

Wikipedian-in-Residence Hay Kranen:

‘Wikidata kan goudmijn zijn voor onderzoekers’

Hoe haal je de banden tussen Wikipedia en de Nederlandse erfgoedwereld aan? Wikipedia's Hay Kranen trok een jaar geleden in bij het Nationaal Archief en de Koninklijke Bibliotheek. En nu zwaait deze eerste Nederlandse Wikipedian-in-Residence weer af.

Steven Claeysens



Hay Kranen foto KB

Hij bracht Wikipedianen naar bibliotheek en archief en gaf aan archivissen en bibliothecarissen meer dan tien Wikipedia-workshops. Maar hij deed nog veel meer. Kranen is trots op wat er is bereikt. “Gaandeweg ontdekte ik dat ik er op een tweeledige manier in stond: een evangeliserende en een pragmatische. Het eerste is beter geslaagd, dan het laatste.”

90.000 views

“We hebben veel gedigitaliseerd Publiek Domein-materiaal van het NA en de KB toegevoegd aan Wikimedia Commons.” En uit alles blijkt dat van de vele scans van Nederlands historisch erfgoed duchtig gebruik gemaakt wordt. Enthousiast: “Het interessantst vond ik het VOC-materiaal uit de Atlas of Mutual Heritage. Een kaart uit deze collectie sierde intussen al de hoofdpagina van de Duitse Wikipedia en kreeg daar in één dag maar liefst 90.000 views.”

FOCUS

‘Een harde boodschap, maar wel de realiteit’

Dat enorme bereik raakt aan de kern van Kranens evangeliserende missie: “In elke lezing of cursus voor medewerkers wees ik erop dat maar één procent van de Nederlandse bevolking gebruikmaakt van een bibliotheek voor het opzoeken van feitelijke informatie. Mensen zoe-

ken informatie op Wikipedia of andere digital native-projecten, niet in de bibliotheek. Die boodschap kwam bij sommigen toch nog hard aan, maar het is wel de realiteit.” Die digitale realiteit is volgens Kranen nu wel doorgedrongen. Daarnaast zijn er ook de nodige plannen gesmeed om de verse banden tussen Wikipedia en de bibliotheek en het archief te bestendigen. Bovendien gingen intussen ook bij andere Nederlandse erfgoedinstellingen Wikipedian-in-Residence aan de slag.

Interessante dingen

Wat wetenschappers met de Wiki-projecten kunnen? “Veel. Ik ontmoet regelmatig onderzoekers die

interessante dingen doen op basis van ons materiaal. Alle teksten en data hebben een open licentie. Wikimedia Commons, bijvoorbeeld, is een gigantische, constant groeiende verzameling van meer dan 22 miljoen encyclopedisch relevante afbeeldingen en andere media. De metadata is jammer genoeg nog niet gestructureerd, maar daar gaat in de toekomst verandering in komen.”

15 miljoen gegevens

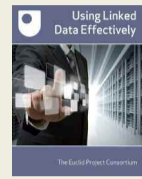
Vooral over de jongste telg van de Wiki-familie raakt Kranen moeilijk uitgepraat. “Wikidata kan een goudmijn zijn voor onderzoekers. In een paar jaar tijd hebben we al meer dan 15 miljoen gegevens verzameld en dat aantal zal de komende jaren erg hard groeien. Wikidata bestaat volledig uit gestructureerde data. Daar kan je ontelbare queries op loslaten. Mijn favoriete voorbeeld is de vraag naar alle burgemeesters ter wereld die niet mens zijn. Dan leer je dat in een klein plaatsje in Alaska een kat de burgemeesterssjerp draagt. Waar anders ontdek je dit?” Ook complete datasets kunnen sinds kort batch-gewijs worden toegevoegd. “Onlangs zijn zo alle Nederlandse rijksmonumenten mét geo-coördinaten geüpload. Het zou toch mooi zijn als we ook de volledige nationale bibliografie kunnen toevoegen en deze dan met crowd sourcing kunnen verbeteren en verrijken.”

