

Toepassingen van taal- en spraaktechnologie in Nederland

Dit document biedt een overzicht van het taal- en spraaktechnologiesveld in Nederland, gezien vanuit een zestal maatschappelijke toepassingsgebieden. Het document is in zekere zin een aanvulling op "Het Nederlands in taal- en spraaktechnologie: prioriteiten voor basisvoorzieningen" van Daelemans en Strik (red.) uit 2002, dat de leidraad vormde voor het STEVIN-programma, gefinancierd door de Nederlandse en Vlaamse overheid en uitgevoerd onder auspiciën van de Nederlandse Taalunie. Zes jaar later is STEVIN al deels gerealiseerd, en plukt het gebied de vruchten. Andere onderzoeksprogramma's van NWO (IMIX, CATCH, Vernieuwingsimpuls) en van SenterNovem (IOP-MMI) hebben in dezelfde periode substantiële onderzoekslijnen helpen ontwikkelen. Verschillende nieuwe bedrijven zagen het licht, en een groeiende groep van MKB's, deels verenigd in Stichting NOTaS, richtte zich op de toepassing van taal- en spraaktechnologie in een groeiende variëteit van diensten. Het gebied als geheel is door STEVIN en NOTaS inmiddels een stevig netwerk geworden waarin de leden van zowel kennisinstellingen als industrie, goed op elkaar voort weten te bouwen, en dat de naam van Nederland op het gebied van de taal- en spraaktechnologie ook buiten de landsgrenzen hoog houdt.

De prioriteitenstelling van Daelemans en Strik diende om focus, slagkracht en eenheid in het veld te bevorderen. We spreken inmiddels over een sterk netwerk van academische en industriële R&D-groepen. In dit document wordt de blik naar buiten gericht. In welke toepassingsgebieden wordt taal- en spraaktechnologie momenteel succesvol toegepast, in welke gebieden zou dit kunnen maar gebeurt dit nog niet, en waar liggen op dit moment de grootste mogelijkheden?

1. Inleiding

De mens heeft van nature de neiging om natuurlijke taal te gebruiken bij het communiceren van kennis en informatie. Hoewel moderne ICT-technologieën met schermen, knoppen en toetsenborden enthousiast omarmd worden, blijven er complexe vragen en verlangens die altijd nog beter in woorden en zinnen uitgedrukt worden. Het is daarom van groot belang dat computers de stap zetten van hun huidige grotendeels "stille" staat naar die van pratende en vooral begrijpende, intelligente systemen.

De beste technologie van dit moment is daadwerkelijk in staat om de meeste stappen die nodig zijn om natuurlijke taal en spraak te genereren en te begrijpen¹, succesvol te zetten – mits het domein of onderwerp van gesprek beperkt en vooraf bekend is. Toepassingen waarbij TST succesvol wordt toegepast zijn ondermeer te vinden bij:

- het plannen van reizen,
- het routeren van informatievragen,
- het genereren van berichten over persoonlijke financiën, het weer, verkeer, en sport,
- het automatisch beantwoorden van veel gestelde vragen
- het uitfilteren en classificeren van specifieke typen emailberichten of documenten.

Intussen wordt in academisch en industrieel onderzoek een nieuwe generatie van applicaties klaargestoomd waarin de volgende stap gezet wordt naar toepassingen met

¹ In de functionele zin van het woord: het interpreteren en vervolgens succesvol beantwoorden van een in "gewoon Nederlands" gestelde vraag.

zowel een hogere flexibiliteit als een grotere robuustheid. Met deze nieuwe generatie applicaties moet het mogelijk worden om:

- vragen in bredere domeinen of zelfs in het algemeen te beantwoorden,
- gepersonaliseerde informatie uit heterogene bronnen te filteren en samen te vatten,
- sprekeronafhankelijk spraak te herkennen,
- spraak ook in lawaaiige omstandigheden te herkennen,
- vloeiende spraak te synthetiseren,
- teksten automatisch van en naar het Nederlands te vertalen.

In vergelijking met de meeste Europese landen investeert Nederland zowel op academisch als industrieel gebied, zeer sterk in de ontwikkeling van nieuwe taal- en spraaktechnologieën. Nederland is op het gebied van taal- en spraaktechnologie al enige decennia een bovengemiddelde speler in het internationale onderzoeksveld.

Kijkend naar succesvolle projecten van de afgelopen jaren, kan worden vastgesteld dat taal- en spraaktechnologie zich een duidelijke plaats hebben verworven in een aantal maatschappelijk en economisch relevante toepassingsgebieden zoals zorg, e-government, cultureel erfgoed, (serious) gaming en vrijetijdsbesteding, veiligheid, media mining, (self-)service en onderwijs. We zetten deze gebieden (waaruit we ook de typen afnemers generaliseren) af tegen de academische en industriële R&D die Nederland rijk is. Uit het overzicht blijkt dat voor elk van deze toepassingsgebieden verschillende soorten afnemers zijn:

- De (rijks)overheid
- De semi-overheid (door de overheid gesubsidieerde afnemers zoals erfgoedinstellingen, omroepen en scholen)
- Commerciële afnemers (grote zoals banken en verzekeraars, maar steeds vaker ook MKB bedrijven).

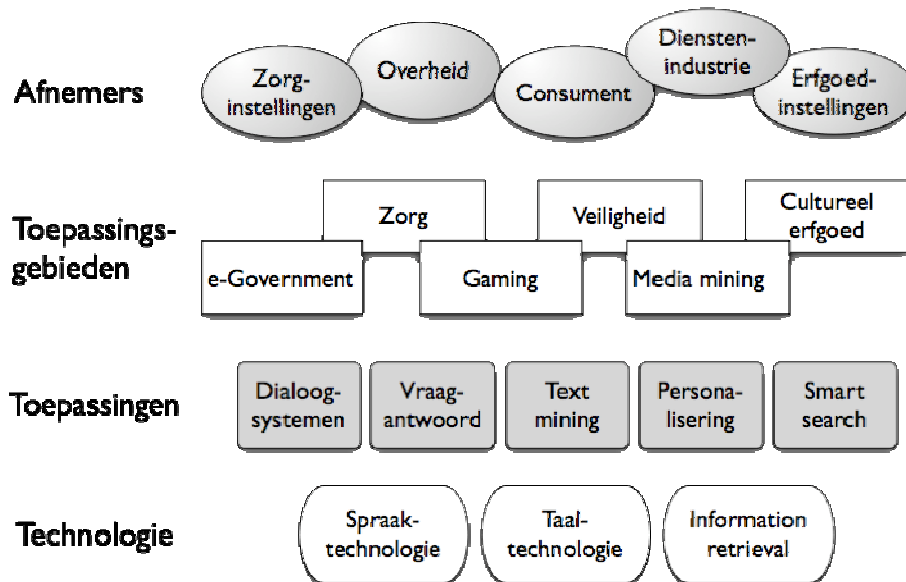
Vooraf in deze laatste categorie schuilt een enorme potentie: als taal en spraaktechnologie eenmaal commercieel succesvol kan worden ingezet dan kan er een enorme schaalvergroting optreden.

2. Toepassingsgebieden, afnemers, en technologieën

Bij de zes maatschappelijke toepassingsgebieden waarin taal- en spraaktechnologie de laatste jaren successen heeft geboekt horen ruwweg vijf afnemersgroepen: de gebruiker, de overheid, de dienstenindustrie, zorginstellingen en erfgoedinstellingen. Centraal in al deze toepassingen staat de gebruiker die in toenemende mate snel, goedkoop en altijd toegang wil hebben tot de verschillende diensten. Service wordt steeds meer gezien als iets dat er "automatisch" bij hoort en gratis is, en als iets dat altijd toegankelijk is, dus ook na

sluitingstijd en in het weekend. Deze ontwikkeling ziet men duidelijk in de meer commercieel georiënteerde diensten, maar in toenemende mate ook in de meer maatschappelijk georiënteerde diensten. Om tegemoet te komen aan deze stijgende druk vanuit “de gebruikers” wordt Taal en Spraaktechnologie steeds vaker ingezet om daar waar mogelijk de mens met intelligente systemen te ondersteunen of soms zelfs te vervangen.

Voor al deze toepassingsgebieden worden systemen gebouwd die onder de motorkap toepassingen draaien als dialoogsysteem, vraag-antwoord-systemen, en smart search. De onderliggende technologie wordt, in feite historisch, meestal onderscheiden in drie technologieën: spraaktechnologie, taaltechnologie, en information retrieval. “Spreekende computer”-helpdesks die dialogen aangaan en vragen beantwoorden, zoals de demonstrator van het IMIX-project die medische vragen kan beantwoorden (“Is de bof besmettelijk?”) kunnen alleen bestaan door de verregaande integratie van alle drie de technologieën.



