

DIXIT

TIJDSCHRIFT OVER TOEGEPASTE TAAL- EN SPRAAKTECHNOLOGIE

Een half leeg glas wijn...

Over nut en waarde van onderzoek

Cendris start met spraakherkenning



Verloedering en
voortgang van
de taal

Wij werken aan de derde generatie Intranet



De Handboog 9, 5283 WR Boxtel, Nederland
Tel: +31(0)411 610802 | Fax: +31(0)411 611027
E-mail: info@knowledge-concepts.com | Web: www.knowledge-concepts.com

INHOUD

Cendris start met spraakherkenning

DUTCHEAR heeft aan Cendris een spraaktechnologieplatform geleverd waarmee het mogelijk is om contact center-taken te automatiseren, bijvoorbeeld het identificeren van klanten. De eerste dienst waarvoor Cendris dit platform inzet, is het automatisch opvragen van een postcode. Volgend op het succes van deze eerste dienst is het Cendris spraaktechnologieplatform beschikbaar gesteld voor al haar klanten. Een interview met Freek Weijtens, manager ICT van Cendris Customer Contact.

4

Verloedering en voortgang van de taal

Wanneer je naar een oud Polygoon-journaal luistert, hoor je meteen hoezeer het Nederlands in een halve eeuw veranderd is. Niet alleen de woordenschat verandert, maar ook de uitspraak, de spelling en zelfs, in mindere mate, de zinsbouw. Het is het tragische lot van ouders door de eeuwen heen zich te moeten ergeren aan nieuwerwets taalgebruik van hun kinderen. Theo van den Heuvel over onvermijdelijke taalverandering.

8

Nut en waarde van onderzoek

Als vanouds blijven nut en waarde van het onderzoek aan universiteiten een heikel onderwerp van discussie. Het spreekt voor zich dat de meetlat die gehanteerd wordt voor het meten van het nut en het stellen van de waarde van een bepaald onderzoeksresultaat niet volkomen objectief gekozen kan worden. De criteria voor nuttig en waardevol onderzoek zijn afhankelijk van de doelen van het onderzoek. Khalil Sima'an gooit de knuppel in het hoenderhok.

14

Rubrieken

Column	7
Opinie	11
Uitgelicht	13
TST in bedrijf	16
NOTaS-nieuws	18

Achterom kijken door vooruit te zien

Eén jaar DIXIT, een mooi moment om eens even terug te kijken. De toekomst beschouwen vanuit het verleden. Niet meer dan logisch, want bouwen doe je op een gedegen fundament. En, pretentius als we zijn, we menen dat we het fundament hebben gelegd voor een magazine dat nu kan gaan uitgroeien tot een serieuze spreekbuis van de TST-gemeenschap.

Op termijn wel te verstaan. Omdat DIXIT er nadrukkelijk voor kiest de stem te zijn van zowel de wereld van fundamenteel onderzoek als de wereld van commerciële toepassingen van taal- en spraaktechnologie, is de weg naar boven niet de meest eenvoudige. Of, met andere woorden, waar de Stichting NOTaS inmiddels hard op weg is om hét platform te worden waar onderzoek en toepassing elkaar ontmoeten, daar moet DIXIT nog hard werken om die natuurlijke symbiose vorm te geven en naar de lezers toe uit te stralen.

Vorm en uitstraling dus. Dat betekent in redactionele termen dat de inhoud en presentatie moet worden aangescherpt. Het is de bedoeling om vanaf volgend jaar DIXIT strak te gaan opdelen in vaste rubrieken die elk TST-gebied recht doet. Natuurlijk blijft de *toepassing* van TST de rode draad, maar niet meer dan dat. TST mag dan wel heel volwassen lijken, maar is dat natuurlijk nog lang niet. Elke nieuwe toepassing stuit vaak meteen op een muur aan onvolkomenheden of onmogelijkheden. Of een vat vol nieuwe kansen. Kortom, het volgen van *onderzoek* naar mogelijkheden die kunnen leiden tot commerciële toepassingen of bestaande problemen uit de weg kunnen ruimen, blijft een belangrijke taak. Een taak die overigens weer naadloos aansluit bij één van de doelstellingen van NOTaS, namelijk het samenbrengen van kennis en toepassing.

Maar DIXIT wil meer. Alleen maar kond doen van succesvolle toepassingen en lopend onderzoek doet de specifieke mogelijkheden van een magazine tekort. Meningingen, opinies, discussie en reacties laten een blad leven, maken van dode letters iets levends. Vandaar dan ook de oproep om redactie en auteurs van replek te dienen. Of van informatie te voorzien die hen klaarblijkelijk is ontgaan.

In dit nummer doet redacteur en universitair docent Khalil Sima'an in zijn opiniërende artikel 'Een half leeg glas wijn...' bijvoorbeeld de om replek vragen de uitspraak dat 'Nederland klein is voor taaltechnologie, omdat er te weinig grote spelers zijn'. Het zou ontbreken aan de drie noodzakelijke M's: middelen, motivatie en menselijk kapitaal. Bepaald geen opbeurende woorden voor een land dat pretendeert een kennis-economie te zijn dan wel te willen worden.

Enfin. Het laatste woord is aan de lezer!

Pierre Pieterse, hoofdredacteur DIXIT

Pierre@dialogoorguitgevers.nl

Automatische adresidentificatie in contact center

Cendris start met spraakherkenning

DUTCHEAR heeft aan Cendris een spraaktechnologieplatform geleverd waarmee het mogelijk is om contact center-taken te automatiseren, bijvoorbeeld het identificeren van klanten. De eerste dienst waarvoor Cendris dit platform inzet is het automatisch opvragen van een postcode. Volgend op het succes van deze eerste dienst is het Cendris spraaktechnologie platform beschikbaar gesteld voor al haar klanten. Een interview met Freek Weijtens, manager ICT van Cendris Customer Contact.

Z

o rond de donkere dagen voor Kerst en Oud en Nieuw is het voor velen een jaarlijks terugkerende activiteit: het versturen van de beste wensen aan alle vrienden, familie en kennissen. Om te zorgen dat de wensen nog voor Kerst in de brievenbus belanden is het verstandig om de juiste postcode te gebruiken. De nieuwe spraak-

| HANS JONGEBLOED |

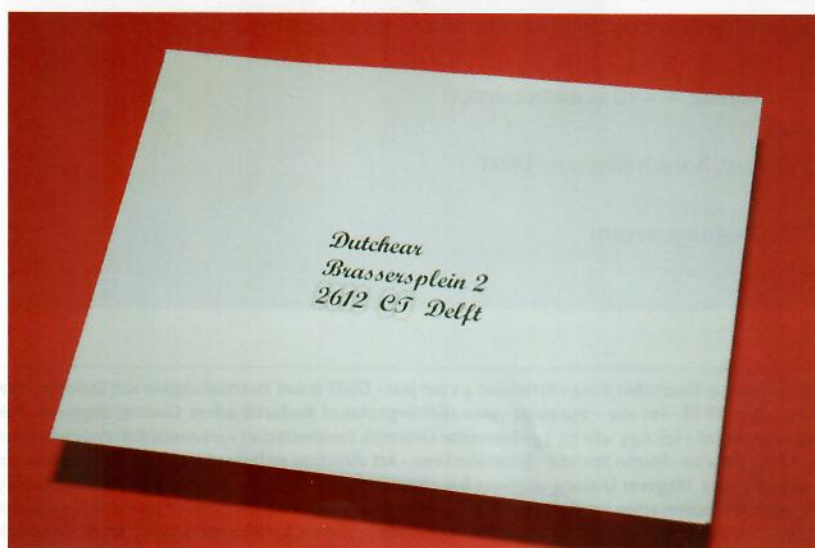
gestuurde dienst van Cendris helpt hierbij een handje. Via de 'postcodelijn' spreken bellers de plaatsnaam en straatnaam in en voeren het huisnummer in. Automatisch krijgen ze dan de juiste postcode te horen. Cendris hoeft door deze automatisering veel minder medewerkers op deze dienst in te zetten, waardoor er meer medewerkers beschikbaar komen voor andere diensten.

