

Visualisatie van datastromen

In aflevering vier van de serie over big data gaat Klaas Jan Mollema in op datavisualisatie. Waar een databestand op zichzelf staand lastig te interpreteren is, spreken visualisaties van data juist tot de verbeelding. Welke visualisatievormen en tools kunnen hierbij worden ingezet?

Klaas Jan Mollema MSc

In de eerste drie afleveringen spraken we over de ontwikkelingen die plaatsvinden aan de kant waar de meeste gebruikers geen weet van hebben: de serverkant. Deze aflevering is gericht op het eindproduct: de gebruikersvriendelijke toegang tot de data door middel van visualisatie-technieken. Visualisatie van informatie maakt onduidelijke data duidelijk en kan zelfs informatieproblemen oplossen.

‘Design is about solving problems and providing elegant solutions. Information design is about solving information problems’

– David McCandless TED 2010

Het visualiseren van informatie gebeurt al eeuwen. Holbewoners waren 20.000 jaar, dan klinkt eeuwenlang weer beetje weinig] Holbewoners maakten hun overwinningen via muurschilderingen kenbaar, kerken werkten met visuele representaties van bijbelse verhalen en atlanten maken sinds jaar en dag de wereld voor ons toegankelijk. De ontwikkeling van

rekenkracht, druktechniek en nieuwe visualisatie-technieken hebben dat – samen met de dataficatie van de samenleving – in een stroomversnelling gebracht.

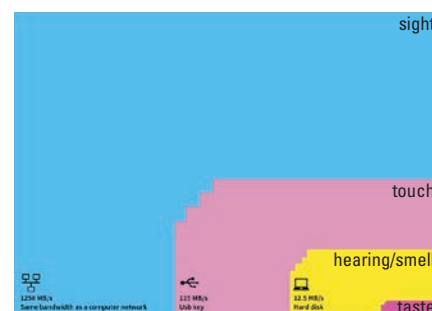
Datavisualisatie versus infographics

Visualisaties kennen een grote diversiteit aan vormen. Het belangrijkste verschil zit in de manier waarop het verhaal verteld wordt. Zo is een infographic een op zichzelf staande grafische uiting en biedt een wandeling door data en conclusies in de vorm van een verhaal. Een datavisualisatie is een ingeladen hoeveelheid gegevens, waar een interactieve visualisatie voor ontwikkeld is en waar je exploreren doorheen kunt gaan.

De vorm is heel bepalend voor de wijze waarop de inhoud van de informatie gecommuniceerd wordt. Een slechte visualisatie kan er – soms doelbewust – toe leiden dat de gebruiker verkeerde conclusies trekt. Het is een veelgebruikte manipulatie-techniek in nieuwsmedia en jaarverslagen.

Waarom datavisualisatie?

Datavisualisatie is een ideale manier om complexe concepten of hoeveelheden data uit te leggen, want de vorm ervan sluit goed aan op onze zintuigen. De Deense



Afbeelding 1. Tor Norretranders, Bandwidth of senses

fysicus Tor Norretranders vertaalde de bandbreedte van de zintuigen naar computertermen (afbeelding 1). Zicht is volgens Tor het snelste zintuig. In een infographic geeft hij met een klein wit vlakje rechtsonderin aan dat we maar 0,7 procent van de waarneming bewust binnenkrijgen. Bij het kijken naar gevisualiseerde informatie combineren we ons snelste zintuig – zicht – met taal, waardoor we de informatie nog makkelijker kunnen opnemen.

Kenmerken van goed informatieontwerp

Bij functioneel informatieontwerp is er gebruik gemaakt van betrouwbare data, waar op basis van de juiste analysetechnieken conclusies uit getrokken zijn. Die conclusies zijn vervolgens in een vorm gegoten die past bij de doelgroep van de

Tools

Voor het het vormgeven van informatie hebben we keuze uit een zeer groot scala aan tools, die variëren in mogelijkheden en kwaliteit. Maar kijk je specifiek naar bigdatavisualisatie, dan wordt de lijst beperkter. De reden: niet alle tools kunnen met grote hoeveelheden gegevens omgaan. Verder zijn er steeds grote verschuivingen in de aanbieders van dergelijke tools, zoals Gartner elk jaar laat zien in haar Magic Quadrant rondom tools voor businessanalyse (afbeelding 4). Terwijl de spelers Qlikview en IBM SPSS lange tijd de markt hebben aangevoerd, heeft Microsoft in twee jaar tijd een groeispurt gemaakt naar de positie van marktleider door sterk te investeren in de (vooral nog) gratis tool PowerBI.