3. Inventarisatie

Een kwantitatieve stand van zaken van taal- en spraaktechnologie in Nederland is, vanuit vogelperspectief, ongeveer de volgende.

1. In tien kennisinstellingen (UvA, UU, UvT, RUG, RU, VU, UL, TU Eindhoven, TU Delft en de UT) en bij semi-overheidsinstelling TNO werken in totaal circa 100 onderzoekers aan taal- en spraaktechnologie en information retrieval, verdeeld over faculteiten of departementen Informatica, Letteren, of Geesteswetenschappen. Het merendeel van de onderzoekers is promovendus of postdoc-onderzoeker. In verschillende vormen bieden de kennisinstellingen Bachelor en Master-onderwijs aan in de taal- en spraaktechnologie, informatiewetenschappen, of kunstmatige intelligentie.
2. Circa 40 Nederlandse MKB-bedrijven opereren binnen de genoemde toepassingsgebieden. Veel van de bedrijven vinden hun oorsprong in onderzoek van kennisinstellingen, en rekruteren een belangrijk deel van hun werknemers uit de studenten (middels stages en afstudeerprojecten) en junior-onderzoekers van de kennisinstellingen.

Met een zelfde vogelperspectief vatten we voor ieder van de zes toepassingsgebieden samen wat er tot op heden is bereikt.

1. **Zorg.**

- a. Er is een grote en groeiende vraag naar patiënteninformatie; computersystemen kunnen een deel van de informatiebehoefte opvangen.
- b. Specifieke patiëntengroepen hebben beperkingen in het omgaan met taal en spraak. Hulpmiddelen worden ontwikkeld die ondersteunend werken in dagelijkse communicatiebehoeften en de patient helpen gebruik te maken van moderne ICT-middelen.
- c. De zorgsector leunt voor een belangrijk deel op documenten: van handgeschreven rapporten tot bijsluiters. Deze teksten bevatten kennis en informatie die op efficiënte manieren ontsloten kan worden met technieken uit de taal- en spraaktechnologie en information retrieval.
- d. Een steeds grotere groep ouderen wil zelfstandig blijven wonen maar dan wel onder controle staan van een of meerdere medische instellingen. Controle op afstand door intelligente, dikwijls TST-gebaseerde systemen waarbij mensen pas worden ingeschakeld als er "iets" niet klopt, worden dan ook steeds vaker toegepast.

2. **e-Government.**

- a. Gemeenten en andere locale overheden willen hun burgers efficiënt alle informatie aanbieden die ze kunnen (en ook wettelijk moeten) geven, ook buiten openingstijden, via internet, en via de telefoon. Verschillende soorten automatische helpdesks zijn voor dit doel ontwikkeld.
- b. Organen binnen de overheid en de rechtswereld produceren enorme hoeveelheden documenten, die stuk voor stuk een groot belang hebben. In

toenemende mate wil men kunnen zoeken in deze immense en groeiende stapel, en wil men samenvattingen genereren; met een oog op privacy en anonimiteit. Bij al deze aspecten assisteren volautomatische systemen.

- c. De democratische besluitvorming die altijd al openbaar was, kan in toenemende mate vanuit de luie stoel thuis bekeken en beluisterd worden. Burgers willen daarbij zelf bepalen welke delen van bv een parlementaire zitting of gemeenteraadsvergadering voor hen interessant is. Dit vereist de mogelijkheid om in de vergaderingen te kunnen zoeken naar "relevante" passages.

3. **Cultureel erfgoed.**

- a. Steeds meer Nederlands cultureel erfgoed komt digitaal beschikbaar doordat het of digitaal geproduceerd wordt of gedigitaliseerd wordt. Met taal en spraaktechnologie en information retrieval methoden wordt het onderzoekers mogelijk gemaakt om complexe onderzoeksvragen te stellen aan automatische "knowledge bases" die zijn geëxtraheerd uit het ruwe digitale materiaal. Ook het lekenpubliek kan nu zoeken in erfgoedarchieven die vroeger slechts in hun unieke papieren vorm in beperkte omstandigheden konden worden geraadpleegd.
- b. Een groeiend deel van het gearchiveerde materiaal bestaat uit gedigitaliseerde audio- en video-opnamen met spraak. Met spraakherkenning kan het geluidssignaal voorzien worden van een extra laag "automatische ondertitels"; de technologie hoeft niet perfect te werken om toch al zeer bruikbare "audio retrieval" te kunnen doen. Een sprekend voorbeeld zijn de toespraken die koningin Wilhelmina tijdens de oorlog voor Radio Oranje hield. Het oplijnen van de in de oorlog uitgeschreven teksten met de soms zeer moeilijk verstaanbare originele geluidsopnamen maakt het mogelijk om via Internet toegang te krijgen tot gezochte passages van alle toespraken van koningin Wilhelmina.

4. **(Serious) gaming en vrijetijdsbesteding.**

- a. In het onderwijs is computer-assisted language learning een beproefde methode van taalonderwijs. Door gebruik te maken van feedback en ook spelelementen kunnen mensen van alle leeftijdsgroepen ook buiten normale klassituaties een taal leren.
- b. Ook de besteding van vrije tijd kent zijn informatiemarkten (welke goede restaurants zijn er hier) en -niches (hoe kan ik snel een taxi bellen) waarbij het domein zo specifiek is dat gesproken dialoogsystemen over de telefoon ontwikkeld kunnen worden die deze informatie bieden.
- c. Gekoppeld aan moderne GPS-systemen, kunnen gebruikers in een groot aantal talen uitleg krijgen over toeristische bezienswaardigheden bij bv stadswandelingen.
- d. Naarmate serious games de geschetste werkelijkheid dichter benaderen, ontstaat steeds meer de behoefte om ook de dialogen tussen bv de gebruiker en de systemen, zo natuurlijk mogelijk te laten verlopen. TST-gebaseerde dialoogsystemen waarbij de gebruiker praat met "het systeem" ipv ermee te communiceren middels toetsenbord en muis, worden daarom steeds vaker gebruikt. Voorbeelden hiervan kan men zien bij trainingsprogramma's van het Amerikaanse

leger of bij flightsimulators in de commerciële luchtvaart waarbij gebruiker praat (opdrachten, vragen) met de systemen via een headset.

5. **Veiligheid.**

- a. Veel informatiestromen worden weliswaar waargenomen en opgeslagen, maar de menselijke expertise en tijd ontbreekt vaak om deze informatiestromen te onderzoeken op de aanwezigheid van afwijkende en uit veiligheidsoogpunt belangrijke fenomenen. Automatische middelen kunnen grootschalige informatiestromen snel verwerken en de saillante passages voor nadere inspectie apart zetten. Het kan daarbij gaan om spraak over telefoonlijnen en in ruimtes met een microfoon, emails, chat, en ander internetverkeer, en ook grote "off line" documentenverzamelingen.

6. **Media mining.**

- a. Terwijl de informatie die beschikbaar wordt gesteld over de oude media en internet zo sterk groeit dat de opslag ervan op afzienbare tijd problematisch wordt, en zoeken in deze informatiestromen een probleem van een nauwelijks te bevatten schaal is, komen er in snel tempo steeds meer typen informatiestromen bij (blogs, podcasts, vodcasts), in allerlei mengvormen van media. Tekst en spraak zijn de meest menselijke communicatiemiddelen en blijven logischerwijs de hoofdrol spelen in de mens-mens communicatie. Willen systemen instaat zijn deze informatie te ontsluiten, dan zullen ze de menselijke communicatie ten minste deels moeten kunnen "nadoen" en "begrijpen". Het omzetten van spraak-naar-tekst, het vertalen van teksten van de ene naar de andere taal, het associëren van het ene document met het andere en het samenvatten van documenten zijn voorbeelden van technieken die nu al toegepast worden bij de ontsluiting van (grote) multimedia archieven.
- b. De consument drukt zijn eigen profiel op de verzameling media die hij beheert op zijn diverse "computers" (inclusief mobiele telefoons en PDA's), en wil graag ondersteund worden in die persoonlijke voorkeuren door gepersonaliseerde informatiediensten, die bijvoorbeeld binnenkomend nieuws filteren op interessantheid voor de gebruiker, en die suggesties doen voor media die gerelateerd zijn aan zojuist genoten tekst, muziek of video.