Waarom is Cendris overgegaan op spraaktechnologie?

Freek Weijtens, manager ICT van Cendris Customer Contact: "De kwaliteit van spraakherkenning heeft inmiddels een hoog niveau bereikt. Bovendien is spraakherkenning betaalbaar geworden. Het is dus een logische verbreding van ons dienstenportfolio. Cendris kan de dienst al dan niet gecombineerd met ondersteuning van een contact center-agent aanbieden. In de toekomst zullen wij steeds meer gebruik maken van spraaktechnologie. Het is een kwalitatieve en voor de opdrachtgever interessante manier om contact center activiteiten te automatiseren."

Welke uitdaging gaat de spraaktechnologie aan?

"Tot voor kort werden de postcodes uitgegeven door medewerkers van Cendris. Hierbij werden de gegevens van de klant geïnterpreteerd en ingevoerd in een databasezoeksysteem. Dat dit nog niet zo'n simpele taak is als het op het eerste gezicht lijkt, wordt duidelijk als je na gaat denken over allerlei rariteiten die in Nederlandse plaatsnamen en straatnamen zitten. Dankzij intelligente mechanismen in het systeem zorgt de spraaktechnologie toch voor de juiste afhandeling. In de volgende paragrafen volgt een uitstapje langs de rariteiten in Nederlandse plaats- en straatnamen."



Wat is de uitdaging bij plaatsnamen?

“Nederland kent minstens 2450 officiële plaatsnamen en daarnaast nog honderden dorpjes, gehuchten, enzovoort. Iedereen kent wel de grote steden zoals Amsterdam en Rotterdam, maar wie heeft van Geelbroek gehoord? Natuurlijk is voor kleine plaatsen de kans heel klein dat ooit iemand het opvraagt, maar het systeem moet het wel (her)kennen.

De voertaal van de dienst is Nederlands. Er zijn echter diverse plaatsen die er niet meteen Nederlands uitzien, zoals Retranchement of Balgoy (ook de spellingchecker krijgt het hier moeilijk). Verder zijn er plaatsnamen die op meerdere manieren kunnen worden uitgesproken, denk aan ‘Den Haag’, officieel aangeduid als ‘s Gravenhage’. Ook weet lang niet iedereen dat ‘Gorinchem’ uitgesproken wordt alsof er ‘Gorcum’ stond.

Het omgekeerde komt ook voor, met dank aan allerlei fusies en gemeentelijke herindelingen in de politiek, dat twee plaatsen één naam krijgen. Iemand die alleen de oorspronkelijke plaatsnaam kent zal natuurlijk gewoon ‘Driebergen’ zeggen in plaats van het officiële ‘Driebergen-Rijsenburg’.

Veel plaatsnamen zijn al heel oud, al van voor de Middeleeuwen. In die tijd was men nog lang niet zo mobiel als tegenwoordig, zodat de plaatsnaam ook alleen lokaal als onderscheid diende van de andere nabij gelegen dorpen. Dit leidde nogal eens tot het bedenken van dezelfde plaatsnaam in een heel andere deel van Nederland. Het waarschijnlijk bekendste voorbeeld is ‘Hengelo’. Men heeft dan ook al snel geleerd het expliciet over ‘Hengelo Gelderland’ te hebben.

Bij de bouw van de plaatsnaamherkenning is zoveel mogelijk rekening gehouden met bovenstaande fenomenen, zodat bellers vrij zijn in hun woordkeuze voor de plaatsnaam. Waar nodig kan het systeem een aanvullende vraag stellen om zeker te zijn welke plaatsnaam bedoeld wordt. Bovendien wordt de gebruiker steeds exact op de hoogte gebracht met welke plaatsnaam het systeem aan de gang gaat.”

Wat is de uitdaging bij straatnamen?

“Ondanks het feit dat alle straatnamen uit het postcodeboek in het Nederlandse taalgebied liggen, zijn lang niet alle straatnamen van Nederlandstalige oorsprong. Veel

straatnamen zijn genoemd naar beroemdheden met een niet-Nederlandse achternaam (denk aan de ‘Schubertlaan’) of achternamen die hun historische spelling behouden hebben (zoals de ‘Van Speykstraat’). Voor het correct automatisch voorlezen en herkennen van deze straten is het onvermijdelijk om intelligente software en/of handmatige bewerkingen uit te voeren om voor de correcte uitspraak te zorgen.

Nederland is zeker creatief in het bedenken van straatnamen; totaal zijn er meer dan 120.000 verschillende namen, met meest uiteenlopende eigenaardigheden. Echte uitdagingen voor het voorlezen van

Automatische klantidentificatie in het contact center

De spraaktechnologie beschreven in het artikel maakt het mogelijk om op basis van plaatsnaam, straatnaam en huisnummer automatisch te bepalen welke klant een bedrijf aan de telefoon heeft. Hiermee kan op basis van de klanthistorie uit een CRM-systeem de routing van het telefoongesprek afhankelijk gemaakt worden van het type klant. Belangrijke klanten gaan direct door naar hun eigen (persoonlijke) accountmedewerker. De contact center-medewerker krijgt de klantgegevens direct op zijn scherm. Ook zouden klanten naar diverse *self service*-systemen doorgeleid kunnen worden. Nog efficiënter is om ook het onderwerp aan de beller te vragen en automatisch te herkennen, zodat direct bij de routing hiermee rekening kan worden gehouden. Het inspreken van de gevraagde gegevens en het onderwerp kost veel minder tijd dan de juiste keuzes in een toetskeuze menu maken. Dit bespaart de beller tijd en frustratie en de contact center-medewerker kan het gesprek sneller en klantgerichter afhandelen.

de straatnamen komen bij het automatisch goed leggen van de klemtoon. Soms is de klemtoon gericht op het duidelijk onderscheid maken tussen twee bijna gelijke straatnamen, zoals bij ‘Kerkstraat’ en ‘Kerkplein’; de nadruk ligt dan op de tweede lettergreep.

Hoe kijken gebruikers en de markt tegen spraaktechnologie aan?

“De gebruikers van de postcodelijn blijken prima met de dienst overweg te kunnen. De bellers komen niet in een wachtrij en zijn zeer snel geholpen. Cendris is dan ook zeer tevreden met het systeem. In de contact center-wereld zijn grote partijen bezig met spraaktechnologie, zoals Cendris, *De Telegraaf*, de NS, PCM (zie vorige uitgave van DIXIT) en KPN. Hiermee heeft spraaktechnologie een eigen plek verworven binnen het telefonisch klantcontact. I

Hans Jongebloed is senior consultant spraaktechnologie bij DUTCHEAR

The right answer at once!

Met de zelflerende Selfservice Suite van AskNow bent u in staat om al uw digitale klantcontacten via e-mail en web, snel, efficiënt en uniform af te handelen.

De toegepaste state-of-the-art technologie- en zoals taaltechnologie, machine learning en semantiek zorgen voor:

- Betere resultaten;
- Minimaal onderhoud;
- Gebruiksvriendelijkheid.

Selfservice
Agent Support
Kennismanagement
Email Response management



Meer weten?
Kijk op www.asknow.nl

Bayes

Onlangs verblijdde de protestantse kerk van Utrecht ons met het bericht dat zij jaarlijks ruim 400.000 vrijwilligersuren in het maatschappelijke leven van die stad investeert. Het financiële equivalent van dit welzijnswerk wordt becijferd op zo'n 8 miljoen euro per jaar. Nadat ik de rondedans die dit soort positieve berichten over de kerk doorgaans bij mij oproept, succesvol had afgerond, bedacht ik me dat de sociale verdiensten van de kerk zich ook wel op andere vlakken manifesteren. Er zijn nogal wat geestelijken geweest die zich verdienstelijk hebben gemaakt voor de maatschappij door naast hun eigenlijke roeping nog wat bij te beunen.

