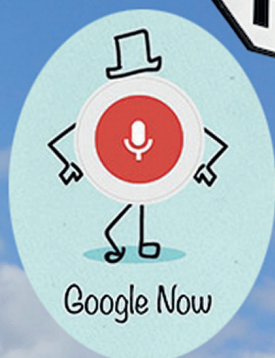
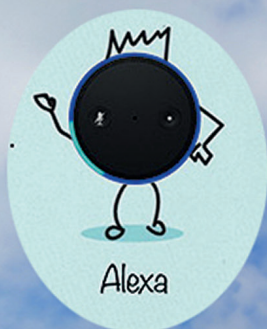


# DIXIT

TIJDSCHRIFT OVER TAAL- EN SPRAAKTECHNOLOGIE

## TST en Kunstmatige Intelligentie





# INHOUD

Algemeen	Pagina
Voorwoord	2
Colofon	2
Inleiding 'Gezocht: intelligentie voor taal- en spraaktechnologie'	3
Directory	39
<b>Successen</b>	
Deep Neural Networks in Automatische Spraakherkenning	4
Hoe een computer zelf leert vertalen	7
Smart Voices	9
Artificial intelligence leert computers lezen	11
<b>Next steps</b>	
Taaltechnologie verbeteren met ...	13
Computer, hoe lang heb ik nog?	15
Syntetische literatuur:	
Neurale netwerken als coauteur	18
Spreek2Schrijf	20
Automatisch samenvatten en antwoorden genereren	22
<b>Uitdagingen</b>	
De impact van neurale netwerken op TST	24
Nieuwe dilemma's met oude en nieuwe online data	26
Deep Learning revolutie - en wat nu?	28
Gevaaren van Artificiële Intelligentie	30
Reinventing customer service	33
Kunstmatige Intelligentie in Clariah	36



# COLOFON

**DIXIT:** Tijdschrift over toegepaste taal- en spraaktechnologie – 14e jaargang, editie *TST en Kunstmatige Intelligentie*. DIXIT is een uitgave van Stichting NOTaS, Postbus 31070, 6503 CB Nijmegen. Tel. 024-3512108 – E-mail info@notas.nl – www.notas.nl **Redactieadres:** Stichting NOTaS, Postbus 31070, 6503 CB Nijmegen **Redactie:** Arjan van Hessen: a.j.vanhessen@utwente.nl, Henk van den Heuvel: h.vandenheuvel@let.ru.nl, Kirsten de Haan: kirstendehaan@dedicon.nl, Suzan Verberne: s.verberne@liacs.leidenuniv.nl, Anne Wijnen en Marieke den Os: info@notas.nl. **Hoofredactie themanummer:** Boris de Ruyter: boris.de.ruyter@philips.com; Advertenties: Stichting NOTaS – info@notas.nl, 024-3512108 **Abonnementen:** Voor een gratis abonnement kunt u zich wenden tot een van de NOTaS-deelnemers **Vormgeving en druk:** Leonard B2B Communicatie **Verantwoording:** DIXIT is een uitgave van Stichting NOTaS. Overname van de artikelen is alleen toegestaan met bronvermelding en na toestemming van Stichting NOTaS. Stichting NOTaS en de bij deze uitgave betrokken redactie en medewerkers aanvaarden geen aansprakelijkheid voor mogelijke gevolgen die zouden kunnen voortvloeien uit het gebruik van de in deze uitgave opgenomen informatie.

# Voorwoord

Toen ik ergens diep in de vorige eeuw wiskunde en informatica studeerde, hadden we al vakken rond kunstmatige intelligentie. 'Expertsystemen' heette dat toen.

Die systemen waren gebaseerd op uitgebreide boomstructuren en netwerken van regels, locale optima, knooppunten en gewichten. Terugkijkend met de 'kennis van nu': veel logica, weinig data.

Hoe anders kijken we nu aan tegen hetzelfde probleem. Als je zegt dat je een 'rule based' systeem hebt dan word je meewarig aangekeken. Big Data is the name of the game. Een systeem is nu zo goed als de big-dataset die het tot zijn beschikking heeft.

En data hebben we nu in overvloed dankzij de oneindige beschikbaarheid van informatie in onze samenleving. Met de digitalisering in bijna ieder vakgebied, is de beschikbaarheid van data geëxplodeerd. Vroeger zouden we daarover meewarig onze schouders ophalen. Al die data, dat is nog lang geen informatie...

Maar met de komst van AI is de wereld opeens anders. We trainen systemen op die enorme berg data, zoals een kleuter spreken leert door het eindeloze en gevarieerde aanbod van haar ouders en anderen om haar heen. En die AI systemen blijken verrassend slim...

Als iemand beweert op een knikpunt te staan en een paradigmaverschuiving voorziet, dan ben ik altijd de eerste die daar een flinke dosis scepsis tegenover zet. Vanuit onze nietige plek in historisch perspectief is het ongelofelijk lastig om te bepalen of iets een paradigmaverschuiving is, als je aan het begin van die verschuiving staat.

Ik geloof dat het deze keer anders is. We staan wel degelijk op zo'n knikpunt. De impact van AI op ons leven zou wel eens van dezelfde orde kunnen zijn als de introductie van het internet. Wat denkt u?

Hopelijk helpt deze DIXIT in uw beeldvorming!!

Staffan Meij, voorzitter Stichting NOTaS



# Gezocht: intelligentie voor taal- en spraaktechnologie

*In deze DIXIT wordt de rol en de potentie van kunstmatige intelligentie voor taal- en spraaktechnologie besproken. Gezocht of vergezocht? Een beetje van beide. Uit de verschillende artikelen in deze DIXIT-editie blijkt dat er reeds concrete voorbeelden zijn van de meerwaarde die kunstmatige intelligentie met zich meebrengt voor de taal- en spraaktechnologie. Er is ook de behoefte aan futuristische ontwikkelingen van deze toepassingen.*

Naarmate ik meer word omgeven met geavanceerde technologische toepassingen, stijgt mijn verlangen naar een meer natuurlijke interactie met deze intelligente systemen. De ontwikkelingen in de kunstmatige intelligentie lijken een antwoord te geven op mijn verlangen. Toch zijn er een paar prangende uitdagingen bij het toepassen van kunstmatige intelligentie in de taal- en spraaktechnologie die ik kort wil aanhalen.

## De objectiviteit van kunstmatige intelligentie

Omdat intelligente systemen getraind moeten worden in het gebruiken van taal en spraak, wordt gebruikgemaakt van kennis uit onze omgeving. Voor de kunstmatig intelligente systemen lijkt die kennis te bestaan uit feiten, terwijl het niet zal verbazen dat veel van de gebruikte kennis vooroordelen bevat. Zo heeft James Zou van de Stanford University een studie gedaan naar het interpreteren van informatie op webpagina's middels kunstmatige intelligentie. Uit deze studie blijkt bijvoorbeeld dat het systeem de term 'programmeur' eerder met mannen ging associëren dan met vrouwen. Wat is het probleem zou je kunnen vragen? Immers, het kan best zo zijn dat er nu eenmaal meer mannen programmeur zijn dan vrouwen en dus is de associatie gebaseerd op een feit en niet op een vooroordeel. Wel, het wordt problematisch indien deze aldus getrainde technologie wordt ingezet om op een intelligente manier vacatures op maat te zoeken, waarbij de vacatures voor programmeur hun weg niet vinden naar vrouwelijke werkzoekenden. Al zijn er (e.g. Arvind Narayanan, Princeton University) die beargumenteren dat dergelijke effecten geen probleem zijn van de kunstmatige intelligentie (gezien deze een getrouwe weergave geeft van wat er in de maatschappij speelt), het is duidelijk dat belangrijke spelers in het aanwenden van kunstmatige intelligentie (e.g. Google met het Word2Vec systeem) wel degelijk aandacht hebben voor het aanpakken van maatschappelijke vooroordelen in het trainen van systemen.

## De gevaren van kunstmatige intelligentie

Er is veel te doen rondom de mogelijke gevaren van kunstmatige intelligentie. Het doembeeld dat technologie ons ooit zal overstijgen baart de mens al decennia ernstige zorgen. Grappig is dat dit doembeeld een eindeloos vertrouwen in de technische mogelijkheden veronderstelt. Immers, als kunstmatige intelligentie zo geavanceerd kan zijn dat het de mens in zijn vermogen overstijgt, dan heeft het ook de potentie om slecht te worden. Naarmate er meer toepassingen van kunstmatige intelligentie hun weg naar ons dagelijkse leven vinden worden de gevaren sterker benadrukt. Zo heeft onlangs Elon Musk (CEO Tesla) het nodig geacht om alarm te slaan over de toepassingen van kunstmatige intelligentie, door te stellen: "Als eenmaal kunstmatig intelligente robots ons afslachten op straat dan is het te laat". Dat klinkt behoorlijk dramatisch en Elon breekt dan ook een lans voor het reglementeren van kunstmatig intelligente toepassingen. Of reglementering de oplossing is weet ik niet. Er is juist een beweging op gang om kunstmatige intelligentie voor iedereen beschikbaar te maken (e.g. de werkgroep OpenAI waar Elon Musk voorzitter van is). Het is in elk geval een vraagstuk dat we in gedachten moeten houden bij het toepassen van kunstmatige intelligentie in de taal- en spraaktechnologie.

Het is duidelijk dat we niet verblind mogen zijn door de fascinerende mogelijkheden die kunstmatige intelligentie met zich meebrengt, en dat we steeds op een maatschappelijk verantwoorde wijze moeten handelen bij het aanwenden van deze technologie. Veel leesplezier en inspiratie met deze editie van DIXIT!



**Boris De Ruyter**  
Philips Research  
Gastredacteur



# Deep Neural Networks in Automatische Spraakherkenning

Automatische Spraakherkenning (ASR) is de omzetting door een computer van spraak naar tekst, van gesproken input naar geschreven output. Sinds een paar jaar zit automatische spraakherkenning in de lift en is het commercieel interessant. In elke smartphone zit wel een spraakherkenner. Het is een spannende ontwikkeling; automatische spraakherkenning komt dankzij Deep Neural Networks steeds dichterbij menselijke spraakherkenning.

Spraakherkenningssystemen worden steeds breder ingezet: in de auto om handsfree navigatie en telefonie mogelijk te maken; in ziekenhuizen voor het snel op papier zetten van een radiologische diagnose en voor medische documentatie; bij informatiesystemen voor het invoeren van zoektermen via spraak; in cursussen op de PC voor het aanleren van een tweede taal door middel van gesproken uitspraak oefeningen. En er zijn nog vele andere toepassingen.

## Hoe het werkt

Wat een ASR-systeem doet lijkt enigszins op wat wij als mensen doen als we naar een spreker luisteren. Wij horen een spraakgeluid, eventueel met een hoop achtergrondlawaai. De akoestische verwerking van het geluid in het oor en de neurale processen langs zenuwbanen en in de hersenen wekken bij de luisteraar de indruk van het horen van 'woorden achter elkaar'. Deze omzetting van het spraaksignaal naar een rijtje van woorden lijkt bij mensen bijna automatisch te gaan en zonder moeite te verlopen, maar uit psycholinguïstisch onderzoek blijkt dat het een heel complex proces is waar we eigenlijk nog maar weinig details van weten.

Net als de spraakherkenning bij mensen is de automatische omzetting van spraak naar tekst niet vanzelfsprekend. De vergelijking tussen mensen en computers gaat uiteraard mank als het gaat om de beschrijving van de processen. De hersenen zijn heel anders van structuur dan een PC met een paar harde schijven. Maar op een wat abstracter niveau kun je wel vergelijken: een conventionele automatische herkenner heeft drie dingen nodig, vergelijkbaar met dingen die luisteraars ook gebruiken.

## 1. Lexicon

Ten eerste gebruikt de herkenner een uitspraakwoordenboek (lexicon) met per woord een fonetische transcriptie hoe dat woord in de regel wordt uitgesproken, zoals: tafel – t a f @ l aardappel – a r t A p @ l

## 2. Fonetische relatie

Een andere belangrijke kennisbron is een beschrijving van hoe de fonetische symbolen t, a, f, @, l, r, A, p feitelijk klinken, dat wil zeggen, de relatie tussen deze symbolen en hoe ze klinken in het spraaksignaal. Vele factoren hebben invloed op de fonetische realisatie, zoals de context waarin ze voorkomen, spraaktempo, humeur van de spreker, de ruimte waarin de spreker staat. Die fonetische realisatie kan bovendien per spreker anders zijn, en kan behoorlijk verschillen van regio tot regio. Deze klank-informatie zit in een 'akoestisch model', en is bijna altijd statistisch van aard.

## 3. Woordvolgorde

Ten derde heeft de herkenner informatie nodig over welke woordvolgordes in taal voorkomen, en de waarschijnlijkheden die daarbij horen. Bijvoorbeeld "Piet zit op de fiets" is waarschijnlijker als uitkomst van een spraakherkenner dan "Piet de fiets zit op". De informatie over woordvolgordes heet een 'taalmodel'.

Lexicon, akoestisch model en taalmodel worden samen gebruikt om een automatische spraakherkenner te bouwen. De taak van die herkenner is dan om bij een akoestische input de 'best bijpassende woordenrij' op te zoeken.

## Zeventig jaar geschiedenis

In de afgelopen decennia hebben automatische spraakherkenners een enorme ontwikkeling doorgemaakt. In de jaren vijftig werd door Bell Labs in de VS een systeem ontworpen voor het herkennen van gesproken cijfers ('zero', 'one', ..., 'nine'). Dat systeem was toen een doorbraak, maar in onze ogen erg primitief. Het kon maar tien losse woorden herkennen.

Tien jaar later werden iets complexere systemen ontwikkeld door Rai Reddy in Stanford. Het lexicon was nog steeds zeer beperkt, en in die tijd moest elk woord afzonderlijk worden uitgesproken. Herkenning van een

lopende vloeiende zin was toen nog veel te complex. In dezelfde tijd werd het Viterbi-algoritme ontwikkeld. Dat algoritme zorgde ervoor dat spraakherkenning kon plaatsvinden op basis van een nieuwe representatie van het spraaksignaal, namelijk als een rijtje van 'spectrale featurevectoren', 100 vectoren per seconde.

## DARPA

Weer tien jaar later, in de jaren zeventig, werd in de VS het DARPA-programma gestart, met als ambitieus doel het herkennen van 1000 woorden. Deelnemers waren onder meer BBN, IBM, CMU en Stanford. Het DARPA-programma was niet in elk opzicht een succes maar leidde wel tot een aantal belangrijke innovaties op het gebied van de featurevectoren en het woordzoekalgoritme. Tegelijkertijd werkten Jim en Janet Baker bij CMU aan een nieuwe representatie van woorden, door gebruik te maken van Hidden Markov modellen. Deze modellen maakten het mogelijk om de akoestiek, de woordenlijst en het taalmodel onder te brengen in een overkoepelende structuur. Ze maakten nieuwe woordzoektechnieken mogelijk die verbeteringen waren ten opzichte van het Viterbi-algoritme.

Dat leidde midden jaren '80 tot het Tangora systeem door IBM (onder leiding van Jellinek), gebaseerd op een combinatie van: statistische akoestische modellen, zogenaamde Hidden Markov modellen, n-gram taalmodellen en een lexicon van ongeveer 20 duizend woorden. In die tijd werden ook gaandeweg herkenners ontwikkeld voor andere talen dan het Amerikaans Engels, bijvoorbeeld voor het Italiaans.



Foto: Jason Rosewell op Unsplash.com

## Google

In de jaren negentig zagen we de eerste commercieel succesvolle spraakherkenning. Dragon Dictate van Kurzweil Intelligent Systems is daar een van de vroege voorbeelden van. In die tijd bouwden Rabiner bij AT&T en Huang bij CMU spreker-onafhankelijke, continue-spraakherkenners met een lexicon dat de woordenschat van een gemiddelde luisteraar gaandeweg begon te overtreffen. En rondweg 10 jaar geleden begonnen Google en Apple aan hun opmars in automatische spraakherkenning.

Tegenwoordig zijn de commerciële bedrijven (zoals Apple, Google, Microsoft, Nuance) de belangrijke spelers op het gebied van verbetering van spraakherkenning.

## Deep deeper deepest

Ook qua algoritmen vinden voortdurend verbeteringen plaats, zoals we hierboven al zagen. Het meest recente voorbeeld is de introductie van Deep Neural Networks (DNNs). Vergeleken met de conventionele methoden laten DNNs grote verbeteringen zien, in de orde van 20 tot 50 procent vermindering van het aantal fout herkende woorden. DNNs stellen ons in staat met methoden uit de Machine Learning problemen op te lossen die met





conventionele methoden moeilijk oplosbaar waren.

DNNs hebben ook een nadeel: ze hebben voor de training heel veel gegevens nodig. Voor spraakherkenning wil dat zeggen dat de resultaten voor talen met zeer grote spraakcorpora (met name Amerikaans Engels) het meest overtuigend zijn. Een van de onderzoeksrichtingen tegenwoordig is hoe het voordeel van DNNs behouden kan blijven bij het gebruik van kleinere spraakcorpora, zodat talen als het Nederlands en Zuid-Afrikaans er ook van kunnen profiteren.

**LSTMs en CTCs**

Er zijn allerlei typen netwerken. Een nu zeer bekend type artificieel netwerk is het 'long short-term memory model' (LSTM). Dit type model is voor het eerst beschreven door Sepp Hochreiter en Jürgen Schmidhuber in 1997. Dankzij zijn structuur kan een LSTM verbanden leren tussen gegevens die in tijd relatief ver uit elkaar liggen. Dat kan belangrijk zijn voor het herkennen van taal, bijvoorbeeld in de zin: "Vanavond bel ik mijn buur die aan de overkant van de straat woont op", waarin 'op' een logisch einde van de zin is aangezien 'bel ik' al eerder voorbij is gekomen.

Naast nieuwe netwerkstructuren komen ook nieuwe classificatiemethoden in zwang, zoals Connectionist Temporal Classification (CTC). Dat is een bepaalde manier van classificatie van gebeurtenissen die in een bepaalde volgorde plaatsvinden (zoals spraakklanken en woorden in een spraaksignaal), met behulp van een netwerk. In 2015 introduceerde Google een zeer succesvolle spraakherkenner die gebruik maakt van een LSTM, getraind in



combinatie met CTC. Toch is CTC niet in alle opzichten ideaal, want wat het modelleren van volgordes betreft is deze methode gebaseerd op dezelfde beperkende statistische aannames waar ook de klassieke methoden van automatische spraakherkenning onder lijden.

