

Van Gompel winnaar CLARIN Young Scientist Award 2015

‘Ik ontwikkel alleen open source’

Maarten van Gompel ontving dit najaar de **CLARIN Young Scientist Award**. *Erica Renckens*

De prijs wordt jaarlijks toegekend aan een veelbelovende jonge onderzoeker die bijdraagt aan het bouwen van taalbronnen, het ontwikkelen van tools en het delen van kennis. Hans Bennis (Meertens Instituut) en Walter Daelemans (Universiteit van Antwerpen) droegen Van Gompel voor. In hun aanbevelingsbrief noemen ze hem ‘een begaafd programmeur en onderzoeker’ die tools ontwikkelt ‘die aan de basis liggen van belangrijke ontwikkelingen’.

Die tools zijn met name FoLiA en CLAM. Van Gompel werkt daar al aan sinds hij na zijn master Human Aspects of Information Technology in Tilburg betrokken raakte bij onder andere het SoNaR-project. Het SoNaR-corpus bevat meer dan vijfhonderd miljoen woorden aan Nederlandse teksten.

Taalkundige annotaties

“FoLiA is een bestandsformaat voor geannoteerde corpora,” vertelt Van Gompel in zijn werkkamer bij de Radboud Universiteit. “Bij de ontwikkeling van SoNaR ontstond de noodzaak voor een goed formaat waarin taalkundige annotaties vastgelegd konden worden. Inmiddels



JONG TALENT

Van Gompel: “De toepassingen die ik ontwikkel, zijn altijd open source. Alleen zo kan onderzoek replicbaar zijn” foto Wieke Hoeke

wordt FoLiA ook door andere corpora gebruikt, zoals Basilex en Nederlab.”

Toch is de tool nog altijd niet ‘af’. Van Gompel: “Ik ontwikkel nog steeds software voor FoLiA, zoals FLAT, waarmee een ge-

bruiker heel makkelijk annotaties in het FoLiA-formaat kan toevoegen. En Frog voert taalkundige analyses uit met FoLiA als outputformaat.” Van Gompel typt razendsnel complexe commando’s terwijl hij zijn werk laat zien. Kan de gemiddelde onderzoeker wel met zijn programma’s uit de voeten? “Daarvoor heb

ik CLAM ontwikkeld,” legt hij uit. “Dat vervangt de command-line en creëert automatisch een overzichtelijke gebruikersinterface. Die interface is een webservice, waardoor ook machines hem kunnen aanspreken. Een gebruiker kan zo ook een zelfontworpen interface gebruiken die weer via CLAM met de software communiceert.”

Alle codes vrij

Voorlopig is de jonge onderzoeker nog druk bezig met het afronden van zijn proefschrift, waarin hij de kwaliteit van automatische vertalingen probeert te verbeteren door te kijken naar de context. Maar daarna wil hij het liefst weer software ontwikkelen die onderzoekers nodig

hebben. Van Gompel: “Ik probeer altijd een zo generiek mogelijke oplossing te vinden voor problemen. Zo kun je voorkomen dat mensen dubbel werk doen.” De toepassingen die Van Gompel ontwikkelt, zijn daarnaast altijd open source. “Daar ben ik heel principieel in. De code is altijd vrij te gebruiken door andere ontwikkelaars. Dat is de enige manier waarop onderzoek replicbaar kan zijn.”

De CLARIN Young Scientist Award bestaat uit een geldprijs van 500 euro en een certificaat. Van Gompel: “Ik werk veel vanuit huis, dus dat certificaat heeft daar een mooi plekje aan de muur gekregen.”

clarin.eu

‘Open source software voor onderzoekers’

Safe harbour-principe en Right to be forgotten uitgelegd

Issues privacy en open data '16

Hoe verhouden privacy en open data zich tot elkaar? Welke issues spelen op dit moment? *Heiko Tjalsma*

Al jaren heerst een steeds sterkere drang om alle (onderzoeks)data zo onbeperkt mogelijk ter beschikking te stellen. Aan de andere kant wordt, óók onder invloed van de Europese Unie, de privacybescherming steeds strenger. Problematisch is dit wanneer personen het onderzoeksonderwerp zijn, zoals in de sociale wetenschappen, de medische wetenschappen en de humaniora. Dit spanningsveld tussen open access en privacybescherming werd ook tijdens de Amsterdam Privacy Conferentie (oktober 2015) geconstateerd. Twee issues, van belang voor wetenschappelijk onderzoek, worden in dit artikel besproken.

