr, op YouTube of in weblogs was. Voor Twitter bestaan diensten, zoals Twubs, waar je een *hashtag* kunt registreren en waarmee je vervolgens selectief in real-time alle berichten met een bepaalde opgegeven tag uit de Twitterstroom kunt laten filteren. Ook bestaat er op dit moment een hulpmiddel, Whathashtag, waarmee je een indruk kunt krijgen welke hashtags het meest worden toegekend bij berichten over een bepaald onderwerp.

Onder de noemer *semantic tagging* worden ook wel systemen ontwikkeld die, op basis van de inhoud van documenten of webpagina's, proberen suggesties te geven voor toe te kennen tags. Een voorbeeld daarvan is Faviki dat aan webpagina's die je wilt taggen, die woorden en begrippen ontleent die ook voorkomen in een uit de Wikipedia gedestilleerde lijst begrippen (Faviki, 2009). Uit dat rijtje voorstellen kun je kiezen, zodat je wel gedwongen wordt om vastgelegde termen te gebruiken. Er zijn ook systemen die proberen te bepalen welke al eerder toegekende tags vermoedelijk een verwante betekenis hebben, op basis waarvan ze die tags clusteren of zelfs als synoniemen beschouwen. Een voorbeeld daarvan is de social bookmarking site Zitag (2008). Zowel Faviki als Zitag maken bovendien gebruik van het *Common Tag* format waarbij tags verwijzen naar wel gedefinieerde concepten waaraan zelfs metadata gekoppeld kunnen zijn en waar met een URL naar verwezen kan worden (Common, 2009). Zo worden tags toch weer in een gecontroleerd vocabulaire geperst.

Brede en nauwe folksonomies

Systemen met tags worden wel onderscheiden in nauwe en brede (*narrow / broad*) folksonomies (Peters, 2009). In een nauwe folksonomy kent de maker of eigenaar van een op internet geplaatst object daar alleen zelf tags aan toe. Een voorbeeld van een dergelijk systeem is de YouTube videosite. In dat geval is dus zeker geen sprake van "social" tagging. In sommige gevallen is het mogelijk dat andere geregistreerde gebruikers die tags nog verder aanvullen. In dat geval spreekt men wel van een *extended narrow folksonomy*. Ook dan wordt een bepaalde tag toch maar één keer aan een object toegekend. Een voorbeeld van deze categorie is de fotosite Flickr. Bij beide *nauwe* varianten staat dus het object centraal.

Bij een brede folksonomy kan elke deelnemer vrij eigen tags aan elk object toekennen. Daar staat dus de gebruiker centraal. Het systeem registreert elke tag

van elke gebruiker afzonderlijk, ook als vijftig verschillende gebruikers allemaal dezelfde tag aan een bepaald object toekennen. Dat biedt dus de mogelijkheid om te zien welke termen het meest aan een object zijn toegekend, dus wat vermoedelijk het centrale onderwerp of de belangrijkste invalshoek daarvan is. Tevens biedt dit meer vrijheid dan bij een nauwe folksonomy, om ook andere invalshoeken gerepresenteerd te krijgen, dan alleen de misschien wat eenzijdige blik van de oorspronkelijke eigenaar van een object. Voorbeelden van brede folksonomies zijn social bookmarking sites als Delicious en Connotea en de sociale catalogiseersite voor boeken LibraryThing.

Hoewel het verschil tussen deze twee manieren van taggen in de praktijk misschien van ondergeschikt belang lijkt, maakt het wel degelijk verschil voor de te bieden functionaliteit bij het zoeken en het presenteren van resultaten. Welk soort statistieken van veel gebruikte en populaire tags je kunt afleiden, is bijvoorbeeld nogal verschillend. Ook bij analyse van gebruikte tags, bijvoorbeeld voor het automatisch afleiden van semantische relaties daartussen, kan het nogal wat uitmaken of je met een brede of een nauwe folksonomy te maken hebt.

Analyse van gebruikte tags

Onderzoek aan tagging beperkte zich aanvankelijk vooral tot kwaliteitsanalyses: in hoeverre eenduidige termen worden gebruikt, in hoeverre er problemen zijn met synonymie, in hoeverre de tags vooral zijn bedoeld voor persoonlijk (her)gebruik en niet gericht op publiekelijke bruikbaarheid, wat de specificiteit van toegekende tags is, voor welke soorten eigenschappen tags zijn toegekend, of voldoende tags zijn gebruikt om alle aspecten van een document, een foto of een video te beschrijven - dus wat de indexeerdiepte is - enzovoort (Macgregor, 2006; Golder, 2006; Chung, 2009). Vaak leidt dat tot de conclusie dat tagging meer een aanvulling is op klassieke ontsluiting, dan een vervanging daarvan (Guy, 2006). Ook wordt wel gekeken naar de verdeling van de frequentie van toekenning van tags. Die blijkt meestal een soort Zipf-verdeling te volgen: een beperkt aantal zeer veel gebruikte tags en een lange staart (*long tail*) van heel weinig of misschien zelfs maar eenmalig toegekende tags (Peters, 2009).