**End-to-end**

De meest recente innovatie is 'end-to-end' spraakherkenning. De traditionele aanpak van spraakherkenning, bijvoorbeeld met een Hidden Markov Model, maakt gebruik van de verschillende componenten lexicon, taalmodel en akoestisch model, die afzonderlijk kunnen worden geoptimaliseerd. Een end-to-end spraakherkenner traint die drie componenten gezamenlijk. Dit kan de training en het ontwerpproces van een spraakherkenner aanzienlijk vereenvoudigen. Op Interspeech, de grootste internationale conferentie over spraaktechnologie, zijn de laatste jaren steeds meer end-to-end toepassingen getoond. Het einde van de verbeteringen is nog lang niet in zicht.

**Hoe verder?**

Automatische spraakherkenning komt langzamerhand kwalitatief in de buurt van menselijke spraakherkenning. Maar we zijn er nog niet. Een grote uitdaging in de nabije toekomst is om machines dezelfde flexibiliteit te geven die mensen laten zien bij het verstaan van spraak. Dat betekent een snelle adaptatie aan andere akoestische omstandigheden, aanpassing aan accenten, aan Donald Duck spraak, het moeiteloos begrijpen van Engelse woorden ingebed in Nederlandse zinnen, enzovoorts. Terwijl de bedrijven grote vooruitgang maken aan de technologische kant, zullen de academische instellingen zich meer moeten richten op specifieke onderzoeksvragen. Hoe kan de taalverwerving bij jonge kinderen worden gesimuleerd? Wat gebeurt er precies in ons hoofd bij het luisteren naar spraak, en hoe kunnen wij de kennis interpreteren die op de een of andere manier in een getraind Diep Neuraal Netwerk ligt opgeslagen?



**Hoe een computer zelf leert vertalen**

De laatste jaren is er een nieuw paradigma voor automatische vertaling in gebruik genomen, dat is gebaseerd op kunstmatige of artificiële neurale netwerken. Het heet neurale automatische vertaalsystemen (in het Engels 'Neural Machine Translation' of afgekort NMT). Zo schakelden zowel Google als Microsoft in 2016 over naar NMT, maar ook Systran1 dat oorspronkelijk regelgebaseerd werkte, heeft de overstap gemaakt. Er duiken ook nieuwe spelers op: de ontwikkelaars van Linguee2 (een zoekmachine voor vertalingen) lanceerden eind augustus hun eigen neurale vertaalmachine, en DeepL3, die naar eigen zeggen nog betere vertalingen aflevert dan Google Translate.

Artificiële neurale netwerken worden in heel uiteenlopende toepassingen gebruikt: voor het voorspellen van beurskoersen, het online aanbieden van reclame, beeld- en spraakherkenning, de ontwikkeling van zelfrijdende auto's en dus ook automatisch vertalen.

Neurale automatische vertaalsystemen zijn net zoals hun voorgangers – de statistische systemen – datagebaseerd: aan de hand van enorm veel voorbeeldvertalingen leert een neuraal systeem hoe het moet vertalen. Artificiële neurale netwerken zijn geïnspireerd op het menselijke brein en zijn heel goed in het herkennen van patronen in data.

**Taal herkennen**

Er zijn verschillende soorten neurale netwerken. Voor de herkenning van beelden wordt een ander soort neuraal netwerk gebruikt dan voor de verwerking van taal. Een beeld verschilt namelijk op een belangrijk punt van woorden en/of zinnen. Een beeld kun je opdelen in een raster van

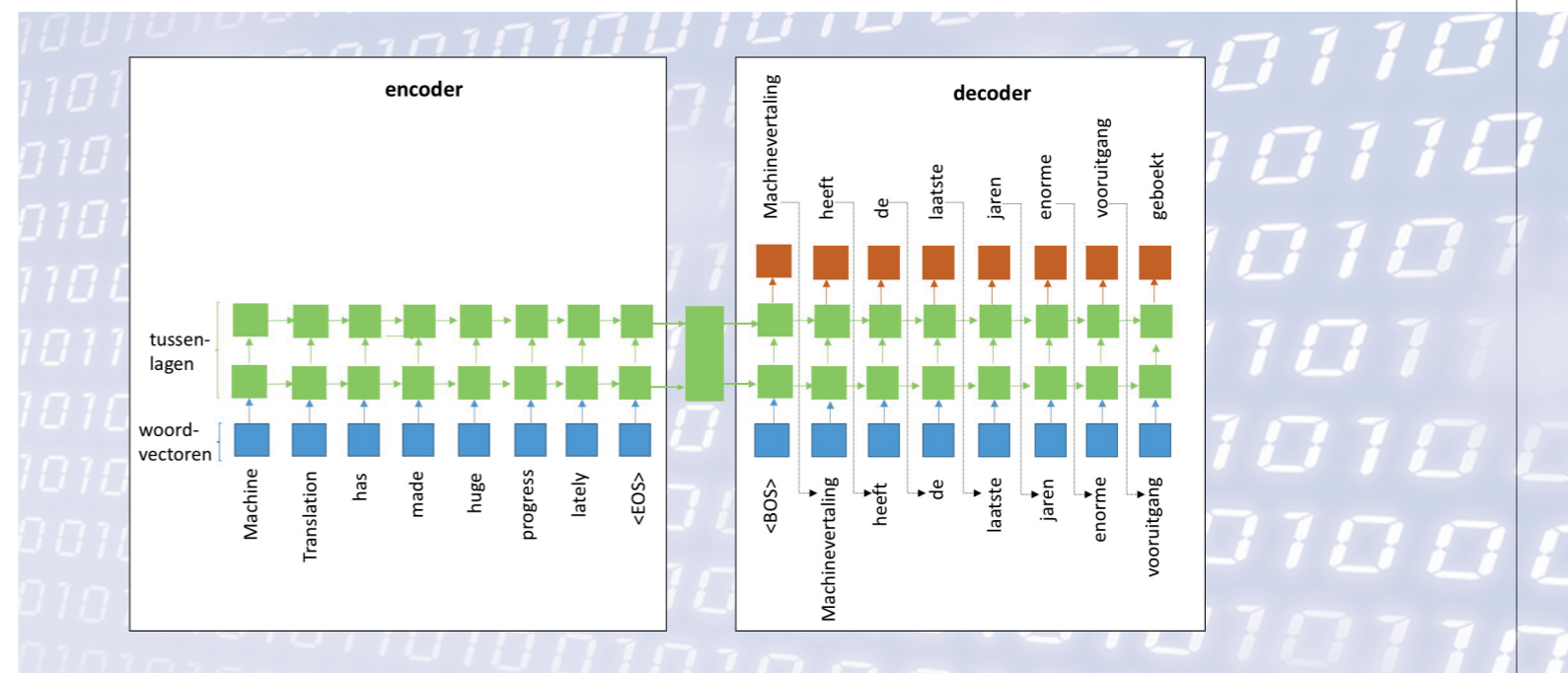
pixels, en de kenmerken van alle pixels kun je tegelijkertijd gebruiken als invoer voor een neuraal netwerk. Met woorden of zinnen lukt dit niet zo goed. Taal is immers van nature sequentieel: letters volgen elkaar op in een woord en woorden volgen elkaar op in een zin. De context van die letters en die woorden is heel belangrijk voor het modelleren van taal. Daarom wordt er in toepassingen die werken met talige data, zoals in spraakherkenning en automatisch vertalen, gebruik gemaakt van een speciaal type neurale netwerken: de recurrente neurale netwerken of RNN's.

RNN's kan je beschouwen als een aaneenschakeling van eenvoudigere neurale netwerken waarbij de uitvoer van het neuron van de vorige stap gebruikt wordt als extra invoer. Door deze speciale architectuur kunnen RNN's context onthouden en gebruiken.

**Woordvectoren**

Een van de redenen waarom NMT succesvol is, heeft te maken met de manier waarop

Lieve Macken  
Universiteit Gent





woorden worden voorgesteld. Elk woord wordt voorgesteld als een vector van getallen. Zo'n vector kan je beschouwen als een groep coördinaten die elk woord positioneert in een meerdimensionale ruimte. Het achterliggende idee is dat woorden die semantisch gelijkaardig zijn, dichter bij elkaar liggen in die ruimte en dat woorden die niet gerelateerd zijn verder van elkaar liggen. Als je met twee dimensies werkt heeft dat niet zoveel nut. Want je zou bijvoorbeeld 'hond' en 'kat' in elkaars nabijheid willen, maar ook 'hond' en 'blaffen', 'hond' en 'puppy' en uiteraard ook 'hond' en 'honden'. Om al die verschillende relaties vast te leggen heb je dus meer dimensies nodig. In realiteit worden 100 tot 1000 dimensies gebruikt. Deze woordvectoren worden gebruikt als invoer in het NMT-systeem.

De meeste NMT-systemen zelf bestaan uit een 'encoder' en een 'decoder'. Figuur 1 visualiseert het hele proces. De encoder bouwt gaandeweg een representatie op van de hele zin in de brontaal. Die analyse wordt bijgehouden als vectorrepresentatie in verschillende tussenlagen. De analyse van de volledige zin wordt doorgegeven aan de decoder die woord voor woord een vertaling genereert.

Doorgaans is de kwaliteit van neurale systemen beter dan die van statistische. Omdat NMT-systemen in staat zijn om rekening te houden met de context van de hele zin, bevatten de meeste geproduceerde doelzinnen minder grammaticale fouten. Een woord uit de doelzin wordt gegenereerd op basis van de analyse van de volledige bronzin en de al geproduceerde linkercontext van de doelzin.

#### Vertaling vergelijken

In het volgende voorbeeld tonen we de uitvoer van zowel de SMT-versie als de nieuwe NMT-versie van Google Translate. Zoals je kan zien zijn de grammaticale problemen uit het SMT-systeem (probleem met woordvolgorde en congruentie) opgelost in de NMT-versie.

Engelse bronzin: It is not our intention to finance a whole range of medical services,

but to really help a certain group of patients through, for example, our fistula project.

**SMT:** Het is niet onze bedoeling om een hele reeks van medische diensten te financieren, maar om echt te helpen een bepaalde groep patiënten door middel van, bijvoorbeeld, onze fistel project.

**NMT:** Het is niet onze bedoeling om een hele reeks medische diensten te financieren, maar om echt een bepaalde groep patiënten te helpen door bijvoorbeeld ons fistelproject.

Een tweede belangrijk voordeel is dat gebruikte woordvectoren het mogelijk maken om beter te generaliseren en dus ook de beschikbare trainingsdata optimaler kunnen gebruiken.

#### Uitdagingen

NMT-systemen lossen echter niet alle problemen op. Polyseme woorden blijven ook voor neurale vertaalsystemen nog moeilijk. Alle bovenvermelde NMT-systemen vertalen stumble nog als struikelen in de volgende zin: Scientists stumbled across a mechanism that could unlock how memory works.

De grootste uitdaging van NMT-systemen (naast de specifieke hardware-vereisten en de veel langere tijd die nodig is om een systeem te trainen) is misschien nog wel dat het hele proces totaal niet transparant is. Zoals eerder vermeld worden woorden omgevormd tot woordvectoren (sequenties van cijfers dus). In feite werkt het hele neurale netwerk enkel met cijfers. Zo'n systeem is natuurlijk ontzettend moeilijk om te debuggen en het is quasi onmogelijk om na te gaan waar en waarom een NMT-systeem een bepaalde fout maakt.

Het verbeteren van de transparantie van NMT-systemen is een onderzoeksrichting waar de komende jaren waarschijnlijk interessante ontwikkelingen zullen plaatsvinden.

1. <https://demo-pnmt.systran.net/production#/translation>
2. [www.linguee.com](http://www.linguee.com)
3. [www.deepl.com/translator](http://www.deepl.com/translator)

## Smart Voices

### AI: De toekomst van TTS

Het gebruik van tekst-naar-spraak synthese (text-to-speech, TTS) is de laatste jaren in een stroomversnelling geraakt. De huidige generatie TTS stemmen wordt gebruikt in mobiele telefoons (denk aan Siri op de iPhone), in GPS-navigatie, als voorleeshulp op websites en in het onderwijs voor mensen met leerproblemen of taalproblemen en voor anderstaligen. De meeste commerciële TTS aanbieders gebruiken een methode die unit selection wordt genoemd, maar het gebruik van AI is momenteel erg in opkomst. In dit nummer over AI en TTS mag dat dan ook niet ontbreken.

### Unit selection TTS

Ieder TTS programma bestaat uit twee componenten: een taalkundige component en een spraakgenererende component. De taalkundige component leidt af hoe de woorden in een zin moeten worden uitgesproken. Daarbij wordt bijvoorbeeld rekening gehouden met woorden die op meer dan één manier kunnen worden uitgesproken (homografen) alsook met cijfers en afkortingen die omgezet moeten worden naar woorden. De uitvoer is dan een reeks van symbolen die de klanken van de taal representeren. Deze reeks wordt doorgegeven aan de spraakgenererende component. Er zijn ook symbolen voor lettergreepgrenzen, klemtoon en leestekens zoals komma's en vraagtekens.

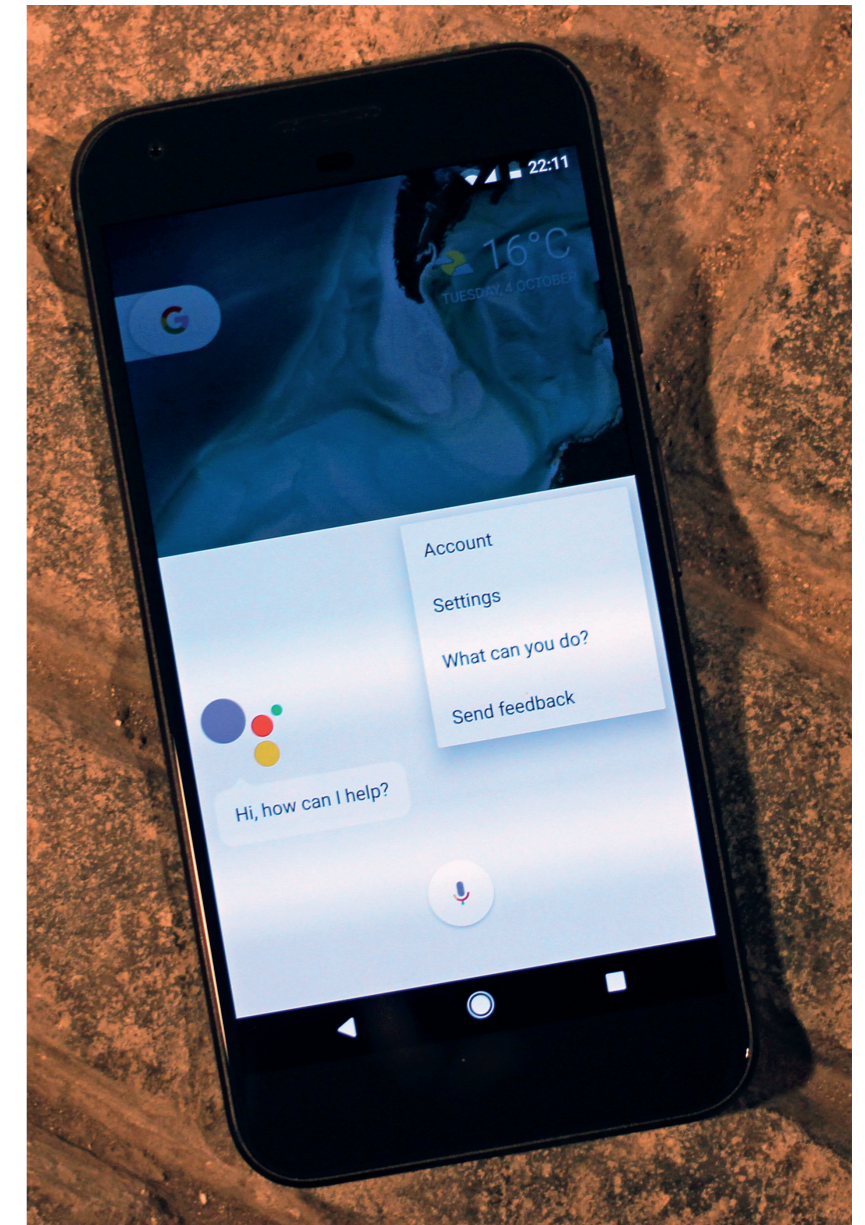
Bij unit selection TTS wordt door de spraakgenererende component in een grote spraakdatabase gezocht naar stukjes spraak die zo goed mogelijk aan de gewenste symbolische beschrijving voldoen. Met een professionele spreker wordt 10+ uur spraak opgenomen in een opnamestudio. Om extreme toonhoogteverschillen te voorkomen gebeurt dit in een neutrale spreekstijl. De teksten die worden ingesproken bevatten veel klanken in veel verschillende contexten. Door slim te zoeken in de zo verkregen spraakdatabase, op basis van complexe kostenfuncties, kan spraak van een hoge kwaliteit worden gegenereerd met weinig 'glitches' (toonhoogtesprongen of afwijkingen in klankkleuren op lasnaden). Hierdoor kost het veel tijd en moeite om een nieuwe stem te maken en de neutrale spreekstijl beperkt het toepassingsgebied.

### DNN TTS

In de laatste jaren is het AI onderzoek sterk toegenomen, ook op TTS gebied. Computers zijn tegenwoordig in staat om heel veel data te verwerken en daarin automatisch patronen te ontdekken. Voordat unit selection TTS in zwang kwam waren er al parametrische synthesesystemen die op basis van handmatig geschreven regels de taalkundi-

ge symbolen omzetten naar de akoestische eigenschappen van een spreker. Die werden vervolgens met een vocoder omgezet naar spraak. Die vereenvoudigde regels leverden een erg robotachtig klinkende stem op. Bovendien waren de regels afgeleid van één bepaalde spreker en zou het heel veel tijd kosten om ander akoestisch materiaal te analyseren en daar regels voor te maken.

