



- [Spring naar de content](#)
- [Spring naar het hoofdmenu](#)

## Het web is plat...

23 februari 2011

Het web bestaat grotendeels uit gekoppelde documenten met ongestructureerde informatie. De koppelingen zijn handig, maar niet erg betekenisvol. [Semantische webstandaarden \(http://en.wikipedia.org/wiki/Semantic\\_Web\)](http://en.wikipedia.org/wiki/Semantic_Web) zijn bedoeld om het web meer structuur en profiel te geven. Ze maken het mogelijk om entiteiten, zoals personen, landen, liedjes of bestemmingsplannen, te beschrijven en te koppelen. Ze zorgen daarmee voor samenhang tussen en vindbaarheid van informatie. Ze maken het web rond. Maar wat moet de Nederlandse overheid ermee?

Informatie op het web is vaak door computers niet goed interpreteerbaar. Daardoor is de informatie ook minder goed vindbaar en toegankelijk voor mensen. Als je bijvoorbeeld zoekt naar afbeeldingen van [Drenthe \(http://www.google.nl/images?hl=nl&rlz=&q=drenthe\)](http://www.google.nl/images?hl=nl&rlz=&q=drenthe), dan krijg je zowel plaatjes van de provincie als van de voetballer te zien. Een ander voorbeeld is [Boston \(http://www.youtube.com/watch?v=TJfrNo3Z-DU\)](http://www.youtube.com/watch?v=TJfrNo3Z-DU), dat de naam is van verschillende steden maar ook van een rockband.

Semantische webstandaarden kunnen dit soort problemen oplossen. Tim Berners-Lee, de grondlegger van het web, is de grote gangmaker achter deze standaarden. Aan de hand van een zakje chips legde hij vorig jaar, tijdens een [overheidsconferentie \(http://www.youtube.com/watch?v=ga1aSJXCFe0\)](http://www.youtube.com/watch?v=ga1aSJXCFe0), de werking uit. De standaarden hebben niet alleen de aandacht van techneuten en academici. Google, Yahoo!, Facebook, Amazon en BBC maken er bijvoorbeeld gebruik van. Buitenlandse overheden, zoals de [Engelse \(http://data.gov.uk/linked-data\)](http://data.gov.uk/linked-data) en [Amerikaanse \(http://www.data.gov/semantic\)](http://www.data.gov/semantic), zetten ook steeds meer in op deze standaarden bij het publiceren van overheidsdata. Volgens recent onderzoek wordt de semantische webstandaard RDFa op 430 miljoen webpagina's gebruikt en [groeiende het gebruik \(http://tripletalk.wordpress.com/2011/01/25/rdfa-deployment-across-the-web/\)](http://tripletalk.wordpress.com/2011/01/25/rdfa-deployment-across-the-web/) in 2010 met 510%.

Een belangrijk mechanisme binnen semantische webstandaarden zijn zogenaamde triples. Het concept is afgeleid van hoe veel natuurlijke talen zijn opgebouwd: **subject – predicaat – object (http://www.youtube.com/watch?v=ldl0m-5zLz4)**. Het subject is datgene wat je beschrijft. Het predicaat geeft aan welk kenmerk van het subject beschreven wordt. Het object is de waarde van het kenmerk. Voorbeelden zijn: 'Piet [s] kent [p] Ben [o].', 'Ben [s] is geboren in [p] Maastricht [o]', 'Maastricht [s] ligt in [p] Limburg [o].'

Bart Knubben (Bureau Forum Standaardisatie)



Bart Knubben houdt zich binnen Bureau Forum Standaardisatie bezig met het selectieproces en de criteria voor de lijsten met open standaarden. Hij heeft zich de afgelopen jaren gespecialiseerd in de elektronische overheid, open source software & open standaarden en projectmanagement. Bart Knubben studeerde bedrijfseconomie (informatiemanagement) en rechten (afstudeerrichting auteursrecht).

[Alle weblogberichten van deze auteur](#)

Ieder subject, predicaat of object heeft een unieke identifier (URI), waardoor er gericht naar kan worden verwezen. Een voorbeeld van een unieke identifier voor in het GBA ingeschreven personen is het BSN. Op het web worden vaak URL's als unieke identifiers gebruikt.

Aan de hand van triples kunnen entiteiten met elkaar in verband worden gebracht. Een voorbeeld zijn de relaties tussen Ferrari, Fiat en Maserati, die [DBpedia \(http://bit.ly/heEHuu\)](http://bit.ly/heEHuu) op basis van gestructureerde Wikipedia-informatie laat zien. De verbanden bieden ook de mogelijkheid om gericht (via zogeheten 'faceted search') naar bepaalde informatie te zoeken, zoals bijvoorbeeld [Nederlandse wetenschappers die zijn geboren in de 19de eeuw \(http://dbpedia.neofonie.de/browse/rdf-type:Scientist/birthDate-year~:1800~1900/nationality:Netherlands/\)](http://dbpedia.neofonie.de/browse/rdf-type:Scientist/birthDate-year~:1800~1900/nationality:Netherlands/).

Natuurlijk zijn semantische webstandaarden niet zaligmakend. Ze bieden bijvoorbeeld geen garantie dat de semantische informatie kwalitatief hoogwaardig is. Daarvoor zijn modelleertechnieken, zoals contextuele verbijzondering waarop het [Essence-consortium \(http://www.novay.nl/okb/projecten/essence/7781\)](http://www.novay.nl/okb/projecten/essence/7781) voortbouwt, relevant. Dit soort, veelal complementaire, technieken is gericht op precoördinatie; vóóraf worden semantische relaties stelselmatig afgestemd.[\*]

Aan de andere kant bieden semantische webstandaarden wel een open platform voor eenieder om wereldwijd betekenis en samenhang in informatie aan te brengen. De opzet staat toe dat de partijen kunnen voortbouwen op elkaars werk. Het web groeit daardoor eigenlijk uit tot een wereldwijde database. Geen relationele database met tabellen, maar een netwerk van data (oftewel [Linked Data \(http://en.wikipedia.org/wiki/Linked Data\)](http://en.wikipedia.org/wiki/Linked_Data)).