Afbeelding 4 - Semantisch geordende tagcloud uit LibraryThing. Selectie van tags waarvan op basis van statistiek is afgeleid dat ze gerelateerd zijn aan de tag 'russian literature'. De clustering van termen per regel is gebaseerd op co-occurrence analyse. De volgorde van de regels en van de tags op elke regel is zodanig dat sterk overeenkomende tags zo dicht mogelijk bij elkaar staan. Door de plaatsing van de tag "c19" naast "19th-century" wordt meteen duidelijk wat deze tag betekent. (Wartena, 2010).

Een nadeel van tags ten opzichte van hiërarchisch gestructureerde ontsluitingstechnieken als thesauri en taxonomieën is het ontbreken van relaties tussen tags waartussen betekenisverwantschap bestaat. Daarom wordt volop geëxperimenteerd met automatische mechanismen om de betekenis van tags en mogelijke semantische relaties tussen tags af te leiden. Ook worden systemen ontwikkeld die verband proberen te leggen met bestaande thesauri of ontologieën. Hierbij worden vaak methoden toegepast die zijn gebaseerd op statistische analyse van aan dezelfde objecten toegekende tags, vaak nog verfijnd door analyse van het daarbij samen voorkomen (*co-occurrence*) van verschillende tags. Afbeelding 4 geeft een voorbeeld van een op basis van dergelijke technieken geordende tagcloud van in LibraryThing aan boeken toegekende tags (Wartena, 2010). De in de woordenwolk getoonde termen zijn het resultaat van een statistische analyse van de tags die iets met het begrip 'russian literature' te maken hebben. Daarbinnen zijn de termen geclusterd op basis van een *co-occurrence* analyse. De termen op dezelfde regel hebben onderling meer met elkaar te maken dan met termen op andere regels. Toepassing van dergelijke technieken kan gebruikers bij hun zoekproces ondersteunen, zoals afbeelding 5 illustreert. Het inperken of uitbreiden van een zoekactie met enkele van de gepresenteerde tags, kan de precisie en (vooral) de vangst van zoekacties in zo ontsloten materiaal sterk verbeteren. Door toepassing van dit soort technieken kunnen systemen proberen om, met behoud van de vrijheid van degene die tags toekent, toch een deel van de nadelen van een folksonomy te compenseren.



Afbeelding 5 - Zoekresultaten en tag cloud voor een zoekvraag naar "british history", in een data sample van LibraryThing. Deze wolk is op dezelfde manier gegenereerd als die in afbeelding 4. Merk op dat de gezochte tag niet voorkomt bij het eerste zoekresultaat. (Wartena, 2010).

In dit artikel genoemde diensten op internet

Bibsonomy	http://www.bibsonomy.org/	Social bookmarking site voor wetenschappelijke publicaties
CiteULike	http://www.citeulike.org/	Social bookmarking site voor wetenschappelijke publicaties
Connotea	http://www.connotea.org/	Social bookmarking site voor wetenschappelijke publicaties
Delicious	http://www.delicious.com/	Social bookmarking site
Diigo	http://www.diigo.com/	Social bookmarking site
Faviki	http://www.faviki.com/	Social bookmarking site die op basis van semantische analyse van pagina's suggesties doet voor daaraan toe te kennen tags
Flickr	http://www.flickr.com/	Site voor het uploaden van foto's
Last.fm	http://www.last.fm/	Muzieksite waarop je persoonlijke collecties kunt aanleggen
LibraryThing	http://www.librarything.com/	Site voor het catalogiseren en beschrijven van persoonlijke boekcollecties
StumbleUpon	http://www.stumbleupon.com/	Zoekmachine waar gebruikers interessante websites en -pagina's moeten aanmelden
Twitter	http://www.twitter.com/	Microblogging dienst waar berichten van maximaal 140 tekens geplaatst kunnen worden
Twubs	http://twubs.com/	Dienst om hashtags aan Twitter-berichten te laten toevoegen en selectief op die tags te filteren
Whathashtag	http://whathashtag.circulorojo.es/	Hulpmiddel om te zien welke hashtag veel voor een onderwerp wordt gebruikt
YouTube	http://www.youtube.com/	Site voor het uploaden van videomateriaal
Zigtag	http://www.zigtag.com/	Social bookmarking site die op basis van semantische analyse van pagina's suggesties doet voor daaraan toe te kennen tags

Gebruikte literatuur:

- Alexander, F. (april 2009). Folksonomies: business use. In *FUMSI*. Geraadpleegd op: 1 maart 2011, van: <http://web.fumsi.com/go/article/manage/3791>
- Becker, P., Linden, M. van der, Magrijn, H., & Sieverts, E. (2010). *Organiseer je informatie; aan de slag met thesauri, taxonomieën, tags en topics*. Leidschendam: Biblion Uitgeverij. ISBN 978-90-5483-954-5
- Belice Baltussen, L., Brinkerink, M., & Oomen, J. (2010). *Evaluatieverslag Waisda? Taggen van audiovisueel archiefmateriaal in spelvorm*. Hilversum: Beeld en Geluid. Geraadpleegd op: 27 maart 2011, van: http://maartenbrinkerink.net/temp/waisda/Rapport_Waisda_DEF.pdf
- Chung, E. & Yoon, J. (2009). Categorical and specificity differences between user-supplied tags and search query terms for images: an analysis of Flickr tags and Web image search queries. *Information Research*, 14(3). paper 408. Geraadpleegd op: 1 maart 2011, van: <http://informationr.net/ir/14-3/paper408.html>
- Common (2009) - *Common Tag*. Geraadpleegd op: 1 maart 2011, van: <http://www.commonitag.org/Home>
- Faviki (2009) - *Social bookmarking tool using smart semantic Wikipedia (DBpedia) tags*. Geraadpleegd op: 1 maart 2011, van: <http://www.faviki.com/pages/welcome/>
- Golder, S., & Huberman, B.A. (2006). Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2),198-208. Geraadpleegd op: 1 maart 2011, van: <http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf>
- Guy, M., & Tonkin, E. (2006) Folksonomies: tidying up tags? *D-Lib Magazine*, 12(1). Geraadpleegd op: 1 maart 2011, van: <http://dlib.org/dlib/january06/guy/01guy.html>
- Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005) Social Bookmarking Tools (I): A General Overview. *D-Lib Magazine*, 11(4). Geraadpleegd op: 1 maart 2011, van: <http://dlib.org/dlib/april05/hammond/04hammond.html>
- Lund, B., Hammond, T., Hannay, T., & Flack, M. (2005). Social Bookmarking Tools (II): A Case Study: Connotea. *D-Lib Magazine*, 11(4). Geraadpleegd op: 1 maart 2011, van: <http://dlib.org/dlib/april05/lund/04lund.html>
- Macgregor, G., & McCulluch, E. (2006). Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review*, 55(5), 291-300. Geraadpleegd op 1 maart 2011, van: <http://strathprints.strath.ac.uk/2335/1/strathprints002335.pdf>
- Oomen, J. & Arroyo, L. (2011). Crowdsourcing in the cultural heritage domain: opportunities and challenges. In: *5th International Conference on Communities & Technologies. Brisbane, Australia - 29 June – 2 July 2011*. Geraadpleegd op: 27 maart 2011, van: <http://www.cs.vu.nl/~marieke/OomenAroyoCT2011.pdf>
- Oomen, J., Belice Baltussen, L., Limonard, S., van Ees, A., Brinkerink, M., Aroyo, L., Vervaart, J., Asaf, K., & Gligorov, R. (2010). Emerging Practices in the Cultural Heritage Domain - Social Tagging of Audiovisual Heritage. In: *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, April 26-27th, 2010, Raleigh*,

NC: US. Geraadpleegd op: 27 maart 2011, van:
http://journal.webscience.org/337/2/websci10_submission_23.pdf

- Peters, I. (2009). *Folksonomies: indexing and retrieval in web 2.0*. Berlin: De Gruyter/Saur. ISBN 978-3-598-25179-5
- Schifanella, R., Barrat, A., Cattuto, C., Markines, B. & Menczer, F. (2010). Folks in folksonomies: social link prediction from shared metadata. In: *Third ACM International Conference on Web Search and Data Mining. New York City, NY, USA - February 03 - 06, 2010*. Geraadpleegd op: 1 maart 2011, van: <http://hal.archives-ouvertes.fr/docs/00/42/98/86/PDF/wsdm141-schifanella.pdf>
- Shirky, C. (2010). *Cognitive surplus: Creativity and generosity in a connected age*. New York: Penguin Press. ISBN 978-1-594-20253-7
- Sieverts, E. (2009). Ik tagde, jij tagde, *Informatie Professional*, 13(4), 13. Geraadpleegd op 1 maart 2011, van:
<http://www.library.uu.nl/medew/it/eric/taggen.html>
- Sigurbjörnsson, B. & Zwol, R. van (2008). Flickr Tag Recommendation based on Collective Knowledge. In: *Proceeding of ACM International World Wide Web Conference, 2008*, 327-336. Geraadpleegd op: 1 maart 2011, van:
<http://www2008.org/papers/pdf/p327-sigurbjornssonA.pdf>
- Strohmaier, M., Körner, C., & Kern, R. (2010). Why do users tag? Detecting users' motivation for tagging in social tagging systems. In: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. Geraadpleegd op: 1 maart 2011, van:
<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/download/1497/1892>
- Vander Wal, T. (2004). Feed on this. In: *Weblog "Off the top", 3 October 2004*. Geraadpleegd op: 1 maart 2011, van:
<http://www.vanderwal.net/random/entrysel.php?blog=1562>
- Wartena, C. (2010). Zoeken met tags; tagvarianten semantisch getemd. *InformatieProfessional*, 14(3), 20-23.
- Zigtag (2008). *How does Zigtag differ from other social bookmarking websites?* Geraadpleegd op: 27 maart 2011, van: <http://zigtag.com/corp/faq/008.jsp>