**Esther Kladders**  
**ReadSpeaker**





Nu kunnen we door middel van AI dit proces automatiseren en dankzij het sterk toegenomen rekenvermogen van computers kunnen we veel gedetailleerder het akoestisch signaal beschrijven. We gebruiken hiervoor deep neural networks (DNN). Om een DNN TTS stem te maken, trainen we twee DNN modellen: een duurmodel en een akoestisch model. Het duurmodel leert een taalkundige beschrijving van klanken om te zetten naar de duur van die klanken. Het akoestisch model leert de taalkundige beschrijving samen met de klankduren om te zetten naar akoestische eigenschappen die dan met een vocoder worden omgezet naar spraak.

**Voordelen**

Het voordeel van DNN TTS is dat er betrekkelijk weinig trainingsmateriaal nodig is, en het dus minder geld kost om opnames te maken en te annoteren. Bovendien worden ook de dynamische eigenschappen van de akoestiek geleerd en via de vocoder gesynthetiseerd, zodat er geen glitches zijn. Dat wil echter niet zeggen dat de kwaliteit perfect is. De natuurlijkheid blijft nog steeds wat achter bij unit selection TTS, maar het is voor een aantal applicaties al goed bruikbaar, bijvoorbeeld in robots, spelletjes en cartoons, waar de luisteraar andere verwachtingen heeft ten aanzien van de natuurlijkheid van de spraak. Door de flexibiliteit van DNN TTS is het mogelijk om meer expressiviteit in de stem te brengen, door meer expressieve spraak op te nemen en die ook in de taalkundige eigenschappen te beschrijven. Het is ook mogelijk om DNN-modellen te trainen voor sprekerconversie: we hebben straks slechts een kleine verzameling akoestische eigenschappen nodig van een nieuwe spreker, om een bestaande DNN-TTS-stem te laten klinken als ware die van een nieuwe spreker.

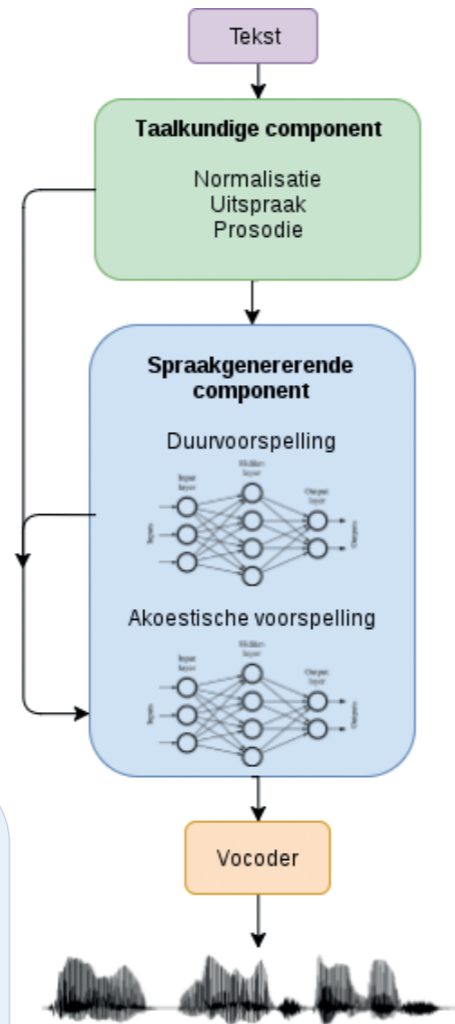
**DNN TTS bij ReadSpeaker**

Voorlopig is er nog veel onderzoek nodig om de kwaliteit van de DNN TTS verder te verbeteren. Het onderzoek bij ReadSpeaker spitst zich toe op een aantal gebieden, zoals:

- Met welk soort DNN architecturen kunnen we het beste de DNN TTS trainen?
- Kunnen we compacte DNN modellen maken die een snelle voorspelling van akoestische eigenschappen mogelijk maken?

- Hoe kunnen we de akoestische databases voor verschillende talen en sprekers het beste annoteren met eigenschappen zoals nadruk en frasering?
- Hoe kunnen we vanuit de invoertekst het beste deze eigenschappen voorspellen? Hoe kunnen we verschillende spreekstijlen produceren?

DNN TTS is in volle opgang en wij zijn erg enthousiast over de mogelijkheden die dat biedt voor toekomstige TTS-producten. Naast robots zullen we meer toepassingen gaan zien waar gepersonaliseerde TTS stemmen levendige conversaties kunnen voeren met verschillende gesprekspartners zoals kinderen, ouderen, of tweede-taal-lerenden.



# Artificial intelligence leert computers lezen

*Omdat Aegon een grote waarde ziet in de toepassingen van data analytics, heeft Aegon sinds een aantal jaren een eigen analytics afdeling. Op deze afdeling onderzoeken we data vanuit alle hoeken van het bedrijf. Dat levert inzichten op die ons helpen de dienstverlening te verbeteren. Tot een jaar geleden was er echter één soort informatiebron die we stevast links lieten liggen: de teksten. Want hoe haal je op een efficiënte manier informatie uit iets wat voor een computer niets meer is dan een brij aan letters en leestekens? Gelukkig kunnen we met behulp van artificial intelligence de teksten verwerken zodat we ook deze bron van informatie kunnen inzetten om onze dienstverlening te verbeteren.*

**Mensentaal**

Een mens snapt dat als een klant ons mailt met de vraag zijn verzekering stop te zetten, dit hetzelfde betekent als de vraag om zijn verzekering te beëindigen. Daarnaast begrijpen we dat 'beëindigen', 'beëindigd' en 'beëindig' allemaal vervoegingen zijn van hetzelfde werkwoord. Voor de computer zijn dit echter allemaal verschillende woorden. Dit maakt het moeilijk om de tekst te analyseren. Zelfs een simpele telling maken van het aantal keren dat een concept voor-

komt in de tekst, is nog niet zo gemakkelijk, omdat hetzelfde concept op meer dan één manier kan worden uitgedrukt.

Een belangrijk onderdeel van tekstanalyse is dan ook ervoor te zorgen dat synoniemen en werkwoordvervoegingen door de computer herkend worden. Voor de synoniemen zou je gebruik kunnen maken van woordenboeken die op het internet beschikbaar zijn. Deze sluiten echter niet altijd aan bij het jargon binnen de verzekeringswereld. Zo missen deze woor-

**Mariët Hebbink**  
**Aegon**

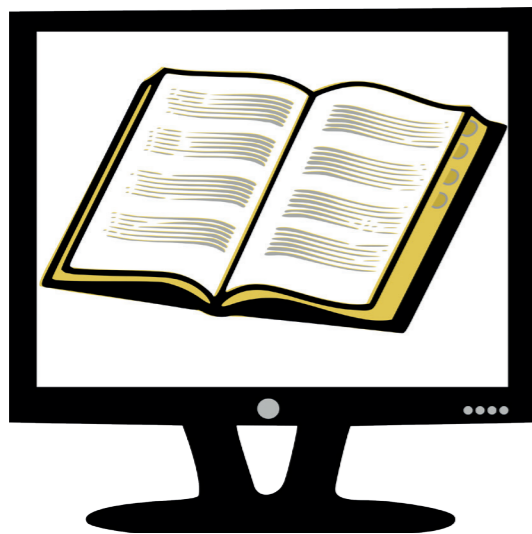




denboeken bijvoorbeeld de afkortingen van onze producten of namen van concurrenten. Voor het herkennen van werkwoordvervoegingen kun je gebruik maken van zogenaamde 'stemming-algoritmes'. Deze zorgen ervoor dat vervoegingen teruggebracht worden naar de stam van het werkwoord. Onze ervaring is dat deze algoritmes nog niet optimaal ontwikkeld zijn voor de Nederlandse taal. Met name voltooide deelwoorden worden nog niet goed herkend.

### Artificial intelligence

Tekstanalyse op basis van bestaande woordenboeken en stemming-algoritmes werkt dus niet altijd goed genoeg. Door gebruik te maken van artificial intelligence kunnen we ons eigen 'computerwoordenboek' maken dat is toegespitst op taalgebruik binnen ons bedrijf. Hierdoor kunnen we de computer de essentie uit een tekst laten halen. Het maken van het woordenboek doen we met behulp van neural networks. We doen dit als volgt: we verzamelen eerst zoveel mogelijk teksten die representatief zijn voor het taalgebruik binnen het bedrijf. Vervolgens worden al die teksten opgesplitst in de woorden die erin voorkomen, en bekijkt een algoritme de context van ieder woord. Die context wordt gevormd door een aantal woorden voor en na het betreffende woord. Elk woord wordt opgeslagen met een representatie van zijn contexten. Deze representatie heet word embedding. Door de context van al die woorden met elkaar te vergelijken, kan het algoritme herkennen welke woorden vaak in dezelfde context voorkomen. Vervolgens begrijpt het algoritme dan dat deze woorden synoniemen zijn, of vervoegingen zijn van hetzelfde werkwoord. Zo ontstaat een wiskundige functie die voor de computer een beetje werkt als een woordenboek. Mooie paradox is dat mensen dit computerwoordenboek (de word embeddings) dan weer niet kunnen lezen.



Het idee van het gebruik van context voor het vinden van synoniemen is al oud, en wordt in de taalwetenschappen 'distributionele semantiek' genoemd. Dankzij de inzet van neurale netwerken en de ontwikkeling van efficiënte toolkits om deze netwerken te implementeren (Word2vec en Glove zijn de bekendste), heeft de distributionele semantiek een grote vlucht genomen de laatste jaren. Verderop in deze DIXIT komen nog meer toepassingen van word embeddings aan bod, onder meer in het volgende artikel van Barbara Plank.

### Toepassingen

Met gebruik van word embeddings kan de computer structuur aanbrengen in de enorme brij van woorden, wat het makkelijker maakt om de tekst vervolgens te analyseren. Hoe je de tekst precies gaat analyseren hangt af van het doel waarvoor je de tekst wilt gebruiken. Een van de manieren waarop wij tekstanalyse toepassen is bijvoorbeeld voor het bouwen van een algoritme dat binnenkomende e-mails routeert. Eerst hebben we het 'woordenboek' ingezet om de mails te structureren op onderwerp. Vervolgens konden we een model bouwen dat iedere e-mail aan het juiste postvak koppelt. Dit model bespaart ons niet alleen veel handmatig werk, maar zorgt er ook voor dat e-mails sneller op de juiste plek belanden, waardoor we onze klanten sneller kunnen helpen.

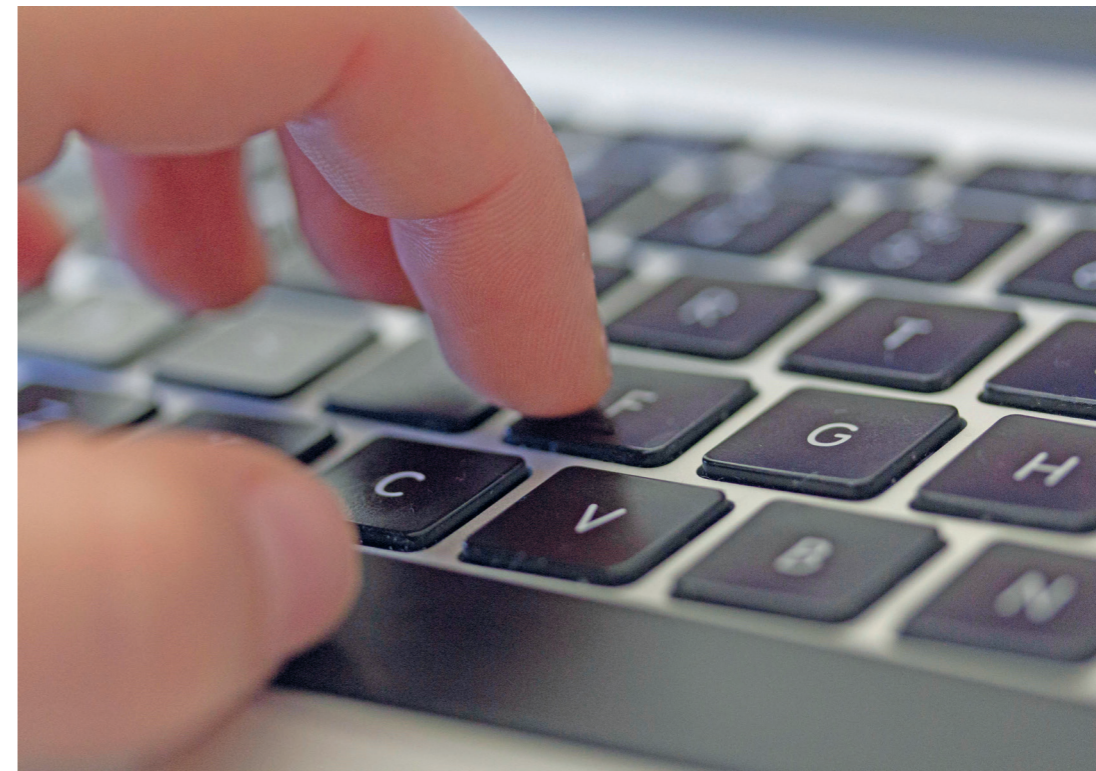
Een andere manier waarop we tekstanalyse inzetten is om te analyseren waarover klanten contact opnemen. We analyseren hiervoor de contactbeschrijving die agents bij de contacten achterlaten. Je zou de agents natuurlijk ook zelf kunnen laten turven, maar dat is vrij arbeidsintensief en foutgevoelig. Als we beter weten waarover klanten contact opnemen, dan kunnen we veel gericht onze informatievoorziening en dienstverlening verbeteren.

### Waar ligt de grens?

Dit zijn slechts enkele voorbeelden waarbij je tekstanalyse kunt inzetten; er zijn er natuurlijk nog veel meer! Denk aan automatische verwerking van documenten voor productaanvragen, of chatbots op websites. Je kunt hierin heel ver gaan. Wij geloven er echter niet in dat een computer alle taken van een mens over kan nemen. Computers kunnen wel een advies geven waardoor processen veel sneller kunnen verlopen. Uitzonderingen zullen altijd door een mens beoordeeld moeten worden. Robots kunnen steeds meer, maar uiteindelijk bepalen mensen nog altijd hoe ze ingezet worden.

## Taaltechnologie verbeteren met fortuitous data en deep multitask learning

*Ongetwijfeld is de komst van deep learning een van de redenen voor de snelle vooruitgang van AI. Het jaar 2015 was het jaar waarin de 'deep learning tsunami' (Chris Manning, Stanford) het onderzoeksveld van taaltechnologie met volle kracht raakte. Moderne AI-systemen die feature-extractie automatiseren (deep neural nets) worden elke dag met succes toegepast, bijvoorbeeld in automatisch vertalen, object-identificatie of spraakherkenning.*

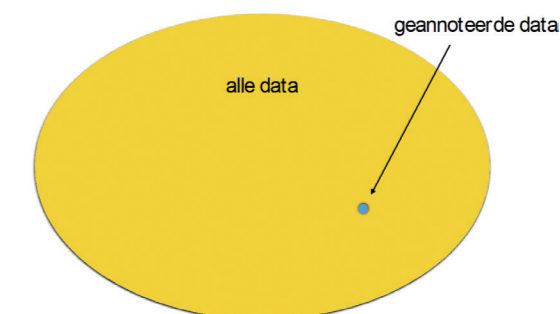


**M**Maar zelfs in onze tijd waarin de toevoer van data nooit slaapt en bijvoorbeeld een half miljoen Twitter-berichten per minuut verstuurd worden, bestaat er een grote asymmetrie. Er zijn voldoende data beschikbaar, maar moderne AI-systemen hebben een enorme hoeveelheid aan handmatig gecorrigeerd en geannoteerd materiaal nodig, zowel om de systemen te trainen als om ze te testen. Het gaat niet alleen om het spraaksignaal, maar ook om de tekst die gezegd werd; niet alleen om de afbeeldingen, maar ook om wat er te zien is op de afbeelding. Dergelijke geannoteerde data bestaan nauwelijks (Figuur 1). Het knelpunt is dus niet de hoeveelheid beschikbare data, maar de beschikbaarheid van de juiste, geannoteerde data.

### 'Fortuitous data'

Om dit data sparsity probleem aan te pakken is mijn onderzoek erop gericht om aanvul-

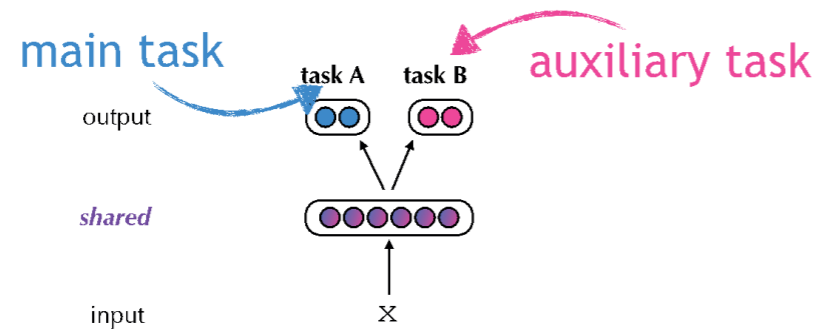
lende bronnen te benutten en deze 'fortuitous data' te koppelen aan Machine-Learning-methoden om natuurlijke taalverwerking te verbeteren. Het idee van fortuitous data komt voort uit het principe multitask learning. Bij multitask learning kunnen meerdere taken



*Figuur 1: Het knelpunt is niet de hoeveelheid aan beschikbare data, maar de beschikbaarheid van geannoteerde data.*

**Barbara Plank**  
**Faculteit**  
**Letteren,**  
**Rijksuniversiteit**  
**Groningen**





Figuur 2: Het idee van deep multitask learning (schematisch)

tegelijk en parallel opgelost worden, terwijl verschillen en overeenkomsten gebruikt worden om beter te leren. Net zoals wij mensen dat doen: bij het leren van een nieuwe taal gebruiken we bijvoorbeeld opnieuw wat we al weten van andere talen (Engels), en bij het leren besturen van een motorfiets gebruiken we opnieuw wat we al weten over het besturen van een auto.

