

Wat is het verschil tussen een goede en een slechte roman? Misschien dat de computer er ooit achter komt. In de alfa- en gamma-wetenschappen valt er digitaal nog veel te ontdekken.
Warna Oosterbaan



Prof. Sally Wyatt leidt de e-humanities groep van de KNAW

Het voordeel van die *e* in e-humanities is dat hij zoveel kan betekenen: *electronic, enhanced, experimental*. Al die termen komen aan bod als je met Sally Wyatt praat over dit vakgebied. Wyatt (1959), van oorsprong Canadese, woont sinds een jaar of tien in de lage landen en spreekt vlekkeloos Nederlands. Ze is van opleiding econoom en een expert op het gebied van de maatschappelijke gevolgen van technologische veranderingen. Onder haar leiding is een jaar geleden het e-humanities programma van de KNAW van start gegaan.

Maar wat zijn e-humanities?

We spreken Wyatt in het opvallende gebouw aan de rand van Amsterdam waar ook het Meertens-instituut is gehuisvest. Ooit was dit modernistische beton het hoofdkantoor van Coca-Cola. Nu staat in de hal vlak bij de entree als een kolossaal relikwie een houten bureau met talloze laatjes, deurtjes en aflegplanken. Het is *Het* bureau, waaraan P.J. Meertens, de vroegere directeur van het Instituut voor Dialectologie, Volks- en Naamkunde zijn werk deed, en dat vereeuwigd werd in de befaamde romancyclus van J.J. Voskuil.

Patronen

De e-humanities behoren tot een heel andere wereld. “In het algemeen bedoelen we daarmee het gebruik van digitale *tools* en technieken in de alfa- en gammawetenschappen”, zegt Wyatt. “En eigenlijk zijn er twee richtingen. Je kunt denken aan het gebruik van digitale technieken om data te analyseren,

om patronen zichtbaar te maken in grote hoeveelheden onderzoeksgegevens. Een andere richting is e-humanities als een onderwerp binnen de wetenschapscultuur: hoe maken onderzoekers gebruik van digitale *tools*?

Hoe beïnvloeden internet, e-mail en de tekstverwerker de wetenschapsbeoefening?” Wyatt en haar staf hebben van de KNAW vijf jaar de tijd gekregen. “Ons doel is de samenwerking tussen verschillende instituten en universiteiten te bevorderen, en om van elkaar te leren.”

Want hoe gewoon de computer in het leven van de modale historicus, socioloog, taalkundige of musicoloog inmiddels ook is, dat je er mee kunt dan een artikel schrijven of e-mailen is bij hen veel minder bekend. Dat literatuurwetenschappers er de authenticiteit van een manuscript – heeft Shakespeare dit toneelstuk werkelijk geschreven? – mee kunnen vaststellen en dat archeologen met behulp van digitale technieken een aannemelijke simu-

latie van een prehistorisch dorp kunnen maken is inmiddels al bewezen. Maar dat je verborgen structuren en patronen uit historische bronnen kunt lichten, dat je onverwachte samenhangen en ontwikkelingen in liedteksten kunt zien en dat je wellicht ooit kunt aantonen dat het verschil tussen een goede en slechte roman niet alléén een kwestie van slechte kritiek is – dat zijn óók beloften van de e-humanities.

Volksliedjes

Voorlopig ligt binnen het project van de KNAW het accent op drie projecten waarbinnen steeds drie onderzoekers zijn aangesteld. Het eerste heet *Tunes & Tales*. In samenwerking met drie universiteiten, het Meertens Instituut en de Fryske Akademy wordt in een groot corpus van volksliedjes en volksverhalen gezocht naar de ontwikkeling van de vormen en de thema's in de tijd. Zo is misschien een model te maken van de wijze waarop een orale traditie zich ontwikkelt. En, ook niet onbelangrijk, zo kan een begin worden gemaakt met de auto-