Persoonsdata over zee

Het *safe harbour*-principe gaat over het doorgeven van persoonsdata naar de Verenigde Staten. Recent is dit onrechtmatig verklaard door het Europese Hof van Justitie. Dit maakt de uitwisseling

van persoonsdata van en naar de VS onmogelijk, ook voor onderzoekers. Er wordt, met hoge prioriteit, aan een nieuwe regeling gewerkt.

Recht om te worden vergeten

Het *right to be forgotten* gaat om de mogelijkheid voor individuen om hun geschiedenis, speciaal op internet, te wissen. Dit principe staat centraal in de door de EU voorgestelde nieuwe privacy-wet, de General Data Protection Regulation (GDPR). Deze wet gaat in de plaats komen van de huidige nationale wetten, zoals de Nederlandse Wet Bescherming Persoonsgegevens (WBP). De komst van de GDPR, en vooral het bijna absoluut genomen *right to be forgotten*, heeft tot grote ongerustheid bij onderzoekers geleid. Gevreesd werd voor een enorme verslechtering bij het gebruik van persoonsdata in wetenschappelijk onderzoek. Medio december 2015 is er eindelijk overeenstemming bereikt tussen de Europese bestuursorganen, naar verwachting wordt de wet in het voorjaar van 2016 aangenomen. Een voorlopige eerste conclusie is dat de wet er uiteindelijk minder

alarmerend uitziet dan eerder leek. Er blijven uitzonderingen mogelijk voor onderzoek- en archiefdoeleinden. Voor actuele informatie zie http://ec.europa.eu/justice/data-protection/reform/index_en.htm.

Informed consent essentieel

Voorlopig blijft de huidige praktijk van kracht, gebaseerd op de WBP. Essentieel is en blijft informed consent: het geven van toestemming door geïnformeerde proefpersonen of patiënten. Juist hier komen nadere aanvullende nationale regelingen. Hoe de nieuwe EU-wet precies zal uitwerken, zal pas over enige jaren duidelijk worden.

Meer weten over de GDPR en de gevolgen daarvan voor onderwijs en onderzoek? SURF heeft de afgelopen jaren een aantal praktische handleidingen gepubliceerd. Meer informatie staat op de website van SURF.

Heiko Tjalsma is juridisch adviseur bij DANS. <https://www.surf.nl/themes/beveiliging/beleidsondersteuning-privacy/index.html>

KORT

LingOA biedt gratis online publicaties

De redacties van vijf taalwetenschappelijke tijdschriften, waaronder *Lingua*, hebben hun traditionele uitgever de rug toegekeerd. Voortaan publiceren zij hun artikelen via Ubiquity Press onder de voorwaarden van de zelf opgerichte stichting LingOA. Dit houdt in dat wetenschappers betalen voor hun publicaties, die vervolgens online vrij toegankelijk worden. Dankzij een garantie van VSNU, NWO en KNAW hoeven auteurs de eerste vijf jaar de publicatiekosten niet zelf te betalen. De verwachting is dat na deze periode het publicatiemodel wereldwijd zal zijn veranderd van abonnementen naar open access. De komende tijd zullen naar verwachting meer linguïstische tijdschriften zich aansluiten bij LingOA. (ER) lingoa.eu

Nieuwe cursusronde RDNL start in mei

Essentials 4 Data Support is een introductiecursus voor diegenen die onderzoekers (willen) ondersteunen bij het opslaan, beheren, archiveren en delen van hun onderzoeksdata. In mei gaat een nieuwe ronde van de cursus van start. De cursus

rdnl

bestaat uit twee groepsbijeenkomsten met begeleiding door coaches, presentaties van experts, online cursusmateriaal en opdrachten. Research Data Netherlands (RDNL) wil met deze cursus een bijdrage leveren aan de professionalisering van en de afstemming tussen datasupporters. Kijk voor inschrijving of de online-only variant van de cursus op de cursus-site van RDNL. (HB) datasupport.researchdata.nl

Subsidie Summer Program ICPSR

DANS biedt ook dit jaar een tegemoetkoming in de kosten voor deelname aan het Summer Program in Quantitative Methods of Social Research 2016 van het Inter-university Consortium for Political and Social Research (ICPSR) in de Verenigde Staten. Er is dit jaar één subsidie beschikbaar van € 2.000. Alleen researchmasterstudenten en PhD's van instellingen die participeren in het Nationale Lidmaatschap van het ICPSR kunnen een aanvraag indienen. De uiterste datum voor het indienen van de aanvraag is 1 april. Meer informatie over de aanvraagprocedure en het lidmaatschap van ICPSR staat op de site van DANS. (MW) dans.knaw.nl