Multitask learning is direct toepasbaar in deep neural nets. De samengestelde aard van het netwerk maakt het gemakkelijk om extra output modes (voor meerdere taken) toe te voegen.

Samen met het succes van automatische feature-extractie, door representaties geleerd uit grote hoeveelheden ongeannoteerde data, resulteert dit in een krachtige en flexibele methode. Deze kan worden toegepast om uit verschillende bronnen te leren.

### Syntactische analyse

Mijn onderzoek was in eerste instantie gericht op het verbeteren van automatische syntactische analyse, de taak om zinsdelen te identificeren (zoals werkwoord-constituenten, naamwoord-constituenten en dergelijke). Fortuitous data zijn data die oorspronkelijk bedoeld waren voor een andere taak, het is dus niet expliciet geannoteerd materiaal. Wat zijn voorbeelden van dergelijke bronnen? En hoe kunnen we uit deze verschillende bronnen leren?

### Toetsaanslagen

Een bron van data die ik benut heb is menselijk gedrag, bijvoorbeeld in de vorm van logs van toetsaanslagen op de computer. De tijd van toetsaanslagen wordt gemeten en ook wanneer welke toets wordt aangeslagen kan worden geregistreerd. Logs van toetsaanslagen worden vooral gebruikt in onderzoek naar schrijfvaardigheden of computer security. Ik vroeg me af of ik dergelijke hulpdata ook kon gebruiken om taaltechnologie te verbeteren. In recent onderzoek heb ik deze bron toegepast op syntactische analyse. Het idee

was dat toetsaanslagen een soort syntactisch signaal bevatten. Door het signaal uit de metingen van toetsaanslagen af te leiden heb ik automatisch geannoteerde data verzameld, en deze gebruikt als aanvullende informatie in een deep multitask learning model (Figuur 2). De resultaten tonen aan dat deze toetsaanslagen een signaal bevatten dat syntactische analyse kan verbeteren.

### En verder

Andere voorbeelden van 'fortuitous data' bronnen zijn het gebruiken van URL's voor het verbeteren van tagging (bv. Entity recognition), of het gebruiken van images in automatisch vertalen. Dit zijn maar de eerste, veelbelovende stappen naar een toekomst om het data sparsity probleem van AI aan te pakken.

Meer informatie over het onderzoek van Barbara is te vinden op: [www.let.rug.nl/~bplank](http://www.let.rug.nl/~bplank)

### Verwijzingen voor de geïnteresseerde lezer

1. Chris Manning. Computational Linguistics and Deep Learning. Computational Linguistics Journal, MIT Press, 2015
2. DOMO. Data never sleeps 5.0 <https://www.domo.com/blog/data-never-sleeps-5/>
3. Barbara Plank. What to do about non-standard (or non-canonical) language in NLP. In KONVENS 2016 <https://arxiv.org/abs/1608.07836>
4. Barbara Plank. Keystroke dynamics as signal for shallow syntactic parsing. The 26th International Conference on Computational Linguistics (COLING). Osaka, Japan. <https://arxiv.org/abs/1610.03321>
5. Barbara Plank, Dirk Hovy, Ryan McDonald and Anders Søgaard. Adapting taggers to Twitter with not-so-distant supervision. In COLING 2014
6. Desmond Elliott and Ákos Kádár. Imagination improves Multimodal Translation. ArXiv, 2017

## Computer, hoe lang heb ik nog?

"Hoe lang heb ik nog?" is een vraag waar iedere ongeneeslijk zieke patiënt graag een antwoord op wil. Correcte prognosticatie is belangrijk. Wanneer arts en patiënt het levenseinde op tijd zien aankomen, hebben zij genoeg tijd om voorkeuren en wensen te overleggen met naasten, en om te bepalen welke zorg nog zinvol en gewenst is. Ook kunnen er dan op tijd bepaalde procedures in gang worden gezet, zoals opname in een hospice en het opstellen van een verklaring voor niet reanimeren.

Het is echter moeilijk en tijdrovend om de levensverwachting in te schatten. Het doet ook een zwaar beroep op het subjectieve oordeel van de arts. Het is bekend dat dokters in slechts 20% van de gevallen de levensverwachting binnen acceptabele marges correct schatten. En dat is begrijpelijk. Hoe een ziekte zal verlopen wordt beïnvloed door tientallen, vaak zelfs honderden factoren, zoals het samenspel van symptomen, allerlei kenmerken van de patiënt zelf, zijn of haar leefgewoonten en bijvoorbeeld de medische familiegeschiedenis. Artsen hebben geen perfect geheugen en de werkdruk is te hoog om telkens complete medische dossiers door te nemen om alle mogelijk relevante factoren op te sporen. Hierdoor is het voor artsen vrijwel onmogelijk om alle belangrijke informatie mee te nemen bij het voorspellen van de levensverwachting - al is het maar voor één patiënt.

### Beslisboom

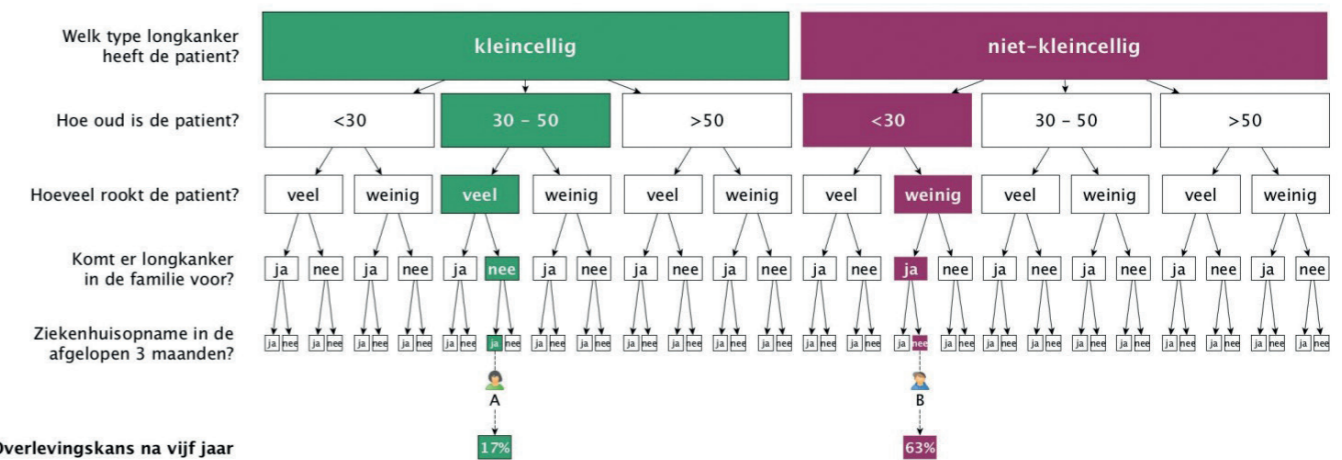
Bestaande hulpmiddelen zijn vaak beslisbomen waarin een aantal factoren zijn opgenomen waarvan bekend is dat zij de levensverwachting beïnvloeden.

Onderzoek wijst uit dat zulke hulpmiddelen vaak slecht functioneren omdat zij te veel generaliseren. Bij bestaande hulpmiddelen

is generaliseren echter noodzakelijk: als alle relevante informatie meegenomen zou worden, zouden beslisbomen zo groot en complex worden dat ze niet meer bruikbaar zouden zijn. Ook wijst onderzoek uit dat te veel afstemming op de patiënt eveneens nadelig werkt: prognoses van artsen zijn slechter naarmate zij de patiënt langer kennen en meer details over hem/haar weten. Bovendien is de gemiddelde huisarts niet geoefend in het herkennen van een naderend levenseinde. Van de tweeduizend patiënten waar de gemiddelde huisarts jaarlijks mee te maken heeft, overlijden er ongeveer vijftien per jaar. Zelfs als een huisarts geen andere patiënten zou hebben om zich over te bekommeren, dan nog levert zo'n kleine, uiteenlopende groep van sterfgevallen niet genoeg data om systematische en bruikbare patronen te herkennen.

### AI

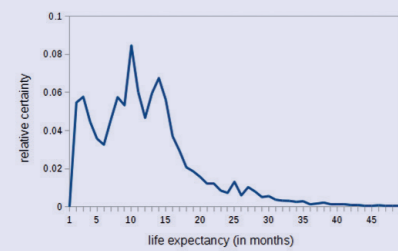
Computers bieden een interessant alternatief voor menselijke beoordeling. Ze kunnen in een mum van tijd getraind worden op honderduizenden sterfgevallen, hebben een perfect geheugen en worden niet beïnvloed door hun verstandhouding met de patiënt. Technieken uit de gebieden van machine learning en automatische tekstanalyse worden steeds populairder in het medische domein. Ze heb-



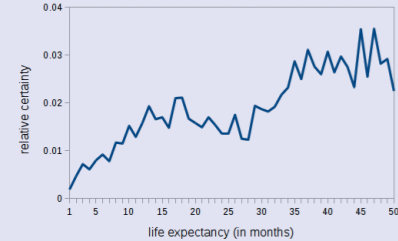
Figuur 1: Versimpeld voorbeeld van een beslisboom om de levensverwachting te bepalen voor mensen met longkanker. Het voorbeeld toont de levensverwachting voor patiënten met verschillende combinaties van kenmerken: een patiënt van type A (groen) en een patiënt van type B (paars).

Merijn Beeksma  
CLST,  
Radboud  
Universiteit

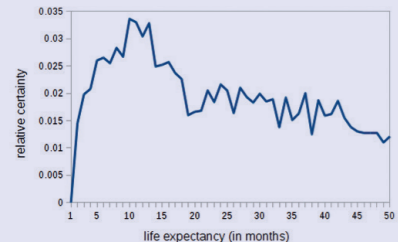




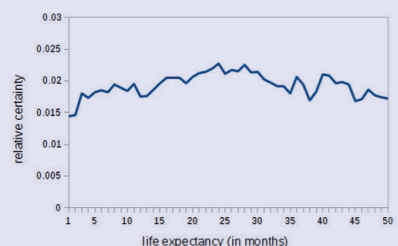
Figuur 2: Onderbroken pieken van verhoogd risico in maanden 3, 11 en 15



Figuur 3: Een lange periode van verhoogd risico in maanden 35-50



Figuur 4: Afgbakende periode van continu verhoogd risico rond maand 12



Figuur 5: Geen duidelijke periode van verhoogd risico

ben veel toepassingen, zoals het ondersteunen van klinische beslissingen en het detecteren en diagnosticeren van ziektes.

#### Neurale netwerken

Geïnspireerd door de verschillen tussen mensen en computers, voeren wij onderzoek uit naar het voorspellen van de levensverwachting met neurale netwerken. Kunstmatige neurale netwerken zijn zelflerende algoritmes die – in meer of mindere mate – op een vergelijkbare manier werken als het menselijke brein. Neurale netwerken zijn in staat om complexe patronen te herkennen die voor mensen niet voor de hand liggend lijken. Werken met klinische data is namelijk uitdagend: vaak zijn gerelateerde stukken informatie weken of maanden van elkaar gescheiden, zoals bepaalde klachten en de bijbehorende diagnose. Relevante informatie kan zelfs in een ander dossier staan dan dat van de patiënt in kwestie, bijvoorbeeld bij erfelijke aandoeningen. Om deze problemen aan te pakken, maken wij gebruik van een long-short term memory (LSTM), een neuraal netwerk dat speciaal ontworpen is om belangrijke informatie gedurende vele tijdstappen te bewaren, en af te leveren wanneer dat nodig is.

#### LSTM

We trainen ons LSTM, of model, op elektronische patiëntendossiers, die worden samengesteld en bijgehouden door de huisarts. Het dossier bevat diagnostische codes, testuitslagen, metingen, voorgeschreven medicatie en andere vormen van data, in de vorm van gestructureerde medische codes. Patiëntendossiers bevatten daarnaast nog een belangrijke bron van informatie: notities van huisartsen en briefcontacten tussen de huisarts en andere specialisten. Getraind op deze data voorspelt het model hoe groot het risico van overlijden is in de toekomstige maanden. De voorspellingen van het model kunnen heel verschillend zijn: één of meerdere duidelijke pieken van verhoogd risico, een langere periode van verhoogd risico of juist géén duidelijk verhoogd risico. De periode waarvoor het model een voorspelling doet is variabel. In de grafieken (figuur 2 t/m 5) doet het mo-

del een voorspelling voor de komende vijftig maanden.

#### Nauwkeurige voorspellingen

Getraind op iets meer dan duizend sterfgevallen overtreft ons model de nauwkeurigheid van dokters bij het voorspellen van de levensverwachting voor nieuwe patiënten. Waar voorspellingen van artsen in 20% van de gevallen kloppen, haalt het model een nauwkeurigheid van 29%. Desondanks is het niet de bedoeling dat de computer binnenkort de dokter vervangt. In onze opzet is de huisarts allesbehalve overbodig. De voorspellingen zijn gebaseerd op de diagnoses die gesteld zijn door de huisarts, op de resultaten van testen die door de arts zijn aangevraagd, uitgevoerd en gedocumenteerd, en op diens notities en de brieven over de patiënt.

#### Brieven

Dit laatste type data, de tekstuele data, bevat observaties die niet te vatten zijn in medische codes. Brieven en notities bevatten nuances over de gedocumenteerde diagnoses, beschrijven details of schetsen juist een algemeen beeld van de mentale en fysieke staat van de patiënt. Ze bieden context en achtergrondinformatie. Ze bieden ook ruimte voor twijfels, verdenkingen en ideeën over de medische klachten. Ze beschrijven kwesties die in isolatie niet belangrijk lijken, maar die de arts opvallend vindt voor een bepaalde patiënt in een bepaalde context. Het model baseert de voorspellingen dus op de data die geleverd worden door de arts, die op zijn beurt op basis van de voorspelling van het model kan inschatten of het behandelplan aangepast moet worden.

#### Mens en machine

Het belang van de symbiose tussen mens en machine moet dan ook niet worden onderschat, vooral wanneer er beslissingen worden gemaakt die van levensbelang zijn. Computers zijn in staat sommige taken even goed en soms zelfs veel beter en efficiënter uit te voeren dan een mens ooit zou kunnen. Waar mensenwerk foutgevoelig is, excelleren computers in het foutloos uitvoeren van instructies. Mensen zijn daarentegen empathisch en kunnen in een oogwenk zien of het goed of slecht met iemand gaat. Zonder dat daar een woord over gesproken hoeft te worden. Hoewel het zinvol is de vaardigheden van computers te vergelijken met die van mensen en te proberen deze laatste te overtreffen, is het minstens zo belangrijk om verschillen tussen beide te erkennen. We moeten zoeken naar de optimale balans tussen de kwaliteiten van mens en machine. Alleen dan zijn we echt in staat technologie te ontwerpen die ons naar nieuwe hoogten doet stijgen.

# READSPEAKER LOOPT VOOROP IN TEKST-NAAR-SPRAAK



## 200+



#### STEMMEN

AI 200+ stemmen in 50+ talen in de **ReadSpeaker** portfolio en investering in AI belooft stemmen die nóg natuurlijker en expressiever zijn

#### LANDEN

Als wereldwijde marktleider heeft **ReadSpeaker** klanten in 65 landen, kantoren in 14 landen en 24/7 klantondersteuning

## 65



## 10.000+



#### KLANTEN

In 30+ sectoren, waaronder onderwijs, overheid, telefonie, IoT, media, transport, zorg en entertainment werkt **ReadSpeaker** aan de toekomst van tekst-naar-spraak

#### JAREN ERVARING

Met een volledig aanbod van SaaS en licensing oplossingen en een toegewijd R&D team is **ReadSpeaker** een erkend expert op het gebied van tekst-naar-spraak

## 18



**ReadSpeaker**

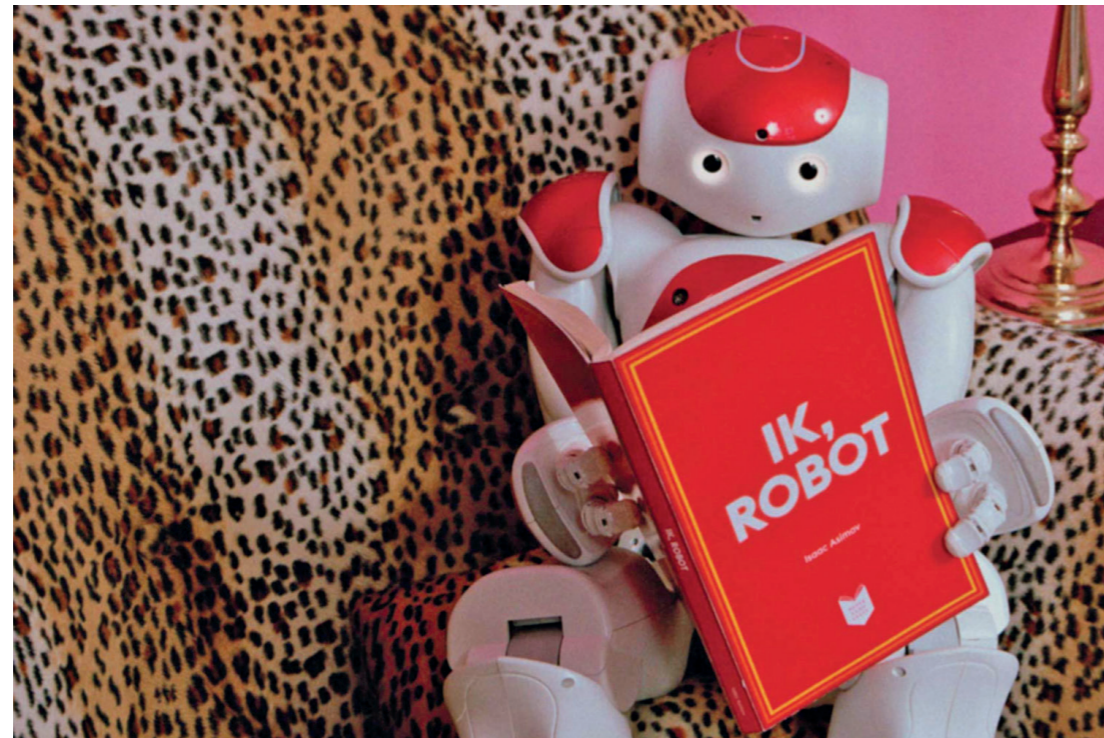
A HOYA SPEECH COMPANY

WWW.READSPEAKER.COM



# Synthetische literatuur: neurale netwerken als coauteur

Ronald Giphart doet een literair experiment door samen met een intelligente robot een verhaal te schrijven. Hij gebruikt zowel een oude tekstgenerator als nieuwe recurrente neurale netwerken. Het resultaat lijkt veelbelovend.