Ik begaf me naar een boekenkast. En allengs rolden een bonte stoet grootheden aan mijn geestesoog voorbij. Allereerst geestelijken die de politiek in gingen. Om er een paar van eigen bodem te noemen: Abraham Kuyper (AR), Ferdinand Domela Nieuwenhuis (SDAP), Pieter Zand (SGP), Ab Harrewijn (Groen Links). Verder zijn daar componisten als Charles Gounod, Antonio Soler, en niet te vergeten Antonio Vivaldi. Ook schoten me ook nog wat sporthelden te binnen, maar ik was nog net alert genoeg om me te bedenken dat hier van een maatschappelijke verdienste eigenlijk geen sprake is, hooguit van maatschappelijke schade als gevolg van blessures. Wel zijn er nog beroemde schrijvers te noemen als Marnix van St. Aldegonde, Guido Gezelle, Nicolaas Beets, en, zo u wilt (maar ik niet): Antoine Bodar. Ook onder filosofen vindt men vele theologen zoals Augustinus, Thomas van Aquino, en Kierkegaard.

Voor ons interessanter zijn de geestelijken die zich in de wetenschap bekwaamden en daar belangwekkende zaken ontdekten. Sir Isaac Newton (1643-1727) is eigenlijk alleen als natuurkundige bekend geworden, terwijl hij veel meer theologische dan natuurkundige verhandelingen heeft geschreven. Verder denk ik nog aan Carl Linnaeus (1707-1778), de augustijner monnik Gregor Mendel (1822-1884) en aan de jezuïet Pierre Teilhard de Chardin (1881-1955).

Terwijl ik al mijmerend aardig aan het indutten was, want niet gestoord door de vierkoppige lawaaignerator die mijn kroost heet, schrok ik plotseling op bij de gedachte aan Thomas Bayes. Daar had ik een kerkelij-

ke rakker bij de oren die op vele terreinen van de wetenschap diepe sporen heeft nagelaten, mijn eigen vakgebied, de spraaktechnologie, niet uitgezonderd. Hij is de held van deze *DIXIT*. Thomas Bayes (1702-1761) was een dominee in Tunbridge Wells, zo'n 50 km ten zuidoosten van London, tot 1752. Na zijn pensionering bleef hij daar wonen en ontwikkelde daar zijn kanstheorieën die hij beschreef in het artikel 'Essay Towards Solving a Problem in the Doctrine of Chances', dat pas drie jaar na zijn dood uitkwam omdat een vriend het toevallig tussen een stapel nagelaten papieren tegenkwam.

Bayes' regel, die de meeste van ons wel eens tegen het lijf zijn gelopen, zegt hoe groot de kans op A is, gegeven dat je B weet: $P(A|B) = P(A) * P(B|A) / P(B)$. De bijbehorende termen zijn: posterior prob. = prior prob. * likelihood / evidence. Bayes' regel werd voorwerp van grote controverse; zijn statistiek is namelijk niet predictief van aard zoals gangbaar was (en is) maar inferentieel. Dat wil zeggen dat opgedane ervaringen uit het verleden worden geaccumuleerd om de kans op een toekomstige gebeurtenis af te leiden. Als wij bijvoorbeeld een spraakherkenner willen maken die bepaalde woorden moet gaan herkennen gegeven een set akoestische modellen (de posterior probs), dan zegt Bayes' regel ons dat we dat kunnen doen door de akoestische modellen te trainen door ze veel observaties van die woorden aan te bieden (dan bepalen we de likelihoods). Typerend voor zo'n Bayesiaanse machine is dat-ie zich door training zelf uit het moeras kan trekken (en omgekeerd door een foute training verder in de stront kan duwen). In het wetenschapsgebied van de taaltechnologie wordt Bayes' theoreem ook gebruikt voor bijvoorbeeld automatisch vertalen, en het maken van SPAM-herkenners voor e-mail-berichten.

Ik zie een deel van genoemd vierkoppige gedaante naderen en schrik op uit mijn mijmeringen. Er is nog net genoeg tijd om u naar een instructief document te verwijzen waarin staat hoe u Bayes' theoreem zelfs kunt gebruiken om televisiequizen te winnen: <http://www.ai.rug.nl/ki/2002/hoorcollege/bayes-rules.pdf>. Bayes moest eens weten... Maar gegeven dat hij dood is, is die kans ook volgens zijn eigen regel nihil.

Bayes moest eens weten... Maar
gegeven dat hij dood is, is die kans
ook volgens zijn eigen regel nihil.

Henk van den Heuvel is
verbonden aan het CLST van de
Radboud Universiteit
Nijmegen

Verloedering en voortgang van de taal

Taal van de toekomst

Wanneer je naar een oud Polygoon-journaal luistert, hoor je meteen hoezeer het Nederlands in een halve eeuw veranderd is. Niet alleen de woordenschat verandert, maar ook de uitspraak, de spelling en zelfs, in mindere mate, de zinsbouw. Het is het tragische lot van ouders door de eeuwen heen zich te moeten ergeren aan nieuwerwets taalgebruik van hun kinderen. Theo van den Heuvel over onvermijdelijke taalverandering.

| THEO VAN DEN HEUVEL |

Misschien is de zin 'Hun hebben meer gedist als ons' over 30 jaar wel volstrekt grammaticaal en lacht ons nageslacht wel om de constructies die nu onze dictees sieren. Er is een voortdurend krachtenspel tussen aan de ene kant onderwijs, vaak behoudend in taalregels, en aan de andere kant mode en 'waan van de dag'. De engelse ziekte grijpt om zich heen (zie kader). Taal is zo diep geworteld in onze cultuur dat het moeilijk is om geen sterke gevoelens te hebben bij dit soort veranderingen. Maar, wat wij er ook van vinden, taalverandering is een gegeven en wij, taalgebruikers, kunnen er maar het beste zo goed mogelijk op reageren. Daarvoor hebben we de juiste hulpmiddelen nodig.

NIEUWE TAAL EN HET SCHRIJVEN

Elk jaar dringen honderden woorden tot de taal door. Vele daarvan behoren tot een



specifiek vakgebied en zullen dus niet het jaar daarop in de Dikke Van Dale terug te vinden zijn. Wie er prijs op stelt om correct te schrijven volgens de actuele spelling- en stijlvoorschriften, doet er goed aan af en toe recente woordenboeken en taaladviesgidsen te raadplegen. Naast papieren producten zijn er tal van elektronische. Dat kunnen aparte programma's zijn, maar ook webdiensten of *plug-ins*, modules die in een bestaand programma kunnen worden gehangen. Tien jaar geleden waren er veel journalisten die het einde van het boekentijdperk voorspelden. Het liep anders. Het is nog steeds zo dat de meeste uitgevers vooral omzet genereren met papieren producten. Dat is wat vreemd, als je bedenkt hoezeer het computergebruik is toegenomen. Wellicht is er een verklaring te vinden als je bedenkt hoe de woordenboeken worden gebruikt. Op het werk en bij de opleiding worden naslagwerken gebruikt bij het lezen en het schrijven van documenten. Het lezen kan achter de computer of van papier. Mensen zijn soms bereid dat lezen te onderbreken om even een woordenboek uit de boekenkast te pakken en open te slaan. Het ligt niet voor de hand om een computer aan te zetten om een woord op te zoeken. Bij het schrijven staat de computer al aan. Dan nog is het openen van een aparte woordenboekapplicatie mensen vaak al een handeling te veel. De beste oplossing bij het schrijven is dus naadloze integratie, dus onmiddellijk toegankelijk binnen de gebruikte kantoorapplicatie. Denk aan een woordenboek dat je kunt openen door op

Engelse ziekte

De gewoonte om samenstellingen als losse woorden te schrijven wordt 'Engelse ziekte' genoemd, omdat het in het Engels regel is. Deze Engelse ziekte rukt op en is met de spellingcontrole niet altijd te herkennen, omdat de delen vaak goedgegeschreven woorden zijn. Zo betekent 'mijn bouw' echt iets heel anders dan 'mijnbouw'.

een suggestie van de speller te klikken met de muis. Of denk aan het lezen van een bestand en het met een muisklik open slaan van een woordenboek bij het woord onder de muis.