Appendix: Projecten gegroepeerd naar hoofdthema's

Zorg

Partners	Dedicon
Projectnaam	Synthetisch gesproken kranten
Thema	Zorg
Omschrijving	Blinden en slechtzienden die <u>niet</u> met een computer kunnen omgaan, blijven verschoten van een dagelijkse krant. Deze groep, over het algemeen senioren >65 jaar, is relatief groot. Als alternatief hiervoor een dagelijkse editie uitbrengen die voorgelezen is met de menselijke stem, is een zo goed als onmogelijke operatie. Spraaksynthese biedt voor dit hiaat een welkome oplossing. Voor deze groep heeft Dedicon een productiesysteem ontwikkeld die met behulp van spraaksynthese dagbladen en tijdschriften in gesproken vorm uitbrengt. De synthetisch gesproken dagbladen en tijdschriften worden op een DAISY cd-rom uitgebracht. Deze cd-rom kunnen zonder computer op een DAISY-speler afgespeeld worden.
Status	Dit project is in 2004 geïmplementeerd, maar inmiddels niet meer operationeel. De kwaliteit van de synthetische vertolking was onvoldoende. Dedicon zou graag een dergelijk systeem willen verbeteren en opnieuw activeren. Hiervoor helaas, tot nu toe, geen financiering gevonden

Partners	Dedicon
Projectnaam	Synthetisch gesproken studieboeken
Thema	Zorg/Onderwijs
Omschrijving	Regelmatig vragen leesgehandicapte scholieren of studenten naar een gesproken versie van een bepaald studieboek, die nog niet in onze collectie beschikbaar is. Dit boek hebben ze echter binnen 1 of 2 weken nodig, want er is een toets of examen op komst. Hiervoor een voorlezer vinden die à la minuut en binnen de gewenste tijd een geheel boek kan komen inlezen is onmogelijk. Dit studieboek in synthetisch gesproken vorm uitbrengen is onder dergelijke tijdsdruk de enige oplossing. Hiervoor heeft Dedicon een systeem ontwikkeld die een luisterboek, met een afluistertijd van 10 uur, in 2 uur met een computerstem kan produceren.
Status	Dit project is in 2005 geïmplementeerd, maar inmiddels (ook) niet meer operationeel. De kwaliteit van de synthetische vertolking was onvoldoende. Dedicon zou graag een dergelijk systeem willen verbeteren en opnieuw activeren. Hiervoor helaas, tot nu toe, geen financiering gevonden.

Partners	Dutchear - TNO ICT - OPG
Projectnaam	ReceptenConnect
Thema	Zorg
Omschrijving	Veel medicijnen die mensen gebruiken, betreffen herhaal recepten. Om het bestellen hiervan te vereenvoudigen is een systeem gemaakt waarbij mensen de

	naam van hun apotheek, hun achternaam en geboorte datum en de medicatiennaam invoeren. Het systeem controleert vervolgens deze gegevens en stuurt het recept door naar de apotheek. De apotheek zorgt vervolgens dat de medicatie thuisbezorgd wordt.
--	---

Partners	CLST, TW Nijmegen, Sint Maartenskliniek
Projectnaam	AwareShoe
Type Contract	EZ-project (IOP-MMI) 15 juni 2007 (duur: 3 jaar)
Thema	Zorg
Omschrijving	In dit project maken we een schoen voor diabetici. De schoen is voorzien van sensors die signalen doorsturen over de conditie van de voet. Op basis van die informatie krijgt de patiënt instructies over passende vervolgacties.

Partners	Telecats
Projectnaam	Automatisch opvragen van brochures van het Koningin Wilhelmina fonds
Thema	Zorg
Omschrijving	Het KWF wilde bellers de mogelijkheid bieden om 24 uur per dag 7 dagen per week folders aan te vragen. Tijdens kantooruren moesten bellers er tevens voor kunnen kiezen te worden doorverbonden met een medewerker van het KWF. Tevens wilde het KWF niet verantwoordelijk zijn voor het beheer van het systeem. Bij het aanvragen van een folder of brochure komen de voordelen van spraakherkenning duidelijk naar voren. Door simpelweg de code of de naam van de folder in te spreken wordt de gewenste folder geselecteerd. Zijn de code en de naam van de folder niet bekend dan kan de beller de gewenste folder zoeken door te zeggen wat het onderwerp is waarover hij een folder wil ontvangen. Het systeem noemt de geschikte folders voor de beller op en de beller kan een keuze maken uit de voor hem geselecteerde folders. Postcode en huisnummer kunnen vervolgens worden ingesproken zodat het adres waar de folder(s) naar toe gestuurd moet worden, bekend is. Op deze manier heeft het KWF geen omkijken naar het bestellen van folders en brochures.
Status	Opgeleverd in het najaar 2004

Partners	Polderland Language & Speech Technology bv, BSL, Vilans
Projectnaam	Pinkhof Medische spellingscontrole / Schrijfhulp Zorg en Welzijn
Thema	Zorg
Omschrijving	Er is bijna geen sector te bedenken waarin zoveel terminologie wordt gebruikt als de zorgsector. Deze terminologie moet zowel voor de professionals als voor de gemiddelde Nederlander eenvoudig te gebruiken zijn. De Pinkhof Medische spellingscontrole en de Schrijfhulp Zorg en Welzijn zijn beide producten die dit mogelijk maken in het schriftelijk taalgebruik. Naast een uitgebreide spellingscontrole met medische terminologie heeft de gebruiker beschikking over een woordenboek/thesaurus voor zorg en welzijn en de manier om medische termen eenvoudig te vinden.

Partners	Polderland Language & Speech Technology bv, IFourC
Projectnaam	Handschriftherkenning
Thema	Zorg
Omschrijving	Ziekenhuizen en andere zorginstellingen worden steeds verder papierloos. Een volledige invoering van elektronische patiëntendossiers staat niet meer ver van ons af. In voorbereiding daarop worden vele archieven al gedigitaliseerd en de patiëntengegevens via tablet PC aan de arts beschikbaar gesteld. De arts werkt echter nog graag met geschreven aantekeningen. Op tablet PC's (Windows) is op dit moment geen Nederlandse handschriftherkenning beschikbaar. Polderland en IFourC werken samen om dit te realiseren.

Partners	Polderland Language & Speech Technology bv, KNOV, hetwep
Projectnaam	Zoektechnologie KNOV-site
Thema	Zorg
Omschrijving	De terminologie die zorgprofessionals gebruiken is niet altijd bekend bij de gemiddelde Nederlander. Dit wordt extra duidelijk als die persoon een zorginformatievraag heeft. Dat is bijvoorbeeld het geval rond zwangerschap. Een handige informatiesite is dan de site van de Koninklijke Nederlandse Organisatie van Verloskundigen. Die site is nu nog voorzien van standaard zoektechnologie waardoor de geïnteresseerde veel informatie misloopt. Polderland zorgt er in samenwerking met hetwep voor dat de zoektechnologie wordt uitgebreid met taaltechnologie waardoor de informatievraag van de geïnteresseerde beter wordt beantwoord.

Partners	inTAAL
Projectnaam	TOP benoefening
Thema	Zorg
Omschrijving	De behoefte aan oefenprogrammatuur voor mensen met afasie is groot. Afasie is een talige handicap die gevolgen kan hebben voor het vinden van woorden, het begrijpen, lezen en schrijven. Binnen de doelgroep van ongeveer 40.000 personen is de behoefte aan oefensoftware groot. In samenwerking met de Afasievereniging Nederland ontwikkelt inTAAL een thuisoefenprogramma rond het benoemen van woorden. De gebruikers krijgen een foto gepresenteerd al dan niet ondersteund met tekst of spraak en zij moeten zeggen wat de betekenis van die foto is. De gebruiker kan ook zelf foto's invoeren. Op deze wijze kunnen mensen met afasie zelfstandig thuis oefenen met een voor hen ontwikkeld oefenprogramma.

Partners	Telecats
Projectnaam	Gesproken FAQ's
Thema	Zorg
Omschrijving	CZ Actief in Gezondheid is met ruim 3 miljoen verzekerden één van de grootste zorg-verzekeraars van Nederland. CZ heeft een sterk sociaal karakter zonder winstoogmerk.