**Folgerit Karsdorp Meertens Instituut Enrique Manjavacas, Ben Burtenshaw en Mike Kestemont Universiteit Antwerpen**

De Stichting CPNB organiseert ieder jaar de campagne Nederland Leest! ter promotie van het Nederlandse boek. Het thema van 2017 is 'Robotica', waarbij aandacht besteed wordt aan de maatschappelijke mogelijkheden en uitdagingen die gepaard gaan met de toenemende rol van artificiële intelligentie in ons dagelijks leven. Een centraal onderdeel van de campagne is de heruitgave van de beroemde sciencefiction-verhalenbundel 'Ik, Robot' van Isaac Asimov (1950).

## Ik, Robot

Het boek van Asimov is in het Nederlands vertaald door Leo Zeldersin 1966. De bundel bestaat uit negen korte verhalen, verschenen in het magazine Astounding Science Fiction tussen 1940 en 1950. Het is vooral beroemd om de 'Drie Wetten van de Robotica' van Asimov. Om het debat over de rol van artificiële intelligentie in literatuur aan te zwengelen, wil het CPNB aan de heruitgave van 'Ik, Robot' een tiende verhaal toevoegen. Dit is geschreven door Ronald Giphart, bekend van romans zoals 'Phileine zegt sorry' en 'Ik ook van jou'. Ronald Giphart schrijft het verhaal

samen met een computer. Hoe gaat deze samenwerking in z'n werk?

## Tekstgeneratie

De voor dit experiment ontwikkelde machine bestaat uit twee onderdelen: een systeem voor tekstgeneratie en een interface waarmee een auteur de interactie kan aangaan met het tekstgeneratiesysteem. Tekstgeneratie is al decennia lang een centraal onderzoeksprobleem in de artificiële intelligentie, waarin onaderzocht wordt hoe automatische systemen taal (zinnen, paragrafen en zelfs hele verhalen) kunnen genereren. Het generatiesysteem dat wordt gebruikt in het huidige literaire experiment is gebaseerd op zogeheten 'statistische taalmodellen' waarvan de oorsprong in begin jaren '80 ligt. Het basisidee van die modellen is heel simpel. Stel, Giphart typt de letter e. Wat zou de volgende letter zijn? Een n is erg waarschijnlijk, evenals opnieuw een e, maar een z is zeer onwaarschijnlijk. Na twee e's verschuiven de waarschijnlijkheden, en is de kans op een n veel groter dan een e (het Nederlands kent immers geen woorden die beginnen met drie

e's). Als nu de volgende 'letter' een spatie is, wordt het lidwoord een gevormd. Uiteraard kunnen we ook langere stukken tekst genereren, zoals zinnen en paragrafen.

## RNN

Dit simpele idee vormt ook de basis voor een nieuw soort taalmodellen die gebaseerd zijn op recurrente neurale netwerken (RNN). Deze zijn in de afgelopen jaren erg populair geworden. Deze modellen zijn ongelooflijk effectief en kunnen – mits er genoeg trainingsmateriaal voorhanden is – voor allerlei doelen worden ingezet. Denk aan het genereren van Shakespearesonnetten, Wikipedia pagina's en – met wisselend succes – kookrecepten. (Uit een aantal experimenten bleek een sterke voorkeur te bestaan voor recepten met chocolade in vreemde combinaties, zoals blijkt uit gegenereerde titels als 'Chocolate Pickle Sauce', 'Chocolate Chops & Chocolate Chips' en 'Chocolate Chicken Chicken Cake').

Voor de literaire tekstgenerator is zo'n RNN getraind op duizenden Nederlandstalige romans geschreven door meer dan 1600 verschillende auteurs. Het getrainde taalmodel is in staat om ingevoerde zinnen af te maken of suggesties te doen voor vervolgzinnen. Deze suggesties zijn echter vrij algemeen van toon. Om de tekstgeneratie wat meer kleur te geven, is een reeks extra modellen getraind op specifieke auteurs, zoals Gerard Reve, Kristien Hemmerechts en Isaac Asimov. Deze modellen genereren tekst in de stijl van specifieke auteurs. Op een beginnetje als 'Erika hield niet van...' reageert het systeem getraind op Kristien Hemmerechts met '...de gedachte dat haar moeder zou sterven'. Op dezelfde opening reageert Asimov's model met '...de melkweg', en een model getraind op het werk van Giphart doet de suggestie '...kinderen. Toen ze dit had gezegd, zei ze: 'Misschien kunnen we er meer van maken.' Deze voorbeelden laten mooi het verschil tussen de verschillende stijlen zien.

## AsiBot

Om de interactie met het systeem voor tekstgeneratie efficiënter, intuïtiever en makkelijker te maken, is een interface ontwikkeld waarmee Giphart zijn verhaal kan schrijven. Deze interface, AsiBot gedoopt, heeft veel weg van een gewone tekstverwerker met een aantal extra knoppen om het taalgeneratiesysteem aan te sturen. Zo is er een 'creativiteitsknop' waarmee de creativiteit (of conservativiteit) van de gegenereerde tekst gecontroleerd kan worden. Daarnaast is er een aantal knoppen



© nito www.fotosearch.com Stock Photography

waarmee specifieke taalmodellen bevroegd kunnen worden. Die modellen zijn getraind op bepaalde auteurs (zoals Gerard Reve of Kristien Hemmerechts) of combinaties van auteurs (bijvoorbeeld Asimov plus Nescio). Bij een druk op een van deze knoppen reageert het gekozen taalmodel op de voorafgaande tekst en komt het met een reeks suggesties hoe de tekst verder kan gaan. Deze suggesties kunnen worden toegevoegd aan de tekst en ook, waar nodig, worden aangepast door de menselijk auteur. Vervolgens kan opnieuw tekst worden gegenereerd of kan de auteur besluiten zelf een stuk te schrijven. Op deze manier ontstaat een interactieve uitwisseling van ideeën tussen mens en machine in een co-creatief proces.

## Is dit de toekomst?

Het verhaal dat Giphart samen met deze machine heeft geschreven wordt in november 2017 gepubliceerd. De echte evaluatie van het systeem zal pas dan door de eventuele lezer plaatsvinden. Het systeem kent evenwel nog veel beperkingen. Het kan bijvoorbeeld nog geen lange lijn in een verhaal plannen of vasthouden, en het leert niet tijdens het schrijven. Toch heeft het abstracte idee van een neurale netwerk als co-auteur veel potentie, ook buiten dit literaire experiment. Zo kunnen modellen ontwikkeld worden die kunnen helpen bij allerlei schrijftaken, zoals sollicitatiebrieven, nieuwsberichten, enzovoort. Die modellen kunnen door het doen van concrete suggesties een writer's block doorbreken, ook als de suggesties niet worden overgenomen. Is dit de toekomst? Volgens AsiBot wel, die deze vraag beantwoordde met: 'Ik heb het hele verhaal gelezen,' zei de robot kalm. 'Het was een schitterend geheel.'



# Spreek2Schrijf

CLST en Telecats doen samen een haalbaarheidsstudie naar de mogelijkheid om de gesproken taal in de plenaire bijeenkomsten van de Tweede Kamer, automatisch om te zetten in schrijftaal zoals die nu in de Handelingen wordt gebruikt. Zij doen dit in opdracht van de Dienst Verslag en Redactie van het Nederlandse Parlement, de DVR.

## Spreek- versus Schrijftaal

Spraakherkenning is in de afgelopen drie jaar heel veel beter geworden. De belangrijkste oorzaak hiervan is het gebruik van een techniek uit de AI, genaamd Recurrent Neural Networks (RNN). In een aantal herkenningsexperimenten die we recent gedaan hebben, bleek een foutpercentage van minder dan 5% haalbaar, waarbij moet worden opgemerkt dat het om heel duidelijke, goed gearticuleerde, correct gesproken spraak ging. Maar toch... dit is iets dat een aantal jaren geleden beslist niet mogelijk was.

Deze goede spraakherkenning roept ook weer nieuwe vragen op. Kun je ook automatisch de sprekerswisselingen bepalen en dus herkennen wie wanneer spreekt? En kun je de herkende tekst omzetten in een beter leesbare tekst, die uit grammaticaal correcte zinnen bestaat in plaats van onafgeronde zinnen met uhm, af- en onderbrekingen en haperingen, die zo karakteristiek zijn voor gesproken taal? Wij mensen spreken immers anders dan we schrijven. Een gesproken zin als "ik ben gisteren... o nee, dat was eergisteren en ook samen met Piet, naar de, eh het huis geweest", is voor de meeste luisteraars volstrekt duidelijk. Ook geschreven begrijp je wel wat er bedoeld werd, maar welgevormd is het niet.

Daarover gaat het project Spreek2schrijf. Zou het niet mogelijk zijn om de gesproken tekst

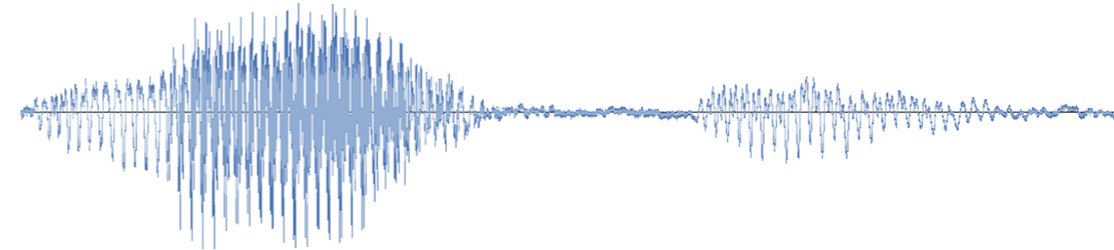
om te zetten in geschreven tekst, waarbij de opeenvolgende woorden worden omgezet in zinnen met een hoofdletter en een punt, zonder de inhoud van de gesproken tekst geweld aan te doen?

Dit is geen makkelijke klus omdat gesproken tekst vaak interne verbeteringen kent ("ik ben, o nee wij zijn..."). Maar ook een niet-perfect resultaat zou de mensen die van de gesproken tekst een leesbare geschreven tekst moeten maken, al enorm kunnen helpen.

## Doel

Het doel van het Spreek2Schrijf-project (S2S) is uit te zoeken, in hoeverre de letterlijke, schriftelijke weergave van hetgeen er in de plenaire sessies van het Nederlandse Parlement gezegd is (bron), omgezet kan worden in schrijftaal (doel), die de manier van schrijven zoals die nu gebruikelijk is, zoveel mogelijk benadert.

Omdat automatische spraakherkenning (ASR) gebruikt zal worden om de spraak in tekst om te zetten, moeten we ervan uitgaan dat het bronmateriaal fouten zal bevatten. Een bijkomend doel van het onderzoek is dan ook om te zien in hoeverre de door de spraakherkenner gegenereerde teksten in schrijftaal overgezet kunnen worden. Hiermee zal menselijke correctie van die



schrijftaal een stuk efficiënter gedaan kunnen worden.

## Data

Sinds september 2014 worden alle vergaderingen in de plenaire zaal van de Tweede Kamer volledig opgenomen. De door de DVR gemaakte teksten (de zogeheten Handelingen) worden automatisch opgelijnd. Hierdoor kunnen de video's van de plenaire debatten automatisch worden ondertiteld en kan er op de gesproken woorden worden gezocht. Deze spraakopnamen worden op dit moment niet door de automatische spraakherkenner gehaald. Om voldoende materiaal voor het S2S-project te krijgen, zullen zoveel mogelijk debatten alsnog door de herkenner gehaald moeten worden.

Er zijn ongeveer 1 miljoen gesproken woorden nodig, naast de parallelle teksten die door de DVR zijn geproduceerd. Een korte beschouwing van de huidige ondertitelbestanden laat zien dat er tussen de 9K en 10K per uur gesproken worden. Dat houdt in dat we minimaal 100 uur AV-materiaal door de herkenner moeten halen om op de benodigde 1M woorden te komen.

## Vertaalmodule

De gesproken en de geschreven teksten worden in S2S beschouwd als twee verschillende talen. Het ligt dan ook voor de hand om, net als bij het vertalen van Nederlands naar Italiaans, automatische vertaalsoftware in te zetten. In S2S gaan we hiervoor gebruikmaken van Moses<sup>1</sup>: een statistisch vertaalsysteem. De vergaarde parallelle teksten, met enerzijds de uitvoer van de spraakherkenner en anderzijds de formele teksten van de DVR, dienen hier als leer materiaal waaruit een statistisch vertaalmodel geleerd wordt. Dit model bestaat uit frasen (woorden of zinssneden) uit de spreektaal, gepaard met frasen uit de schrijftaal die als vertaling gezien kunnen worden. Daarbij wordt voor elk paar een bepaalde waarschijnlijkheid berekend. Het model wordt uiteindelijk ingezet om nieuwe teksten in spreektaal om te zetten naar schrijftaal. Voor andere aanpakken van machinaal vertalen verwijzen we graag naar de artikelen van Lieve Macken (p. 7) en Sander Wubben (p. 22) in dit DIXIT-nummer.

## Twee strategieën

### Correctie van de herkenning

Het is niet zo dat de door de politici gesproken en door de DVR geschreven teksten geheel verschillen. In de meeste gevallen volgt de DVR de gesproken tekst. Wellicht kunnen via een betrouwbaarheidsscore in de oplijn-routine die plekken waar spraak en schrift afwijken, automatisch gedetecteerd worden. Als dat lukt dan hoeft alleen de tekst die afwijkt herschreven te worden. Als dat werkt, hebben we 'perfecte' herkenning (=spreektaal) waarop het spreek2schrijftaal algoritme (S2S-module) losgelaten kan worden.

### Correctie van de schrijftaal

Een andere optie is om de imperfecte spraakherkenningsresultaten met het Spreek2schrijftaal algoritme om te zetten in (imperfecte) schrijftaal en die schrijftaal vervolgens te corrigeren. Het is zelfs mogelijk dat consequente fouten in de ASR-output door de S2S-module automatisch gecorrigeerd worden (als dezelfde fouten in het trainmateriaal van de S2S-module voorkwamen).

In het S2S-onderzoek zullen we beide opties in samenspraak met de DVR bekijken (wat werkt het best?).

## Conclusie

S2S is een mooi voorbeeld van publiek-private samenwerking van twee NOTaS deelnemers: het CLST en Telecats. Als blijkt dat de gekozen aanpak inderdaad leidt tot een verbetering van het werkproces van de DVR (het maken van de Handelingen op basis van wat er gezegd werd in de Kamer), dan biedt de gekozen aanpak enorm veel mogelijkheden voor andere projecten waarbij gesproken spraak in een verslag moet worden omgezet. Maar het is waarschijnlijk niet zo dat de software 1-op-1 ingezet zal kunnen worden. De spraak in de Tweede Kamer is nu eenmaal anders dan die in een willekeurige gemeenteraadsvergadering, of bestuursvergadering van een hondenclub. Wel zal hopelijk blijken dat het idee om dit te zien als een vertaalprobleem tot een succesvolle oplossing kan leiden.

<sup>1</sup> <http://www.statmt.org/moses/>

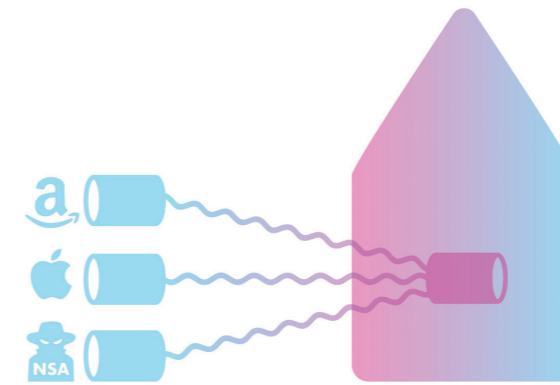






# De impact van neurale netwerken op taal- en spraaktechnologie

Commerciële toepassingen van taal- en spraaktechnologie hebben in korte tijd grote sprongen voorwaarts gemaakt. Een belangrijk voorbeeld is het verschijnen van persoonlijke digitale assistenten (PDA's) zoals Alexa (Amazon), Siri (Apple), Cortana (Microsoft) en Google Assistant (Google). PDA's worden gebruikt in commerciële spraakgestuurde producten zoals Amazon Echo en Google Home. Ze vormen ook de basis voor het aansturen van allerlei andere apparaten zoals thermostaten (Nest) en koelkasten (LG). PDA's brengen ons dus dichterbij het zogenaamde Internet of Things (IoT). Voor velen van ons is het uitvoeren van dagelijkse taken op basis van gesproken commando's een mooi vooruitzicht.



bedrijven volgden snel met software die van vergelijkbare technologie gebruikmaakt. Wat was de doorbraak die het mogelijk maakte om daadwerkelijk bruikbare PDA's te ontwikkelen? Het antwoord is de grote impact die neurale netwerken (NN's) de afgelopen jaren hebben gehad binnen de kunstmatige intelligentie.

Het ontwikkelen van netwerken van digitale componenten om het brein te modelleren was al een onderzoeksonderwerp in de jaren '40 van de vorige eeuw. NN's zijn in staat om patronen te ontdekken waarmee ze specifieke problemen kunnen oplossen. Ze zijn voortdurend in ontwikkeling geweest en hebben geleid tot toepassingen op het gebied van bijvoorbeeld beeldherkenning, medische diagnose en het genereren van tekst.

### GPU

Op het gebied van spraakherkenning en het genereren van spraak waren NN's echter lange tijd inferieur aan andere technieken. In de laatste jaren is deze situatie enorm veranderd, mede door algoritmische verbeteringen maar voornamelijk door het gebruik van veel grotere datasets en de beschikbaarheid van veel snellere chips (voornamelijk graphical processing units; GPU's). Verrassend genoeg waren we dertig jaar geleden al op de juiste weg maar hadden we niet de data en hardware om modellen optimaal te trainen! Nu we wel zo ver zijn: Wat zijn eigenlijk de maatschappelijke implicaties van deze technologie?

