

BiographyNED scheept orde in chaos

Kan een analytische tool voor het Biografisch Portaal ons verleden ordenen door mensen aan gebeurtenissen te koppelen? Els Kloek en Susan Legêne

In feite is ons verleden chaos: een opeenhoping van ontelbare gebeurtenissen, waar geschiedschrijvers iets zinnigs van proberen te maken. Het duiden van het verleden is zowel het uitdagende als het aantrekkelijke van de geschiedwetenschap, vinden historici. De vraag is: kunnen moderne technieken (*tools*) worden ontwikkeld bij pogingen om orde in die chaos te scheppen, bijvoorbeeld door verbindingen te tonen tussen mensen en gebeurtenissen? Of kunnen we technieken ontwikkelen om in de gegevens uit het verleden netwerken, patronen of relaties te ontdekken die het eenvoudige menselijke oog nooit kon ontwaren?

Dankbaar gebruik

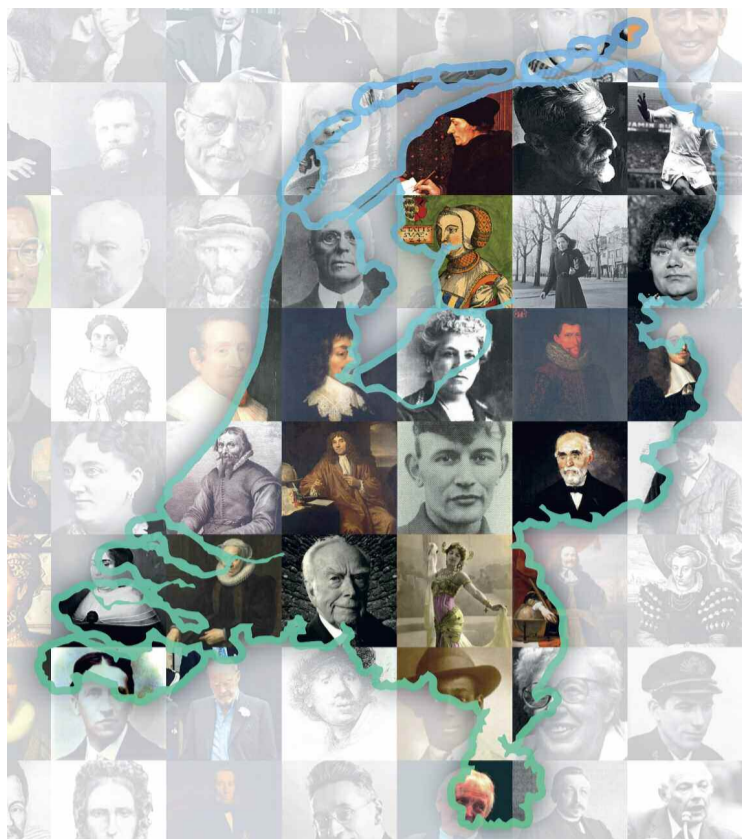
Dit zijn vragen waar het onderzoeksprogramma BiographyNED, onlangs gehonoreerd door het nieuwe Netherlands eScience Center (NleSC, zie rubriek Focus), zich de komende drie jaar mee wil gaan bezighouden. Het project maakt daarbij dankbaar gebruik van de tienduizenden biografieën die eerdere geschiedschrijvers bijeenbrachten in gespecialiseerde naslagwerken. Het e-Science Centre wil interdisciplinair onderzoek bevorderen dat

effectief gebruik maakt van bestaande data en technieken en een bredere inzet van ICT technologieën mogelijk maakt. Die doelstelling sluit goed aan bij BiographyNed, waarin computerwetenschappers en historici van de VU, het Huygens ING en e-Science Centre zullen samenwerken.