	<p>Om het groeiend aantal standaard vragen per telefoon adequaat te kunnen beantwoorden, werd een openvraag-spraakherkenningssysteem geïmplementeerd. Bellers kunnen hiermee in gewoon Nederlands aangeven wat ze van CZ willen weten. De meest gestelde standaardvragen worden beantwoord via vooraf opgenomen voice-prompts. Andere vragen worden door medewerkers beantwoord. Nieuw is dat bellers geen kennis hoeven te hebben van de menustructuur.</p> <p>De applicatie wordt gebruikt wanneer een beller in een wachtrij dreigt te komen. Het systeem zegt: Welkom bij CZ. Om u van dienst te zijn, vragen we u om uw vraag aan CZ in te spreken. Het systeem tracht de vraag te herkennen (interpreteren). In het systeem zitten tal van FAQ's. De vraag wordt, indien mogelijk, direct beantwoord. Verder krijgt de beller afhankelijk van de vraag of de situatie de mogelijkheid te worden teruggebeld op een door de beller aan te geven tijd of om te worden doorverbonden met een medewerker.</p> <p>Enkele belangrijke voordelen voor de klanten en CZ zijn:</p> <ul style="list-style-type: none"> • Veel informatie zit in de spraakherkenner; • Geen lange menu's • Een groot percentage selfservice; • Betere spreiding van de gesprekken; • Snelle routing naar de juiste agentgroepen; • Prioritering van gesprekken mogelijk (bijvoorbeeld overlijden); en • De vraag die de klant heeft wordt bij doorverbinden al naar de agent ingefluisterd door het systeem
Status	Opgeleverd in het voorjaar 2007

Partners	Telecats, FocusCura
Projectnaam	Telemonitoring COPD-patiënten
Thema	Zorg
Omschrijving	<p>Telecats heeft op een COPD-telemonitoringapplicatie gemaakt voor het project 'Zorg waarmee je thuis kunt komen!' van de Zorggroep Zuid Gelderland en het Canisius-Wilhelmina Ziekenhuis. COPD (Chronic Obstructive Pulmonary Disease: chronische bronchitis en longemfyseem) is een aandoening van de longen en luchtwegen. De COPD-telemonitoringapplicatie wordt gebruikt door en voor patiënten met COPD, specialisten, longverpleegkundigen en thuiszorgverpleegkundigen. Het is een onderdeel van de herinrichting van het proces na ziekenhuisontslag. □Patiënten vullen via de telefoon of Internet geregeld een gevalideerde vragenlijst in waaruit blijkt hoe ze zich de afgelopen periode hebben gevoeld. De vragenlijsten worden direct doorgestuurd naar het beveiligde telemonitoringdossier van de betreffende patiënt. Uit de ingevulde vragenlijst volgen 4 scores die automatisch worden berekend aan de hand van de CCQ-methodiek. Met behulp van een speciale telemonitoringmodule hebben betrokken longartsen en longverpleegkundigen de mogelijkheid om scores van patiënten in te zien, trends in de scores te analyseren en inzage te krijgen in de gezondheidsstatus van de patiënt. Indien de score daartoe aanleiding geeft, neemt de longverpleegkundige contact op met de patiënt. Om de betrokkenheid van de patiënt bij het eigen zorgproces te stimuleren, heeft de patiënt zelf ook de mogelijkheid om online de eigen scores in te zien en wordt er uitgebreide uitleg gegeven over de betekenis van scores en de aandoening</p>

	COPD in het algemeen. □ Het project is genomineerd voor de verkiezing 'Goede Voorbeelden' van de NPCF (Nederlandse Patiënten Consumenten Federatie).
Status	Opgeleverd in het voorjaar 2007

Partners	Dutchear - Eykenburg
Projectnaam	SpraakgestuurdRapporteren
Thema	Zorg
Omschrijving	In de zorgsector (in dit geval ouderenzorg) moet steeds meer gerapporteerd worden door de medewerkers die zorg aan cliënten hebben geleverd. Om dit proces te versimpelen heeft Dutchear een systeem gebouwd, waarbij met een PDA met RFID geregistreerd kan worden bij welke cliënten de (ambulante) medewerker is aangekomen en vertrokken. Vervolgens kan bij vertrek met diezelfde PDA een systeem gebeld worden en kan ingesproken worden welke zorgtaken uitgevoerd zijn en hoe lang er per taak aan besteed is. De gegevens uit de PDA acties worden gebruikt om te bepalen welke zorgtaken (verschillende tarieven) er doorberekend moeten worden. □

Partners	Dutchear - SPING - VIZIRIS
Titel project	IRIS - Individueel Reis Informatie Systeem
Thema	Zorg/Veiligheid
Omschrijving	Voor blinden en slechtzienden is alleen reizen lang niet altijd mogelijk, omdat er begeleiding tijdens de reis nodig om op grote stations de juiste weg te vinden. Om dit voor blinden en slechtzienden makkelijker te maken is een systeem gebouwd dat gebeld kan worden. Na het inspreken van de naam van het station kan de beller een aantal keuzes maken (van/naar perron, van bus of taxi of lopend), op basis waarvan bepaald wordt welk route gelopen moet worden. Vervolgens wordt de routebeschrijving "u komt langs een muur, houd links aan, u komt nu bij het loket langs" voorgelezen. Indien nodig kan het voorlezen op elk moment worden gepauzeerd en kunnen fragmenten herhaald worden. Uit praktijkproeven met de doelgroep zijn hele enthousiaste reacties gekomen. Meestal konden mensen dankzij dit systeem zelf hun weg vinden op een voor hun onbekend groot station (Den Haag - Amersfoort - Utrecht)

Partners	Dutchear - Almende - Humanitas (Rotterdam)
Projectnaam	Luister - Spraakgestuurde zorgverlening
Thema	Zorg
Omschrijving	In de zorgsector wordt steeds vaker een beroep gedaan op vrijwilligers om bepaalde zorgtaken uit te voeren. Om zo snel mogelijk te bepalen welke soort zorg en daarmee welke klasse medewerkers (vrijwilligers, alfa hulpen, verpleging) ingeschakeld moet worden is een spraakgestuurd systeem ontwikkeld dat op basis van de zorgvraag van de persoon (bijvoorbeeld "mijn verband moet verwisseld", "ik wil een rondje wandelen") bepaald welk soort zorg er moet worden gevraagd. Op basis van de zorgvraag en de agenda's van medewerkers EN op basis van voorkeuren uit het verleden tussen cliënten en verzorgers/vrijwilligers (zelf lerend systeem), wordt een match gemaakt en wordt de cliënt doorverbonden

	met de juiste persoon.
--	------------------------

Partners	Irion en Stichting Wereldconsument
Projectnaam	Wikifair
Thema	Zorg
Omschrijving	Stichting Wereldconsument spant zich in voor een grotere bewustwording van de consument, door consumenten te wijzen op de effecten van hun koopgedrag op factoren als milieu, arbeidsomstandigheden, gezondheid, en ander duurzame aspecten. Een voorbeeld: de gevaren van producten voor de gezondheid van de consumenten zelf staan vaak wel vermeld in de bijsluiters, maar niet de gevaren die bij de productie en transport worden gecreëerd. Het project Wikifair beoogt een combinatie van wiki-technologie (denk aan Wikipedia) en taaltechnologie toe te passen in een portaal waar consumenten duurzaamheidsinformatie kunnen uitwisselen: de wiki zorgt ervoor dat duurzaamheidsgegevens over producten kunnen worden verzameld tegen zeer lage kosten, de taaltechnologie zorgt ervoor dat alle informatie zodanig wordt ontsloten dat consumenten in hun eigen taal en woordgebruik alle relevante informatie snel kunnen vinden.

Partners	Irion en Ministerie van VWS
Projectnaam	VWS Classification
Thema	Zorg
Omschrijving	De inhoud van de website van het ministerie van VWS is slechts ontsloten met een full text zoekstelsel van Verity. Om de toegankelijkheid voor burgers te vergroten heeft het ministerie de op taaltechnologie gebaseerde classificatietechnologie van Irion ingebouwd, zodat er voor burgers een grotere vrijheid is in het formuleren van hun zoekvragen, en de accuraatheid van de antwoorden (mate waarin de antwoorden relevant zijn) sterk wordt verbeterd.

Partners	Irion & TNO
Projectnaam	Arbouwportaal
Thema	Zorg
Omschrijving	In een proefproject heeft Irion samen met TNO een landelijke zoekmachine gebouwd waarin zowel burgers als experts informatie kunnen vinden op het gebied van arbeidsomstandigheden en gezondheid.