### Ongeoorloofd winkelen

Een schattig voorbeeld van de gevaren van stemgestuurde systemen is het volgende. Een jong meisje uit Texas vroeg aan de Echo Dot van haar ouders: "Can you play dollhouse with me and get me a dollhouse?" Dit resulteerde tot verrassing van haar ouders in de zending van een poppenhuis van \$170. Wat nog zorgelijker was: Toen een nieuwslezer op televisie dit verhaal rapporteerde en zei:

"I love the little girl saying, 'Alexa ordered me a dollhouse'", probeerden de Echo's van veel kijkers poppenhuizen te bestellen. Gezond verstand en daadwerkelijk bewustzijn van hun omgeving is iets waar het PDA's nog steeds aan ontbreekt.

### Luisterende PDA's

Er is ook zorg over waar PDA's naar luisteren in privésituaties en of de overheid gebruik mag maken van dit soort opnames. Bijvoorbeeld, het activeren van de luistermodus van Amazon's Alexa door het zeggen van "Alexa" zal een opname van 60 seconden starten. In een recente moordzaak heeft men dit soort opnames proberen op te vragen. Zeker als je de naam 'Alexa' hebt staat het aanschaffen van deze PDA gelijk aan het laten afluisteren van je hele leven. Aan de andere kant kunnen PDA's levens redden door hulpdiensten te zenden in noodgevallen en bieden ze mensen een vorm van kameraadschap. Ouders van kinderen met autisme hebben bijvoorbeeld de ontwikkelaars van Siri bedankt voor het eindeloze geduld en respect waarmee PDA's de oneindige stroom vragen van hun kinderen beantwoorden. Het zijn dit soort nuttige toepassingen van PDA's die de volgende zorg des ter urgenter maken: veel groepen zouden PDA's niet effectief kunnen gebruiken omdat hun accent niet begrepen wordt.

### Minderheidsstemmen

De stemmen van veel PDA's zijn vrouwelijk (wat op zichzelf al rolbevestigend werkt), met een accent dat lijkt op standard Midland American English. Echter, als je zelf een sterk accent hebt (bijvoorbeeld Schots of Iers), dan kan dit standaard accent juist als irritant worden ervaren. Het is in ieder geval zo dat het moeilijk is om robuuste spraakherkenning te ontwikkelen voor minderheidsgroepen met een sterk accent. Stel je de frustratie voor op het moment dat je steeds meer apparaten niet effectief kan besturen vanwege een te sterk accent. Ontwikkelaars moeten zich ook bewust zijn van andere biases die PDA's kunnen ontwikkelen. Zo was Tay, een chatbot ontwikkeld door Microsoft, op non-actief ge-

steld nadat 'trolls' het geleerd hadden om racistische tweets te verspreiden. Onlangs zijn ook in China twee chatbots uit de roulatie gehaald omdat ze neerbuigend spraken over het communisme. Kort gezegd, spraakherkenning en het genereren van spraak in afwezigheid van gezond verstand en moraliteit resulteren vaak in onverwacht en ongewenst gedrag.

### Conclusies

Concluderend wijzen deze problemen op een noodzaak voor ontwikkelaars om zich bewust te zijn van de sociale context waarin zij hun spraakgestuurde apparaten creëren en beschikbaar maken. Men dient niet enkel na te denken over hoe apparaten gebruikt worden, maar ook hoe apparaten misbruikt kunnen worden door eigenaars, onkundige of kwaadwillende derden en de overheid. Speciale aandacht zou gegeven moeten worden aan de mogelijke uitsluiting van mensen door de keuze voor de interface, stem of gebruikte taal. Ondanks deze aandachtspunten zullen spraakgestuurde apparaten steeds verder verweven raken in ons dagelijks leven. In de toekomst zou dit zelfs kunnen leiden tot het verdwijnen van banen (denk aan callcenters). Maar laten we ze tot die tijd gebruiken voor het aangenamer maken van ons leven: "Siri, is er een pizzatent in de buurt?"



George Kachergis & Marcel van Gerven  
Donders Centre for Cognition,  
Radboud Universiteit

Afbeeldingen:  
Elisa Chaudet

### Evoluërende technologie

Spraakherkenning heeft een lange historie waarin gedurende lange tijd slechts kleine stappen werden gezet. In 1952 had men een systeem ontwikkeld dat gesproken getallen (1 tot en met 10) kon herkennen. Tien jaar later had men een systeem ontwikkeld dat 16 woorden kon herkennen. Na tien jaar onderzoek kon men dus zes woorden meer herkennen. Het is zeker zo dat het herkennen van spraak zeer uitdagend is gegeven de complexiteit, variabiliteit en ambiguïteit van gesproken taal. Mensen gebruiken dagelijks duizenden woorden uit een vocabulaire van ongeveer 60.000 woorden. Het segmenteren van deze continue geluidsstroom in individuele woorden is zeer lastig. Hidden Markov modellen (HMM's) die de overgangskansen van sequenties van spraakelementen schatten zijn redelijk succesvol gebruikt in commerciële dicteersoftware (bijvoorbeeld Dragon NaturallySpeaking, 1997). Echter, tot een aantal jaren terug was het niet mogelijk om de transcriptieaccuraatheid van die van mensen te benaderen (zo'n 89-94 procent).

### Siri

Apple's introductie van Siri in 2011 voor de iPhone luidde een nieuw tijdperk in voor toepassingen van spraakherkenning. Andere



## Nieuwe dilemma's met oude en nieuwe online data

*Sinds een aantal jaren ben ik lid van de Ethische Toetsingscommissie voor de Geesteswetenschappen van de Radboud Universiteit. Deze commissie toetst of voorgenomen onderzoeksprojecten op een verantwoorde manier omgaan met mensen en de gegevens van mensen. Onderzoeksprojecten mogen proefpersonen niet hinderen of schaden in hun privacy. Voor allerlei soorten onderzoek, zoals experiment en interview, hebben we beschreven hoe deze standaard plaatsvinden. Een standaard voor onderzoek met grootschalige online data in het kader van AI en taal/spraaktechnologie is er nog niet.*

**José Sanders**  
CLST,  
Radboud  
Universiteit

Afbeeldingen:  
Pixabay

Die beschrijving ontbreekt niet omdat het niet voorkomt – onderzoek met grootschalige online data komt juist steeds vaker voor – maar omdat het helemaal nog niet standaard is. Bij elk voorkomend project bekijken we wat relevante aspecten zijn en zo worden 'werkendeweg' standaardeisen ontwikkeld.

### Eigenaar van gegevens

Aanvankelijk was bij AI-onderzoekers de grootste zorg meer juridisch dan ethisch van aard. Het ging om de vraag van wie online data eigenlijk zijn. Bij gegevens die op het internet staan en voor onderzoek worden gebruikt zijn zowel het auteurs- als het reproductierecht van belang. Voor corpus-opbouw moesten overeenkomsten worden ontwikkeld (zoals in het SoNaR-corpus).

Men vroeg zich af of ook nog eens afgewogen moest worden of ongerief voor personen, of de bescherming van privacy van personen, bij onderzoek naar internetdata een punt van aandacht moest zijn. De tot de persoon van de auteur herleidbare gegevens uit online data zijn immers via het internet algemeen beschikbaar gesteld door of na-

mens deze persoon. Mogelijk ongerief is dan toch al veroorzaakt door het aanwezig zijn op het internet en niet door het onderzoek.

### Identiteit achterhalen

Gaandeweg zijn we daar wel wat genuanceerder over gaan denken. AI is het juridisch geen probleem, het kan toch moreel lastig zijn om gegevens die gebruikers in het verleden geplaatst hebben, voor onderzoek te gebruiken, zonder zich rekenschap te geven van de gevolgen die dit zou kunnen hebben. Dit geldt vooral voor het grootschalig openbaar maken in publicaties en publiek beschikbare corpora, met name als de gegevens achteraf gezien schadelijk kunnen zijn voor mensen (de gebruiker zelf of de mensen over wie de gegevens gaan). Dan moeten tot deze personen herleidbare details worden verwijderd. Dat zijn niet alleen namen en plaatsen maar ook andere elementen die het mogelijk maken om de identiteit van mensen te achterhalen.

Zo'n eis geldt in mindere mate als het gaat om mensen die toch al gekozen hadden voor een openbaar leven, zoals politici en andere in massamedia optredende perso-

nen. Verondersteld mag worden dat zij zich ervan bewust zijn dat wat zij in media uiten, beschikbaar blijft. Dit geldt ook voor mensen die doelbewust een groot online publiek nastreven, zoals bloggers en vloggers. Maar overige gebruikers van online media zijn zich daar vaak minder van bewust en gebruiken online media als onderlinge interactiekanalen. Dat is zeker aan de orde als gebruikers ten tijde van het plaatsen van informatie minderjarig waren.

### Minderjarigen

Bij het actief verzamelen van online gegevens, zoals geluidsfragmenten van dialecten of tekstfragmenten van Whatsapp, staan onderzoekers voor een dilemma. Hoe kun je online controleren a) wat de leeftijd van deelnemers is en b) of hun ouders of wettelijke vertegenwoordigers toestemming geven voor deelname? Er wordt immers geen identiteit van proefpersonen opslagen. De gelijktijdige eisen om zowel anonimiteit van opgeslagen data te garanderen als deze te controleren voor toestemming in geval van minderjarigheid, zorgen voor een paradox. Bij het online werven van data voor corpora en in surveys wordt daarom geëist dat:

1. in het online protocol de leeftijd wordt ingevoerd en
2. dat bevestigd wordt dat één van de ouders/vertegenwoordigers bij het invullen aanwezig is, meeleest en naleest, en
3. die ouder/vertegenwoordiger expliciet zelf toestemming geeft voor onderzoekgebruik en opslag van wat de jongere aanlevert.

Omdat deze zaken niet achteraf gerepareerd kunnen worden, is het belangrijk om ze voorafgaand aan het plaatsen van oproepen om online informatie goed te regelen. Leeftijdinformatie en toestemming moeten eerst binnengehaald worden. Waterdicht zijn deze maatregelen niet, maar ze maken aan gebruikers wel expliciet duidelijk wat wel en niet de bedoeling is. Zo krijgen zij daarin een grotere alertheid; in die zin werkt het in elk geval preventief en dat is misschien op dit moment wel het hoogst haalbare.

### Gegevens uit het verleden

Het komt ook voor dat in het verleden verzamelde mediagegevens, bijvoorbeeld uit radio of krant, nu in grote doorzoekbare corpora op het internet beschikbaar worden gesteld voor verder onderzoek. Dat levert een extra dilemma op. In de periode vóór het internet zullen mensen zich niet hebben gerealiseerd dat hun ingezonden brief of radio-interview in plaats van éénmalig, eindelijk beschikbaar zou worden gesteld.



Achteraf om toestemming vragen is ondoenlijk. Met zulke gegevens moeten we daarom voorzichtig zijn. Openbaarmaking van gegevens die achteraf gezien ongerief kunnen veroorzaken bij nog levende bronnen of bij nabestaanden, moet net als geldt bij historisch onderzoek, voorkomen worden door anonimisering.

De vraag is vervolgens wie moet inschatten wat ongerief kan veroorzaken: de onderzoeker? En is het corpus niet te groot om dit voor alle fragmenten vast te stellen? Als dat niet mogelijk is, zijn er goede argumenten om zo'n corpus niet openbaar te maken, maar alleen ter beschikking te stellen aan onderzoekers. Als die er een deel van willen gebruiken in een publicatie, kan dat stuk in elk geval van herleidbare gegevens worden ontdaan. Ook hier kunnen gegevens van destijds reeds openbaar bekende mensen uitgezonderd worden.

### Relativering

Onmiskenbaar veranderen internet en online media de samenleving in de manier waarop wij omgaan met media-informatie. Er zijn zoveel online data dat het misschien gaandeweg minder belangrijk wordt hoeveel en wat van ons ergens online staat. Jongeren staan daar nu vaak al relativerender in dan ouderen. Dat kan betekenen dat we over een aantal jaren minder eisen zullen stellen aan anonimiteit en toestemming, simpelweg omdat van iedereen wel ongerief veroorzakende data te vinden zijn en we het daarom collectief opgeven om ons daar druk over te maken. Maar zo ver zijn we nog niet, en daarom blijft ethiek een extra uitdaging voor iedereen die zich met de verzameling en opslag van online data bezighoudt.





# Deep Learning revolutie - en wat nu?

Deep Learning heeft in de afgelopen 5 jaar een revolutie teweeggebracht in de kunstmatige intelligentie, daarbij de taal- en spraaktechnologie meeslepend in zijn stroom. Wat zijn de uitdagingen waar we nu voor staan? Op welke vlakken bieden diepe neurale netwerken nu nog geen oplossing?

Suzan Verberne  
Leiden Centre of  
Data Science

De techniek van kunstmatige neurale netten bestond al lange tijd: in de jaren '80 waren de belangrijkste delen van de algoritmiek al ontwikkeld. Maar pas tussen 2009 en 2012 werden grote stappen gezet, zowel in doorontwikkeling van de diepe netwerken als in het beter gebruikmaken van de inmiddels beschikbare computerkracht. De kwaliteit van beeldherkenning maakte een grote sprong. Zo groot dat zelfrijdende auto's nu al uitgebreide testritten maken in steden in Europa, Amerika en Azië. Dit werd 10 jaar geleden niet voor mogelijk gehouden.

## Deep learning voor automatisch vertalen

Ook in de taal- en spraaktechnologie had deep learning een groot effect. De kwaliteit van automatisch vertalen en spraakherkenning maakte grote sprongen. Op 27 september 2016 kondigde Google aan dat dankzij de zogenoemde LSTM's voor een aantal taalparen vertalingen van menselijke kwaliteit mogelijk waren<sup>1</sup>. (LSTM's zijn neurale netwerken voor taaldata, waarbij het netwerk een groot stuk van de voorgaande tekst in zijn geheugen kan houden.)

Het bedrijf DeepL haalde eind augustus 2017 de krant met het nieuws dat hun vertaalsysteem nog beter is dan Google Translate<sup>2</sup>. Je zou bijna denken dat automatisch vertalen een opgelost probleem is. Hoe kan het dan dat de ervaring van gebruikers vaak heel anders is? Dat een vertaling van Google Translate toch nog veel fouten bevat?

## Kwaliteit

Ik stelde die vraag aan mijn collega Arianna Bisazza, een specialist in automatisch vertalen. Ze legde uit dat automatisch vertalen inderdaad soms een opgelost probleem lijkt, als je de nieuwsberichten mag geloven. Dat komt doordat de gemeten kwaliteit afhangt van de evaluatiemethoden die gebruikt worden. Het type data is een belangrijke factor: een formele tekst automatisch vertalen gaat veel beter dan een informele tekst, en volledige alinea's automatisch vertalen gaat veel beter dan korte zinnen. Dus: als je naar formele, complete teksten kijkt, dan lijkt automatisch vertalen een opgelost probleem. Maar gebruikers van Google Translate

en DeepL willen vaak ook informele (web) teksten laten vertalen, en dan geeft het systeem helemaal niet altijd de gewenste uitvoer.

## Verklarende modellen

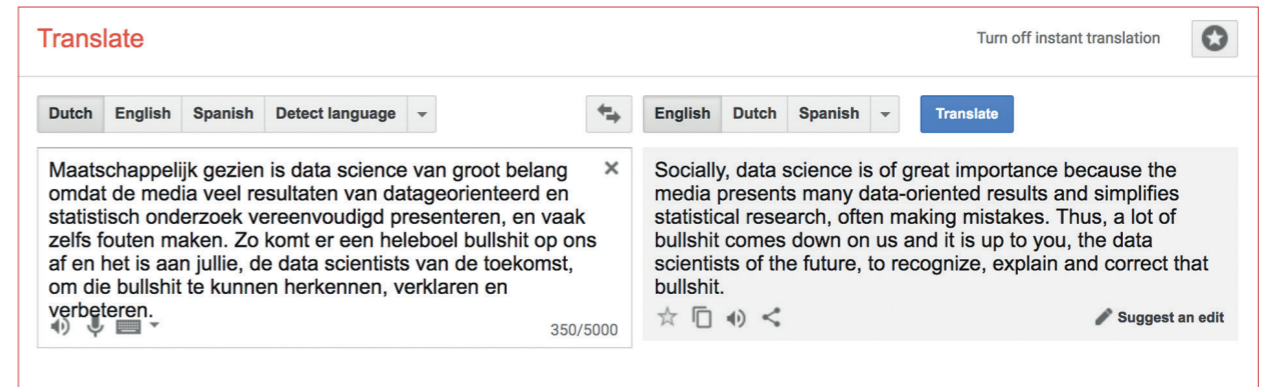
Arianna gaf aan dat er een grote behoefte is aan de ontwikkeling van 'verklarende modellen', die kunnen laten zien wat het neurale netwerk precies leert en hoe de geleerde kennis wordt toegepast op de te vertalen tekst. Wat zijn de taalkundige constructies die niet goed gaan? Wat doet de computervertaler anders dan de menselijke vertaler? Voor dergelijke analyses is kennis van linguïstiek nodig.

## Grote stappen

Niet-letterlijk taalgebruik, ironie en subtiliteiten zijn de grote uitdagingen voor automatisch vertalen. Op Twitter (@renepijl) kwam ik het volgende voorbeeld tegen: "Ga toch heen met je Brexit" werd door DeepL vertaald als "Go ahead with your Brexit." Oeps, nét niet helemaal goed. Informeel en figuratief taalgebruik is erg lastig voor vertaalsystemen!

Een redelijk formele, lange tekst daarentegen gaat behoorlijk goed. Het voorbeeld in Figuur 1 is van mijzelf, toen ik een cursusbeschrijving aan het vertalen was van het Nederlands naar het Engels. De vertaling is goed naar mijn smaak, al zouden sommige formuleringen subtiel anders kunnen. Met een taalkundige bril bekeken is het best indrukwekkend dat Google Translate de laatste bijzin in de juiste Engelse volgorde zet: "om die bullshit te kunnen herkennen, verklaren en verbeteren". Het algoritme moet de hele bijzin in zijn geheugen hebben om de Engelse vertaling correct te kunnen geven.

