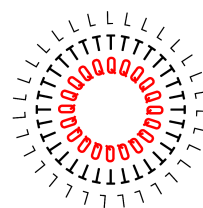


Europese dag van de talen, 26 september

Resultaten van het QTLeap project voor een hybride technologie voor betere automatische vertaalsystemen



In de afgelopen maanden zijn er duizenden vluchtelingen gearriveerd in Europa. Een van de belangrijkste onderdelen van hun integratie is het leren van de taal van hun gastland. De moeilijkheid en de omvang van deze taak benadrukt nogmaals dat een van de grootste uitdagingen van Europese integratie nog altijd niet is opgelost, namelijk dat de taalgrenzen de fysieke grenzen vergroten.

Een nieuwe taal leren is niet eenvoudig. Kinderen leren zowel hun eigen als buitenlandse talen spelenderwijs of door naar volwassenen te luisteren en hun taal als het ware te kopiëren. Volwassenen hebben hier echter veel meer moeite mee. Zij kunnen zich de grammaticale regels en de woordenschat van een vreemde taal alleen eigen maken door er veel moeite en tijd in te steken.

Online vertaalsystemen leren taal op eenzelfde manier als kinderen. Zij maken gebruik van statistische methodes om van grote hoeveelheden tweetalige teksten waarschijnlijkheden te leren die beschrijven hoe bepaalde woorden en woordsequenties vertaald moeten worden. Taal is echter meer dan slechts een opeenvolging van woorden omdat het grammaticale structuren bevat. Sommige automatische vertaalsystemen maken gebruik van deze informatie. Deze zijn echter weinig flexibel en worden daarom beperkt gebruikt. Een voorbeeld hiervan is de Spaanse krant *la Vanguardia* die elke dag met een dergelijk systeem vertaald wordt in het Catalaans.

Op het internet zijn diverse hulpbronnen beschikbaar zoals varianten van Wikipedia die door de computer gelezen kunnen worden. Deze bronnen verschaffen kennis over semantiek, de betekenis van woorden, en bevatten structurele kennis van de wereld. Ze “weten” bijvoorbeeld dat Parijs de hoofdstad is van Frankrijk maar ook dat er een stad ligt in Texas met dezelfde naam.

De afdeling computertaalkunde van de Rijksuniversiteit Groningen, werkt mee in het QTLeap project van de Europese Unie. Dit project heeft als doel een hybride technologie te ontwikkelen voor automatische vertalingen. Eerder genoemde statistische methoden worden verrijkt met taalkundige kennis en semantische informatie van het internet zodat de kwaliteit van de vertalingen groter wordt.

De eerste prototypes zijn veelbelovend maar er is nog veel ruimte voor verbetering. De Rijksuniversiteit Groningen is verantwoordelijk voor generatiecomponent van het systeem. Dit betekent dat er onderzoek gedaan wordt naar het genereren van vloeiende zinnen in de doeltaal die dezelfde betekenis hebben als de oorspronkelijke zin in de brontaal. Ook voor deze component van het systeem wordt gebruik gemaakt van grammaticale structuren en semantische informatie van het internet.

Onze kinderen vragen zich vandaag de dag waarschijnlijk af hoe wij ooit ons huiswerk konden maken zonder de hulp van Wikipedia. Op dezelfde manier zullen onze kleinkinderen in de toekomst zich wellicht afvragen hoe de mensen die in 2015 naar Europa kwamen konden communiceren voor de ontwikkeling van een universeel vertaalsysteem.

Voor meer informatie en contactgegevens kunt u terecht op:

Website: <http://qt leap.eu/>

Facebook: <https://www.facebook.com/qt leap>

Twitter: <https://twitter.com/QT Leap>

LinkedIn: https://www.linkedin.com/company/qt leap-project?trk=EML_cp-admin

Mendeley: <https://www.mendeley.com/sign-in/>

[prof.dr. Gertjan van Noord, g.j.m.vannoord@rug.nl](mailto:g.j.m.vannoord@rug.nl)

[Dieke Oele, d.oele@rug.nl](mailto:d.oele@rug.nl)