Partners	ViaTaal, Offis, KTH, Prosyst, Telefonica
Projectnaam	Hearing @ Home
Type Contract	EU-IST 1-12-06 tot 30-10-09
Thema	Zorg
Omschrijving	TV settopbox voor communicatie en domotica t.b.v. slechthorenden □HaH ontwikkelt een settopbox voorzien van de volgende technologie: □-telefonieondersteuning via Synface-talking head (lipleesondersteuning)□-in-house

	alarmering □-audioverbetering (voor individu en groep)
--	--

Partners	ViaTaal, LOEM, Sporthochschule Keulen, IfAP, Bronnikov Algorithms, TU Delft
Projectnaam	Automatisch spraakafzien
Thema	Zorg: Audio Visuele Spraakherkenning
Omschrijving	Ontwikkeling Geautomatiseerd Spraakherkenningssysteem via o.a. inzet hogesnelheidsopnames en neurofysische modellering en videosegment co-articulatie

Partners	ViaTaal, Radboud Universiteit, TU Delft
Projectnaam	Linguabytes
Type Contract	Onderzoekscontract Dr. Phelps-stichting en zorgfondsen
Thema	Zorg: Speelleeromgeving voor kinderen met sensomotorische problematiek
Omschrijving	Het speel-leerhulpmiddel bevat verhalen, die aansluiten bij de belevingswereld van de peuter. Begrippen uit deze verhalen worden vervolgens uitgediept en verbreed en systematisch in oefeningen aan het kind aangeboden. Deze begrippen kunnen worden ondersteund met verschillende symboolsystemen voor ondersteunde communicatie, zoals Picture Communication Symbols (PCS), Bliss en Bèta-symboolsystemen. In het speel-leerhulpmiddel worden beeld, bediening en computer geïntegreerd, zodat het aansluit bij de leef- en belevingswereld van kinderen (meer speelgoed dan computer) en bij de sensomotorische vaardigheden van de peuter.

Partners	ViaTaal, Nederlands Gebarencentrum, Handicom
Projectnaam	Lees & Teken
Thema	Zorg: Uitontwikkeling SignPS-programma
Omschrijving	Het uitontwikkelen van het programma SignPS. Het maken van tekeningen van NGT-gebaren met het programma SignPS.

Partners	ViaTaal, KPN, Tolknet
Projectnaam	Sign Interpretation
Thema	Zorg: Animated Agents
Omschrijving	Ontwikkeling van gebarende avatar als o.a. gebarentolk in o.a. mobiele toepassingen

Partners	ViaTaal, AnnieS, Dovenschap, KPN
Projectnaam	SMS+
Thema	Zorg:

	Interactive texting
Omschrijving	De mobiele teksttelefonie geeft op basis van de moderne telecomprotocollen de mogelijkheid van character-by-character texting. In een proef met doven en slechthorenden is gebleken dat dit de communicatie sterk bevordert t.o.v. chatten by MSN of SMS. Onderzocht wordt of deze nieuwe vorm van communicatie ook voor niet slechthorenden mogelijkheden biedt.

Partners	Viataal, EAC, KEGG, speciaal onderwijs
Projectnaam	Laccess
Thema	Zorg: Lexicale Database met multimedia toegang voor mensen met een handicap
Omschrijving	Een leersysteem dat kinderen die moeite hebben met het aanleren van taal en geletterheid ondersteund bij het vergaren van kennis over de betekenis van woorden. Hierbij worden o.a. visueel-grafische representaties van taal gebruikt en gesynthetiseerde spraak.

Partners	ViaTaal, Handicom, Nederlandse Gebarencentrum
Projectnaam	Linguasign
Thema	Zorg: Leren van taal via een gebarende avatar
Omschrijving	Kinderen leren sneller een tweede taal door de ondersteuning van een gebarende avatar. Een project voor normaal horende kinderen die een tweede taal leren.
Type Contract	Opdracht verschillende goede doelen fondsen

e-Government

Partners	Noterik
Projectnaam	BestuurOnline
Thema	e-Government
Omschrijving	<p>De opkomst van Breedband Internet heeft vele nieuwe vormen voor het uitzenden van beeld en geluid op Internet mogelijk gemaakt.</p> <p>BestuurOnline richt zich op het online uitzenden van politieke vergaderingen zoals Provinciale Staten en Gemeenteraadsvergaderingen. BestuurOnline maakt het mogelijk tegemoet te komen aan de wet 'Openbaarheid van Bestuur'. Daarnaast wordt BestuurOnline ook ingezet op momenten dat een gemeente of provincie aandacht wil besteden aan specifieke thema's. Een voorbeeld hiervan is de installatie van een nieuwe burgemeester of een politiek vragenuurtje met het College van B&W. De toeschouwers kunnen via een interactieve chat vragen stellen aan het College van B&W en op deze manier betrokken worden bij het politieke besluitvormingsproces. In samenwerking met de regionale omroepen kan het beeld en geluid ook via radio en televisie worden uitgezonden. Intelligente indexatie van de synchroon toegevoegde metadata, maakt het mogelijk om in de uitzendingen te zoeken. In de nieuwe versie (2008) zal automatische spraakherkenning gebruikt worden om ook de daadwerkelijk gesproken tekst te indexeren.</p>

Partners	ILLC (UvA), Samenwerking met Dublin City University (DCU), Dublin, Ierland
Projectnaam	PRIORS en ATTEMPT
Thema	e-Government
Omschrijving	<p>De noodzaak voor het vertalen van teksten tussen paren van talen behoeft geen introductie.</p> <p>In dit projekt ontwikkelen wij in samenwerking met DCU (projekten PROSPECT en ATTEMPT – Andy Way) verschillende vertaalsystemen die gebruikmaken van statistiek over meer dan platte sequenties van woorden, bijvoorbeeld, syntactische structuur. Deze systemen worden zodanig gebouwd dat de snelheid en accuraatheid van het vertalen bruikbaar zou kunnen zijn voor echte applicaties. Hoewel het werk binnen het projekt is niet beperkt tot bepaalde talen, hebben wij tot op dit moment werkende vertaalsystemen voor Arabisch->Engels en Duits->Engels.</p>

Partners	ILLC (UvA)
Projectnaam	U-DOT, Automatisch Vertalen van Europese Parlementaire Vergaderingen
Thema	e-Government
Omschrijving	<p>De vergaderingen van het Europese Parlement worden vertaald in alle 23 officiële talen van de unie. Voor zowel burgers als politici is het van belang dat parlementstukken in de eigen taal kunnen worden gelezen. Dit kostbare en tijdrovende vertaalwerk (door 750 vertalers) kan voor een aanzienlijk deel worden geautomatiseerd. Wij ontwikkelen een systeem (U-DOT, "Unsupervised Data-</p>

	<p>Oriented Translation”) dat aan de hand van statistische en grammaticale generalisaties over grote hoeveelheden voorbeeldvertalingen van eerdere parlementsvergaderingen, automatisch leert vertalen tussen 11 talen van de EU. Het systeem genereert halffabricaten die zoveel mogelijk lijken op door mensen gemaakte vertalingen. Deze halffabricaten kunnen vervolgens door professionele redacteur-vertalers worden uitgewerkt tot definitieve vertalingen. Op deze manier kan een grote tijdsbesparing tot stand worden gebracht, al dienen de redacteur-vertalers eindverantwoordelijk te blijven voor de tekst.</p>
--	---

Partners	Telecats, Carp Technologies
Projectnaam	RechtSprakHerkenning
Thema	e-Government
Omschrijving	<p>Rechtbanken moeten steeds vaker de geluidsopnamen in de rechtszaal volledig uit schrijven hetgeen tot een forse vermeerdering van de werklast leidt. Bovendien is het zo dat huidige opnamen niet geïndexeerd zijn waardoor ze niet doorzoekbaar zijn. Automatische spraakherkenning lijkt hier de oplossing, maar helaas is 100% correcte herkenning nog niet mogelijk. Wel is voor de vaste sprekers met de nodige tuning 90% haalbaar. Omdat de opnamen toch dikwijls volledig worden uitgeschreven is het transcript meestal beschikbaar. Oplijning van het transcript met de geluidsopnamen zorgt ervoor dat van ieder woord precies bekend is wanneer het werd uitgesproken. Daardoor kan in de tekst gezocht worden en kunnen de gevonden fragmenten met een muisklik beluisterd worden. Om het zoeken te verbeteren wordt taaltechnologie ingezet zodat foutgetypte zoekwoorden(feits ipv fiets) of “verkeerd” gekozen woorden (fiets ipv rijwiel) toch het gewenste zoekresultaat kunnen geven. Ook zal een gesproken samenvatting gemaakt worden waarmee een luisteraar die alleen de meest relevante audiofragmenten beluisterd snel een idee van het geheel kan krijgen</p>
Status	Dit project is in 2007 gestart en zal voorjaar 2008 worden beëindigd. Afhankelijk van een aantal factoren (resultaat, prijs) zal er mee verder gewerkt gaan worden.

