



Het semantisch web: wat is het en hoe werkt het?

Nastasia Vanderperren, PACKED vzw

Karel de Grote opzoeken op internet is eenvoudig. Je komt meteen terecht op zijn Wikipediapagina. Zoeken op ‘de kinderen van Karel de Grote’ is voor een zoekmachine veel complexer. Zonder het semantisch web gaat de zoekmachine hierbij op zoek naar teksten waarin ‘Karel de Grote’ en ‘kinderen’ vermeld worden. Je zal hoofdzakelijk records of pagina’s krijgen over Karel de Grote; records van zijn vele kinderen zullen grotendeels ontbreken.

Een zoekrobot kan geen verbanden leggen en zal de vraag niet correct interpreteren. Je zal de namen van zijn kinderen moeten kennen om de juiste records te vinden. Dankzij het semantisch web kunnen zulke vragen wel beantwoord worden.

Het semantisch web vormt een onderdeel van wat men vaak aanduidt als web 3.0 en bestaat uit een web van linked data. Web 1.0 was de fase in de ontwikkeling van het web waarin gebruikers enkel online informatie konden bekijken. In web 2.0 kunnen ze daarentegen zelf ook bijdragen toevoegen aan de inhoud van webpagina’s. In deze vorm is het web niet meer dan een verzameling documenten die door links met elkaar verbonden zijn. De inhoud (semantiek) van de documenten blijft voor webapplicaties onduidelijk. Web 3.0 is de trend waarbij men aan data een extra betekenislaag toekent zodat ze makkelijker kunnen worden gedeeld en hergebruikt. Computers kunnen verbanden leggen tussen verschillende informatiebronnen en complexe vragen beantwoorden. Concrete toepassingen zijn DBpedia en Wikidata, waarbij informatie in Wikipedia gestructureerd beschikbaar en machineleesbaar is en gekoppeld wordt met externe linked open datasets.

PERSISTENTE URI

Verbanden worden gelegd door de relaties tussen de verschillende informatiebronnen in metadata vast te leggen. Om een zoekmachine een verband te laten leggen tussen Karel de Grote en zijn kinderen, kunnen we dit doen door dit als eigenschap te definiëren. We definiëren de eigenschap ‘kind’ en geven hieraan als waarde de persistente URI (zie META 2013/5) die verwijst naar een kind van Karel de Grote, en doen dit voor al zijn kinderen. Op dezelfde

Karel de Grote < https://www.wikidata.org/wiki/Q3044 >	Vader < https://www.wikidata.org/wiki/Property:P22 >	Pepijn de Korte < https://www.wikidata.org/wiki/Q81212 >
Karel de Grote < https://www.wikidata.org/wiki/Q3044 >	Moeder < https://www.wikidata.org/wiki/Property:P25 >	Bertrada van Laon < https://www.wikidata.org/wiki/Q266751 >
Karel de Grote < https://www.wikidata.org/wiki/Q3044 >	Kind < https://www.wikidata.org/wiki/Property:P40 >	Pepijn de Gebochelde < https://www.wikidata.org/wiki/Q302212 >
Karel de Grote < https://www.wikidata.org/wiki/Q3044 >	Kind < https://www.wikidata.org/wiki/Property:P40 >	Lodewijk de Vrome < https://www.wikidata.org/wiki/Q43974 >

Een voorbeeld van een RDF-triple.

manier kan je ook bijv. zijn vader, moeder en echtgenotes vastleggen. Bij zijn kinderen definieer je dan de eigenschap ‘vader’, die als waarde de URI van Karel de Grote krijgt. Ook de relaties worden in URI’s uitgedrukt. Op deze manier krijg je een dataset waar de verschillende data met elkaar verbonden zijn. Deze kan je nog verder verrijken door ze te linken met externe linked open datasets, zoals de Getty Vocabularies, DBpedia, Geonames, Wikidata, etc.

Voor het semantisch modelleren van data wordt door het World Wide Web Consortium het gebruik van RDF¹ gepromoot. Dit is een standaardmodel voor het coderen van metadata. Het model gaat uit van drie onderdelen:

object-eigenschap-waarde. Het object is de bron die beschreven wordt (Karel de Grote), die een bepaalde eigenschap heeft (kind) en waarvan die eigenschap een bepaalde waarde heeft (Lodewijk De Vrome). RDF kan in verschillende formaten uitgedrukt worden, zoals XML, JSON-LD² en Turtle³.

Wanneer je nu via een SPARQL-endpoint⁴ ingeeft dat je op zoek bent naar de kinderen van Karel de Grote, krijg je een verwijzing naar de records of pagina’s van de verschillende kinderen. Semantische technieken maken het mogelijk om nieuwe verbanden te leggen en maken je data beter doorzoekbaar. ■

1 <https://www.w3.org/TR/REC-rdf-syntax/>

2 <http://json-ld.org/>

3 <https://www.w3.org/TR/turtle/>

4 Een SPARQL-endpoint stelt gebruikers in staat om een kennisbank te bevragen m.b.v. de SPARQL-taal. Het resultaat ontvangt men dan meestal in één of meerdere machineleesbare formaten. Een SPARQL-endpoint is meestal ontworpen als een machinevriendelijke interface naar een kennisbank; de bevraging dient door een machine/computer te kunnen gebeuren en niet door menselijke gebruikers.