

De digitale kloof tussen de eerste en derde wereld is enorm. De World Wide Semantic Web-werkgroep wil daar met Linked Data-technieken verandering in brengen. René van Horik

Ontwikkelingslanden kennen een informele economie: arbeidsintensieve werkmethode, lage investeringsbehoeften en weinig mechanisatie. Straathandelaars, ambachtsslui en kleine boeren vormen het grootste deel van de beroepsbevolking. Een directe informatiestroom tussen deze groepen onderling kan het welvaartsniveau doen stijgen – in contrast met ontwikkelde landen, waar de Linked Data-informatiestroom *top-down* via overheidsportalen verloopt.

De werkgroep World Wide Semantic Web werkt aan projecten die het delen van data ook in de derde wereld mogelijk maken door gebruik te maken van bestaande Linked Data-benaderingen. Christophe Guéret (DANS) is een van de drie bestuursleden van de werkgroep en is betrokken bij verschillende projecten.

Rekening houden

De westerse benaderingen voor het delen van data zijn niet één op één over te nemen in de derde wereld. “We moeten de ideeën over data-delings ‘down-scalen’ van grote gecentraliseerde systemen voor grote gebruikersgroepen naar kleinere gespecialiseerde gereedschappen die rekening houden met de specifieke realiteit en context van de gebruikers”, aldus Guéret. “Dit is het juiste moment om te handelen. Omdat de meeste Linked Data-technolo-



Boeren in Mali tonen de mobieltjes waarmee ze via het spraaksysteem Radio Marché de lokale marktprijzen opvragen foto's Bruno van Moerkerken

Welvaart derde wereld dankzij ‘datadelen’

gieën nog in ontwikkeling zijn, kan er nog invloed worden uitgeoefend.” Er is eigenlijk maar een klein aantal

FOCUS

De data moeten lokaal betekenis hebben



aanpassingen nodig om ook minder ontwikkelde gebieden te laten profiteren van de voordelen van het delen van gestructureerde data, informatie en kennis via een netwerk. De systemen moeten zich richten op een kleine lokale infrastructuur, gebruik maken van multimodale interfaces op de data en toegespitst zijn op data die dicht bij de burgers staan

dan de overheidsbronnen die wereldwijd openbaar gemaakt worden. “De data moeten lokaal betekenis hebben en moeten toegankelijk zijn in lokale talen en dialecten”, vertelt Guéret.

Lokale marktprijzen

Omdat veel boeren in ontwikkelingslanden beperkt geletterd zijn, maakt het project VOICES bijvoorbeeld gebruik van spraaksystemen. Lokale boeren kunnen bellen naar Radio Marché, een systeem dat lokale marktprijzen doorgeeft. “Het grootste deel van de wereldbevolking heeft een mobiele telefoon”, aldus Guéret. “Dit is een voorbeeld van een systeem waar innovatieve, context-afhankelijke interfaces voor datatoegang worden gebruikt.”

Ook de opkomst van nieuwe, goedkope hardware met een laag ener-

gieverbruik, zoals de \$100-laptop (tegenwoordig XO-laptop) of de RaspberryPi, biedt mogelijkheden voor nieuwe ICT-diensten voor ontwikkelingslanden. “Met het Entity Registry System (ERS), bijvoorbeeld, kan iedereen Linked Data produceren en delen, ook zonder toegang tot het web en zonder het gebruik van krachtige servers.”

Meewerken

Bij het World Wide Semantic Web zijn ontwikkelaars uit verschillende landen en van verschillende instituten en organisaties aangesloten. “Het is een openbare groep, dus ik wil geïnteresseerden vooral oproepen contact met ons op te nemen en mee te werken aan onze projecten.”

worldwidesemanticweb.org/mvoices.eu/

2,5 miljoen Spinoza-premie voor Piek Vossen

Piek Vossen, hoogleraar Computatonele Lexicologie aan de Vrije Universiteit, heeft dit jaar de Spinozapremie ontvangen. De Spinozapremie van 2,5 miljoen euro is de hoogste onderscheiding in de Nederlandse wetenschap en wordt toegekend aan onderzoekers die tot de absolute wereldtop behoren.

Vossen combineert taalkunde en informatica om de context van taaluitingen te herkennen, zodat computers taal zo goed mogelijk kunnen begrijpen. Daarmee maakt hij *word-nets* – spinnenwebben van woorden

die met elkaar verbonden zijn op basis van hun betekenis. Deze maken het mogelijk om op een systematische manier betekenis en cultuur in taal te onderzoeken.