Partners	Irion, MOOI Informatiebeheer en de Gelderse Milieufederatie
Projectnaam	Wikihier
Thema	e-Government
Omschrijving	<p>In Nederland bestaan vele organisaties die zich bezighouden met het voorlichten van burgers inzake milieu: Ministerie van VROM, de provinciale milieufederaties, allerlei gesubsidieerde milieuorganisaties (zoals Milieudefensie) en vele stichtingen en verenigingen (bijvoorbeeld de Vogelvereniging, IVN, WNF, etc.). Het grote probleem voor de burger is daarbij, dat deze niet precies weet met welke vraag hij naar welke organisatie zou moeten gaan, een situatie die nog wordt bemoeilijkt doordat veel vragen door verschillende organisaties oanders worden beantwoord. Een centrale wiki, één ingangspunt voor de burger, moet daaraan een einde maken, en Wikihier heeft tot doel dit te realiseren. Dit wordt momenteel gerealiseerd door middel van een combinatie van een partieel gesloten wiki, waar experts op het gebied van natuur en milieu in wiki-vorm hun informatie kunnen bijhouden en unificeren, en een op taaltechnologie gebaseerd dialoogsysteem, waarmee zowel burgers als experts informatie kunnen opvragen.</p>

Partners	Carp Technologies
Projectnaam	Samenvatten Kamerstukken Tweede Kamer
Thema	e-Government
Omschrijving	De informatiedienst van de Tweede Kamer maakt van alle kamerstukken en andere documenten samenvattingen voor de bewindspersonen. Dit tijdrovende karwei ondersteunen wij met een geautomatiseerd samenvat-systeem dat een halffabricaat samenvatting produceert dat vervolgens door een professionele redacteur wordt uitgewerkt tot definitieve samenvatting. Bij het ontwikkelen van dit systeem is gebruik gemaakt van een grote hoeveelheid voorbeeldsamenvattingen, zodat de halffabricaten zoveel mogelijk overeenkomen met de handgemaakte samenvattingen, waardoor er een aanzienlijke tijdsbesparing kan worden gerealiseerd. Ook hier geldt dat het belangrijk is dat een redacteur verantwoordelijk is voor de inhoud van de samenvatting. Immers, bewindspersonen moeten er op kunnen vertrouwen dat de samenvatting alle relevante informatie bevat.

Partners	Dutchear - IRION - Gemeente Gilze en Rijen
Projectnaam	GemeenteConnect
Thema	e-Government
Omschrijving	Gemeentes willen burgers steeds beter informeren. Echter hebben (vooral kleine) gemeentes niet altijd de capaciteit om telefonisch alle vragen te beantwoorden. Daarom is er een systeem ontwikkeld dat kennis bevat over alle standaard vragen en producten van een gemeente. Burgers die de gemeente bellen met een vraag worden zo goed mogelijk door het systeem geholpen om antwoord te vinden op hun ingesproken vraag. Als het alsnog een te specialistische/complex vraag betreft kan worden doorverbonden met een medewerker van de gemeente. Een bijkomend voordeel van een dergelijk systeem is, dat het ook werkt buiten de openingstijden van de gemeente (denk aan 's avonds nog een vraag stellen over een nieuw rijbewijs).

Partners	Carp Technologies
Projectnaam	Automatisch anonimiseren rechterlijke uitspraken
Thema	e-Government
Omschrijving	Om jurisprudentie toegankelijk te maken voor advocaten is het van belang dat in Nederland uitspraken van rechters openbaar worden gemaakt. Voordat ze worden gepubliceerd moeten ze wel eerst worden geanonimiseerd. Daarbij worden alle privacygevoelige passages vervangen door een geschikte vervangingsterm. Denk dan aan bijvoorbeeld "slachtoffer", of "kentekennummer". Omdat dit een zeer arbeidsintensief proces is, was het onmogelijk om alle uitspraken te publiceren. In opdracht van de Raad voor de Rechtspraak hebben wij een systeem ontwikkeld voor strafrecht, dat automatisch arresten anonimiseert en gebruikt wordt als hulpmiddel. De anonimiseer-applicatie geeft automatisch met kleurmarkeringen aan welke privacygevoelige elementen de uitspraak bevat en de kleur geeft de rol aan. Een redacteur hoeft de anonimisering alleen nog maar te controleren en kan eventueel nog wijzigingen aanbrengen. Deze controlestep is belangrijk, want het anonimiseren van een uitspraak is een

	gevoelige kwestie, en de verantwoordelijkheid daarvoor kan je niet bij een machine leggen. Met dit systeem is de tijd die is gemoeid is met het anonimiseren met een factor drie te verminderen.
--	--

Partners	Carp Technologies
Projectnaam	Effectief zoeken in gemeentelijke informatie
Thema	e-Government
Omschrijving	<p>Veel gemeentes bieden tegenwoordig een Digitaal Loket aan, waar burgers informatie kunnen vinden over allerlei gemeentelijke producten en diensten. Maar dikwijls is het niet eenvoudig voor de burger om de juiste informatie te vinden, omdat er een taalbarrière is tussen het ambtelijk taalgebruik in de informatiepagina's en de taal van de gewone burger. Om deze barrière te overbruggen hebben wij een geavanceerd zoekstelsel ontwikkeld dat onder andere bij de gemeente Enschede wordt toegepast. Dit zoekstelsel overbrugt de taalbarrière door de burger in staat te stellen zijn of haar zoekvraag te formuleren in een gewone Nederlandse zin. Het maakt daarbij niet uit welke woorden worden gebruikt; het zoekstelsel maakt een betekenisanalyse van de zin en zoekt automatisch het meest relevante informatieproduct. Zo zal een zoekvraag als "ik heb ruzie met de buurman over de schutting" het informatieproduct "geschillen bij erfbegrenzing" opleveren. De zoekmachine is slim en weet dat een ruzie hetzelfde is als een geschil, en dat schutting een ander woord is voor erfbegrenzing. Uiteraard is de zoekmachine ook uiterst vergevingsgezind bij typefouten in allerlei moeilijke termen. Uit onafhankelijk onderzoek van TNS-NIPO blijkt dat maar liefst 87% van de websitebezoekers uiterst tevreden is over de zoekfunctie en aangeeft de juiste informatie zeer gemakkelijk te hebben gevonden.</p>

Partners	Carp Technologies
Projectnaam	Trefwoordtoekenning Kamervragen en moties Tweede Kamer
Thema	e-Government
Omschrijving	<p>Om het terugvinden van Kamervragen en moties te faciliteren worden deze door de informatiedienst van de Tweede Kamer voorzien van een aantal trefwoorden. Deze worden gekozen uit een lange lijst van enkele duizenden termen. Het labelen van alle Kamervragen en moties kost veel tijd en uitzoekwerk, omdat handmatig in de lijst van beschikbare trefwoorden een passende term moet worden gezocht. Om dit proces te ondersteunen hebben wij een trefwoordtoekenningsstelsel ontwikkeld specifiek voor deze informatie van de Tweede Kamer. Ons toekenningsstelsel geeft de redacteur voor elke Kamervraag en motie suggesties voor passende trefwoorden. De redacteur beslist vervolgens welke er definitief worden toegekend.</p>

Partners	Polderland Language & Speech Technology bv, Sabel Communicatie
Projectnaam	Elektronische Eindredacteur
Thema	e-Government
Omschrijving	Voor de overheid is het in toenemende mate belangrijk om klantgericht te zijn.

Klantgerichtheid bereik je ondermeer door heldere communicatie met de klant. Communicatie vanuit de overheid is veelal schriftelijk. Denk aan brieven, folders, rapporten, websites, etc. Polderland ontwikkelt schrijfhulpmiddelen op maat. Een voorbeeld daarvan is de Elektronische Eindredacteur. De Elektronische Eindredacteur controleert teksten aan de hand van een beschreven teksthuisstijl en doet suggesties voor verbetering. De Elektronisch Eindredacteur controleert bijvoorbeeld of de tekst lezergericht is (korte, actieve zinnen, weinig bijzinnen, geen ambtelijk taalgebruik, geen afkortingen, etc.) en komt met een uitgebreide spellingcontrole waaraan ook bijvoorbeeld namen van personen, afdelingen en organisaties zijn toegevoegd. Als uitgangspunt wordt de standaard spellingcontrole in Microsoft® Office genomen, ontwikkeld door Polderland Language & Speech Technology. Een eerste versie van de Elektronische Eindredacteur is ontwikkeld voor de Provincie Noord-Brabant in samenwerking met Sabel Communicatie.

Partners	Dutchear - IRION - Mooi - Gemeente Nijmegen
Projectnaam	Gemeentelijke Milieu Telefoon
Thema	e-Government
Omschrijving	Sinds het verdrag van Arhus moet er veel meer milieu informatie openbaar zijn. Daarom is er een systeem ontwikkeld dat kennis bevat over alle standaard milieu vragen van een gemeente/provincie. Bellers die bellen met een milieu vraag worden zo goed mogelijk door het systeem geholpen om antwoord te vinden op hun ingesproken vraag. Als het alsnog een te specialistische/complex vraag betreft, kan worden doorverwezen naar meer informatie op internet. Een bijkomend voordeel van een dergelijk systeem is, dat het ook werkt buiten de openingstijden van de gemeente (denk aan 's avonds nog een vraag stellen over een sloopafval). Het systeem dekt een ruime variatie van milieu informatie af (oppervlakte water, bodemgebruik, vergunningen, dieren, lucht, etc.).