Arianna gaat de komende jaren werken aan het ontwikkelen van verklarende modellen, die kunnen helpen begrijpen wat kunstmatige neurale netwerken anders doen dan mensen en waarom. Ze zegt dat deep learning niet meer weg te denken is uit onderzoek naar en ontwikkeling van methoden voor automatisch vertalen. Hoe zit dat met andere taken in de taal- en spraaktechnologie?



## Deep learning als oplossing voor alle problemen?

Informaticastudenten willen heel graag werken met diepe neurale netwerken. Ze hebben het succes gezien van die methoden voor beeldherkenning. Ik leg studenten uit dat deep learning door de succesverhalen een grote aantrekkingskracht heeft, maar dat het niet voor elk probleem de beste (of zelfs een mogelijk geschikte) oplossing is. Barbara Plank noemt in haar artikel (p. 13) al de beperking van te weinig geannoteerde data. Veel informatie-extractie-problemen lijden daaronder, omdat ze erg specifiek zijn. Voorbeelden die ik recentelijk ben tegengekomen in mijn eigen werk, zijn het vinden van kanker-recidieven in radiologierapporten en het herkennen van persoonsnamen in bankoverboekingen. Voor dergelijke specifieke problemen zijn we aangewezen op

traditionele informatie-extractie-methoden met behulp van een relatief kleine set handmatig geannoteerde voorbeelden.

Op dit vlak moeten deep-learning-methodieken de komende jaren nog grote stappen gaan zetten, met generieke methoden voor specifieke problemen, en met verklarende modellen die ons inzicht verschaffen in de uitdagingen en verbeteringen.

1. <https://research.googleblog.com/2016/09/a-neural-network-for-machine.html>
2. [www.volkskrant.nl/tech/deze-nieuwe-vertaalmachine-gaat-de-strijd-aan-met-google-en-microsoft~a4513761](http://www.volkskrant.nl/tech/deze-nieuwe-vertaalmachine-gaat-de-strijd-aan-met-google-en-microsoft~a4513761)

**Your wake-up call**

**leonard**  
B2B COMMUNICATIE

0162-453203 | [leonard.nl](http://leonard.nl) | [hello@leonard.nl](mailto:hello@leonard.nl)  
WEBSITES | SEO/SEA | SOCIAL MEDIA | ACTIVATIE

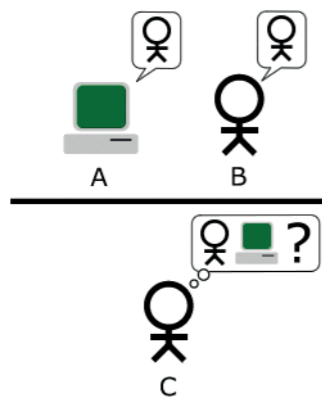


# Gevaren van Artificiële Intelligentie

Als je de artikelen in de pers erop naslaat, dan vind je verschillende soorten 'gevaren van Artificiële Intelligentie ofwel AI' die zich in twee categorieën laten indelen. De een is de terugkerende angst dat 'ze' het van ons gaan overnemen. De ander is dat 'ze' je leven gaan bepalen. En was het 'vroeger' zo dat de 'angsthazen' afgedaan konden worden als fantasisten, die a) niet wisten waar het over ging en b) een veel te rijke fantasie hadden, nu zijn het ook 'geleerden van naam' die waarschuwen tegen het ongebreidelde gebruik van AI.

## Wat is AI?

Maar wat is die artificiële intelligentie nu eigenlijk? Een sluitende definitie is lastig te geven, maar je zou kunnen zeggen dat AI-software de 'cognitieve' functies nabootst die door de meeste mensen geassocieerd worden met de menselijke psyche. Denk daarbij aan het vermogen om te 'leren', 'abstraheren', 'vertellen', 'samen te vatten', 'liegen' en andere, in onze ogen typisch menselijke gedragingen. De ultieme proef voor AI-software is nog steeds de Turing1 test: een test waarbij een mens met 'iets' communiceert en aan de hand van de antwoorden op gestelde vragen, probeert te bepalen of dat 'iets' een mens is of een computer.



## Soorten AI

'Kunstmatige intelligentie (AI) is het domein van de computerwetenschap dat zich toelegt op de ontwikkeling van software/machines die dezelfde taken kunnen nabootsen en uitvoeren als een mens'. Volgens de AI-filosofie kan AI verdeeld worden in twee soorten: zwakke AI en sterke AI.

Onder zwakke kunstmatige intelligentie verstaan we AI-software die gericht is op het afhandelen van een speciale taak. Denk daarbij aan het herkennen van gesproken spraak, het spelen van een spelletje GO, het zelfstandig besturen van een auto en dergelijke.

Onder sterke kunstmatige intelligentie verstaan we daarentegen technologie die kan denken en functioneren zoals mensen. Het

kan menselijk gedrag nabootsen, en niet alleen in speciale omstandigheden, zoals bij Weak-AI.

## Voetbal

Apparaten met sterke AI-software zouden in principe mee kunnen doen in onze samenleving zonder dat het ons zou opvallen dat het hier om machines gaat. Denk daarbij aan een 'Conversational Agent' aan de telefoon waarmee je op dezelfde wijze zou kunnen praten als met een mens van vlees en bloed. Je belt met zo'n agent over de mogelijkheid om een hypotheek af te sluiten (dat kan nu deels al) en je schakelt halverwege moeiteloos over op de beroerde prestaties van het Nederlands elftal (dat kan nu nog niet). Echte Sterke AI komt dan in de buurt van wat veel mensen als een bedreiging voor het menselijke ras zien, maar zover is het (gelukkig?) nog lang niet.

Alle succesvolle AI-software die we de laatste jaren zien (het herkennen van gezichten, het besturen van auto's, het signaleren van een verhoogde kans op autisme bij baby's) is zwakke AI. Het is dikwijls razend slimme software die is ontwikkeld om een bepaalde taak af te handelen. Dat kunnen ze goed, maar probeer niet aan de software die gezichten kan herkennen, te vragen om een auto te besturen. Wij mensen kunnen dit (meestal) wel: moeiteloos van het ene naar het andere switchen.

## Patroonherkenning versus Cognitie

Maar AI kan ook nog op een andere manier worden onderverdeeld; in AI die sterk leunt op het herkennen van patronen en AI die gebruik maakt van talige, menselijke informatie.

Zonder al te specifiek te worden kun je stellen dat bij patroonherkenning je heel veel data aanbiedt waarvan vooraf bekend is wat het is. Bij het herkennen van plaatjes met poezen krijgt de computer heel veel foto's aangeboden waarvan bekend is dat er wel of niet een poes op staat. De informatie op de foto's (de pixels) wordt op allerlei manieren

bewerkt en aan de verschillende lagen van het Neurale Netwerk aangeboden (vandaar de term Deep Neural Networks). Uiteindelijk moeten de gewichten van iedere laag zo worden ingesteld, dat van een nieuwe foto gezegd kan worden of er wel of niet een poes op staat.

Dezelfde benadering werd door Google's Deep Mind gevolgd bij het leren spelen van het spelletje GO. Het aantal beschikbare voorbeelden was te klein om er goed mee te trainen. Toen kregen de ontwikkelaars het briljante idee om twee computers tegen elkaar te laten spelen. Net als bij plaatjes was de invoer simpel: dit leidt tot winst, dit tot verlies.

## Watson

Een andere, zeer krachtige vorm van AI is de cognitieve, meer menselijke kunstmatige intelligentie. Het meest sprekende voorbeeld daarvan is IBM's Watson. Watson 'leest' tekstdocumenten en is in staat om relaties te leggen tussen de woorden in de tekst. Uit een zin als 'Agatha Christie, de schrijfster van detectives, verbleef een paar maanden met haar man, een bekend archeoloog, in Aleppo' leidt dan tot allerlei relaties zoals:

- Agatha Christie –status-> getrouwd
- Agatha Christie –heeft-> man
- Man –is-> Archeoloog
- Agatha Christie –is-> vrouw
- Agatha Christie –is-> schrijfster
- Schrijfster –van-> boeken
- Boeken –soort-> detectives
- Agatha Christie –is geweest in -> Aleppo

Als we dan nog uit Wikipedia de informatie halen dat Aleppo in het Midden-Oosten ligt, dan kunnen we de vraag beantwoorden: 'welke schrijfster verbleef er in het Midden-Oosten?'

Watson is dus in staat om heel snel grote hoeveelheden tekst te lezen en automatisch dit soort relaties te leggen (zogeneten Triples).

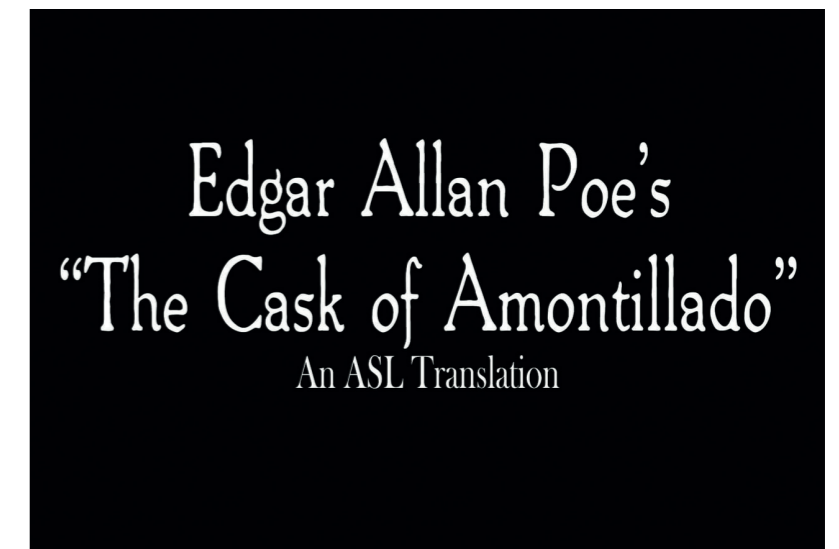
Vervolgens kan Watson redeneren over deze Triples en antwoord geven op vragen. Het is dus duidelijk iets anders dan de hierboven beschreven patroonherkenning, maar we rekenen het net zo goed tot de AI. We mogen verwachten dat beide vormen van AI zich de komende jaren zullen blijven ontwikkelen en ongetwijfeld tot spectaculaire resultaten zullen leiden.

## Gevaar geweken?

Echt sterke AI is er voorlopig nog niet maar is daarmee het gevaar van AI van de baan?

Nee, want ook aan het ongelimiteerd inzetten van zwakke AI kunnen flink wat gevaren kleven. Het gevaar is dat 'ze' de wereld overnemen, terwijl de AI getraind werd met 'verkeerde' data.

Maar wat zijn verkeerde data? Een mooi voorbeeld is het herkennen van foto's door Google. De zoekmachine gaat het internet op en verzamelt alle foto's die een bijschrift hebben. Als in zo'n bijschrift 'poezen', 'poesjes', 'katten', 'kitten' et cetera staat, dan mag je veronderstellen dat er ergens op de foto een afbeelding van een poes te vinden is. Staat die er niet bij, dan is er ook geen poes te zien. Natuurlijk is dat niet altijd waar, want er kan best een foto zijn met Marlies met de poes op schoot en als onderschrift '10de verjaardag van Marlies in Utrecht'. De poes wordt niet vermeld! Of, met alleen Marlies: 'tiende verjaardag Marlies. Ze wil graag een poes, maar dat zien haar ouders niet zitten'. Geen poes te zien, maar wel genoemd.



Maar als er voldoende plaatjes zijn met een poes erop en een bijschrift met het woord poes, dan zou het moeten lukken. En dat lukt want Google kan tegenwoordig goed plaatjes met poezen herkennen.

## Bias

Hetzelfde lukt met plaatjes die als bijschrift 'witte teenagers' of 'zwarte teenagers' hebben. Google kan die plaatjes gebruiken en het algoritme trainen om witte of zwarte teenagers te herkennen. Tot zover is dat geen probleem, maar wat als die plaatjes met zwarte teenagers vooral uit veel beelden van jonge zwarten in een gevangenisomgeving bestaan? En de plaatjes van witte teenagers bestaan vooral uit beelden van ideale schoonzonen/dochteren op het sportveld? Het algoritme 'leert' dat beelden van zwarten in de gevangenis representatief zijn voor zwarte





teenagers. Google en het algoritme valt weinig te verwijten, want ongetwijfeld zitten er verhoudingsgewijs meer jonge zwarten in de gevangenis dan jonge witten. Maar dat wil niet zeggen dat wij, als we zoeken op zwarte teenagers, alleen maar die bepaalde foto's willen zien.

Hetzelfde geldt voor zoeken op CEO. De meeste CEO's zullen nog steeds oudere blanke mannen zijn, maar als wij op CEO's zoeken en alleen maar blanke middelbare mannen voorgeschoteld krijgen, dan zal dat zonder meer ons wereldbeeld beïnvloeden.

#### Hypotheek

Zolang je er bij het zoeken naar plaatjes maar van bewust bent dat dit soort bias kan optreden, is de schade niet zo groot. Maar wat als een op AI gebaseerd algoritme straks gaat besluiten of jij wel of niet een hypotheek krijgt? Stel dat jij een oppassende burger bent die 'toevallig' in een postcodegebied woont waar veel fraude wordt gepleegd. Helaas: de premie gaat omhoog of je krijgt überhaupt geen hypotheek. Dit soort bias is niet alleen voorbehouden aan algoritmes maar treedt ook op bij mensen. Dit bleek uit

de discussie over etnisch profileren bij de politie. Mensen hebben echter de mogelijkheid om over dit soort biases te reflecteren en hun oordeel desnoods aan te passen. Voor algoritmes geldt dit niet. Hierbij laten we buiten beschouwing of het een nu beter is dan het ander.

Het gevaar van het zomaar inzetten van AI zit 'm in het feit dat we van de AI-algoritmes dikwijls niet weten waarom ze bepaalde keuzes maken. Dit gebrek aan transparantie kan er dan toe leiden dat we klakkeloos de suggesties van de computer volgen. We vergeten zelf kritisch te kijken naar de uitslag en de manier waarop die uitslag tot stand is gekomen. Zeker als de algoritmes het vaak 'best goed' doen (want ja, in dat postcodegebied wordt ook vaak fraude gepleegd), dan zullen we ons oordeel in sterke mate door de computer laten bepalen, waarbij we vergeten dat 'vaak' niet hetzelfde is als 'altijd'.

1. <https://nl.wikipedia.org/wiki/Turingtest>
2. [www.differencebetween.com/difference-between-strong-ai-and-vs-weak-ai](http://www.differencebetween.com/difference-between-strong-ai-and-vs-weak-ai)

## Reinventing customer service

*Niet veel mensen realiseren zich dat er in Nederland meer dan 180.000 mensen in meer dan 500 Call Centers werken. Call Centers? Je moet steeds vaker 'Klant Contact Centers' (KCC) zeggen. Doordat steeds meer mensen op verschillende wijze contact opnemen met organisaties (en bedrijven of overheid), is er in toenemende mate sprake van kanaalintegratie. De KSF (KlantenServiceFederatie, een soort NOTaS voor KCC's) vroeg een aantal specialisten uit het veld om een overzicht te maken van de ontwikkelingen die de komende jaren in de KCC's zouden kunnen plaatsvinden.*

Call Centers zijn al niet meer van deze tijd. Nu stuur je een WhatsAppje naar de gemeente en belt diezelfde avond nog of er al een antwoord bekend is. En als je dan toch iemand aan de lijn hebt, verwacht je ook dat hij weet wat je een paar weken eerder gemailld hebt.

Dat is technisch niet heel ingewikkeld, maar in het recente verleden werd nog gewerkt met aparte kanalen (post, mail, telefoon) of die bestonden gewoon nog niet (Skype, WhatsApp, Facebook). Nu gebruiken klanten/burgers privé steeds vaker alles door elkaar en verwachten dat dus ook van organisaties.

De wereld van de KCC's is een overwegend talige wereld. Mensen stellen vragen, medewerkers kennen de antwoorden of zoeken ze op en delen dit vervolgens mee. Spraakherkenning, spraaksynthese, dialoogvoering, antwoordsuggestie en onderwerp-detectie hebben dan ook al heel lang hun plek gevonden in de KCC's.

Wat gaat er veranderen als AI steeds verder doordringt in ons dagelijks bestaan? Hoe gaan social media zich ontwikkelen en wat heeft dat voor effect op de KCC's? Wat zijn de gevolgen van een sterk groeiend Internet of Things (IoT)? We werpen een blik in de nabije toekomst.

#### AI en TST

Het gebruik van steeds betere AI en TST zal ervoor zorgen dat een groot deel van de 'gesprekken' die nu nog door mensen worden afgehandeld, straks door de computers zal worden gedaan. In eerste instantie zal het 'laaghangend fruit' zijn, maar steeds vaker zullen ook meer complexe vragen door computers worden behandeld. Het is voldoende om naar de huidige prestaties van IBM's Watson te kijken om te begrijpen dat de complexiteit van automatisch te beantwoorden vragen zal uitstijgen boven 'hoe laat gaat de winkel open'. Mensen zullen over 10 jaar in eerste instantie met machines communiceren.

klantenservicefederatie

Pas bij noodzaak (te ingewikkeld, emotioneel beladen, communicatieproblemen) of als mensen bereid zijn er extra voor te betalen, zal een mens het gesprek overnemen.

#### Internet of Things

De massale komst van slimme apparaten, permanent verbonden met internet, zal een enorm effect hebben op de manier waarop KCC's gaan werken. Voorlopig zijn de KCC's vooral reactief: mensen bellen pas als er iets niet goed gaat. De bestelling komt niet aan, de gaskachel brandt slecht, de auto heeft een beurt nodig. Doordat apparaten straks permanent hun status uploaden, weet een leverancier ruim van tevoren dat een gashaard een servicebeurt nodig heeft. Een geautomatiseerd systeem kan de eigenaar vervolgens een berichtje sturen: "We zien dat uw gashaard binnenkort een servicebeurt nodig heeft. Onze monteur is volgende week bij u in de buurt. Komt het gelegen als we dan tussen X en Y langskomen?"

