

# NATUURLIJKE TAAL EN TEKSTUELE EXPERTSYSTEMEN

L.A.M. van der Heijden<sup>1</sup>

*Goede agrarische terminologie is noodzakelijk voor intelligente informatieverwerkende systemen in landbouwkundig onderzoek en voorlichting.*

*Binnen het landbouwkundig onderzoek en de landbouwvoorlichting worden steeds meer teksten vastgelegd in een voor de computer leesbare vorm, bijvoorbeeld als bestanden van tekstverwerkers en databases. Dit schept mogelijkheden voor geautomatiseerde voor- of nabewerking van deze teksten met behulp van programma's die gebruikmaken van kennis van agrarische terminologie, zogenaamde tekstuele expertsystemen. Dit artikel poogt een beeld te geven van deze mogelijkheden en aan te zetten tot een gecoördineerde aanpak door landbouwterminologen.*

## Landbouwterminologie

Binnen het landbouwkundig onderzoek zijn diverse commissies en individuen bezig met het opstellen van terminologielijsten voor de verschillende vakgebieden. Deze commissies bestaan meestal uit onderzoekers die in de vakgebieden werkzaam zijn en die trachten met deze lijsten de begrippen in hun vakgebied zo eenduidig mogelijk te definiëren.

In een onderzoek van Van der Burg et al. (1987) is naar voren gekomen dat binnen het Ministerie van Landbouw door een 35-tal groepen gewerkt wordt aan het verzamelen van grote hoeveelheden landbouwtermen (Tabel 1). Daarnaast zijn er nog enkele andere grote institutionele termencollecties: de lijst van trefwoorden van AGRALIN (20.000 Nederlandse termen met UDC-getallen); de Agrovoc-thesaurus van de FAO database Agrindex, bestaande uit 18.000 Engelse trefwoorden; de ongeveer 40.000 Nederlandse termen met equivalenten in Engels, Frans en Duits in EURODICAUTOM, de on-line termenbank van de EG in Luxemburg.

Van enige coördinatie tussen deze terminologische activiteiten is hierbij geen sprake: de doelstelling, de wijze van opslag van de termen (lijsten of bestanden) en de soort informatie die per term wordt opgenomen (definitie, equivalenten in andere talen, tekstvoorbeelden, voorkeursspelling, termrelaties) verschillen sterk van collectie tot collectie. Bij deze individuen of commissies bestaat wel bereidheid om samen te werken, maar het ontbreekt aan een instantie die dit technisch en organisatorisch coördineert.

*Tabel 1. Verdeling naar taal en aantallen termen in de collecties van landbouwtermen binnen de Directie Landbouwkundig Onderzoek*

Taal	Aantal collecties	Aantal termen
Nederlands	32	100365
Engels	18	70365
Duits	14	21185
Deens	2	1200
Frans	13	25085
Spaans	3	600
Andere	8	.

## *Toepassingen van natuurlijke taal in intelligente informatie-verwerkende systemen*

Onder intelligente informatie-verwerkende systemen versta ik in deze context die programma's die intelligente bewerkingen uitvoeren op teksten, een soort tekstuele expertsystemen. Enkele voorbeelden geef ik hieronder, met speciale aandacht voor landbouwteksten.

### *Spellingcorrectie*

Automatische spellingcorrectie, waarbij een programma zelfstandig volgens bepaalde regels spel- en typefouten corrigeert, is nog in ontwikkeling. Dit in tegenstelling tot de interactieve spellingcorrectieprogramma's die nu al bij tekstverwerkers gebruikt worden. Voor beide toepassingen zijn echter uitgebreide woordenlijsten van de vele agrarische vakgebieden een noodzakelijke voorwaarde om deze hulpmiddelen zinvol te kunnen toepassen.

### *Redactionele systemen*

Dit zijn programma's die met behulp van kennis van de grammatica en syntaxis van natuurlijke (wetenschappelijke) taal redactionele bewerkingen uitvoeren op teksten. Hiermee kan in een tekst bijvoorbeeld gezorgd worden voor consistente terminologie, consistent gebruik van SI-eenheden, vermijden van langdradigheid, correctie interpunctie, woordvolgorde en cursiveren van plant- en diernamen. Ook deze programma's zijn nog in ontwikkeling (voor het Nederlands bijv. in het kader van SPIN). Voor PC's zijn er ook enkele programma's voor Engelse teksten (bijv. Grammatik), maar deze maken nog maar summier gebruik van de huidige mogelijkheden van de computerlinguïstiek.