Enkele tools op een rij:

- > *Microsoft Excel*. De klassieker op het gebied van dataverwerking en basale visualisatiemogelijkheden is Microsoft Excel. Zeker met inzet van plug-ins als PowerPivot en PowerQuery levert het een krachtige tool op. De visualisaties zijn echter vrij basic.
- > *Microsoft PowerBI*. Volgens Gartner op dit moment de marktleider in het aanbieden van visualisatietools van data. De gratis (!) tool PowerBI is verkrijgbaar in zowel een lokale variant als een clouddienst en biedt ongekende mogelijkheden op het gebied van bigdata-analyse. Waar Excel zich verslikt in de bestandsgrootte en complexiteit van big data, weet PowerBI daar wel mee om te gaan. De visualisaties zien er goed uit en PowerBI biedt zelfs functies om voorspellende analyses te doen op basis van historische data.
- > *Tableau*. Nog een marktleider op dit gebied is Tableau. Hoewel niet gratis, biedt Tableau het compleetste pakket aan voor data-analyse en bigdatavisualisatie. Voorspellende algoritmes zijn ingebouwd en de visualisaties zien er heel professioneel uit. Overigens kunnen studenten een gratis educatieve versie aanvragen.
- > *Google Data Studio*. De wereld staat niet stil, dus ook het Magic Quadrant van Gartner niet. Zo zou het zomaar kunnen dat volgend jaar niet Microsoft, maar Google rechtsboven in het Magic Quadrant vermeld staat met haar Data Studio. De hiermee gemaakte visualisa-

ties zijn iets basaler dan die met Tableau en PowerBI, maar Google weet zich snel verder te ontwikkelen. Data Studio is alleen beschikbaar als clouddienst.

- > *Kibana*. Deze krachtige visualisatietool voor Elasticsearch wordt veel gebruikt voor de ontwikkeling van *Business Intelligence Dashboards* waarbij big data in het spel zijn.
- > *IBM SPSS*. In de wetenschappelijke wereld is dit een veelgebruikte tool voor statistische analyse van data.
- > *Oracle Big Data discovery*. Ook Oracle biedt een visualisatietool voor big data, waarbij ze een totaalpakket van Hadoop, een data-opschoontool en een visualisatietool aanbieden.

voor iedereen bekende concepten (door bijvoorbeeld een grote hoeveelheid vloeistof te ‘vertalen’ naar de inhoud van een aantal zwembaden). Relateer getallen aan elkaar door op schaal te differentiëren in de grootte van een vorm (zoals Billion-dollar-o-gram, afbeelding 3). Tot slot pas je humoristische elementen toe, zodat de boodschap beter beklijft bij de gebruiker.

Iedereen kan visualiseren

Omdat nieuwe generaties steeds meer in een beeldcultuur opgroeien, is het visualiseren van informatie een vaardigheid geworden die niet alleen meer wordt opgepakt door grafisch ontwerpers. Het



Afbeelding 4. Gartner Magic Quadrant Business Analytics

Hoe te beginnen?

Om te kunnen werken met tools, volgt een aantal tips om te komen tot een goede informatievisualisatie. Allereerst moet je natuurlijk je data goed op orde hebben en weten wat de verschillende waarden en metadata betekenen. Verder analyseer je de doelgroep om een vorm te kiezen die aansluit op de beleving van de gebruiker. Ook hanteer je een goede opbouw en structuur, neem je de gebruiker mee door het verhaal in de data en breng je je boodschap duidelijk over. Gebruik waar nodig metaforen om data te relateren aan

betekenisvol maken van data was al een vaardigheid voor informatieprofessionals. Door te experimenteren met bovengenoemde tools kan deze beroepsgroep zichzelf vrij gemakkelijk aanleren hoe ze die betekenis kunnen visualiseren. Een vaardigheid die zowel de informatieprofessional als zijn werk meer zichtbaar zal maken in de organisatie. <

Klaas Jan Mollema MSc. (www.zijlmo.nl) is specialisatiecoördinator Business Data Management aan de opleiding Informatica van Hogeschool Leiden.