Ook op kantoor verandert de taal: het cliëntenbestand wijzigt, er komt een afdeling bij, nieuwe functies worden geïntroduceerd, er verdwijnt een product. Dit betekent nu vaak dat medewerkers wat meer rode kringeltjes op hun scherm krijgen. Ze kunnen die woorden aan hun systeem leren door ze in het gebruikerswoordenboek (*custom dictionary*) toe te voegen, maar dat moet dan elke medewerker voor zichzelf doen. Laten we hopen dat ze het op dezelfde manier doen, want anders blijft verwarring regeren en zijn de documenten ook moeilijk terug te vinden in het elektronische archief. Gelukkig kunnen we dit voor een organisatie centraal oplossen, zodat consequente spelling, zeker van zulke cruciale zaken als product- en klantennamen, gegarandeerd is.

HET NIEUWE SCHRIJVEN

Een gebied waar taalverandering erg zichtbaar wordt is het chatten en SMS-en. Met name bij het laatste is het verwonderlijk tot welke duimacrobatiek de gebruikers zich laten verleiden. Het kan hoogstens een kwestie van tijd zijn voor er gemakkelijker bruikbare bedieningsmogelijkheden worden ingevoerd. Hierbij moet men zich realiseren dat dit soort communicatiemiddelen voor bijvoorbeeld doven geweldige mogelijkheden bieden.

Ondertussen blijft het onderwijs zich richten op de heersende ideeën over spelling, grammatica en stijl. We leren onze kinderen goed te spellen en mooie opstellen te maken en oudere kinderen en jongvolwassenen leren soortgelijke zaken voor vreemde talen. Taaltechnologie schiet te hulp. Woordenboeken voor taalleerders

Dyslexie

Er is de laatste jaren veel meer zicht gekomen op het probleem van de dyslexie. Ongeveer 10% van de mensen heeft daar in meer of mindere mate last van. De instrumenten die we gebruiken om spellingskeuzes te volgen kunnen wellicht ook een rol spelen bij het vroegtijdig vaststellen van de symptomen van dyslexie, zodat zo vroeg mogelijk aangepaste lesprogramma's kunnen worden aangeboden.

zijn anders dan 'gewone' woordenboeken: ze geven andere informatiesoorten, bevatten meer voorbeelden en de keuze van de woordenschat is aangepast. Een spellingcontrole voor kinderen en leerders van een tweede taal is ook anders dan een gewone, omdat hij met andere fouten rekening moet houden. Een kind zou kunnen schrijven 'loopte' in plaats van 'liep'. Dit is niet meer alleen een kwestie van spelling. Hoewel je een kind niet wil straffen met een rood kringeltje voor het gebruiken van een moeilijk woord, zul je in een kinderspellingcontrole de suggesties moeten afstemmen op de juiste woorden-



schat. Er zijn instrumenten die het spelling- en schrijfgedrag van kinderen volgen. De kennis die deze metingen opleveren kan worden ingezet in nieuwe leermiddelen. Met behulp hiervan is er nog wel hoop dat de schrijvers van morgen nog door de lezers van vandaag gelezen kunnen worden. |

Nieuwe woorden

Niet alle nieuwe woorden zijn leenwoorden. Veel ontspruiten aan de creatieve geesten van mensen als Marten Toonder (denkraam, minkukel), Wim T. Schippers (gekte, pollens) of Kees van Kooten en Wim de Bie (regelneef, soppen). Iemand heeft voor het balkje tussen de boodschappen op de kassaband het woord 'beurtbalkje' bedacht als alternatief voor het aanzienlijk onvriendelijkere 'kassabanddivider'. Het is natuurlijk prachtig dat taal ons zoveel ruimte biedt voor creativiteit. Taalverandering is niet altijd slecht.

Theo van den Heuvel is managing director van Polderland Language & Speech Technology

New

The first ever British English electronic spellchecker
with the unrivalled authority of Oxford,
fully integrated in Microsoft® Word

Oxford Spellchecker & Dictionary

Up to date and comprehensive: checks your spelling against more than 340,000 words from the *Oxford Dictionary of English*, including proper names, and technical and specialist terms

Fully integrates with Microsoft® Office 2000, XP, or 2003 — replacing the Microsoft® standard spellchecker

Dictionary link to the *Oxford Dictionary of English* allows you to check definitions, and to look up words independently

Definitive Oxford spellings: user settings can standardize alternative (including American) spellings to the preferred Oxford spelling, and warn about commonly confused words.

Wilt u Oxford Spellchecker & Dictionary bestellen of wilt u meer informatie?
Bezoek www.polderland.nl of neem contact met ons op via info@polderland.nl of 024-3522866.

OXFORD
UNIVERSITY PRESS

Language & Speech Technology

polderland



| KHALIL SIMA'AN |

Een half leeg glas wijn...

Als vanouds blijven nut en waarde van het onderzoek aan universiteiten een heikel onderwerp van discussie. Het spreekt voor zich dat de meetlat die gehanteerd wordt voor het meten van het nut en het stellen van de waarde van een bepaald onderzoeksresultaat niet volkomen objectief gekozen kan worden. De criteria voor nuttig en waardevol onderzoek zijn afhankelijk van de doelen van het onderzoek. Als het doel is het bouwen van 'marktgerichte' systemen, zullen de criteria anders zijn dan wanneer het onderzoek tot doel heeft het exploreren van een tot nog toe onopgelost probleem. Bovendien komt het vaak voor dat de waarde (voor bijvoorbeeld de economie) van een bepaald resultaat in eerste instantie onderschat wordt.

In dit verband wil ik terug komen op het hoofdredactioneel van DIXIT 2004/1 ('We drinken een glas...'). Pierre Pieterse schetst een realiteit die moeilijk valt tegen te spreken. De samenwerking tussen bedrijven en universiteiten op het vlak van taaltechnologie zou miniem zijn. De kennis-economie en kennismaatschappij, die zo centraal zijn gesteld in de plannen van de regering, zouden evenmin worden bereikt omdat het onderzoek niet goed aansluit bij behoeftes van de markt. Verder stelt hij vast dat er een gerede kans bestaat dat de verdeling van de onderzoeksgelden door NWO en EZ 'als vanouds opnieuw hun doel volkomen missen'. Hoewel ik deze schets van de situatie wat taal- en spraaktechnologie in Nederland betreft onderschrijf, wil ik toch even stil staan bij zijn conclusie betreffende het doel van de financiering van

het onderzoek en de rol van de universiteiten in de economische ontwikkeling in de kennis- en technologie-maatschappij.

Onderzoekers zijn niet helderziend

Laat ik voorop stellen dat ik volledig eens ben met de stelling dat er weinig initiatieven zijn voor samenwerking tussen bedrijven en universiteiten op het gebied van taal- en spraaktoepassingen. In tegenstelling tot Pierre Pieterse wil ik hier de schuld (als er van schuld sprake is) niet bij NWO of de overheid leggen, maar hoofdzakelijk bij de bedrijven zelf. Waarom? Simpelweg omdat NWO en de universiteiten niet helderziend zijn, of *kunnen* zijn. De initiatieven tot onderzoek dat aansluit bij businessplannen van bedrijven moeten komen van de bedrijven zelf. De onderzoeksgelden van NWO en de universiteiten zijn hoofdzakelijk bedoeld voor onderzoek dat *mogelijk* kan leiden, op onbepaalde termijn, tot economische ontwikkeling. Om de link tussen onderzoek en toepassingen te stimuleren, heeft NWO een specifieke stichting in het leven geroepen. Bij de Technologiestichting (STW) kunnen universiteiten en bedrijven gezamenlijk toepassingsgericht onderzoeksprojecten aanvragen. Ik citeer: 'Technologiestichting STW stimuleert excellent technisch-wetenschappelijk onderzoek aan de Nederlandse universitaire onderzoeksinstituten en bevordert de toepassing van de resultaten van het onderzoek door derden'. Onlangs, tijdens een bijeenkomst van de Taalunie, vernam ik van de vertegenwoordiger van STW, dat het aantal aanvragen vanuit ons vakgebied de nul nadert!

