

Bij de afdeling geschiedenis van de UvA ging jarenlang een wisseltrofee rond onder de deelnemers aan een onderzoek naar verzuiling. De trofee was voor degene die de oudste vindplaats vond van het woord ‘verzuiling’ – een maatschappelijk verschijnsel dat in de eerste helft van de 20ste eeuw opkwam. Toen de Koninklijke Bibliotheek in 2010 de historische krantensite lanceerde, dacht de jonge onderzoeker Bram Mellink de trofee snel te scoren door verzuiling in te tikken. Hij vond meer dan 60 oudere vindplaatsen – de oudste in 1781, en dus een flagrant anachronisme. Uit de originele teksten bleek dat alle oudere dateringen leesfouten waren van het computerprogramma dat de tekens uit de kranten moest herkennen. In de teksten staan woorden als vervulling, vervuiling, verzaking. Deze anekdote deed de ronde tijdens een expertmeeting op 12 september, georganiseerd door de leerstoelgroep politieke geschiedenis van de Radboud Universiteit Nijmegen. De bijeenkomst draaide om de vraag hoe ‘e-Humanities’ binnen het vakgebied kunnen worden ingezet. E-Humanities of digitaal geesteswetenschappelijk onderzoek is een enigszins problematisch begrip, omdat nog niet is uitgekristalliseerd wat eronder wordt verstaan en wat onderzoekers eraan (kunnen) hebben. De voordelen van digitale bronnen zijn evident: de computer kan relevante begrippen opzoeken zodat je de teksten niet meer integraal hoeft te lezen. Waarbij onderzoekers voortdurend de kwaliteit van de bron in de gaten houden, zo blijkt uit het voorbeeld van verzuiling. Digitale bronnen vragen om computationele analyses, en daar ligt het knelpunt. Dit vergt technische en statistische kennis die veel geesteswetenschappers zelf niet bezitten, en waarvoor ze dus moeten samenwerken met technici.

COLUMN

Computer spit door politiek

NICOLINE VAN DER SIJS



Huib Zuidervaart van het Huygens ING gaf tijdens de expertmeeting een mooi voorbeeld van de problemen die computeranalyses kunnen opleveren. Binnen het project Geleerdenbrieven hebben onderzoekers brieven van 17de-eeuwse natuurwetenschappers verzameld, te vinden op het in juni gelanceerde ePistolarium (zie bit.ly/16l8bxt). Met zogenoemde topic modeling-software lieten de onderzoekers de computer uitrekenen welke brieven over hetzelfde onderwerp gaan. De computer tekende prachtige plaatjes van aan elkaar gerelateerde brieven. Maar toen de onderzoekers de originele bronnen bekeken, bleken die plaatjes niet door de data te worden ondersteund. Inmiddels hebben ze met behulp van een ander topic modeling-programma wél relevante verbanden binnen de geleerden correspondentie weten te leggen. Computeranalyses werken het best als het materiaal van tevoren is voorbereid. Maar dat kost tijd, waaraan onderzoekers toch al gebrek hebben. Gelukkig zijn politiek-historici bevoorrecht. Ze beschikken over grote verrijkte digitale bestanden: alle Handelingen (Kamerverslagen, Kamerstukken en Kamervragen) van 1814 tot heden zijn te vinden op StatenGeneraalDigitaal.nl en Overheid.nl en verrijkt binnen het Political Mashup-project.

Onlangs hebben Jaap Kamps en Maarten Marx van de UvA een omvangrijke subsidie van NWO Creatieve Industrie gekregen voor ExPose (Exploratory Political Search). Met deze subsidie gaan zij de gedigitaliseerde Nederlandse parlementaire geschiedenis gereedmaken voor een nieuwe manier van (onder)zoeken, die zij noemen: verkennend zoeken. Hiervoor koppelen ze bronnen aan elkaar op basis van ankerpunten als tijd, plaats, spelers

en onderwerpen. Een voorbeeld. In de Handelingen is vaak sprake van de voorzitter, of de woordvoerder van de PvdA. Dit verwijst dan naar een eerdere passage waar gemeld is wie de voorzitter of woordvoerder is. Kamps en Marx hebben een computerprogramma geschreven dat automatisch iedere verwijzing relateert aan de bedoelde persoonsnaam, die weer is gekoppeld aan de partij die deze op dat moment vertegenwoordigt. Met deze gegevens kunnen allerlei netwerken worden vastgesteld: welke politici reageren regelmatig op elkaar, hoeveel spreektijd eisen de verschillende partijen over een bepaald onderwerp op, wat verwezenlijken partijen van hun verkiezingsbeloftes.

Uit de Nijmeegse discussies kwam naar voren dat veel historici nog niet goed weten hoe ze e-Humanities het beste kunnen inzetten in hun onderzoek en tot welke nieuwe onderzoeksvragen dit kan leiden. Die problematiek is internationaal. Eind vorig jaar publiceerde het New-Yorkse onderzoeksbureau Ithaka S+R het rapport ‘Supporting the Changing Research Practices of Historians’ (bit.ly/18QJYJu). Zij concluderen dat de geschiedwetenschap zich in een overgangperiode bevindt: terwijl het aantal nieuwe digitale bronnen enorm toeneemt, maakt slechts een klein aantal onderzoekers gebruik van nieuwe methoden van computeranalyse. De twee belangrijkste digitale onderzoeksmethoden die Amerikaanse historici toepassen, zijn geografische informatiesystemen, waarmee netwerken in ruimte en tijd kunnen worden afgebeeld, en textmining: het slim zoeken naar begrippen in digitale bestanden. E-Humanities bieden echter meer mogelijkheden. Om die beter te kunnen benutten, beveelt Ithaka S+R aan om met name jonge historici beter te scholen in digitale onderzoeksmethoden. Die aanbeveling kan zo voor Nederland worden overgenomen.