

Google en het semantische web

Door: Daniel Bos

Begin mei 2009 heeft Google aangekondigd bezig te zijn met het verrijken van de snippets die bij zoekresultaten getoond worden. Google heeft aangegeven hierbij gebruik te maken van onder meer Microformats en RDFa, beide standaarden die ontwikkeld zijn in het kader van het semantische web.

Het huidige internet bestaat uit een enorm aantal HTML documenten, vol met gestructureerde gegevens, die grotendeels ontoegankelijk zijn voor applicaties. De reden hiervoor is dat computers niet in staat zijn zelfstandig de semantische context van de gegevens te bepalen. Hierdoor kunnen zoekmachines wel de gegevens, maar niet de betekenis van deze gegevens vinden.

Semantische context

Wanneer zoekmachines de semantische context van gegevens kunnen bepalen, kunnen zij deze gegevens veel beter ordenen, en de zoekers van dienst zijn met relevantere resultaten. Wanneer in een document bijvoorbeeld de titel van een film genoemd wordt, ziet een zoekmachine dat het document iets met de film te maken heeft. Wat de zoekmachine niet weet is of het een recensie betreft, of misschien een quote uit de film of een samenvatting. Als binnen dit document links staan naar andere documenten, ziet een zoekmachine wel dat er naar deze documenten verwezen wordt, maar weet niet of zo'n link nu verwijst naar een vergelijkbare film, de homepage van een van de acteurs of een webshop waar de film gekocht kan worden.

Het semantische web wil hier verandering in brengen, door standaarden te specificeren die aangeven wat de betekenis is van een stuk tekst of een verwijzing. De twee standaarden met de meeste aanhang zijn RDFa en Microformats.

RDFa

RDFa staat voor *Resource Description Framework – in attributes* en is een toepassing van de RDF standaard op HTML documenten. RDF is een model dat de betekenis van gegevens weergeeft door deze te groeperen in zogenaamde *triples*. Deze triples bestaan uit een subject, een predicaat en een object. In de bewering "De lucht heeft de kleur blauw" bijvoorbeeld is het subject "de lucht", het predicaat "heeft de kleur" en het object "blauw".

RDFa definieert hoe deze relaties binnen een HTML document aangegeven kunnen worden. Om bijvoorbeeld aan te geven dat een stuk tekst binnen een document een recensie van een bepaalde film betreft, kan gebruik gemaakt worden van de volgende code:

```
<div xmlns:v="http://rdf.data-vocabulary.org" typeof="v:review">
  <span property="v:itemreviewed" about="http://url.van/film">
    Titel van film
  </span>
</div>
```

Microformats

Omdat het gebruik van RDFa in sommige gevallen nogal ingewikkeld is, de standaard erg rigide is, en het uitbreiden daarvan een tijdrovend proces is, zijn de microformats in het leven geroepen. Microformats bieden de mogelijkheid om zelf uitbreidingen te maken, zonder afhankelijk te zijn van standardisatie organisaties. Er is wel een Wiki met aanbevelingen, die als defacto-standaard dienst doet.

Google maakt al enige tijd gebruik van een van de microformats, namelijk de rel-nofollow tag, die aangeeft dat een document wel verwijst naar een ander document, maar dit andere document niet noodzakelijk aanbeveelt. Websites maken hier gebruik van om aan te geven dat links die in reacties geplaatst worden niet door de pagina aanbevolen worden. Google zal in dat geval ook geen PageRank overdragen via deze links.

Naast de rel-nofollow tag zijn er nog veel meer microformats. Om bijvoorbeeld de adresgegevens van een bedrijf op de website te tonen, kan gebruik gemaakt worden van de volgende code:

```
<div class="vcard">
  <div class="fn">Daniel Bos</div>
  <div class="org">Indenty B.V.</div>
  <div class="tel">053 8507505</div>
  <a class="url" href="http://www.indenty.nl/">www.indenty.nl</a>
</div>
```

Door gebruik te maken van de class tags in de elementen begrijpt een zoekmachine wat daarmee bedoeld wordt. Daarnaast zijn er ook microformats voor het specificeren van locaties, gebeurtenissen en sociale relaties.

Zoekmachine optimalisatie en semantische opmaak

Google heeft aangegeven geleidelijk meer en meer gebruik te gaan maken van semantische opmaak binnen pagina's. Het grootste deel van de pagina's op het internet zal hier echter geen gebruik van maken. Het is daarom niet zo dat pagina's die geen gebruik maken van deze standaarden geweerd zullen worden uit de zoekresultaten.

Aan de andere kant biedt het semantische web wel mogelijkheden voor het optimaliseren van pagina's. Google zal namelijk graag gebruik maken van de gegevens die op deze manier verzameld kunnen worden, en het toevoegen van deze gegevens aan pagina's belonen met bijvoorbeeld aangepaste snippets bij zoekresultaten. Bij een productpagina kan bijvoorbeeld de prijs van het product getoond worden, bij een recensie kan de gegeven score worden getoond met sterretjes en locatie gegevens kunnen gebruikt worden om het bedrijf zichtbaar te maken in Local Search of Google Maps.

Naar verwachting zal de semantische context van pagina's op termijn een steeds grotere invloed hebben op de algoritmen van zoekmachines, doordat de betekenis van de gegevens op het internet steeds begrijpbaarder wordt. Het verdient daarom de aanbeveling om zoveel mogelijk gebruik te maken van tenminste de verschillende microformats en, in dien mogelijk, van RDFa.

Bronnen

- <http://radar.oreilly.com/2009/05/google-announces-support-for-m.html>
- <http://googleblog.blogspot.com/2009/05/more-search-options-and-other-updates.html>
- <http://www.w3.org/TR/rdfa-syntax/>
- <http://en.wikipedia.org/wiki/Rdfa>
- <http://en.wikipedia.org/wiki/Microformats>