Kunstmatige kloof?

Waarom worden er zo weinig aanvragen betreffende taaltechnologie ingediend bij STW? Zou de kloof tussen onderzoek en toepassingen te groot zijn? Wordt deze kloof kunstmatig groot gehouden, of is het daadwerkelijk het geval dat de *state of the art* binnen de taalkunde, computerlinguïstiek en zelfs de spraakherkenning te abstract is, en zelfs te *zwak*, om tot daadwerkelijke toepassingen te leiden? Zelf geloof ik dat het antwoord aangaande de laatste vraag ergens in het midden is. De '*state of the art*' binnen het onderzoek levert basis-kennis over de verwerking van taal die gebruikt zou kunnen worden in taaltechnologische toepassingen. Maar onze huidige kennis en begrip van hoe menselijke taalverwerking werkt is op dit moment (daadwerkelijk) te basaal om complexe mens/machine-communicatie aan te kunnen. Gegeven deze stand van zaken wordt het belachelijk om te eisen dat het onderzoek wordt gericht naar korte-termijntoepassingen waar de markt om vraagt. Want deze korte-termijntoepassingen zullen binnen de kortste keren worden ingehaald door diepgaander onderzoek.

Natuurlijk kunnen we niet verlangen van de bedrijven zich geduldig te tonen en te wachten tot het onderzoek iets oplevert dat economisch van belang kan zijn. In deze situatie, waarin een kloof bestaat tussen 'vraag naar' en 'aanbod van' kennis, ontstaat de behoefte om een strategie te kiezen waarbij samenwerking in onderzoek naar toepassingen tussen bedrijven en universiteiten meer kans krijgt. Hoewel de middelen hierbij van essentieel belang zijn, is een 'ontmoetingsplek' op dit moment nog belangrijker. Deze ontmoetingsplek moet uitgaan van een soort vraag/aanbod-situatie waarin de universiteiten tentoonstellen wat ze te bieden hebben, en bedrijven tentoonstellen wat ze nodig hebben. NWO zorgt op dit moment, op kleine schaal, voor dit soort ontmoetingsplekken, maar toegegeven, dat is nog te weinig om de kloof te dichten.

'Di•gi•taal'

**Taalsoftware op
elke werkplek**

**ook voor
intranet**



van Dale

IN BEDRIJF

www.vandale.nl/kantoren

De lat hoger leggen

Ik meen dat Nederland klein is wat taaltechnologie betreft, omdat er geen (of te weinig) grote spelers zijn op het gebied van toegepast onderzoek onder de bestaande bedrijven. De VS blijft wat dat betreft koploper: bedrijven als IBM hebben heel lang de boventoon gevoerd in toegepast onderzoek (bijvoorbeeld spraaktechnologie). Zonder de drie M's - Middelen, Motivatie, en Menselijke kapitaal - op exact en technisch gebied, zal het de bedrijven ontbreken aan de juiste mix om verder te komen op het gebied van taaltechnologie. Terwijl de middelen en de motivatie in overvloed aanwezig zouden kunnen zijn bij sommige Nederlandse bedrijven, kan er meer gedaan worden aan de kennis van experts op dit gebied. Wat de universiteiten betreft, zie ik een belangrijk aanknopingspunt: deze moeten opleidingen aanbieden die de lat op *state of the art*-niveau leggen, en niet slechts een beperkte vorm van taalkunde. Dit is op dit moment, in sommige gevallen, niet het geval.

Het is bijvoorbeeld niet van deze tijd dat een taalkunde-opleiding absoluut geen raakpunten heeft met de bestaande realiteit waarin een sterke link wordt gerealiseerd met aangrenzende gebieden, zoals de computerlinguïstiek en *machine learning*. Veel studenten in Nederland krijgen de kans niet eens om kennis te maken met de probabilistische taalmodellen en *machine learning*-algoritmen die op dit moment zo centraal staan in *state of the art*-onderzoek naar spraakherkenning en taalkundige taken zoals *parsing*. Hier kan meer aan worden gedaan: de universiteiten hoeven in principe hun onderzoeksvoorkeur niet te wijzigen, maar zullen de student wel eerlijk de kans moeten bieden om zich te ontplooiën in meer toegepaste richtingen.

Khalil Sima'an is universitair docent natuurlijke taalverwerking en statistische leeralgoritmen en is verbonden aan de Universiteit van Amsterdam

Uitgelicht!

Organize of Organise?

Met trots heeft het Nederlandse bedrijf Polderland Language & Speech Technology B.V. haar nieuwste product gelanceerd: de **Oxford Spellchecker and Dictionary**. Het softwareproduct, ontwikkeld in samenwerking met Oxford University Press, is de allereerste spellingcontrole die op de Oxford-standaard is gebaseerd.

De Oxford Spellchecker and Dictionary is een ideaal taalhulpprogramma bij het schrijven en redigeren van Brits Engelse teksten. Het product biedt een unieke combinatie van spellingcontrole en woordenboekfunctionaliteit en wordt op eenvoudige wijze gebruikt binnen Microsoft® Word®. Met de meest recente editie van de gezaghebbende Oxford Dictionary of English als basis, kunnen meer dan 200.000 woorden op spelling en betekenis worden gecontroleerd. Daarnaast omvat de spellingcontrole 25.000 eigennamen, geografische namen en medische termen.

Alhoewel er geen officiële standaard bestaat, zoals we voor het Nederlands het groene boekje hebben, wordt de Oxford-spelling herkend als dé stan-

daard voor het Brits Engels. Met de Oxford Spellchecker and Dictionary kunnen teksten nu voor het eerst worden gecontroleerd volgens deze standaard. Is het 'pedal' of 'peddle', 'organise' of 'organize'? De Oxford spelchecker and Dictionary geeft hier antwoord op. Als extra functionaliteit kunnen gebruikers zich laten attenderen op woorden die bijna identiek worden gespeld, maar verschillende betekenissen kunnen hebben. Geef je iemand nu een 'compliment' of een 'complement'?

Debbie Kenyon-Jackson, managing director van Polderland is duidelijk over het belang van de Oxford Spellchecker and Dictionary. "Als Britse expatriate in Nederland sta ik steeds versteld van het hoge niveau van het gesproken Engels van Nederlanders, terwijl het niet hun moedertaal is. Toch blijft het een uitdaging om correct Brits Engels te schrijven. De Oxford Spellchecker and Dictionary is een onmisbaar hulpmiddel voor iedereen die correct Brits Engels volgens de Oxford standaard wil schrijven."

De Oxford Spellchecker and Dictionary is verkrijgbaar als CD-rom en als netwerkversie. Voor meer informatie kunt u terecht op www.polderland.nl/oup.htm

ATA-congres zeer succesvol

Op zaterdag 6 november vierde de **Nederlandse Association of Translation Agencies, ATA**, hun 10 jaar bestaan. 130 leden en genodigden hoorden interessante sprekers over de razend snelle ontwikkelingen in ICT-toepassingen toegespitst op hun vak. Er werden presentaties gegeven over uiteenlopende taal- en spraaktechnologische oplossingen door diverse leveranciers, inclusief Polderland Language & Speech Technology, die deze dag gebruikte om de nieuwe Oxford Spellchecker en Dictionary te promoten. Er was ook veel belangstelling voor de DIXIT en NOTaS hoopt dan ook binnenkort meer ATA-leden onder onze lezers te kunnen tellen.