matische classificering van het corpus – en dat is weer erg handig om de collectie toegankelijk te maken voor etnomusicologisch onderzoek. Al even ambitieus is het project *The Riddle of Literary Quality*, waaraan de Universiteit van Amsterdam, de Fryske Akademy en het Huygens-ING meedoen. “Kan een computer zien of een roman goed of slecht is? Natuurlijk zijn sociale en culturele factoren erg belangrijk om de ontvangst van een literaire tekst te verklaren”, zegt Wyatt. “Maar we nemen aan dat formele aspecten ook een rol spelen. Het gebruik van moeilijke woorden, een gecompliceerde grammatica, het gebruik van adjectieven, etcetera. Dat kan een computer betrekkelijk gemakkelijk analyseren.”

Het derde project is ook een poging in een groot aantal bronnen structuur te zien. In CEDA_R proberen het Internationaal Instituut voor Sociale Geschiedenis in Amsterdam, data-instelling DANS en de Amsterdamse Vrije Universiteit met behulp van historische economische gegevens uit verschillende landen verbanden te zien tussen macro-economische veranderingen, de levens van individuele burgers, politieke systemen, arbeidsmarkten en welvaart. Gegevens uit Nederlandse volkstellingen zullen worden gebruikt om de fundamentele te leggen voor een ‘semantisch web’ dat een antwoord kan geven op de vraag hoe al die factoren elkaar beïnvloeden.

Voor de projecten vormen Sally Wyatt en haar medewerkers het centrale punt. In de ruime kamers waar ooit de Coca-Colabestuurders hun frisdrankstrategie bepaalden, komen nu twee keer per week de onderzoekers bij elkaar om te vergaderen, te brainstormen en gegevens uit te wisselen. Want hoe *e* de wetenschap inmiddels ook is, zonder persoonlijke contacten kun je het nog steeds moeilijk stellen.

<http://ehumanities.nl>

CLARIN krijgt ERIC-status

Taaltechnologieproject CLARIN is erkend als European Research Infrastructure Consortium (ERIC), een rechtspersoon voor onderzoeksinfrastructuur. EC-voorzitter Barroso zal de CLARIN ERIC binnenkort ondertekenen. Nederland en acht andere landen en organisaties zijn vooralsnog de stichtende leden van de CLARIN ERIC.

Binnenkort starten ook de projecten die gehonoreerd zijn binnen de derde oproep van CLARIN-NL. Geïnteresseerden mochten voorstellen indienen om bestaande data en toepassingen uit de geesteswetenschappen naar CLARIN-standaarden om te zetten, specifieke onderzoekers werden uitgenodigd een voorstel in te dienen om de spreiding over de verschillende geesteswetenschappen te garanderen. Vierentwintig onderzoekers hebben een voorstel ingediend. In totaal zijn dertien projecten gehonoreerd. (ER)

Wurdboek in Taalbank INL

Het volledige *Wurdboek fan de Fryske Taal* (WFT) is opgenomen in de Geïntegreerde Taalbank van het Instituut voor Nederlandse Lexicologie (INL). De benodigde aanpassingen en integratie in de Taalbank zijn financieel mogelijk gemaakt door CLARIN-NL. De Fryske Akademy begon in 1938 aan het Wurdboek, in 2011 is het vijftiengste en laatste deel gepresenteerd. Het WFT bevat zo'n 115.000 lemma's met hun betekenis, uitspraak, dialectische varianten, vervoegingen, zegswijzen en etymologie. De nieuwste woorden in het overzichtswerk stammen uit 1975 – het woord kompjüter staat er nog net in. (ER)

<http://gtb.inl.nl>

Het overzicht toont een aantal databestanden die recent voor onderzoekers beschikbaar zijn gekomen bij CBS, CentERdata, Huygens ING (Huygens Instituut en Instituut voor Nederlandse Geschiedenis) en DANS.