Meer mogelijkheden

Het onderzoeksprogramma BiographyNed neemt het Biografisch Portaal van Nederland als uitgangspunt. Dit portaal, dat in 2010 werd gelanceerd, verbindt een groot aantal online en/of gedigitaliseerde biografische naslagwerken en andere datasets met elkaar. De aldus verzamelde biografische informatie werd in verschillende tijden verzameld en geordend. Sommige biografische collecties gaan over specifieke groepen, zoals kunstenaars of predikanten, andere hebben een meer algemene noemer als 'bekende Nederlanders' of 'vrouwen'. Het Biografisch Portaal verbindt momenteel deze gegeven via een beperkt aantal metadata. Doel van BiographyNed is de vergroting van de mogelijkheden van het portaal voor historisch onderzoek. Welke geschiedenis duikt op uit die meer dan 100.000 persoonbeschrijvingen en portretten?

Meer specifiek wil BiographyNed gaan onderzoeken of uit het Biografisch Portaal gegevens over mensen en gebeurtenissen, hun reizen en verblijfplaatsen en hun netwerken aan elkaar kunnen worden ge-



63.000 Nederlanders op biografischportaal.nl illustratie Jelle Hellinga

Historici en informatici werken aan BiographyNED

Prof.dr. Guus Schreiber, Informatica, Faculteit der Exacte wetenschappen, VU; **Prof.dr. Piek Vossen**, Taal en Communicatie, Letteren Faculteit VU; **Prof.dr. Susan Legêne**, Geschiedenis, Letteren Faculteit VU; **Dr. Els Kloek**, Biografisch Portaal, Huygens ING; **Dr. Jelle Gerbrandy**, Biografisch Portaal, Gerbrandy Software

koppeld. Als het portaal dit met ICT inzichtelijk kan maken, vinden we misschien antwoord op een breed scala aan vragen over elitevorming in Nederland. Wat waren de knooppunten waar mensen met macht en invloed elkaar tegenkwamen, in welke netwerken functioneerden zij? Met dergelijke vragen als sturing ontwikkelt BiographyNed een 'project demonstrator' die een semantische laag over de zeer uiteenlopende data van het huidige Biografisch Portaal heenlegt. Dat

begint met een pilot gebaseerd op een kwalitatieve selectie van personen, waarvoor ondermeer de virtuele portretgalerij die het Rijksmuseum ontwikkelt, het uitgangspunt vormt. Wie geschilderd werd zit in de regel immers ook opgeslagen in het Biografisch portaal. Het project zal in het najaar van 2012 beginnen; vacatures voor twee post-docs zijn door de VU in mei gepubliceerd.

www.biografischportaal.nl

GELEZEN

Users, narcissism and control tracking the impact of scholarly publications in the 21st century

door Paul Wouters en Rodrigo Costas

In dit rapport analyseren de auteurs zestien tools en methodologieën om de impact van wetenschappelijke publicaties te meten. Voorbeelden zijn Google scholar en Microsoft Academic research. Met hulp van deze tracking tools is te volgen hoe onderzoeksuitkomsten gelezen, geciteerd, gebruikt en ingezet worden. Dit inzicht biedt nieuwe kansen en uitdagingen voor de beoordeling en kwaliteit van de wetenschap. Het rapport bevat een aanzet tot een samenhangend onderzoeksprogramma waarin meetinstrumenten worden gerelateerd aan reeds bestaande indicatoren voor wetenschappelijke impact en levert hiermee een bijdrage aan de ontwikkeling van beter bruikbaar gereedschap voor de wetenschappelijke wereld. (RvH)

www.surf.nl/nl/publicaties/Pages/Users_narcissism_control.aspx

COLUMN

Gezocht: eenvoudige richtlijnen voor laagdrempelige digitalisering

Onlangs stond er een paginagroot artikel in NRC Handelsblad over e-boeken. Kortste samenvatting: het is geen marginaal verschijnsel meer. Kun je ook zelf e-boeken maken, vroeg NRC Handelsblad zich terzijde af. Ja, "met een gewone scanner, speciale software en veel geduld." Om de lezers van dienst te zijn, legde de redactie in een kadertje in vier stappen uit hoe je zelf een boek kunt digitaliseren. Eerste punt: "Klem een boek met de rug naar boven in een werkbank. Zaag de rug af met een cirkelzaag." De andere punten zal ik u besparen, die waren net zo onzinnig. Want heb je werkelijk een werkbank en cirkelzaag nodig om de bladzijden van een boek los te snijden? Welnee, in principe kom je een heel eind met een liniaal en stanleymes. En als je het professioneler wil aanpakken kun je het