<sup>1</sup> Dr. L.A.M. van der Heijden is hoofd van de sectie Publiciteit, Redactie en Informatie bij het Centrum voor Agrobiologisch Onderzoek (CABO) te Wageningen (tel. 08370-19012; fax 08370-23110; Agronct: CABO::PRI; SURFnet: CABOPRI@HWALHW50).

### *Ruwe vertaalsystemen*

Van al het onderzoek in de gehele wereld vindt slechts 1,5 % (in geld en menskracht) in Nederland plaats. Dit betekent dat wij meer van het buitenland kunnen leren dan het buitenland van ons. De meeste buitenlandse informatie is beschikbaar in de Engelse taal (zo'n 80%), terwijl de rest alleen beschikbaar is in talen, die doorgaans slecht toegankelijk zijn. Goede vertalingen hiervan door menselijke vertalers zijn te kostbaar. Een ruwe vertaling door een computervertaalsysteem is mogelijk en betaalbaar, zeker als de brontekst ook in computerleesbare vorm beschikbaar is. Als voorbeeld kan een computervertaling van een Russische tekst dienen:

#### Origineel (getranslitereerd):

Fiziologitsjeskie osnovy upravleniâ rostom i produktivnost'û rastenij v reguliruemyh usloviah.

#### Ruwe computervertaling:

Physiological principles of control of growth and productivity of plants in controlled conditions.

Als de juiste terminologie aanwezig is, dan leveren dit soort computervertalingen geen enkel probleem op en draagt zij aanzienlijk bij tot de instroom van informatie vanuit het buitenland.

### *Raadplegen van databases via natuurlijke taal*

De landbouwkundige onderzoeker en voorlichter van de toekomst die op zoek is naar bepaalde informatie, zal geconfronteerd worden met een aantal verschillende on-line informatiediensten, elk met hun eigen procedurele zoektaal. Nu zijn deze zoektaalen weliswaar erg efficiënt als men regelmatig in een bepaalde database moet zoeken, maar voor de irreguliere zoeker is het handiger als vragen gesteld kunnen worden zoals:

- Wat is het adres van ir. F.H. Janssen?
- Welke literatuur bestaat er over bereiding van papierpulp uit hennep?
- Hoeveel kg sojaschroot heeft Nederland in 1987 geïmporteerd?
- Wie in Nederland kan mij wat vertellen over melkrobots?
- Hoe zeg ik in het Frans "gemengd bedrijf"?

Deze vragen in "natuurlijke taal" zouden door een interface geanalyseerd moeten worden en daarna omgezet in de betreffende formele zoektaal (eventueel nadat het interface nog wat extra vragen heeft gesteld omdat de menselijke vragensteller niet specifiek genoeg was). Men kan zich zelfs voorstellen dat het interface ook de voor het beantwoorden van de vraag noodzakelijke database(s) selecteert. Ook voor deze toepassing is het nodig dat de agrarische onderzoekwereld terminologisch voldoende juist is beschreven.

### *Semantische filters in E-Mail-systemen*

Wie al wat ervaring heeft opgedaan met E-mail weet dat de hoeveelheid boodschappen die men op die manier kan ontvangen behoorlijk kan oplopen. Om orde in deze informatie-overvloed te scheppen zijn al zogenaamde "semantische filters" in ontwikkeling. Deze selecteren bepaalde boodschappen op hun inhoud en rangschikken ze volgens een bepaald "interesseprofiel" in uw MAIL-BOX.

### *On-line agrarische-termenbank (AGRATERM)*

Woordenboeken zijn onmisbare hulpmiddelen bij schrijven en vertalen van teksten over landbouwkundig onderzoek. Het probleem is alleen dat binnen zich snel ontwikkelende vakgebieden de meeste woordenboeken al weer verouderd zijn op het moment dat ze verschijnen. Een on-line openbaar toegankelijke agrarische-termenbank zou een oplossing kunnen zijn, aangezien die door een netwerk van per vakgebied gegroepeerde terminologen steeds geactualiseerd kan worden.

### **De noodzaak van een netwerk van agrarische terminologen**

Hoe gaan de informatiesystemen van de toekomst er uit zien? Prof. G.M. Nijssen, internationaal expert op het gebied van ontwerp en gebruik van databases, schreef er recentelijk in *Computable* het volgende over: "De volgende generatie informatiesysteemmethodologieën is dan ook gebaseerd op begrippen uit de taal die iedere gebruiker kent. En dat betekent dat de volgende generatie informatiesysteemmethodologieën tevens de laatste generatie is, aangezien de structuren van de natuurlijke taal in de afzienbare toekomst niet zullen veranderen" (Nijssen, 1989). Als wij het binnen landbouw nu maar eens worden over ons taalgebruik, dan zijn wij dus klaar voor de nieuwste methodologieën voor informatiesystemen.