Welcome Customer Care Solutions nieuwe partner AskNow

AskNow, aanbieder van zelflerende software voor het optimaliseren van digitale klantcontacten, heeft met Welcome Customer Care Solutions een partnerovereenkomst getekend. Welcome Customer Care Solutions is een Nederlandse organisatie die zich bezig houdt met het ontwikkelen en in de markt zetten van software op het gebied van klantgerichtheid. Welcome gaat de oplossing van AskNow op basis van ASP in de markt zetten onder de naam Welcome Web Selfservice 'Powered by AskNow'. Het bedrijf richt zich daarbij met name op het MKB.



| KHALIL SIMA'AN |

De Agenda van de Computerlinguïstiek

Een aantal jaar geleden, aan het begin van mijn wetenschappelijke loopbaan, verbaasde ik mij over de grote kloof tussen de taalkunde en de taal- en spraaktechnologie. Toen, komende uit de informaticahoek, begreep ik eerst geen van beiden. Niet omdat ik de gulden middenweg niet kende (dat is te veel gevraagd van een informaticus), maar omdat ik hun taal niet zo goed begreep. Inmiddels weet ik dat het onbegrip tussen beiden mede veroorzaakt is door taalverwarring.

Deze taalverwarring wordt veroorzaakt door conceptuele verschillen. Voor de taalkundige ligt het begrijpen van taalverwerking in het beschrijven van de taalfenomena door een minimaal stelsel van categorieën en regels (een grammatica). Met andere woorden: de taalkundige houdt zich vooral bezig met het beschrijven van welke uitingen grammaticaal zijn en welke niet. Dus: de taalkundige bekommert zich niet om hoe de taal wordt gebruikt in werkelijkheid, dat is immers niet het doel van de taalkunde.

Voor de taaltechnoloog gaat het om een model dat taalverwerking zo correct mogelijk uitvoert. Dat betekent dat het model altijd bedoeld is voor taalgebruik dat in een bepaalde situatie wordt gebezigd: bijvoorbeeld een dialoog over vertrek- en aankomsttijden van treinen. Het model van de taaltechnoloog wordt getest volgens een correctheidsmaat die van belang is voor de taak waarvoor het model is gebouwd. Als het model correcte vertrek- en/of aankomsttijden moet verschaffen, dan gaat het slechts daar om.

Het paradigma van de taalkundige stamt uit de discrete wiskunde en de verzamelingenleer: een grammatica werkt altijd met een eindige verzameling categorieën en regels zoals werkwoorden en zelfstandige naamwoorden. De grammatica beschrijft de taal als een verzameling uitingen die acceptabel zijn (grammaticaal); alle andere uitingen zijn in feite niet acceptabel.

De taaltechnoloog daarentegen moet omgaan met problemen die grammaticaliteitsoordelen te boven gaan: uitingen zijn vaak dubbelzinnig (ambigue) en bevatten 'ruis' (bijvoorbeeld als de spreker begint te stotteren of de uiting opeens onder-

breekt om zichzelf te corrigeren of extra informatie te verschaffen). Het model moet kunnen omgaan met onzekerheid over welke uitingen in de input zouden binnenkomen en wat die zouden kunnen betekenen. Om die onzekerheid het hoofd te kunnen bieden, maakt de technoloog gretig gebruik van waarschijnlijkheden en statistiek uit data; het model van de technoloog wordt zodanig gebouwd dat het 'zo correct mogelijk werkt'. Deze houding gaat uit van een chronisch gebrek aan sluitende kennis over de aard van de uitingen die een mens daadwerkelijk uitspreekt, schrijft en begrijpt.

MINIMAAL VERSUS CORRECT

Deze methodologische verschillen tussen taalkunde en -technologie weerspiegelen een verschil in doelen: minimalisme versus correctheid. Maar betekenen deze verschillen daadwerkelijk dat de twee gezichtspunten onverenigbaar zijn? Ik zal hieronder betogen dat ze tot elkaar veroordeeld zijn! Daarna zal ik het kort toelichten hoe deze combi wordt gerealiseerd in de moderne computerlinguïstiek.

Zoals boven is vermeld, de taalkundige grammatica is te zwak om daadwerkelijke taalgebruik aan te kunnen. Vaak zal de taalkundige veroordeeld zijn tot het vergaren van kennis buiten de taalkunde om, zoals wereldkennis (bijvoorbeeld dat de uiting 'de slang blaft' minder waarschijnlijk is dan 'de hond blaft', omdat honden nou eenmaal blaffen en slangen niet). Een manier om de omvangrijke en dynamische wereldkennis te verkrijgen, is door de statistiek 'los te laten' op combinaties van woorden. Echter, de statistiek is blind voor de complexe overeenkomsten tussen de

Khalil Sima'an is universitair docent natuurlijke taalverwerking en statistische leeralgoritmen en is verbonden aan de Universiteit van Amsterdam

verschillende uitingen in de taal. Aan de oppervlakte lijkt het vaak alsof de twee uitingen 'Jan eet een appel' en 'Marie leest een boek' totaal verschillend zijn. Statistiek losgelaten op woorden is blind voor de overeenkomsten tussen deze twee zinnen. Omdat de taaltechnoloog zijn statistiek uit eindige data verzamelt, zullen er altijd uitingen zijn die buiten de data vallen, hetgeen betekent dat het model moet kunnen omgaan met nieuwe, ongeziene uitingen. Om dat aan te kunnen, moet het model de kracht bezitten om de uitingen die aanwezig zijn in de data te 'generaliseren'. Op grond waarvan moet deze generalisatie gebeuren?

GRAMMATICA VERSUS PLATTE WOORDSEQUENTIES

Uit de korte beschrijving hierboven blijkt dat de taalkundige iets probeert te weten te komen over hoe verschillende uitingen op elkaar lijken in een bepaald opzicht, bijvoorbeeld wat syntactische structuur betreft. De gelijkenis tussen uitingen kan dus benaderd worden door syntactische gelijkenis (bijvoorbeeld frasestructuur). De grammatica wordt geacht de kracht te bezitten om een nieuwe uiting te kunnen vergelijken met bekende uitingen op een diepgaandere manier dan een vergelijking op basis van platte woordsequenties. De diepere structuur van uitingen die in de grammatica wordt samengevat, kunnen we beschouwen als nuttige 'achtergrondkennis' die de taaltechnoloog zou kunnen gebruiken om de (enigszins manke) statistiek te kunnen emanciperen over de limieten van de eindige data. De vraag is dus niet of de taalkundige kennis of de statistiek nodig zijn voor correctere modellen van taalverwerking, maar hoe de taalkundige kennis gecombineerd kan worden met de statistiek, en dat ook nog eens op vruchtbare wijze.

Het combineren van de 'a priori-kennis' (buiten de data om) met statistiek uit data vormt de hoeksteen van verschillende automatische leerparadigma (zoals *Bayesian learning*). De exacte manier waarop kennis en data kunnen worden gecombineerd in één model is het onderwerp voor uitgebreide discussie en experimenteel onderzoek, zowel binnen het vakgebied *machine learning* als binnen de huidige computerlinguïstiek.

GESTRUCTUREERDE VERSUS PLATTE DATA

Terwijl het onderzoek naar spraakherkenning een leidende rol heeft gespeeld in het herintroduceren van de data-gebaseerde, statistische taalmodellen binnen taaltechnologie, lijkt verdere vooruitgang afhankelijk te zijn van de uitbreiding van de woord-gebaseerde taalmodellen met 'diepere kennis', bijvoorbeeld uit de taalkunde. De taalmodellen uit de spraaktechnologie zijn gebaseerd op statistiek over de combinaties van opeenvolgende woorden in typische uitingen. Zodra de uitingen uitgebreid worden met syntactische en semantische structuren, kunnen we spreken van 'gestructureerde taalmodellen' die de overeenkomsten tussen de verschillende uitingen zichtbaar maken op grond van

Uitingen die een mens daadwerkelijk uitspreekt, schrijft en begrijpt

gelijkenis tussen syntactische en semantische eigenschappen van de uitingen. Deze ontwikkeling binnen computerlinguïstiek en taalverwerking boekt reeds (bescheiden) successen zowel in de spraaktechnologie als in taalkundige taken zoals syntactische analyse (*part-of-speech tagging* en *parsing*).