Centraal Bureau voor de Statistiek

Een volledig overzicht van de CBS-bestanden staat op www.cbs.nl/microdata

- Waarde onroerende zaken, 2009-2010
- Productiestatistiek: Autohandel, 2009; Bouwnijverheid, 2009; Commerciële diensten, 2009; Delfstoffenwinning 2009; Detailhandel, 2009; Energie, 2009; Groothandel, 2009; Industrie, 2009; Transport, 2009
- Algemene Nabestaandenwet, 2001-2006
- Bijstandsuitkeringenstatistiek, Registraties, 2001-2006
- Gemeentelijke Basisadministratie Personen, 1995-2011
- Mobiliteitsonderzoek Nederland, 2009
- Integrale veiligheidsmonitor, 2010
- ICT huishoudens en personen, 2008
- Onderzoek verplaatsingsgedrag in Nederland, 2010

SINDS KORT BESCHIKBAAR

CentERdata LISS Data Archive

De LISS-bestanden zijn kosteloos beschikbaar via www.lissdata.nl/dataarchive

Studies LISS panel

- Validating the Dutch SF-6D and EQ-5D Using Pairwise Comparisons and Best-Worst Scaling (Jonker, M., Donkers, B., Bekker-Grob, E. De), april 2011
- Religion and Ethnicity - Wave 4 (CentERdata), januari 2011/februari 2011
- Personality - Wave 4 (CentERdata), mei 2011/juni 2011
- The public's opinion on the control of terrorism: Attitudes and willingness-to-pay - Wave 2 (Wilsen, J. v., Woude, M. v.d.), maart 2011
- The Impact of Style and Rhetoric on the Perception of Right-Wing Populist Leaders (Bos, L.), februari 2011
- Social Integration and Leisure - Wave 4 (CentERdata), februari 2011/maart 2011
- Civic Participation, (Ingen, E. J. v.), mei 2011
- Alcohol Use and Coping with Stress - Wave 1, (Crutzen, R.), januari 2011

- Alcohol Use and Coping with Stress - Wave 2, (Crutzen, R.), april 2011
- Telephone use and regional elections, (CentERdata), maart 2011

Studies Immigrant panel

- Religion and Ethnicity - Wave 1 (CentERdata), januari 2011
- Personality - Wave 1 (CentERdata), mei 2011
- Action Control Scale (ACS-90) (Chasiotis, A., Bender, M., Vijver, F. van de), juni 2011

Huygens ING

De bestanden staan op www.historici.nl

- Weenske gezantschapsberichten van 1670 tot 1720 (2 delen)
- Holland bestuurd. Teksten over het bestuur van het graafschap Holland 1299-1567
- Bronontsluiting voor historisch onderzoek
- Rijkskroniek van Holland (366-1305)
- Waalse kerken 1601-1697
- Overzicht van de door bronnenpublicatie aan te vullen leemten der Nederlandse geschiedkennis

- Dagboek aantekeningen van vice-admiraal F. Pinke, commandant zeemacht in Nederlands-Indië 1914-1916, 1 deel
- Handel op de Oostzee 1122-1499, 6 delen
- De Vroedschap van Amsterdam, 1578-1795, 2 delen
- Relazioni Veneziane. Venetiaansche berichten over de Vereenigde Nederlanden van 1600-1795
- Dagboek Egbert Alting 1533-1594
- De Nederlandse kerk in Londen 1569-1585

DANS EASY

De bestanden zijn kosteloos beschikbaar via <http://easy.dans.knaw.nl>

Sociale Wetenschappen

- Data - Voer voor psychologen, archivering, beschikbaarstelling en hergebruik van onderzoeksdata in de psychologie 2010 (C. Voorbrood en H. van Luijn - DANS)
- Onderzoek Verplaatsingen in Nederland 2010 - OVIN (CBS - Rijkswaterstaat)

Ruimtelijke Wetenschappen

- TOP10NL 2010 (Kadaster)
- Bestand Bodemgebruik 2008 (CBS - Kadaster)