

Taaltechnologen van de VU bouwden een computerprogramma dat relaties legt tussen nieuwsfeiten en op termijn wellicht de toekomst kan voorspellen.

DOOR MARIEKE KOLKMAN
FOTO VU/RIECHELLE VAN DER VALK

Beter zoeken dan Google

Newsreader is eindelijk af. Na drie jaar ontwikkelen, leest het computerprogramma van hoogleraar computationele lexicologie Piëk Vossen helemaal zelf het nieuws. Voor de bouw van de taaltechnologische database werkte Vossens onderzoeksgroep de afgelopen drie jaar samen met universiteiten en bedrijven uit Spanje, Italië en Engeland. “Er zitten nu ongeveer twee miljard beweringen in die in het nieuws gedaan zijn over de autoindustrie”, zegt Vossen. “Dus wie deed wat met wie, waar en wanneer.” De Europese onderzoekspartners kozen voor de autoindustrie omdat die bestaat uit complexe wereldwijde netwerken, en de media er doorlopend over schrijven. Het project begon toen Volkswagen Porsche overnam. Nu het ten einde loopt, haalt Volkswagen met vervuilende motoren opnieuw het nieuws. Vossen: “In de database kun je een gebeurtenis of persoon opzoeken in een bepaalde context. Je wilt bijvoorbeeld weten waarom de oud-ceo van Porsche, Wendelin Wiedeking, is afgetreden. Die

gebeurtenis voer je in en Newsreader laat zien wat vooraf in het nieuws was. Daaruit blijkt dat Wiedeking te veel had geleend om Volkswagen te kunnen kopen. Toen werd het crisis en kocht juist Volkswagen Porsche op. Dus moest hij aftreden.”

Relaties weergeven

Dat kan ik toch ook gewoon in Google opzoeken, denk je nu. Inderdaad, dat kan. Maar dat werkt alleen goed als iemand anders al heeft geschreven over de reden dat Wiedeking aftrad. Google toont domweg alle webpagina's en documenten waar de gezochte termen in staan en kent er verder geen betekenis aan toe. Newsreader onderwerpt het nieuws automatisch aan een flinke taalkundige schoonmaakactie. Allereerst leest het programma nieuws in vier talen: Engels, Nederlands, Italiaans en Spaans. Het herkent ook om wie of wat het gaat in een nieuwsbericht en wat daarover gezegd wordt. Dat gebeurt met technologie die automatisch taal verwerkt en relaties weergeeft tussen de personen, bedrijven en gebeurtenissen. Dan blijven alleen de kale beweringen en relaties over, waar

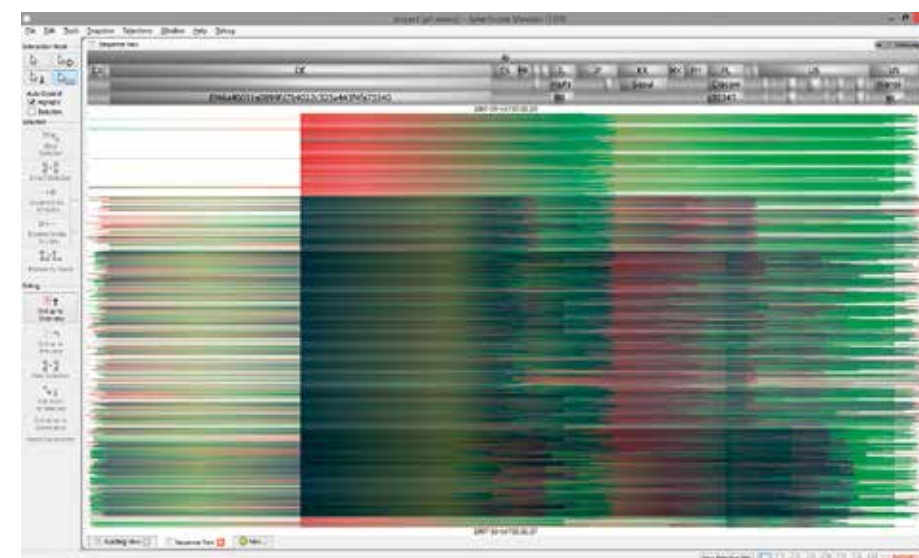
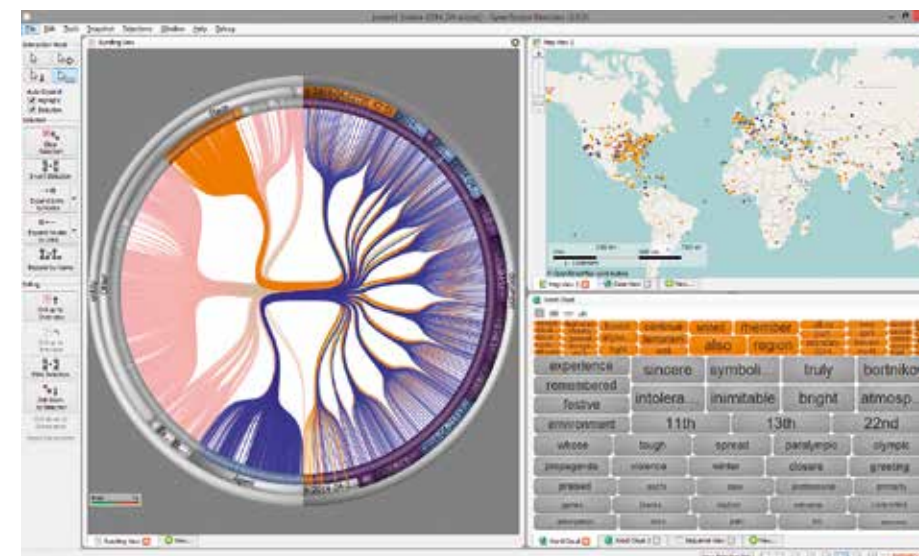