Het spannendste van deze ontwikkeling is dat het onderzoek binnen computerlinguïstiek naar 'gestructureerde taalmodellen' tot nieuwe, theoretische inzichten leidt over de aard van de menselijke taalverwerking en zijn relatie tot andere cognitieve capaciteiten. Zo is het bijvoorbeeld goed denkbaar dat taalverwerking vergelijkbaar is met de manier waarop visuele perceptie plaatsvindt. Hoewel niet direct van belang voor taaltechnologie, is het op een langere termijn essentieel om deze diepere inzichten in menselijke taalverwerking te verkrijgen ten einde te komen tot systemen die complexere mens/machine-interactie aankunnen. Want de taaltechnologie zou de plank volledig missen als taal niet verwordt tot een natuurlijk medium van communicatie tussen mens en machine die aansluit bij andere vormen van menselijke communicatie. |

Taal- en spraaktechnologie in bedrijf: Comfour TeleCom

Comfour Telecom is een telecomaانبieder met een compleet dienstenpakket voor infrastructures (DSL en ADSL), alsmede voor vaste en mobiele telefonie. Deze onderneming werd in 1999 opgericht en de huidige activiteiten bestaan uit het internationaal in- en verkopen van belminuten op de wereldmarkt en het verkopen van deze belminuten aan bedrijven en particulieren in Nederland tegen een concurrerend tarief.

Bellers kunnen door activatie van de carriepreselectcode 1614 direct worden geschakeld op het netwerk van Comfour Telecom. Met behulp van een eigen switch kan het telefoonverkeer wereldwijd worden geschakeld op netwerken van andere telecomaانبieders. Comfour Telecom heeft samenwerkingsovereenkomsten met verschillende internationale carriers en heeft directe verbindingen met onder andere London, Parijs, Frankfurt en New York.

Taal- en spraaktechnologie spelen bij de activiteiten van Comfour Telecom een belangrijke rol. In de infrastructuur van verschillende van hun diensten zijn Taal- en Spraaktechnologie geïntegreerd, namelijk in die van het internationaal calling card platform, het 0800- / 0900-platform en het call center-platform.

Call center-diensten Whisper/hold

Het voice response-systeem beantwoordt de inkomende oproep en speelt een voice prompt af (Music on hold) en belt vervolgens het call center-nummer. Zodra het call center-nummer wordt beantwoord speelt het IVR-systeem een bericht af dat alleen gehoord wordt door de beantwoorder. Nadat het bericht is afgelopen wordt het gesprek doorverbonden, tenzij een van beide of beide verbindingen zijn verbroken. Het bericht dat wordt afgespeeld is afhankelijk van het door de beller gebelde nummer.

Call center-diensten 0800 / 0900 blokkering

Het voice response systeem controleert haar database op A-nummer (het nummer van de beller / CLI) en het B-nummer (gebelde nummer). Indien het A-nummer voorkomt voor het B-nummer dan zal de IVR een bericht afspelen en het gesprek beëindigen.

Keuze uit diverse talen uitgesproken volgens 'linguistic rules'

In augustus 2001 nam Comfour Telecom het technologisch centrum van het Nederlandse RSL Com over, inclusief hun calling card-platform. Dit platform werd en wordt intensief gebruikt door veelal allochtone klanten in Nederland die een telefoonkaart met beltegoed kopen bij een verkooppunt (bijvoorbeeld een sigarenwinkel) en vervolgens voor een goedkoop tarief bellen met mensen in landen als Turkije, Marokko, Tunesië, India, Pakistan, enzovoort.

De klanten kunnen het beltegoed op hun kaart gebruiken door naar een toegangsnummer van Comfour Telecom te bellen. Zij worden dan door het calling card-systeem begroet in het Engels en vervolgens verzocht hun kaartnummer in te toetsen. Daarna kunnen zij kiezen in welke taal ze verder toegesproken willen worden. Het systeem bevat naast Nederlands en Engels onder andere de volgende talen: Arabisch, Turks, Frans, Duits, Fins, Zweeds, Deens, Spaans, Italiaans en Portugees.

Uniek aan dit calling card-systeem is dat het beschikt over zogenaamde *linguistic rules*. Hiermee spreekt het systeem woorden automatisch uit volgens de uitspraak-

regels van de betreffende taal. In de studio zijn slechts de spraakfragmenten opgenomen, die door het systeem worden gecombineerd tot woorden en zinnen. Het systeem zorgt er bijvoorbeeld voor dat een datum op de juiste locale wijze wordt uitgesproken.

Separate spraakherkenningsserver bij het 0800- / 0900-platform

Een ander onderdeel waarin taal- en spraaktechnologie is geïntegreerd is het 0800- / 0900-platform. Allerlei 0800- en 0900-diensten worden door klanten van Comfour Telecom ondergebracht op dit platform, waarbij Comfour Telecom zorgt voor de juiste registratie, routing, afhandeling en afrekening van de gesprekken.

Een voorbeeld van toepassing van taal- en spraaktechnologie in dit platform is 0900-123VALK (0900-1238255) voor Van der Valk-hotels en -restaurants. Uit een IVR-menu kan de beller eerst de gewenste dienst kiezen (bijvoorbeeld het reserveren van een tafel, een kamer of een zaal). Vervolgens kan hij, door eenvoudigweg een plaatsnaam te noemen worden doorverbonden met het door hem gekozen Van der Valk-hotel of -restaurant. Hij mag ook een plaats in de nabijheid noemen. Bijzonder aan deze toepassing is dat bellers gebruik kunnen maken van spraakherkenning.

Uniek aan de toepassing voor Van der Valk is dat voor de spraakherkenning gebruik wordt gemaakt van een separate spraakherkenning-server. Via VOIP worden de te herkennen spraakfragmenten naar de spraakherkenner gestuurd alwaar ze geanalyseerd worden. Het resultaat wordt vervolgens via de netwerkverbinding terugge-

Call center-diensten Antiplaag - harassment check

Het voice response systeem controleert haar database op de combinatie van A-nummer (het nummer van de beller / CLI) en B-nummer (door de beller gekozen nummer). Is de combinatie aanwezig in de database dan kan de IVR een bericht afspelen en het gesprek beëindigen. Is de combinatie niet aanwezig dan zal het IVR het gesprek doorverbinden.

Call center-diensten Antiplaag - Harassment monitor

Het voice response-systeem belt het B-nummer en zet de verbinding op. Het voice response-systeem zal vervolgens het gesprek monitoren op de uitgaande verbinding op een specifieke dtmf (druktoetsen) combinatie. Als deze DTMF-combinatie wordt waargenomen (m.a.w. als deze is ingetoetst door de agent) wordt de uitgaande verbinding verbroken en wordt de combinatie van A-nummer (het nummer van de beller / CLI) en B-nummer (het gebelde nummer) opgeslagen in de database. De volgende keer als de beller het nummer belt wordt een bepaald bericht afgespeeld.

stuurd naar het 0800- / 0900-platform, waar vervolgens de verdere afhandeling of routing van het gesprek plaatsvindt.

Call center-platform

Het laatste onderdeel van de telecom infrastructuur van Comfour Telecom waarin taal- en spraaktechnologie is geïntegreerd is het call center-platform. Het call center-platform kent verschillende applicaties, onder andere routing op basis van menukeuze waarbij het voice response-systeem de beller de mogelijkheid biedt om te kiezen uit n-items. Het voice response-systeem verbindt de beller door met het telefoonnummer dat hoort bij het gekozen item. Op vergelijkbare wijze kan het call center-platform doorverbinden met een specifiek telefoonnummer op basis van tijd/datum, postcode en netnummer van de beller.