Partners	Carp Technologies
Projectnaam	Samenvatten rechterlijke uitspraken
Thema	e-Government
Omschrijving	Om het voor advocaten makkelijker te maken om in jurisprudentie te zoeken maken uitgeverijen van juridische informatie zoals bijvoorbeeld Kluwer samenvattingen van rechterlijke uitspraken. Zo kan de advocaat heel snel de relevante informatie tot zich nemen, zonder dat daarvoor het hele vonnis hoeft te worden gelezen. Om het maken van deze samenvattingen te ondersteunen hebben wij een speciale versie van onze tekstsamenvatter ontwikkeld die volautomatisch samenvattingen genereert van vonnissen. Daarbij wordt de essentie uit het vonnis gepikt, waarbij met name de overwegingen van de rechter van belang zijn. Bovendien is deze samenvatter in staat om zinnen in te korten en samen te voegen door ze te parafaseren. Dat resulteert in een nog beknoptere samenvatting.

Partners	Irion, MOOI Informatiebeheer en Stichting Rioned
Projectnaam	Rioned

Thema	e-Government
Omschrijving	Stichting Rioned speelt een sleutelrol op het gebied van het verzamelen en verspreiden en van informatie over het Nederlandse rioleringensysteem. De informatie is vaak specialistisch van aard, en lastig te vinden met gewone trefwoorden. Door het op taaltechnologie van Irion gebaseerde zoekstelsel te gebruiken wordt de vindbaarheid van deze informatie sterk vergroot.

Partners	Irion, Bureau Taal en Tremani
Projectnaam	Texamen
Thema	e-Government
Omschrijving	Bureau Taal houdt zich bezig met het ontwikkelen en toepassen van methodes om taal begrijpelijker te maken, in het bijzonder de taal van ambtenaren voor burgers. Een recent product van Bureau Taal is de veelgeprezen Grondwet in eenvoudig Nederlands. Om te bepalen hoe moeilijk of makkelijk leesbaar een tekst is, heeft Bureau Taal een speciale standaard ontwikkeld, genaamd Texamen. Met deze standaard kan voor iedere tekst worden bepaald voor welke doelgroepen zij leesbaar is. De meetmethode was vereiste tot voor kort volledig handmatige analyse van teksten, maar Irion heeft in samenwerking met Bureau Taal een stelsel ontwikkeld, dat geheel automatisch de leesbaarheid van een tekst kan bepalen. Met deze automatische methode is het voor overheden heel eenvoudig geworden om te bepalen of een bepaalde tekst wel of niet geschikt is voor burgers.

Cultureel erfgoed

Partners	Carp Technologies
Projectnaam	Zoeken in historisch krantenmateriaal
Thema	Cultureel Erfgoed
Omschrijving	<p>De Koninklijke Bibliotheek is bezig met het digitaliseren van 10 miljoen krantenpagina's die teruggaan tot enkele eeuwen. De bedoeling is om dit historisch krantenmateriaal op artikelniveau te ontsluiten. De moeilijkheid hierbij is dat de Nederlandse taal in de loop der jaren aanzienlijk is veranderd. De zoekmachine zal een vertaalslag maken tussen het hedendaags Nederlands in de zoekvraag, en alle varianten van oud Nederlands daarop in de krantenartikelen. Om de zoekfunctie nog verder te assisteren zal de zoekmachine ook een samenvat-functie bevatten die automatisch samenvattingen genereert van de gevonden artikelen, zodat de gebruiker snel kan zien in hoeverre een artikel relevant is.</p>

Partners	Noterik
Project	Levende Herinneringen
Thema	Cultureel Erfgoed
Omschrijving	<p>Levende Herinneringen is een cultureel erfgoed project dat als doel heeft het behoud en de overdracht van de Indische cultuur. Binnen het project worden via video diepte-interviews actief getuigenissen verzameld over diverse belevingen van Indische identiteit. Ruim 900 oudere Indische personen zullen worden geïnterviewd. Hiermee zullen hun waardevolle herinneringen niet verloren gaan voor toekomstige generaties. Jongere Indische mensen kunnen zelfstandig hun eigen getuigenissen aan het archief toevoegen. Het archief dat hiermee tot stand komt, is voor iedereen ten allen tijden openbaar toegankelijk via Internet. Automatische spraakherkenning gaat gebruikt worden om de daadwerkelijk gesproken spraak te indexeren zodat ook in de interviews gezocht zal kunnen worden. Geïndexeerde "name entities" (plaatsnamen) worden o.a. gebruikt om automatisch links aan te maken met bv GoogleEarth.</p>

Partners	HMI-UTwente, NIOD, Beeld&Geluid
Projectnaam	Radio Oranje
Thema	Cultureel erfgoed
Omschrijving	<p>De radio-opnamen die Koningin Wilhelmina gedurende oorlog in Londen voor Radio Oranje maakte, zijn grotendeels (in wisselende kwaliteit) bewaard gebleven. Ook de letterlijke tekst (uitgeschreven gedurende de oorlog) is bewaard. Deze twee bronnen werden gebruikt om op de, aan Koningin Wilhelmina gewijde website van het NIOD, een zoekapplicatie te maken waarmee in de toespraken van Koningin Wilhelmina gezocht kan worden. Het resultaat is het originele geluidsfragment, de tekst als ondertiteling waarbij het uitgesproken woord onderstreept wordt en een, op basis van de uitgesproken/uitgeschreven tekst, relevante foto uit het NIOD-fotoarchief getoond wordt. De combinatie spraak en tekst maakt het mogelijk om ook passages met een zeer slechte kwaliteit "te verstaan": iets dat anders zeer lastig zou</p>

zijn. □ <http://niod.al-m.nl/nl/thema/10/>

Partners	ILK (Uv Tilburg), Naturalis
Projectnaam	MITCH
Thema	Cultureel erfgoed
Omschrijving	In het onderzoekslab van Naturalis werkt een team van Tilburgse onderzoekers aan wat wel de verborgen schat van Naturalis wordt genoemd: de vele honderden veldboeken en registerboeken die in detail beschrijven waar, door wie en in welke omstandigheden de miljoenen opgeslagen dieren en planten van het museum zijn gevonden. Zonder deze boeken is serieus onderzoek naar de collectie onmogelijk. Werken met de boeken zelf is berucht tijdrovend. Pas met gedigitaliseerde versies van boeken wordt het mogelijk om in korte tijd zittend achter de computer complexe onderzoeksvragen te beantwoorden, zoals: "wat was de verspreiding van deze specifieke familie van gifkickers in de Amazone, van 1800 tot nu, gemeten in periodes van 25 jaar?"

Partners	RU Nijmegen en Polderland.
Projectnaam	D-kwadraat
Type Contract	Digitaliseringsfonds NWO.
Thema	Cultureel erfgoed
Omschrijving	Het project "Digitale databanken en digitaal gereedschap voor de woordenboeken van de Brabantse en de Limburgse dialecten" (D-kwadraat) heeft tot doel de grote hoeveelheden dialectgegevens (ongeveer 3 miljoen records) zoals gepubliceerd in de vele bundels van twee Nederlandse dialectwoordenboeken te digitaliseren. Het gaat om het Woordenboek van de Brabantse Dialecten (WBD) en het Woordenboek van de Limburgse Dialecten (WLD). De momenteel in de dialectgegevens aanwezige taalkundige niveaus zijn het fonetisch-fonologische niveau, de lexicale vorm en de lexicale betekenis. □ Naast het digitaal toegankelijk maken van de dialectgegevens, wil het project D-kwadraat extra morfologische informatie toevoegen aan de bestaande gegevens en digitale onderzoeksgereedschappen ontwikkelen, waaronder eigentijdse cartografische gereedschappen.

Partners	Polderland Language & Speech Technology bv
Projectnaam	Spellingcontrole voor kleine talen en dialecten
Thema	Cultureel Erfgoed
Omschrijving	"Iedere serieuze taal heeft een spellingcontrole!" Spellingcontroleprogramma's en elektronische woordenboeken dragen bij aan behoud van talen en dialecten doordat het sprekers in staat stelt zich niet alleen mondeling in die taal goed uit te drukken, maar ook schriftelijk. Polderland ontwikkelde al vele spellingcontroles (en elektronische woordenboeken) voor kleinere talen en dialecten waaronder het Fries, Samisch, IJslands, Macedonisch, Bosnisch, Drents, etc. Maar ook voor bijvoorbeeld voor de Hebreeuwse en Jiddische woorden in het Nederlands.