Voor elk van de bovengenoemde toepassingen zijn nu of in de nabije toekomst dus uitgebreide agrarische terminologieën nodig. Het onderzoek van Van der Burg et al. (1987) heeft uitgewezen dat binnen het landbouwkundig onderzoek al veel terminologisch werk wordt verricht. De computernetwerkfaciliteiten binnen landbouw maken het (sinds kort) technisch mogelijk moeiteloos terminologische gegevens uit te wisselen. Het enige wat hiervoor nodig is, is een groepje mensen dat zich buigt over enkele organisatorische en technische problemen (bijv. een standard exchange format voor agrarische termen), er een oplossing voor zoekt en deze oplossing voorlegt aan de gemeenschap van agrarische onderzoekers. Iedereen die hierin geïnteresseerd is, wordt uitgenodigd met mij contact op te nemen.

## Literatuur

Burg, J. van der, L.A.M. van der Heijden & J.C. Rigg, *Agraterm: how can we best exploit terminological resources for agricultural science?* TermNet News, no. 17, p. 39-41., 1987

Nijssen, G.M., "De volgende generatie informatiesysteemontwikkelingsmethoden? Behoeftebepaling vraagt geheel eigen methodiek", *Computable*, 13 oktober, p. 31., 1989 □

## VERSLAG THEMAMIDDAG EXPERTSYSTEMEN

---

Vrijdagmiddag 6 oktober vond de tweede themamiddag expertsystemen plaats, georganiseerd door VIAS i.s.m. TFDL/ECIT van het Ministerie van Landbouw en Visserij. Waren er vorige keer ruim 40 mensen aanwezig, nu was het aantal teruggelopen tot 25. De aanwezigen waren over het algemeen echter allen actief bezig met expertsystemen, terwijl de eerste keer het gezelschap ruimer vertegenwoordigd was door een grote groep mensen die alleen deelnamen ter oriëntatie.

De middag bestond uit twee lezingen. De eerste lezing werd gehouden door J. Backus van het proefstation voor de Champignoncultuur, met als titel: "Interactieve media ter ondersteuning van expertsystemen". In deze lezing ging dhr. Backus in op de wijze waarop een beeldplaat ondersteuning kan bieden bij het vraag en antwoordspel dat kenmerkend is voor veel diagnostische expertsystemen. Dhr. Backus ziet de kracht van deze aanpak met name bij de ondersteuning van de dialoog met een gebruiker. Juist daar waar veel tekstuele omschrijvingen noodzakelijk zijn om vragen correct te kunnen stellen, kunnen beelden een goed alternatief vormen.

Om bovengenoemde aanpak te verwezenlijken is gekozen voor expert system shell die de mogelijkheid biedt om de besturing van de beeldplaat te integreren met de knowledge base. Er is gekozen voor KES, omdat het mogelijk is functies met behulp van C routines te herdefiniëren. Door het invoeren van wat dhr. Backus een "meta-taal" noemde, is het mogelijk om bij het dialoog met de gebruiker deze meta-taal te interpreteren en de gebruiker een vraag met bijbehorende beeld(en) van de beeldplaatspeler te tonen.

In de discussie die op het verhaal volgde, werd ingegaan op de vraag wie die beelden moet samenstellen en/of kiezen. Is het de taak van de knowledge-engineer om deze foto's samen te stellen, of moet de expert deze voordragen. Tevens werd gevraagd of dhr. Backus het fotomateriaal ook bij de interviews met de expert gebruikte. Dit was niet het geval. De reactie hierop was dat juist de expert wellicht m.b.v. deze foto's meer duidelijkheid omtrent zijn beslissingen kon verschaffen. Het gevaar van het te snel grijpen naar beelden als ondersteuning is dat de onderliggende beslissingen wellicht niet allemaal naar boven komen. Nog steeds geldt dat de visuele interpretatie van de expert naar een kwalitatieve of kwantitatieve uitspraak op basis van onderliggende kennis plaats vindt.

Uit de discussie werd in ieder geval duidelijk dat wanneer een kennissysteem met behulp van dit medium wordt gebouwd, er duidelijk goed moet worden overwogen hoe het materiaal wordt samengesteld, wie