Daarnaast werkt Vossen aan een viertalige ‘geschiedenisrecorder’, een computerprogramma dat dagelijks financieel-economisch nieuws ‘leest’ en bijhoudt wat, wanneer, waar gebeurt en wie erbij betrokken is. Het programma maakt uit al deze nieuwssnippers één volledig verhaal en houdt verschillen in feiten en meningen van bronnen bij. Dat maakt systematische analyses mogelijk die laten zien welke trends er zijn, wat de verschillende visies zijn en wat de impact van de gebeurtenissen is. (VU Amsterdam)

vossen.info/

Website over wereldwijde ongelijkheid en groei

Erica Renckens

Op 21 juni werd aan het Internationaal Instituut voor Sociale Geschiedenis (IISG) de Clio Infra-website gelanceerd. Via deze website hebben wetenschappers en andere geïnteresseerden toegang tot datasets uit de hele wereld met sociaaleconomische gegevens van de laatste vijf eeuwen, met name van de laatste tweehonderd jaar. Hiermee is het mogelijk onderzoek te doen naar wereldwijde economische groei en ongelijkheid. Met een speciale visualisatietool kunnen de globale



Wereldwijd is ongelijkheid een groot probleem foto Jonathan McIntosh

verschillen inzichtelijk gemaakt worden. De datasets, samengesteld door het IISG, Rijksuniversiteit Groningen, Universiteit Utrecht en Universiteit van Tübingen, worden gearchieveerd door DANS. Het Clio Infra-project is gefinancierd door NWO. clio-infra.eu/

KORT

Luisteren én lezen: de sprekende kaart

Op de website van het Meertens Instituut zijn meer dan duizend uur aan geluidsopnames van dialectsprekers uit alle delen van Nederland als ‘sprekende kaart’ te beluisteren. Recent zijn daar ook, dankzij een subsidie van de KNAW, transcripties aan gekoppeld. In totaal gaat het om meer dan meer dan 12.000 scans. Met de combinatie van geluid en transcriptie kunnen wetenschappers grootschalig fonetisch onderzoek uitvoeren naar bijvoorbeeld variatie in klinkerkwaliteit en klinkerduur, en statistisch onderzoek naar woordfrequentie in het Nederlands. De scans zijn tevens opgenomen in de CLARIN-infrastructuur. (DZ) www.meertens.knaw.nl/soundbites

Aftrap voor CESSDA-ERIC gegeven

De Council of European Social Science Data Archives (CESSDA) bereidt zich voor op een officiële ERIC (European Research Infrastructure Consortium)-status. De tientallen jaren oude samenwerking tussen de Europese sociaal-wetenschappelijke dataarchieven krijgt daardoor een formeel juridisch karakter. Tot nu toe hebben dertien landen, waaronder Nederland, de consortiumovereenkomst ondertekend. Het centrale bureau van de CESSDA-ERIC zal in Bergen gevestigd worden. De komende maanden worden de laatste voorbereidingen getroffen om CESSDA-ERIC volledig operationeel te maken. (HT)

Instituutsaccount geeft toegang tot EASY

Inloggen in het elektronisch archiveringssysteem (EASY) van DANS kan sinds kort via SURF-federatie. De gebruiker hoeft dan niet meer zijn eigen EASY-wachtwoord en gebruikersnaam te onthouden; hij gebruikt die van de instelling waar hij werkt. Van alle bij SURFconext aangesloten instellingen hebben momenteel 21 instituten toestemming gegeven voor federatief inloggen in EASY. Bij de overige is de aanvraag nog in behandeling. (MW)

Data Seal of Approval voor LISS-data

De data van het LISS-panel van CentERdata heeft het internationale datakeurmerk, het Data Seal of Approval (DSA), ontvangen. Het DSA geeft aan dat de data veilig en duurzaam zijn opgeslagen, dat ze grondig zijn beoordeeld op juistheid en dat ze beschikbaar zijn voor analyses. Universiteiten, onderzoeksscholen en individuele wetenschappers kunnen onderzoeksvoorstellen indienen om kosteloos vragenlijsten af te nemen bij het LISS-panel. De data worden later beschikbaar voor alle academici. Het project wordt gesubsidieerd door NWO. (MdB) lissdata.nl/dataarchive