

CatVis: het visualiseren van bibliotheekgegevens

Op het KNVI Jaarcongres belicht Michel Westenberg, Assistent Professor aan de Technische Universiteit van Eindhoven, het zogeheten Visual Analytics for the World's Library Data-project. Het doel van dit project is om een toolkit te ontwikkelen ten behoeve van de visualisatie van bibliotheekgegevens.

Westenbergs collega Thom Castermans (promovendus) licht alvast het project toe.

Door: Edwin Mijnsbergen

Wat is CatVis?

'Het Visual Analytics for the World's Library Data-project, informeel ook wel "CatVis" (www.win.tue.nl/CatVis/) genoemd, heeft als doel om visualisaties van onder andere catalogusgegevens, primair bibliotheekgegevens, beschikbaar te maken voor onderzoekers in de geesteswetenschappen, beheerders van bibliografische data en de gebruikers daarvan. In het bijzonder werken aan het project filosofen van de Universiteit van Amsterdam en verschillende mensen van OCLC mee. Visualisaties zijn een krachtig middel om patronen te vinden in grote hoeveelheden data, maar ook om opvallende of mogelijk foutieve datapunten te vinden. Ze worden in sommige vakgebieden al veel gebruikt, bijvoorbeeld in de medische wereld. Denk dan aan scans. In andere vakgebieden niet, terwijl er wel potentie voor is. Wij vullen dat gat met het CatVis-project.'

Kun je een voorbeeld van zo'n visualisatie geven?

'Neem bijvoorbeeld de visualisatie van data uit WorldCat, die als illustratie bij dit interview dient. WorldCat is een enorme catalogus van bibliotheekgegevens, die wordt beheerd door OCLC. De visualisatie in de achtergrond toont publicaties over de Italiaanse renaissance.'

'We hebben er na een ontwerpproces voor gekozen om voor elke publicatie een symbool op de wereldkaart te tekenen. Dit is simpelweg een klein vierkantje, waarvan de kleur aangeeft wanneer het boek gepubliceerd is en de locatie overeenkomt met die van de uitgever. De kleur is een van vijf tinten paars. Des te donkerder de tint is, des te recenter het werk is gepubliceerd.' 'De grote kracht van de visualisatie is dat we vervolgens symbolen die overlappen samenvoegen, zodat alle symbolen goed leesbaar zijn. We doen dit niet handmatig, maar met een speciaal daarvoor ontworpen algoritme, dat we geïmplementeerd hebben en dat door een computer uitgevoerd wordt. We ontwerpen een soort recept dat door de computer gerealiseerd kan worden, en bewijzen dat het werkt. In dit geval betekent dat dat alle overlap gedetecteerd en opgelost wordt. Daarna implementeren we het zo efficiënt mogelijk in de praktijk, wat een uitdaging op zichzelf is.'

'Deze visualisatie werkt online, ongeveer zoals Google Maps werkt. Het is mogelijk om in en uit te zoomen en de kaart te verschuiven. Bij inzoomen zullen samengevoegde symbolen uit elkaar vallen, zodat meer detail zichtbaar wordt. De visualisatie is te bekijken op glammap.net.'

Wat heeft de 'gewone' gebruiker eraan?

'Patronen op een kaart kunnen in een oogopslag gezien worden. In een tekstuele interface is dit veel moeilijker. Hetzelfde geldt voor afwijkende datapunten. Een voorbeeld van een patroon in bovengenoemde visualisatie is dat te zien is dat er veel gepubliceerd is in Italië.'
'Als een publicatie een verkeerde locatie heeft in de data en daardoor in zee of op de noordpool belandt is dat natuurlijk duidelijk te zien, terwijl het in een tekstuele interface niet direct opvalt.'

Van wanneer tot wanneer loopt het project en wat zijn de voorlopige bevindingen?

'Het project is in september 2015 begonnen en zal eindigen in september 2019. Er is een nieuw type visualisatie ontworpen en er zijn al ideeën voor verder werk. De visualisatie richt zich op het tonen van zoekresultaten en zal nog worden gepubliceerd. Het project focust zich nu op het bouwen van tools die helpen bij het opschonen van data. Denk daarbij aan een interface waarbij gebruikers fouten kunnen vinden en corrigeren. Ook willen we gebruikers in staat stellen om data te verrijken, bijvoorbeeld door datasets te combineren.'

Leent WorldCat zich als catalogus voor dit project?

'WorldCat biedt een schat aan informatie en is in die zin uitstekend geschikt voor ons project. Natuurlijk is WorldCat niet perfect. Dat is echter geen probleem, want een van de doeleinden waar visualisaties geschikt voor zijn, is het inzichtelijk maken van fouten en onvolledigheden in de getoonde data. Een van onze projectdoelen is dan ook om tools te bouwen waarmee de kwaliteit van onder andere WorldCat verbeterd zou kunnen worden.'

TRACK 2

VISUALISEREN VAN DATA EN INFORMATIE

Michel Westenberg: Visualisaties van bibliotheekdata: het project CatVis