

## **Taal- en spraaktechnologie voor het Nederlands: Toptechnologie voor de topsectoren**

Taal- en spraaktechnologie (TST) maakt het mogelijk om natuurlijke taal (geschreven of gesproken) te gebruiken in communicatie met of via computers of andere apparatuur. Voorbeelden van TST zijn computerondersteund vertalen, automatisch samenvatten, intelligente informatie-extractie, (semantische) zoektechnologie, automatische spraakherkenning en tekst-naar-spraak-omzetting.

TST is een breed toepasbare enabling technology, een creatieve technologie die inzetbaar is in iedere sector waar informatie en kennis in taal gevat is/wordt en ontsloten moet worden. Hierdoor kan efficiënter worden gewerkt, en worden innovatieve manieren van communicatie mogelijk:

- **Zorg:** diagnostiek en therapie op afstand, ondersteuning, langer zelfstandig wonen;
- **Onderwijs:** automatische correctie en feedback, leren op maat, afstandleren en automatische toetsing;
- **Veiligheid:** automatische controle van teksten en spraakopnames ten behoeve van opsporing;
- **Sociale Media:** het automatisch detecteren in sociale media van de perceptie en beleving van producten, diensten, bedrijven en maatschappelijke kwesties (sentiment en opinion mining);
- **Gaming:** TST maakt de interactie realistischer en leidt tot een rijkere leerervaring;
- **Musea en Cultureel erfgoed:** ontsluiting van informatie en kennis uit historische teksten en spraakopnames, en uit metadata die in natuurlijke taal weergegeven zijn;
- **Ondersteuning van bedrijfsprocessen:** intelligente informatieontsluiting, business intelligence, efficiënte klantendiensten, computerondersteund vertalen;
- **Mobiele toepassingen:** mobiele apparaten hebben een klein scherm en toetsenbord waardoor spraak de meest aangewezen interactievorm is (spraakherkenning, spraaksynthese), bijv. 'Voice Search', Siri spraakinteractie op iPhone 4S, enzovoort.

Een sterke TST-infrastructuur versterkt de economische en culturele positie van een taal in de moderne digitale informatie- en communicatiewereld en is daarom belangrijk voor haar toekomst. Tegen deze achtergrond en gedreven door het potentieel van bovenstaande toepassingen, werd op initiatief van de Nederlandse Taalunie in 1999 een Vlaams-Nederlands TST-platform opgericht waarin belangrijke spelers uit Nederland (Ministeries ELI en OCW, NWO, Agentschap NL) en Vlaanderen (Vlaamse overheid - Departement EWI, FWO, IWT) besloten tot samenwerking met als doel:

- Het bevorderen van de samenwerking tussen bedrijfsleven, kennisinstellingen en beleidsmakers, het vergroten van de bekendheid van TST en het stimuleren van het gebruik van TST-resultaten door marktpartijen;
- Het inventariseren van behoeftes aan basisvoorzieningen om de digitale taalinfrastructuur voor het Nederlands te versterken;
- Het opstellen van een plan voor beheer, onderhoud en distributie van digitale basistaalmaterialen ontwikkeld met overheidsgeld.

In 2004 vonden de nieuwe plannen voor TST voor het Nederlands hun concrete uitwerking in de **TST-Centrale**, een centraal loket voor TST-materialen voor het Nederlands, en **STEVIN**, een transnationaal programma ter stimulering van TST voor het Nederlands om zo de innovatiecapaciteit van de sector te vergroten. Het STEVIN-programma smeedde de gouden driehoek van beleid, wetenschap en industrie samen via verschillende stimuleringsinstrumenten:

- Subsidies voor samenwerkingsprojecten tussen universiteiten onderling en tussen universiteiten en industrie om een digitale taalinfrastructuur voor het Nederlands te realiseren;
- Subsidies voor strategisch onderzoek op het gebied van taal- en spraaktechnologie;
- Opdrachten aan het MKB om demonstratieprojecten te realiseren met een wervend karakter die het nut en de haalbaarheid van TST-applicaties aantonen;
- Subsidies voor netwerkactiviteiten om kennisoverdracht te bewerkstelligen tussen spelers in het TST-veld;
- Activiteiten ten behoeve van kennisoverdracht, zoals de "Taal in Bedrijf"-dagen.

STEVIN is door een commissie van internationale experts en door de Technopolis Group positief geëvalueerd<sup>1</sup>. Ook de in opdracht van NWO uitgevoerde evaluatie (Broek en Nijssen, 2009) naar de maatschappelijke impact van geesteswetenschappelijk onderzoek toont aan dat STEVIN hoog scoort in termen van disseminatie, betrokkenheid, bruikbaarheid en benutting.