Piëk Vossen: ‘We vonden al iets smeugs rond de bankencrisis’

ook de dubbelingen uitgehaald zijn. Niemand kan zoveel lezen als Newsreader en al helemaal niet alle details onthouden. De data worden vervolgens overzichtelijk gemaakt met visualisaties als interactieve landkaarten of grafieken waarin alle directe relaties tussen bijvoorbeeld personen en bedrijven worden getoond. Zo vind je dus veel sneller en efficiënter wat je zoekt dan via Google.

Criminele netwerken

Het zal alleen nog wel even duren voor de gewone mens Newsreader kan gebruiken voor duiding van het dagelijkse nieuws. Met de huidige 2,3 miljoen nieuwsartikelen over de autoindustrie in de database lopen de onderzoekers hard tegen de grenzen van de technologie aan. “De servercapaciteit die we nu hebben, laat



Newsreader laat het netwerk van interacties tussen mensen en organisaties zien, hun geografische spreiding en de spreiding over de tijd.

Newsreader 100.000 artikelen per dag lezen, terwijl er dagelijks een miljoen artikelen binnenkomen in de nieuwsdatabase van online uitgever Lexisnexis. Je loopt dus direct al eindeloos achter.” Maar op kleinere schaal zijn er al wel toepassingen gemaakt. Vossen werkte bijvoorbeeld samen met historici aan een database voor criminele netwerken in de laatste veertig jaar. “Men heeft het vermoeden dat voor de handel in drugs, mensen en dieren dezelfde criminele netwerken worden gebruikt. De handelaars gaan door de tijd heen misschien andere dingen verhandelen, maar de netwerken blijven in stand. In veertig jaar nieuws zit heel veel informa-

Helaas zijn de voorspellingen nog niet helemaal waterdicht

tie over die netwerken.” Of het vermoeden klopt, kan Vossen alleen nog niet zeggen.

Bankencrisis

Newsreader werd ook al door de Tweede Kamer ingeschakeld voor een analyse achteraf van de parlementaire enquête over de bankencrisis. De computer las daarvoor 800.000 documenten. “We vonden nog iets smeugs toen we gingen zoeken op de mensen die niet waren komen opdagen. Maurice Lippens was destijds de ceo van Fortis, dat ABN Amro kocht, waarna de bank ten onder ging. We ontdekten dat Lippens de Russische zakenman Soeiejman Kerimov had overgehaald om 100 miljoen euro

in Fortis te investeren na de aankoop van ABN Amro. Kerimov, een vriendje van Poetin, is eigenaar van een investeringsbedrijf dat allemaal kleine bankjes opkoopt, Sberbank. Nadat Fortis omgevallen was, kon hij wel fluiten naar zijn geld.”

Momenteel zijn de onderzoekers in gesprek met de Spaanse regering. “Die zoekt naar innovatie. Ze wil weten waar mensen en bedrijven mee bezig zijn en welke projectaanvragen er bijvoorbeeld allemaal gedaan worden.” Newsreader kan wellicht ook wat betekenen in discussies over medische kwesties als vaccinaties. “Op internet bestaat een enorme kluwen aan informatie en opinies, waar je als lezer niet altijd direct uit komt. De computer kan van het hele veld in kaart brengen wat iedereen zegt. Overigens zonder zelf een standpunt in te nemen.”

Lijnen doortrekken

Een leuke bijkomstigheid van het programma is dat het op basis van alle geschiedenis, de toekomst kan proberen te voorspellen. Het trekt daarvoor de lijnen door zoals die op dit moment lopen. Zo zou op basis van het auto-nieuws tot 2007 worden voorspeld dat India een grotere speler in de auto-industrie is geworden in 2015. Dat klopt. Maar helaas zijn de voorspellingen door onvoorziene wendingen niet helemaal waterdicht. “Ook de computer is niet perfect”, zegt Vossen. “Soms stelt die dat dezelfde persoon op hetzelfde moment op twee plekken tegelijk is. Dat kan natuurlijk niet.” Dus er zijn nog altijd mensen nodig om het programma te perfectioneren. 

<http://www.newsreader-project.eu>