



Een mooier voorbeeld is dat van de BBC. Het semantische web wordt namelijk al gebruikt. De BBC had namelijk een probleem, ze wilden graag informatie over hun shows en productie op het internet zetten. Zo kan als iemand wil weten wie de presentator is van, bijvoorbeeld "Top Gear", op de website van BBC kijken om daar achter te komen. Het probleem komt naar voren als de gebruiker dan wil weten welke shows Jeremy Clarkson nog meer presenteert. De BBC kan onmogelijk alle mogelijk combinaties van de informatie op het internet zetten.

Wat ze gedaan hebben is in plaats van tekst en plaatjes aan te bieden, de data aan te bieden, zodat er door de gebruiker zelf de informatie kan worden samengesteld. Ook is deze informatie weer gekoppeld met andere datasets, zodat ook informatie die de BBC niet beschikbaar heeft, zeg de geboortedatum, toch via een koppeling met DBpedia de informatie kan aanbieden.

Een ander voorbeeld is dat van Reuters. Deze kwamen erachter dat hun berichten niet meer het meeste werden gelezen door mensen, maar door computers. Dat moet je je als volgt voorstellen: Reuters komt elke dag met grote hoeveelheden informatie zoals winstwaarschuwingen van bedrijven, of nieuws over nieuwe innovaties, welke door computers worden gelezen, om op basis daarvan beslissingen te nemen, zoals het kopen of verkopen van aandelen op de beurs.

Reuters wilde graag hun klanten van meer informatie voorzien, en hebben een systeem ontwikkeld dat naast het persbericht zelf, ook metadata aanlevert over het persbericht. Zo kan door een klant, door bijvoorbeeld op "Washington DC" te zoeken, toch bij een artikel over het witte huis komen, zonder dat het woord Washington in de zoekopdracht hoeft voor te komen.

### **Wat zijn de belangrijkste obstakels**

Er zijn 4 belangrijke punten waar aandacht aan moet worden besteed. Dat zijn schaalbaarheid, bedrijven, privacy & trust en heterogeniteit.

Schaalbaarheid is een probleem vanwege de gigantische hoeveelheid data er is. Nu al zijn er miljarden tripels aan semantische data, terwijl dat nog maar uit een klein aantal projecten voort is gekomen. De opslag daarvan kan niet door 1 enkele partij gedaan worden, gezien de natuur van het internet. Ons onderzoeksteam is daar overigens ook mee bezig, en werkt aan mogelijke oplossingen.

Het tweede problemen zijn bedrijven. Je moet je voorstellen dat de gemiddelde reactie van een manager bij een bedrijf zou zijn, waarom zou ik mijn data gratis open stellen voor anderen. Neem als voorbeeld amazon: andere bedrijven kunnen nu uitlezen wat de huidige prijzen zijn, en daar standaard onder gaan zitten. Je moet de bedrijven er dus van overtuigen dat het in hun voordeel is om data open te stellen, zodat daar andere projecten uit voor kunnen komen, wat weer in het voordeel is van het bedrijf.

Een ander belangrijk item is privacy & trust. Omdat de gegevens makkelijk toegankelijk worden, en gekoppeld kunnen worden, kan dit ook makkelijker voor de verkeerde doeleinden worden gebruikt. Dit is overigens meer een sociaal probleem, net als de overtuiging van de bedrijfswereld.

Het laatste, en naar mijn idee nog wel het grootste probleem is heterogeniteit. Als voorbeeld, als amazon bij als zijn boeken de auteur vermeld, en door een andere partij een lijst met schrijvers wordt samengesteld, dan kan deze data niet worden gekoppeld. Ook al gaat het (voor mensen) over het zelfde. Computers zien dit niet. Hoewel er veel onderzoek naar gedaan wordt, en we ook wel langzaam steeds dichterbij goede oplossingen komen, moet er nog een hoop gebeuren voordat dit probleem verdwenen is.

### **Hoe moet de semantische informatie er komen (door wie)**

*Om weer het voorbeeld van de dokter te gebruiken. Hoe moeten bijvoorbeeld de openingstijden van de praktijk online komen. Een secretaresse kan waarschijnlijk niet zelf de informatie op een "Web 3.0" correcte manier op het internet zetten.*

Dit klopt, maar dat is ook nog logisch gezien het semantische web nog maar in zijn beginstadium is. Waarschijnlijk komen er projecten die automatische de websites van het midden en kleinbedrijf doorlopen op informatie, die dit dat kunnen omzetten in data.

Waarschijnlijk zal er, naarmate er meer gebruik van wordt gemaakt, vanzelf toepassingen komen waarmee de secretaresse alsnog simpel de informatie in bijvoorbeeld het RDF formaat op het internet kan zetten. Maar dit is dat zal zich vanzelf uitwijzen.

### **Zijn er bepaalde maatregelen die de overheid moet treffen**

De overheid heeft maar 1 belangrijke taak, en dat is openheid. Deze moet zoveel mogelijk data in een open formaat aanbieden zodat andere partijen er mee kunnen werken. Een bekende uitspraak over open informatie is "Data wants to be Free". Er zijn overigens al wel een aantal overheden die hiermee bezig zijn. Zo wordt in spanje door de lokale overheid sommige informatie al in RDF formaat aangeboden. En ook in de verenigde staten zijn ze er mee bezig, maar dan voornamelijk in de hogere overheid. Nu al helemaal met Obama als president. Al zijn speeches, maar ook grote hoeveelheden informatie die de medewerkers van het witte huis produceren is beschikbaar in het RDF formaat. Dit wordt onder andere gekoppeld aan het FoaF project<sup>8</sup>

---

<sup>8</sup> <http://www.foaf-project.org/>