



Het nut van computerlezen

In 2009 noemde de Amerikaanse hoogleraar William Pannacker de *digital humanities* ‘the next big thing’. In de academische wereld zocht de term ‘digitale geesteswetenschappen’ rond sinds 2004. In dat jaar verscheen bij Blackwell A *Companion to Digital Humanities*. Dit handboek bevatte een overzicht van het onderzoek dat binnen de geesteswetenschappen met behulp van computertechnologie plaatsvindt.

Natuurlijk werd de computer al vóór 2004 binnen de geesteswetenschappen ingezet. In de jaren 60 en 70 van de vorige eeuw was Hugo Brandt Corstius in Nederland de pionier van de ‘computertaalkunde’. In de jaren 80 werd ‘alfa-informatica’ of ‘computationele taalkunde’ in de curricula van de meeste universiteiten opgenomen. Men probeerde in deze jaren de computer te leren hoe hij taal kon begrijpen en spreken. De onderzoekers die zich hiermee bezighielden, hadden allemaal een technische achtergrond, vaak gecombineerd met een taalkundige opleiding.

Dat in 2004 gekozen werd voor een naamsverandering – ‘digitale geesteswetenschappen’, maar ook *e-Humanities* – was deels een imago-kwestie. De nieuwe namen benadrukken dat het geesteswetenschap-

pelijk onderzoek centraal staat en computertechnologie daarvoor dienstbaar wordt gemaakt. Op die manier hoopte men dat meer geesteswetenschappers zich tot de nieuwe leer zouden bekeren. Want niet iedereen stond en staat te juichen bij het feit dat technologie in de geesteswetenschappen een steeds grotere rol krijgt.

Die grotere rol is echter onvermijdelijk: in de 21ste eeuw hebben wetenschappelijke instituten en bibliotheken massaal digitale bestanden beschikbaar gemaakt. Onderzoekers moeten daardoor omschakelen: ze kunnen niet langer alleen maar informatie uit teksten destilleren door deze geduldig te lezen. De Digitale Bibliotheek voor de Nederlandse Letteren alleen al bevat vier miljoen pagina’s. Om die aan te kunnen zijn computers nodig: die kunnen in een paar seconden de opkomst en verbreiding van begrippen als ‘pamflet’ of ‘televisie’ in deze *big data* tonen. Literatuuronderzoekers van The Stanford Literary Lab hebben voor dit computerlezen de term *distant reading* gemunt – als tegenhanger van het oude *close reading*.

Distant reading levert vaak veel treffers op. Die moeten op waarde worden geschat. Taalgegevens zijn per definitie *fuzzy* of ruizig: ze kun-

nen niet zonder meer door een computer worden geïnterpreteerd. Als je bijvoorbeeld een woord in oude teksten zoekt, zul je rekening moeten houden met de vele spel- en vormvarianten die het in de loop van de tijd heeft gekend. Als je iets wilt zeggen over de veranderende opvattingen over ‘slavernij’, dan moet je niet alleen bekijken hoe vaak er over dit onderwerp is geschreven, maar ook met welke gevoelswaarde. Je kunt dat doen door de computer te laten tellen hoe vaak ‘slavernij’ met positief dan wel negatief geladen bijvoeglijke naamwoorden is gecombineerd.

In de digitale geesteswetenschappen is enige kennis van statistiek en computertechnologie vereist. Technici hanteren niet voor niets de geveugelde uitdrukking ‘*to err is human, to really foul things up requires a computer*’. Omdat de meeste geesteswetenschappers niet veel kaas gegeten hebben van technologie, zoeken ze samenwerking met technici.

Sommige geesteswetenschappers denken dat de digitale technologie zal leiden tot een wetenschappelijke revolutie. Anderen zien de computer alleen als handig hulpmiddel. En nog weer anderen houden zich verre van iedere kwantitatieve en digitale benadering. Die laatste groep ziet e-

Taalgegevens zijn *fuzzy*: een computer kan ze niet zo maar interpreteren

Humanities vaak als een bedreiging. En dat is niet onbegrijpelijk want overal kiezen politiek en wetenschap voor de inzet van ICT, en sturen daar de geldstromen heen.

In buitenlandse media woeden al heftige discussies over e-Humanities. Op 28 oktober vorig jaar schreef Stephen Marche in de *Los Angeles Review of Books*: ‘*Literature is not Data: Against Digital Humanities*’. Op 4 januari van dit jaar boog een commissie van de Modern Language Association zich over ‘*The Dark Side of the Digital Humanities*’. Er werd met veel scepsis gesproken over de ‘*DH bubble*’: de aanhangers van die zeepebel zouden lijden aan ‘*techno-utopianism*’. Als tegengeluid gaf William Pannacker op 18 februari in *The Chronicle of Higher Education* tien adviezen om het imago van e-Humanities op te vijzelen. Het eerste is: ‘*Stop calling it Digital Humanities*’.

In Nederland zet de Koninklijke Nederlandse Akademie van Wetens-

schappen al enkele jaren in op e-Humanities. In 2010 is de KNAW e-Humanities Group opgericht. Binnen het op te richten KNAW Humanities Center moet technologie een verbindende factor worden en zorgen voor vernieuwing in het onderzoek. De ambitie is hoog: het Center moet tot de internationale top gaan behoren, aldus het KNAW-rapport dat eind vorig jaar verscheen. Twee weken geleden vond op het NIAS een geanimeerde tweedaagse workshop plaats over de mogelijkheden die e-Humanities bieden voor samenwerking met het bedrijfsleven.

Dat de Nederlandse academische wereld veel e-Humanities-activiteiten ontplooit, is nauwelijks naar de buitenwereld doorgesijpeld. In de krantenbank LexisNexis komt ‘digitale geesteswetenschappen’ of ‘e-Humanities’ welgeteld viermaal voor, waarvan tweemaal in een stuk van of over de onvermoerbare pleitbezorger Rens Bod, hoogleraar aan de Universiteit van Amsterdam. Het is te hopen dat kranten snel e-Humanities als *big thing* oppikken. In ieder geval zal ik op deze plaats zo af en toe rapporteren over interessante resultaten. Want ik vind de mogelijkheden en perspectieven van e-Humanities razend interessant en veelbelovend.