Dankzij de inspanningen op TST-gebied beschikken we nu over:

- Een uitgebreid Vlaams-Nederlands netwerk van kennisinstellingen, bedrijven en beleidsinstellingen die betrokken zijn bij TST en die elkaar kennen en vertrouwen;

---

<sup>1</sup> [http://taalunieversum.org/taal/technologie/stevin/documenten/stevin\\_eindevaluatierapport.pdf](http://taalunieversum.org/taal/technologie/stevin/documenten/stevin_eindevaluatierapport.pdf)

- Een sterkere digitale infrastructuur van data en softwaremodules voor het Nederlands die via een centraal loket voor allerlei toepassingen en gebruikers beschikbaar is;
- Een instrumentarium en architectuur voor het opzetten van grensoverschrijdende onderzoek- en ontwikkelingsprogramma's waar goed is nagedacht over moeilijke kwesties als subsidievoorwaarden, IPR en overdrachtprocedures.

Als een 'topsector in het klein' heeft de TST-sector in Nederland en Vlaanderen zich door samenwerking sterk gemaakt, niet alleen door STEVIN, maar ook door samenwerking op andere vlakken:

- In internationaal onderzoek is taalafhankelijke TST van Nederlandse bodem van oudsher invloedrijk, mede door een vroege aanjagersrol van pioniers als Hugo Brandt Corstius, een sterke traditie in het opleiden van jonge onderzoekers, en een leidende rol in Europese infrastructuren voor onderzoek zoals CLARIN;
- Nationaal onderzoek is sterk op samenwerking gericht; de kennisinstellingen hebben samen een doorlopende reeks van grote nationale programma's geformuleerd (OVIS, IMIX, STEVIN).
- MKB en kennisinstellingen actief op het gebied van TST hebben zich georganiseerd in NOTaS, de Nederlandse Organisatie voor Taal- en Spraaktechnologie: <http://www.notas.nl/>.
- TST maakt zijn rol waar in landelijke onderzoeksprogramma's op het gebied van media (COMMIT), erfgoed (CATCH) en geesteswetenschappelijk onderzoek (KNAW Computational Humanities, CLARIN-NL).

Net als in de Creatieve Industriesector vindt bedrijvigheid op het gebied van TST in Nederland en Vlaanderen hoofdzakelijk in MKB's plaats (met Nuance als uitzonderlijke, grote organisatie). Maar de klanten die gebruik maken van TST van Nederlandse bodem zijn wel vaak grote bedrijven: Rabobank, KPN, Philips, Nokia, en Microsoft. Google gaf onlangs een subsidie aan specifiek Nederlands TST-onderzoek in het "Google Digital Humanities Fund" aan onderzoekers van de Radboud Universiteit Nijmegen.

Pareltjes van recente TST:

- *Ondertiteling met spraakherkenning*  
Bij de VRT wordt spraakherkenning ingezet in de respeaking-methode om programma's van ondertiteling te voorzien met aanzienlijke kostenbesparing.
- *Text analytics bespaart tijd in recruitment business* – Textkernel (TST MKB)  
Door het gebruik van intelligente en taalafhankelijke informatie-extractie, web mining en semantische search kan veel tijd bespaard worden door het automatisch converteren van vacatures en CV's naar databases.
- *Detectie van gebeurtenissen die toch niet plaatsvonden* – HiTiMe (NWO CATCH)  
Op basis van een database van alle stakingen die Nederland ooit gekend heeft en het historisch krantenarchief van de Koninklijke Bibliotheek is een tekstclassificatiesysteem ontwikkeld dat nieuwsartikelen kan vinden die sociale onrust en de aanloop naar stakingen signaleren.
- *De monteur van de toekomst* – SAMM (TST MKB)  
Als winnende inschrijving van de Defensie Innovatie Competitie 2009 heeft het SAMM consortium bestaande uit Tedopres, Dutcheer en BlueTea het concept SAMM (System for Asset Maintenance Management) ontwikkeld: een monteur wordt door TST geholpen bij het uitvoeren van taken. Via spraaksynthese en spraakcommando's wordt de monteur stap voor stap begeleid in het uitvoeren van de handelingen. De tool kan gebruikt worden om een nieuwe machine te leren bedienen.
- *Computer helpt politie bij ontmaskering pedofielen* – DAPHNE (Universiteit Antwerpen)  
Technieken uit de computationele stylometrie helpen om leeftijd, sekse, afkomst, persoonlijkheid, opleidingsniveau en andere psychologische en sociologische eigenschappen van de auteur van een tekst te achterhalen. In het DAPHNE-project worden deze technieken gebruikt voor het signaleren van mogelijk valse profielen in sociale netwerken, een belangrijke stap in de detectie van "grooming" door pedofielen.

Referentie

Broek, S.D. en Nijssen, A.J. (2009) Impact Assessment Geesteswetenschappen, Onderzoek uitgevoerd door Research voor Beleid in opdracht van Gebiedsbestuur Geesteswetenschappen, NWO.