best een zogenoemde stapelsnijder kopen, een apparaat om een stapel papier mee te guillotineren. Voor de tekstherkenning werd in de NRC de verkeerde software aangeraden. De redactie bleek zich voornamelijk te hebben gebaseerd op een verouderd filmpje op internet van een goedbedoelende amateur die zo'n acht uur nodig had om één boek te digitaliseren. Hoewel ik vaste medewerker ben van NRC Handelsblad, kon ik het niet laten om in een ingezonden brief te melden dat dit toch allemaal wel een stuk sneller en beter

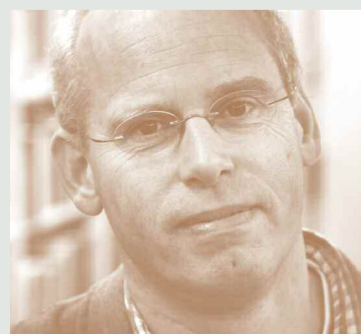


foto Leo van Velzen

kan. Ik scan zelf al jaren en als je het een beetje efficiënt aanpakt, kun je een boek van 220 pagina's binnen 5 minuten volledig digitaal doorzoekbaar maken. Die brief leverde een paar mailtjes op waar ik van schrok. Zo meldde zich een universiteitsmedewerker die de hele dag proefschriften scant. Zoals u wellicht weet zijn veel universiteiten en wetenschappelijke instellingen relevante publicaties aan het digitaliseren om die op internet toegankelijk te maken. Ook deze medewerker sneed eerst de bladzijden los, want dat is nu eenmaal de goedkoopste en vaak ook beste aanpak, maar alleen voor het scannen was hij per proefschrift toch zeker zo'n twintig minuten bezig. Kon dat werkelijk sneller? Ja, met de juiste apparatuur. Iemand anders, eveneens van een

wetenschappelijke instelling, meldde geschrokken dat ze daar kennelijk al heel lang de verkeerde software gebruikten voor tekstherkenning.

Ik zie dit nu al jaren. Ik ken instituten waar mensen al jarenlang publicaties aan het scannen zijn zonder echt te weten hoe dit moet, op de verkeerde apparatuur, met de verkeerde tekstherkenningsoftware.

Is dan niet makkelijk te vinden hoe het wel moet? Metamorfoze beschikt over goede richtlijnen voor digitale conservering en ook bij het Stichting Digitaal Erfgoed Nederland (DEN) is van alles te vinden.

Maar bij mijn weten ontbreekt het aan een actueel, simpel gidsje waarin kort en helder wordt uitgelegd hoe je een efficiënte workflow opzet, welke software en

hardware je het best kunt gebruiken, hoeveel metadata je minimaal nodig hebt, wat je met eenvoudige middelen kunt doen aan digitale duurzaamheid, welke mogelijkheden er zijn als je boeken of tijdschriften niet wil of kunt versnijden, enzovoorts.

Geen digitalisering op het allerhoogste niveau, maar praktisch en laagdrempelig, want vaak is dat al voldoende.

Dat zo'n gidsje er nog niet is, vind ik zorgelijk, want het betekent dat er nu veel tijd en geld wordt verspild. Om nog maar te zwijgen over de soms belabberde kwaliteit van de gedigitaliseerde publicaties.

Ewoud Sanders

Taalhistoricus en journalist. Sanders is vaste medewerker van onder meer NRC Handelsblad en Onze Taal