(Serious) gaming en vrijetijdsbesteding

Partners	Ubisoft, ILK (UvTilburg), CNTS (U Antwerpen)
Projectnaam	My Word Coach
Thema	Gaming
Omschrijving	Nederlandse versie van spel voor Nintendo DS en Nintendo Wii. My Word Coach biedt een verzameling mini-games om spelenderwijs de woordenschat en woordgebruik van de speler te prikkelen en te stimuleren. My Word Coach biedt een scala aan woorden die variëren van zeer bekend tot uiterst zeldzaam. Het spel volgt de vorderingen van de speler op de voet; met het behalen van prestatieniveaus stijgt de moeilijkheidsgraad van alle mini-games.

Partners	CLST (RU Nijmegen)
Projectnaam	Dutch-Capt
Thema	Onderwijs, Gaming
Omschrijving	Systeem met automatische spraakherkenning ten behoeve van feedback CALL systemen (uitspraaktraining Nederlands als tweede taal)

Partners	RU is penvoerder. Andere partners: U Antwerpen, Polderland, UTN
Projectnaam	DISCO
Thema	Onderwijs, Gaming
Omschrijving	DISCO aims to develop and test a prototype of an ASR-based CALL application for training oral proficiency for Dutch as a second language (DL2). The application optimizes learning through interaction in realistic communication situations and provides intelligent feedback on various aspects of DL2 speaking, viz. Pronunciation, morphology and syntax.

Partners	CLST (penvoerder) Univ. Gent, Univ. Utrecht, Nuance, TeleAtlas
Projectnaam	AUTONOMATA TOO
Thema	Vrijetijdsbesteding, Veiligheid
Omschrijving	The aim of this application-oriented research project is to build a demonstrator version of a Dutch/Flemish Points of Interest (POI) information providing business service, and to investigate new pronunciation modelling technologies that can help to bring the spoken name recognition component of such a service to the required level of accuracy. The demonstrator service (running on a PC) will contain a simple user interface and a restricted but realistic database of POI information.

Naam bedrijf	Dutchear
Titel project	taxitaxi

Thema	Vrijetijd/veiligheid
Omschrijving	Als je 's avonds uit een café komt weet je misschien niet precies hoe je de plaatselijke taxicentrale snel kunt bereiken. 0900-TAXITAXI lost dit op met een makkelijk te onthouden nummer en verbind je (na het inspreken van de plaats waar je bent) door met een lokaal taxibedrijf. Dit project is vereerd met een nominatie voor de taxi innovatie prijs van de taxi branche.

Veiligheid

Partners	Dutchear - Politie Utrecht
Projectnaam	Nummerbord Retrieval Tool (Kentekenlijn)
Thema	Veiligheid
Omschrijving	Politieagenten hebben in allerlei situaties de behoefte om informatie over een auto en de eigenaar van deze auto op te vragen. Voorbeelden hiervan zijn: het staande houden van een auto na een verkeersovertreding, het 'verdacht' gedragen van een auto in het verkeer of op een parkeerplek maar ook in situaties waarbij de agenten surveilleren. Snelheid bij het opvragen van deze informatie is van groot belang: het heeft niet veel zin de informatie dat de eigenaar van een auto vuurwapen gevaarlijk is pas te hebben als deze allang gearresteerd is. Deze informatie kon de Politie tot voorkort alleen opvragen bij de 112 meldkamer. Deze heeft echter ook andere, vaak belangrijkere taken, waardoor deze snelheid in het gevaar komt en agenten vaak een aantal minuten moeten wachten op de informatie. Met de nieuwe Kentekenlijn hoeft dit niet meer. Een agent belt het automatische systeem, spreekt het kenteken van de auto in en hoort direct alle gegevens van de auto en de bestuurder. Dit leidt tot een veiligere situatie voor agenten, en een betere controle van de openbare weg en tot het uitschrijven van meer bekeuringen voor bijvoorbeeld het verlopen van een APK.

Partners	ParaBotS
Projectnaam	XBotS
Thema	Veiligheid
Omschrijving	XBotS is een programma voor het grootschalig verzamelen en analyseren van data op het internet. De gebruiker geeft een paar voorbeelden van pagina's waar men in geïnteresseerd is en welk type informatie uit de verzamelde sites moet worden gerapporteerd. Uit de voorbeelden worden tekst-classifiers afgeleid, die worden gebruikt om het zoekproces te richten en om aan te geven in hoeverre verzamelde pagina's lijken op de voorbeeldconcepten van de gebruiker. Naast tekst-classificatie gebruikt XBotS ook informatie-extractietechnieken om bijvoorbeeld adressen en contactinformatie te verzamelen. □De XBotS technologie wordt momenteel gebruikt door overheidsdiensten in 7 verschillende landen. (Zie http://www.parabots.nl)

Partners	Polderland Language & Speech Technology bv
Projectnaam	text/topic extraction
Thema	Veiligheid, e-Government
Omschrijving	Taaltechnologische tools zoals taalidentificatie, lemmatisering, woordklasse tagging, named entity recognition, spelling tolerant search, etc. worden alom gebruikt documenten die "verdacht" zijn te extraheren uit stromen elektronische documenten. Het kan daarbij gaan om filters voor SPAM, maar ook om het herkennen van terroristische boodschappen. Voor dat laatste is vooral kennis van andere/meerdere talen van belang. Polderland levert internationaal taaltechnologische producten voor vele talen.

Media Mining

Partners	CLST, ILK Tilburg, Search Expertise Centre
Projectnaam	À Propos (Pro-active personalization for professional document writing)
Type Contract	IOP-MMI project, 1 jan. 2005 - 1 juli 2008
Thema	Media mining
Omschrijving	À Propos stelt zich tot doel om een pro-actief systeem te ontwikkelen en te testen voor het ondersteunen van professionals die als lid van een werkgroep documenten moeten schrijven door relevante informatie te zoeken en die 'just-in-time' aan te bieden.

Partners	ParaBotS
Projectnaam	Vox Populi
Thema	Media mining
Omschrijving	Zinnen dragen emoties over. Iedere dag worden er miljoenen zinnen geschreven op het internet. Mensen hebben het over elkaar, over producten die ze gebruiken, hoe ze denken over bepaalde bedrijven, politici, politieke partijen, BN'ers, enzovoorts. Elke paar uur bekijkt de Vox Populi duizenden pagina's waarop nieuwsberichten en meningen daarover staan. Door gebruik te maken van software die natuurlijke taal analyseert kan de Vox Populi bepalen wat er gezegd wordt in een zin, en met name of dit positief of negatief is. Door naar alle zinnen te kijken op de gedownloade pagina's kan op deze manier de publieke opinie worden afgeleid. □Zie: http://www.vox-pop.nl , http://www.parabots.nl

Partners	Irion & Reed Business International
Projectnaam	Hoffrog
Thema	Media Mining
Omschrijving	Bedrijven zoeken via het internet andere bedrijven, als leverancier of klant. Een zeer lastig probleem daarbij is het juiste koppelen ("matchen") van bedrijven die iets voor elkaar kunnen betekenen: standaard zoekmachines zijn niet in staat om de juiste content bij elkaar te verzamelen en te koppelen aan een informatiebehoefte. Reed Business International heeft vanuit Australië een wereldwijd systeem opgezet waarin bedrijven zich kunnen profileren en zich beschikbaar kunnen stellen voor koppeling aan andere bedrijven. Een cruciaal probleem daarbij is, dat de meeste bedrijven dermate specialistisch zijn, dat het matchen alsnog zeer moeilijk is, en zonder hulpmiddelen niet goed kan plaatsvinden. Irion zet haar speciale mining technologie in om voor ieder bedrijf geheel automatisch een nauwkeurig en universeel te matchen profiel te genereren op basis van een inhoudelijke taalkundige analyse van de website. Daarbij wordt onder andere gebruik gemaakt van een "folksonomy": een taxonomie, die door gebruikers interactief wordt opgebouwd. Hoffrog draait momenteel in Australië, USA, Engeland, Duitsland en Frankrijk, en Irion heeft voor al deze landen de mining gebouwd en ingeregeld.

Partners	Irion, Carp en MDInfo
-----------------	-----------------------

Projectnaam	EasyInfo
Thema	Media Mining
Omschrijving	Nu de pers zich steeds meer via het internet manifesteert, groeit ook het aantal zogenaamde "digitale knipseldiensten": services die nieuwsberichten via het internet aanbieden aan lezers. Een hoogwaardige nieuwsvoorziening vereist een nauwkeurige classificatie en samenvatting van de content. In het project EasyInfo wordt een nieuwe knipseldienst opgezet, waarin zowel hoogwaardige classificatie van content, als automatische samenvattingengeneratie wordt ingebouwd.