Andersom kan ook Wikidata veel hebben aan onderzoekers, geeft Kranen tot slot mee. “Als onderzoekers interessante datasets hebben die passen in het Wikidata-project, laat ze zich vooral melden. We voegen de data graag toe.”

www.wikidata.org
commons.wikimedia.org

GELEZEN

‘Using Linked Data Effectively’, eBook, diverse auteurs.
Uitgever: The Open University 2013. Gratis via Apple iTunes en euclid-project.eu
René van Horik



Er is groeiende belangstelling voor het semantisch web, waarin het ‘web van documenten’ transformeert naar een ‘web van data’. De onderliggende standaarden en technieken die voor deze transformatie nodig zijn, zijn bekend onder de naam ‘Linked Data’. Recentelijk verscheen het elektronische boek ‘Using Linked Data Effectively’. Dit boek geeft onderzoekers een overzicht van de mogelijkheden van de Linked Data standaarden en technieken. Het boek bestaat uit zes hoofdstukken over aspecten zoals het creëren, het gebruiken en het bouwen van applicaties op basis van Linked Data. Het bevat filmpjes met achtergrondinformatie bij de informatie. Het boek is ontwikkeld door EUCLID (EdUcational Curriculum for the usage of LInked Data), een Europees project waarbinnen kennisoverdracht aan ‘data-beoefenaars’ centraal staat. De visie van EUCLID is dat het gebruik van Linked Data het dagelijks werk van onderzoekers vergemakkelijkt. EUCLID biedt een combinatie van leermiddelen en activiteiten zoals eBooks, webinars en trainingen. Op de website van EUCLID staat meer informatie. euclid-project.eu

COLUMN

In de e-humanities ziet niemand door de bomen het bos meer

Onlangs bezocht ik een bijeenkomst van de Universiteitsbibliotheek Leiden over ‘Digital Scholarship and the Role of the Library’.

Een van de conclusies van de bijeenkomst was dat academische bibliotheken studenten en onderzoekers in de e-humanities beter moeten leren omgaan met digitale hulpmiddelen om patronen in grote datacollecties te ontdekken. De toverwoorden die hierbij horen zullen de meeste lezers van dit blad bekend zijn. Men heeft het niet over computerprogramma’s of software, maar over tools. Die moet je loslaten op een dataset. De meeste datasets zijn zo groot dat onderzoekers ze onmogelijk grondig kunnen bekijken. Maar gelukkig kunnen tools automatisch patronen uit die data vissen – wat distant reading wordt genoemd. Het liefst vindt men onbekende of nieuwe patronen. Alle aanwezigen, een internationaal publiek van zo’n zestig mensen, konden zich goed in die conclusie vinden. Wereldwijd stelt men vast dat veel studenten en

onderzoekers slimmer gebruik willen maken van datasets, maar welke tools werken het best voor bepaalde taken? En hoeveel tools zijn er eigenlijk?

Daar weet niemand. Nergens is een overzicht te vinden van alle tools die zijn of worden ontwikkeld – althans niet voor de geesteswetenschappen. Instituten krijgen nauwelijks geld voor het aanleggen van goede datasets, want subsidiegevers beschouwen dat niet als wetenschappelijk werk. Wel worden er bakken geld gegeven aan het bouwen van nieuwe tools, want daarmee geef je je volgens de subsidiegevers juist in de frontlinie van de wetenschap.

Of er elders in de wereld al een

vergelijkbare tool is ontwikkeld, doet niet ter zake. Leiden heeft onlangs een eigen instituut voor e-humanities opgezet, maar werkt niet samen met Amsterdam of Utrecht, waar ze ook dergelijke centra hebben. In de frontlinie van de geesteswetenschappen knokt ieder voor zich – dat is nou eenmaal de traditie.

Of de tools vaak worden gebruikt, lijkt men niet belangrijk te vinden



foto Leo van Velzen

– het gaat erom dat ze er zijn. Men ziet het ook niet als taak om ze te onderhouden. Ze zijn er, wie ze verder wil ontwikkelen kan dat doen. Dat er soms heel weinig data aan de tool gekoppeld zijn, beschouwt men niet als probleem. In principe zou je er veel data aan kunnen hangen.

In de zaal had dit alles, als ik het goed heb gezien, een enigszins verlamdend effect. Vele van de gedemonstreerde tools waren de aanwezigen onbekend. Ze stelden vragen als: ‘Waarom laat je ons speciaal deze tool zien voor deze functie? Zijn er ook andere beschikbaar?’ Of: ‘Werkt dit ook voor andere talen dan het Engels?’

Ik zag iemand fronsen toen een

onderzoeker vertelde dat er nauwelijks tijd was geweest om een tool (die in twee jaar tijd 700.000 euro had gekost), met data te vullen. Het bekende verhaal: de tool was nu voor iedereen beschikbaar, maar werd niet onderhouden. Gelukkig was er wel subsidie voor een nieuwe tool.

Het beste nieuws vond ik dat de Universiteitsbibliotheek Leiden, in samenwerking met studenten, een nuttig lijstje op internet heeft gezet met tools waarvan is vastgesteld dat ze bepaalde dingen goed doen. Dat andere universiteiten in de wereld al tientallen vergelijkbare lijsten op internet hebben gezet, slechts deels met dezelfde tools, bewijst mijns inziens dat niemand in de e-humanities door de bomen het bos meer ziet.

Ewoud Sanders

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