Binnen de Nederlandse publieke sector lijken er slechts een paar initiatieven te zijn rondom semantische webstandaarden. Zo is [Overheid.nl Web Metadata Standaard \(http://standaarden.overheid.nl/owms/\)](http://standaarden.overheid.nl/owms/) (OWMS) gebaseerd op met name [Dublin Core \(http://dublincore.org/\)](http://dublincore.org/) en ook op RDFa en URI. Het online [Nederlandse Wettenbestand \(http://www.overheid.nl/help/wr/deeplinks\)](http://www.overheid.nl/help/wr/deeplinks) werkt op basis van unieke identifiers, waardoor je duurzaam en gericht kan verwijzen naar wetgeving. En [Geonovum \(http://www.geo-info.nl/download/?noGzip=1&id=17669412\)](http://www.geo-info.nl/download/?noGzip=1&id=17669412) (PDF, 2,4 MB) verkent de mogelijkheden van het semantisch web voor Geo-informatie.

Wie kent andere bestaande initiatieven binnen de Nederlandse overheid? Wat zijn zinvolle nieuwe toepassingen? En welke beperkingen hebben deze standaarden? Ik ben zeer benieuwd naar jullie ideeën over een ronder web...

[\*] Over contextuele verbijzondering zal Marijke Abrahamse van Bureau Forum Standaardisatie binnenkort een blogbijdrage schrijven.

## 4 Reacties

1. *Rutger Lammers* zei op [23 februari 2011 om 10:02](#)

Interessante links in het artikel!

2. *Frank Verschoor* zei op **23 februari 2011 om 17:02**

[QUOTE] 'Informatie op het web is vaak door computers niet goed interpreteerbaar. Daardoor is de informatie ook minder goed vindbaar en toegankelijk voor mensen. Als je bijvoorbeeld zoekt naar afbeeldingen van Drenthe, dan krijg je zowel plaatjes van de provincie als van de voetballer te zien. Een ander voorbeeld is Boston, dat de naam is van verschillende steden maar ook van een rockband. Semantische webstandaarden kunnen dit soort problemen oplossen.' [EINDE QUOTE]

Ik vraag me echt af voor wie dit een probleem is? Als ik zoek naar plaatjes van de provincie Drenthe en Royston staat tussen mijn zoekresultaten, dan sla ik die gewoon over. Of ik voeg 'provincie' toe aan mijn zoekopdracht. Die interpretatie hoeft een computer voor mij niet te doen. Liever niet zelfs.

3. *Bart Knubben* zei op **23 februari 2011 om 21:02**

@Frank: Het voorbeeld 'Drenthe' is een relatief eenvoudig voorbeeld. Bij 'Boston' is het al wat ingewikkelder, omdat er (zoals in het filmpje terugkomt) 27 plaatsen, een band en basketbalclub zijn waarvoor deze naam gebruikt wordt.

Natuurlijk kan je in beide gevallen zelf handmatig filteren. Dat kost wel tijd en er bestaat een risico op fouten; bijvoorbeeld dat je treffers die betrekking hebben op de provincie toch wegfiltert en niet ziet. Dit soort kosten en risico's kunnen in bepaalde gevallen aanvaardbaar zijn (zoals in dit geval blijkbaar voor jou), maar er zijn gevallen waarbij dat minder aanvaardbaar is.

De overheid wil graag geautomatiseerd domein- en organisatie-overstijgend informatie uitwisselen. Semantische verwarringen kunnen daarbij leiden tot kostbare en zelfs gevaarlijke misverstanden. Denk aan verwarrende informatie over gevaarlijke stoffen. Of denk aan de gegevens in de basisregistraties. Voor de schrijfwijzen van straatnamen bestaan er bijvoorbeeld verschillende normen (onder meer BOCO en NEN). Als je verschillende schrijfwijzen voor straatnamen hanteert, dan kan dat leiden tot adresseringsfouten en koppelproblematiek. Binnen het stelsel van basisregistraties is men hier recentelijk tegenaan gelopen. Semantische webstandaarden kunnen erbij helpen om dit soort problemen op te lossen. Ze maken het mogelijk om entiteiten (zoals straatnamen) op een voor eenieder eenduidige en transparante wijze te definiëren en te koppelen.

4. *JoG* zei op **24 februari 2011 om 00:02**

Zie de groep 'Web 3.0: Semantisch Web' op Archief 2.0 -> <http://www.archief20.org/group/web30>

Verder is er een 'Linked Open Data' congres in voorbereiding door enkele mensen in Haarlem, onder wie Ad Aerts en Lukas Koster -> <http://sites.google.com/site/dclod11/>

"... Op 13 mei 2011 vindt de eerste Dutch Culture Linked Open Data event plaats in de Openbare Bibliotheek Haarlem, lokatie Centrum. Het doel van deze dag is het gebruik van linked open data in de culturele en bibliotheeksector in Nederland en Vlaanderen te bevorderen..."

Deze site is gemaakt met [WordPress Mu](#).

Tenzij anders vermeld, valt de content op deze site onder een [Creative Commons licentie](#).

NOiV is onderdeel van