Daarnaast kan het call center-platform het gesprek doorverbinden met een alternatief nummer als het gebelde nummer bezet is of als er niet wordt opgenomen. Als alternatief voor deze optie kan de beller de mogelijkheid geboden worden een boodschap achter te laten (voice-mail). Bijzonder daarbij is dat het bericht vervolgens als e-mail-bericht verstuurd wordt naar het call center, waar een agent het bericht beluistert en benodigde actie onderneemt. Enkele andere specifieke call center-diensten worden beschreven in de kaders.

Comfour Telecom heeft taal- en spraaktechnologie opgenomen in haar eigen platformen c.q. infrastructuur. Daarvoor werkt ze nauw samen met TeleCats. Comfour Telecom werkt ook samen met partijen die taal- en spraaktechnologie-diensten aanbieden, bijvoorbeeld het bedrijf Voice Data Bridge dat de 118 dienst van Comfour Telecom afhandelt.

NOTaS groeit! De afgelopen maanden heeft de organisatie een aantal nieuwe leden kunnen verwelkomen. Daarnaast zijn er nog een reeks aanmeldingen in behandeling. We zijn in het bijzonder blij met de belangstelling uit de klantenkringen en hopen op steeds meer bedrijven van de gebruikerskant.

Terugblik op de Deelnemersvergadering

Op 1 oktober 2004 kwamen de deelnemers bijeen op de Universiteit van Tilburg. Ook een aantal van de nieuwe leden waren voor het eerst aanwezig: de TST-Centrale en DUTCHEAR.

Tijdens deze bijeenkomst werd een terugkoppeling gegeven over de bestuursvergadering. Het bestuur heeft unaniem het besluit aangenomen om niet-Nederlandse bedrijven en kennisinstellingen als deelnemers te verwelkomen. De eerste geïnteresseerde partijen hebben zich in dit kader al gemeld op de STEVIN-dag, de TST-Centralesessie en direct aan het secretariaat van NOTaS.

Op de bijeenkomst werd ook een presentatie gegeven van IRC Nederland. Deze organisatie is een samenwerking tussen Senter en Syntens en maakt deel uit van een uitgebreid Europees netwerk. IRC Nederland stimuleert internationale samenwerking door:

- contacten te leggen tussen bedrijven,
- informatie te verschaffen over buitenlandse vraag- en aanbodprofielen,
- en internationale bijeenkomsten te organiseren.

Meer informatie over de organisatie: www.irc-nl.org.

DIXIT

Op 1 oktober vond ook een DIXIT-redactievergadering plaats. Tijdens deze vergadering is gesproken over een eventuele functionele uitbreiding van de redactie en een nieuw *format* voor 2005. De bedoeling is om het magazine nog herkenbaarder te maken voor de lezers, en om de redactionele expertise per rubriek uit te breiden en nog meer te verdiepen.

Team voor Taal 2005!

Team voor Taal is het Nederlandstalige ontmoetingsplatform op het gebied van TST voor investeerders, managers, bestuurders en wetenschappers. De bedoeling is om mensen en organisaties die TST al toepassen of van plan zijn dat te gaan doen, bij elkaar te brengen. De succesvolle bijeenkomst van Team voor Taal op 21 januari krijgt een prominent vervolg op donderdag 2 juni 2005. Reserveer nu alvast 2 juni 2005 in uw agenda. Meer informatie: m.spitholt@telecats.nl

TST-centrale

1 oktober was een drukke dag, want diezelfde middag had de TST-centrale, in samenwerking met NOTaS en de Taalunie, een kennismakingsbijeenkomst voor bedrijven

en kennisinstellingen. Rond vijftig belangstellenden namen deel aan de bijeenkomst waarin het team van de TST-centrale werd geïntroduceerd en meer informatie werd gegeven over de organisatie zelf. Tijdens het programma gaf de TST-centrale een presentatie over haar doelstellingen, achtergrond en activiteiten. Polderland ging hierna in op de rol van de TST-centrale voor het bedrijfsleven.

Voor meer informatie over de centrale: www.tst.inl.nl

STEVIN update

Na het zeer succesvolle Brokerage Event op 15 september is de eerste *call for proposals* op 2 november gesloten met 19 projectvoorstellen. De programmacommissie is zeer tevreden met de grote opkomst en het animo van de belangstellenden.

Op 15 december komt het TST-bestuur bij elkaar om een definitief besluit te nemen over de toekenningen voor de eerste ronde. Vóór deze datum worden de voorstellen door een internationale beoordelingscommissie geëvalueerd en geprioriteerd. De programmacommissie bespreekt hierop het advies van de internationale commissie en stuurt haar eigen definitieve advies aan het bestuur.

Gezocht: stageplaatsen

Since 2000, the University of Leuven offers a one-year postgraduate program in Speech and Language Technology which can be followed by students with a Master's degree in Computer Science, Linguistics or Electrical Engineering. Besides courses, exercise sessions and seminars, the program includes an internship of 12 weeks, which will start on March 14, 2005. This year we have 13 students. Their CVs and more information about the internships can be found at <http://www.ccl.kuleuven.ac.be/Courses/Internship/index.html>

Currently we are trying to find hosts for each of the thirteen students. For more information contact Frank van Eynde, coordinator of the SLT internships, at: frank@ccl.kuleuven.ac.be

NOTaS

Deelnemers van Stichting NOTaS zijn:

AskNow Solutions – CLST – Comsys – Dutchear – Em@ilco.nl – Human Inference – Institute for Logic Language and Computation (ILCC) – Knowledge Concepts – ORCAvoice – Polderland Language and Speech Technology – Q-go – TST-Centrale, Stichting van Nederlandse Lexicologie – TeleCats – TNO TPD – Universiteit van Tilburg Faculteit Letteren – Universiteit Twente Faculteit Informatica – Van Dale Lexicografie – Vocalibur Language & Speechtechnology

Meer informatie: www.stichtingnotas.nl

Selfservice Telefonie

Interactive Voice Response en Spraakherkenning



Veel organisaties besparen enorme kosten door het inzetten van selfservice telefoniesystemen, terwijl de bereikbaarheid sterk wordt verbeterd.

TeleCats is de partij om zaken mee te doen als we praten over selfservice met behulp van IVR en spraakherkenning. Organisaties die dat dagelijks ondervinden zijn o.a.: De Telegraaf, PCM, Essent, OHRA, SNT, Telfort, het Ministerie van VROM, De Belastingdienst en NS reizigers.

Wilt u weten hoeveel kosten u kunt besparen door de selfservicetelefoniesystemen van TeleCats in te zetten? Maak dan een afspraak met ons.



Internet: www.TeleCats.nl info@TeleCats.nl Telefoon: 053 4889900



Laat uw klanten niet zoeken - maar vragen.

Het is een gemiste kans en u weet het: klanten gebruiken uw website om producten en diensten af te nemen maar vaak kunnen zij de gezochte informatie niet vinden. Het gevolg? Uw klanten blijven bellen en extra kosten veroorzaken. Of erger: ze verlaten u omdat ze niet op tijd worden geholpen. Terwijl het antwoord op hun vraag direct online beschikbaar is.

De Online Marketing & Self Service software van Q-go is gebaseerd op een unieke vorm van Natural Language Processing. Hierdoor kan uw klant intuïtief in eigen woorden zijn vragen stellen op de website. Het resultaat? Uw klant wordt begrepen en direct online geholpen.

Het voordeel voor u? Betere service, tevreden klanten, inzicht in hun vraaggedrag en bovendien aanzienlijke kostenbesparingen. Q-go garandeert dat u uw investering in de software binnen 6 maanden terugverdient. Klanten als Postbank, Rabobank, TPG, La Caixa, Deutsche Telekom en Telefónica maken succesvol gebruik van Q-go.

Wilt u meer weten en een ROI berekening laten maken? Surf naar www.q-go.com of stuur een email naar astrid.smedema@q-go.com. Of bel naar 020 531 3800.



Online Marketing & Self Service