Ook zal het IoT een deel van de zaken die nu nog door mensen zelf gedaan worden, overnemen. Denk bijvoorbeeld aan de Tesla auto die zonder tussenkomst van mensen nieuwe software en (soms) nieuwe functionaliteit krijgt. Bij je computer en telefoon is dit al normaal, maar voor je fornuis, je thermostaat of je sproei-installatie nog niet.

#### Avatars

Ook zijn er per operating systeem of leverancier verschillende gesproken avatars waarmee men zijn/haar wereld kan benaderen (SIRI, GoogleNow, Cortana, Amazon's ECHO, Samsung's Bixby).

Het is de verwachting dat deze intelligente, met spraak en tekst bestuurbare avatars straks de hoofd-apps worden die de overige apps aansturen. Nu nog heb je een app voor

Arjan van Hessen  
Telecats  
Ernst-Jan Kruijze  
Gemeente Utrecht

de NS maar straks zal je je treininformatie benaderen via bijvoorbeeld SIRI. Die zal op haar beurt de noodzakelijke informatie ophalen via een soort hidden NS-app (=functionaliteit), maar de invoer en uitvoer zal via SIRI verlopen. Wil je je bankbetalingen ook via SIRI laten verlopen, dan download je de functionaliteit-app, maar opnieuw: de invoer en uitvoer gaat via SIRI.

### Social Media

Ook op Social Media zien we een interessante ontwikkeling. Nu zijn er nog veel verschillende platformen (Facebook, Twitter, Flickr, LinkedIn) die niet of nauwelijks met elkaar 'praten'. Het is waarschijnlijk dat er straks slechts een paar dominante partijen overblijven. Een social network heeft pas nut als je vriendjes erop zitten, dus hoe meer vriendjes erop zitten, hoe lastiger het wordt om er weg te gaan. Wel zullen deze social networks multi-channel worden (voorzover ze dat al niet

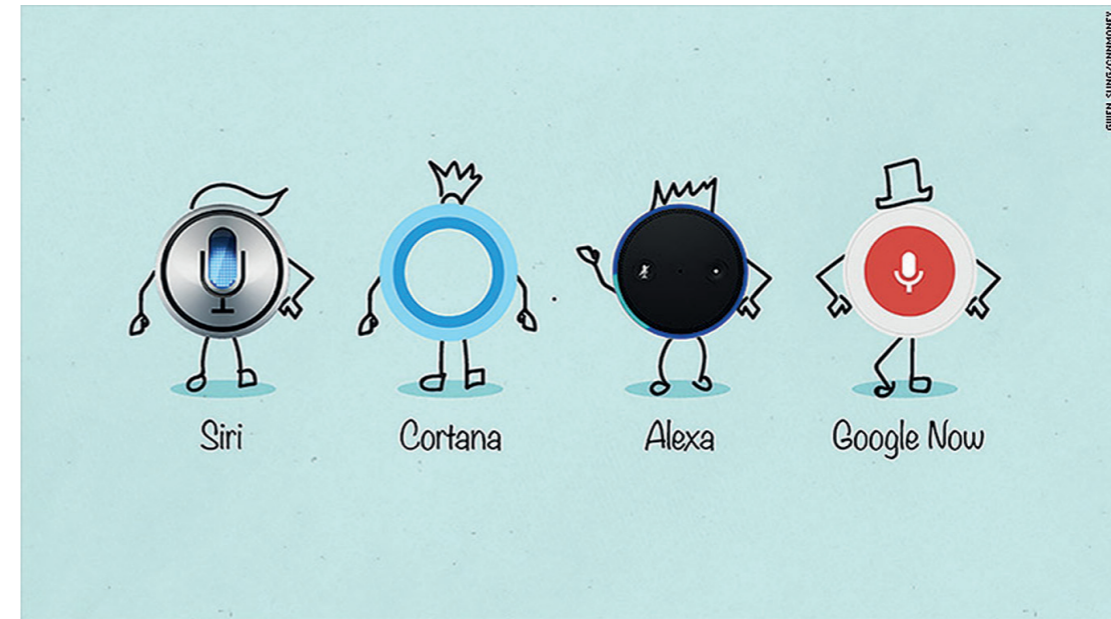
zijn). Neem Facebook/WhatsApp waarmee je al kunt bellen, videostreamen en chatten. Of denk aan het Chinese WeChat dat behalve de WhatsApp-functionaliteiten ook de mogelijkheid heeft om te betalen.

### Kracht van de klant neemt toe

Op het eerste gezicht lijkt de positie van de klanten sterker te worden. Er zijn dikwijls meerdere gelijksoortige producten beschikbaar. Gekoppeld aan de toegenomen mogelijkheden om abonnementen of contracten op te zeggen, is het voor klanten makkelijker geworden om een relatie met een aanbieder te beëindigen. Ook in bestaande klantrelaties zal de afhankelijkheid van de traditionele klantenservice afnemen. Zowel via klantfora (klanten-helpen-klanten) als via AI-gedreven chatbots zal men straks 24/7 op een bevredigende wijze antwoord kunnen krijgen. Pas als niets meer lukt, zullen mensen aangeven dat er contact met ze kan worden opgenomen.



© KlantenServiceFederatie



Organisaties komen dan in de wachtrij van hun klanten!

### Het bedrijf staat (ook) sterker

In tal van situaties hebben consumenten – ook over tien jaar – nog altijd behoefte aan goed en persoonlijk contact met organisaties. Voorwaarde is wél dat betrokken klantcontactmedewerkers volwaardige gesprekspartners zijn en niet met voor de hand liggende standaardantwoorden komen aanzetten. Die waren immers al door de klantfora/chatbots gegeven. Klanten willen empathische en klantvriendelijke adviseurs, optimaal ondersteund door de juiste technologie. Haperende systemen, gebrekkige content en onvolledige klantinformatie zijn funest voor de klantbeleving.

### Technologie als zesde zintuig

Organisaties die dit inzien en medewerkers (en hun klanten) maximaal willen faciliteren, zullen de kracht van slimme technologie als zesde zintuig ontdekken. Over tien jaar hebben we slimme tools tot onze beschikking om alle mogelijke data te verzamelen en real-time te analyseren. Op basis van wat klanten zeggen en hoe, kan de technologie vragen beantwoorden als: 'Wat is de aanleiding van de klant om contact te zoeken?', 'Welke emoties zijn onderliggend aan de interactie?', en 'Hoe groot is de kans dat deze persoon loyaal blijft?'. Klanttevredenheidsonderzoek is niet meer nodig want deze informatie wordt gedistilleerd uit gesproken en geschreven tekst

en gecombineerd met allerlei andere persoonlijke gegevens die voorhanden zijn.

### Klant zijn wordt een gunst

Anno 2017 zijn er al financiële instellingen die op basis van zelflerende software kunnen voorspellen welke klanten over drie maanden hun hypotheeklasten niet meer kunnen opbrengen. Over tien jaar worden dergelijke inzichten ook gebruikt om te bepalen of een persoon überhaupt klant mag worden. Klant worden is straks geen vanzelfsprekend recht. Je kunt worden geweigerd of geconfronteerd met een organisatie die de relatie met jou opzegt.

### Conclusie

De toekomst voorspellen blijft lastig maar we durven met zekerheid te zeggen dat AI en TST een belangrijke rol gaan spelen in het contact tussen klant/burger en bedrijven, organisaties en de overheid. Dat we over tien jaar nog contactcenters hebben die sterk lijken op wat we nu zien, is daarom hoogst onwaarschijnlijk.

Dit artikel is gebaseerd op het 'Reinventing Customer Service', document van Léon Cozijnsen, Daan Gönning, Lotte Willemsen, Ernst-Jan Kruize en Arjan van Hessen



# Kunstmatige Intelligentie in CLARIAH

## Wat is CLARIAH

CLARIAH is een infrastructuurprogramma voor de Geesteswetenschappen. Het doel is om onderzoekers op een gebruikersvriendelijke wijze toegang te geven tot data (resources) en tools (software). Hiermee kunnen zij hun werk beter en sneller doen, maar het kan ook tot nieuwe onderzoeksvragen leiden doordat men patronen ziet die eerder zo niet gevonden konden worden.

CLARIAH zorgt ervoor dat allerlei dataverzamelingen en tools, die in de afgelopen decennia ontwikkeld zijn, beter en makkelijk ontsloten worden voor geesteswetenschappers en dat tools en data interoperabel worden. Daarmee kan het geesteswetenschappelijk onderzoek versneld worden, gebaseerd worden op grotere en rijkere datasets en kunnen wetenschappelijke vragen aangepakt worden waarvoor dat tot voor kort niet mogelijk was. CLARIAH creëert zo het Nederlandse deel van de Europese onderzoeksinfrastructuur voor geesteswetenschappelijk onderzoek in CLARIN en DARIAH. CLARIAH werd voorafgegaan door het CLARIN-NL project (2009-2015) en omvat zelf tot nu toe 3 projecten: CLARIAH-SEED (2014-2015), CLARIAH-CORE (2015-2018) en CLARIAH-PLUS (een aanvraag voor een project dat, indien gehonoreerd, zal lopen van 2019-2024)

## AI en CLARIAH

Allerlei vormen van kunstmatige intelligentie worden gebruikt in CLARIAH, en nog veel meer vormen van kunstmatige intelligentie zouden we willen gebruiken in CLARIAH-PLUS.

Het zou te ver voeren om alle vormen van kunstmatige intelligentie die binnen CLARIAH worden ontwikkeld en/of toegepast hier te

beschrijven. Maar grosso modo kan men stellen dat binnen de taal- en spraaktechnologie steeds vaker gewerkt wordt met een tak van op AI gebaseerde software die leert van gegeven voorbeelden die door onderzoekers gelabeld zijn.

Een illustratief voorbeeld (zij het niet direct uit CLARIAH) is het herkennen van plaatjes in de kranten van de Koninklijke Bibliotheek. Thomas Smits van de Radboud Universiteit doet onderzoek naar de vorming van nationale en transnationale identiteit middels beeldmateriaal in zeven Europese geïllustreerde kranten tussen 1842 en 1870. Het gaat hier om zoveel beeldmateriaal dat het doorzoeken ervan nauwelijks meer met de hand te doen is. Hij ontwikkelde daarom samen met computerwetenschappers software waarmee plaatjes in oude kranten herkend en gelabeld konden worden. Maar als dat kan, kun je ook onderzoeken of bijvoorbeeld plaatjes met mannen te paard vaker in het ene dan in het andere land werden gebruikt, en of dit soort plaatjes vooral in het voorjaar werden gebruikt en zo ja waarom? Door de grote hoeveelheden data die nu beschikbaar zijn, kunnen dit soort onverwachte patronen gevonden en bestudeerd worden.

## Taaltechnologie

Allerlei vormen van taal- en spraaktechnologie vallen tegenwoordig onder kunstmatige intelligentie. Bekende toepassingen zijn:

- Toekennen van woordsoorten aan woorden in een tekst ('part-of-speech taggers', bijvoorbeeld Frog)
- Automatische syntactische ontleding van taaluitingen ('parsers', bijvoorbeeld Alpino)
- Automatische herkenning van namen van verschillende categorieën zoals perso-

nen, locaties of organisaties in een tekst ('named entity recognisers', bijvoorbeeld zoals in NameSpace).

Deze technologieën worden gebruikt voor veel verschillende doeleinden. Eén voorbeeld is het verrijken van grote tekstcorpora (Corpus Gesproken Nederlands, LASSY, SoNaR) en deze beschikbaar maken in zoekapplicaties (zoals GrE TEL, OpenSoNaR, NameSpace). Taalkundigen en literatuurwetenschappers kunnen die gebruiken om hun onderzoek te versnellen en te baseren op een grotere set van data. Verschillende zoekapplicaties stellen een onderzoeker in staat om zijn/haar eigen tekstcorpus op te laden, automatisch te laten verrijken en doorzoekbaar te maken voor deze verrijkingen. Concrete voorbeelden zijn AutoSearch (voor het doorzoeken van een automatisch met woordsoortinformatie verrijkt tekstcorpus) en PaQu (voor het doorzoeken van een automatisch met syntactische structuren verrijkt tekstcorpus).

Dergelijke technologieën worden ook gebruikt om databanken (gestructureerde data), die gebruikt worden in de sociaal-economische geschiedenis, te verrijken en aan te vullen met informatie die verborgen zit in ongestructureerde data zoals lopende natuurlijke-taalteksten. Zo is er bijvoorbeeld met behulp van deze technologieën informatie over beroepen van personen uit lopende teksten geëxtraheerd. Er is ook een plan voor een project binnen CLARIAH om allerlei informatie over films uit filmladders en filmreviews te extraheren voor gebruik in mediastudies.

## Workflow

Al deze technologieën samen zijn gecombineerd in een workflow om gebeurtenissen (events) en hun eigenschappen automatisch te extraheren uit lopende tekst, in het bijzonder uit verschillende teksten die over hetzelfde onderwerp gaan. Die events worden dan gerepresenteerd met technieken van het semantisch web, zoals Linked (Open) Data. Objecten worden verbonden met Linked-Data-representaties van externe kennisbronnen, zoals DBpedia (een Linked Data representatie van Wikipedia), en met kennisbronnen die in CLARIAH zelf aangeemaakt en beheerd worden, zoals databases voor personen, locaties en objecten via ANANSI.

## Spraaktechnologie

Spraakherkenningssoftware wordt gebruikt om audio- en videomateriaal te voorzien van transcripties, op basis waarvan indexen gemaakt kunnen worden om het doorzoeken



van audio- en videomateriaal vele malen efficiënter te maken. Dat is van groot belang voor onderzoekers op het gebied van mediastudies, onderzoekers op het gebied van 'Oral History' en voor taalkundigen die werken met audio of audiovisueel materiaal (bijvoorbeeld voor taaldocumentatie, kindertaalvererving, fonologie en fonetiek). Zoals Louis ten Bosch in het artikel elders in deze DIXIT (p. 4) heeft toegelicht, spelen Deep Neural Networks uit de AI tegenwoordig een essentiële rol in automatische spraakherkenning.

We zouden graag ook andere vormen van spraaktechnologie gebruiken in de CLARIAH-infrastructuur. In een eerdere fase zijn er al experimenten gedaan met het onderscheiden van verschillende sprekers in een audiovisueel bestand ('speaker diarisation'), en met het onderscheiden van spraak van andere akoestische gebeurtenissen dan spraak (zoals lachen, kuchen of muziek). Dit is bedoeld om het annoteren van dergelijke databestanden te versnellen, maar we hebben hiervoor robuustere implementaties nodig.

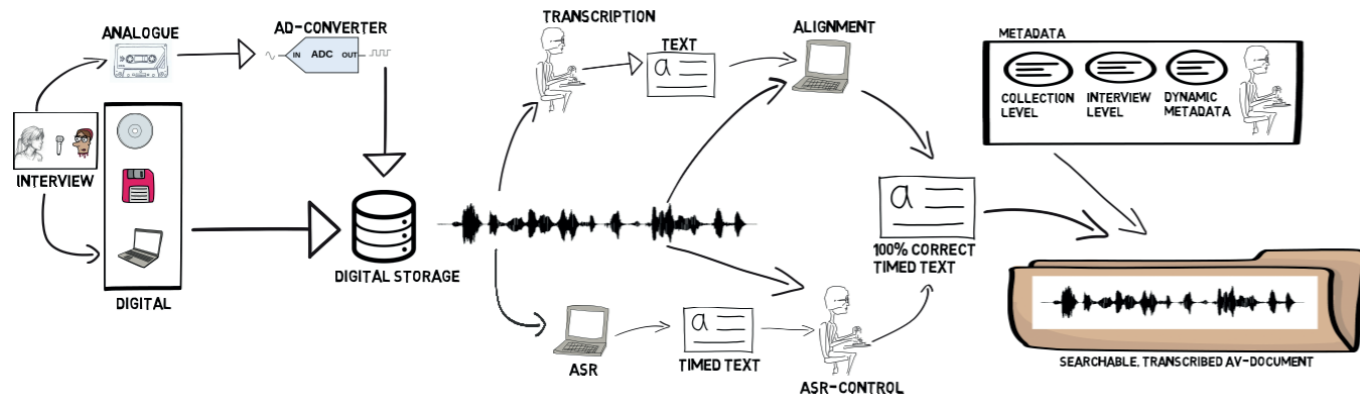
## Videotechnologie

Op het gebied van video is het van groot belang dat annotatie zoveel mogelijk ondersteund wordt door automatische processen. Concrete voorbeelden op de wensenlijst (en in de plannen voor CLARIAH-PLUS) zijn het automatisch herkennen van scène-overgangen, het automatisch herkennen van gezichten, van objecten en zelfs van gebeurtenissen en emoties in bewegend beeld. Voor een deel bestaan hier al oplossingen voor, die echter geïntegreerd moeten worden in de infrastructuur die CLARIAH aan het opbouwen is (voor deze functionaliteit in de CLARIAH Media Suite). Voor een deel ligt deze functionaliteit aan de rand van wat er technisch mogelijk is en zal er nog cutting-edge onderzoek gedaan moeten worden om deze technologieën zo goed te laten worden dat ze een algemeen bruikbaar onderdeel kunnen worden van de CLARIN-infrastructuur.

### Links:

1. CLARIAH: <https://www.clariah.nl>
2. CLARIAH Media Suite: <http://mediasuite.clariah.nl/>
3. Anansi: <https://anansi.clariah.nl/>
4. CLARIN: <https://www.clarin.eu/>
5. DARIAH: <https://www.clarin.eu/>
6. Informatie over en links naar de andere genoemde tools: <http://portal.clarin.nl/clarin-resource-list-fs>

De Oral History Transcription Chain zoals die door de CLARIN-EU en CLARIAH-NL begin 2018 zal worden opgeleverd. Onderzoekers kunnen hun audio-bestanden uploaden en krijgen de herkenningresultaten terug.











# TELECATS

VANZELFSPREKEND



TELEFONIE & CONTACT CENTER OPLOSSINGEN  
MET TAAL- & SPRAAKTECHNOLOGIE

[www.telecats.nl](http://www.telecats.nl)