

"Voor een informatica-omgeving zijn we niet conventioneel"

Anton Nijholt is hoogleraar bij de vakgroep Software Engineering en Theoretische Informatica van de faculteit Informatica in Twente. Zijn interesse in het bouwen van bruggen en verwante zaken deed hem oorspronkelijk besluiten in Leeuwarden aan de HTS Weg- en Waterbouwkunde te gaan studeren. Vervolgens studeerde hij in Delft eerst Civiele Techniek en vervolgens ook Wiskunde en Informatica. Na zijn promotie in Amsterdam werkte hij onder andere in Twente, Nijmegen en Brussel.



Sinds 1989 is Anton Nijholt hoogleraar bij de vakgroep SETI in Twente. Daar leidt hij het bloeiende Parlevink-project, een koepelproject voor een verscheidenheid aan onderzoek op het gebied van computerlinguïstiek en natuurlijke-taalverwerking. Binnen het kader van dit project worden regelmatig workshops georganiseerd in de reeks Twente Workshops on Language Technology (TWLT).

Daarnaast is Anton Nijholt auteur van een groeiend aantal boeken, met de wens een eigen uitgeverij beginnen. Binnenkort kunnen wij in de boekhandel een nieuw boek van hem verwachten over de geschiedenis van de informatica.

Van Weg- & Waterbouw tot Informatica

"Ik denk niet dat je kunt zeggen dat ik één studie gedaan heb. Ik ben begonnen met HTS, Weg- en waterbouwkunde, in Leeuwarden. Bruggen bouwen en dat soort zaken trokken me erg aan. Nog steeds eigenlijk wel. Na de HTS heb ik eerst een poosje gewerkt bij TNO in Rijswijk, bij het Instituut voor Bouwmaterialen en Bouwconstructies. Ik zat daar bij de

rekenafdeling, die pas bezig was met het gebruik van computers voor allerlei sterkteberekeningen: berekeningen aan bruggen, wanden, vloeren, viaducten en dat soort zaken. Daar kreeg ik voor het eerst te maken met computers en volgde ik allerlei cursussen op het gebied van computers, programmeren en programmeertalen.

"In die tijd was er nog niet zoveel op computergebied. Wat betreft het gebruik van computers liep TNO voorop, denk ik. Ze kregen overal vanuit Nederland om bepaalde rekeningen te maken: van Rijkswaterstaat, van gemeenten, van provincies, etcetera. Ze liepen ook voorop wat betreft de manier waarop ze computers gebruikten; onder andere hadden ze op die afdeling een van de eerste time-sharing faciliteiten en was het mogelijk de computers per modem te gebruiken. Heel fijn dat je opeens gebruik kon maken van een computer die ergens anders stond. Dat was in Nederland toen nieuw. In de Verenigde Staten was het een of twee jaar oud en het werd vrij vlot door andere landen overgenomen. We werden toen een keer geïnterviewd door de Haagse

Courant. Het krantenartikel had als titel "Een kwartje per seconde"; zoveel kostte het rekenen op die manier namelijk.

"In die tijd - 1967, 1968 - heb ik al eens programma's geschreven die gebruik maakten van natuurlijke taal om het wat gemakkelijker te maken input aan de computer aan te bieden. De input bestond typisch uit duizenden getallen die met de hand op formulieren moesten worden ingevuld. Getallen getallen getallen. Ik heb wel eens een keer geprobeerd om daar wat structuur in te brengen, zodat men met woorden kon aangeven welke getallen er kwamen. Het computerprogramma las netjes de woorden en concludeerde: 'nu komt dit, nu komt dat'. Heel simpel allemaal, eigenlijk. Aan de andere kant, het was toch nieuw in die tijd.

"In de periode dat ik bij TNO werkte, realiseerde ik mij dat als je iets zelfstandig wilt doen - dat wil zeggen niet alleen als programmeur bezig wilt zijn - dat je dan ook inhoudelijk verder moet gaan dan enkel die HTS-opleiding. Daarom ben ik vervolgens naar Delft gegaan om daar Civiele Techniek te studeren. Om toch nog meer computers en informatica in mijn opleiding te krijgen ben ik later ook Wiskunde gaan studeren. Een aparte informatica-opleiding was er in die tijd nog niet, maar wel kon je binnen de studie wiskunde in de laatste jaren het accent leggen op informatica. Op die manier ben ik in de informaticawereld terechtgekomen.

"Na Delft ben ik met een afstudeerhoogleraar meegegaan naar Enschede. Daar heb ik een jaar gewerkt op het gebied van formele talen en automaten. Vervolgens deed ik promotieonderzoek in Amsterdam, aan de Vrije Universiteit van Amsterdam, bij prof.dr.J.W. de Bakker. Daarna ben ik voor vier maanden naar Hamilton, Ontario, Canada gegaan. Vervolgens een jaartje naar Nijmegen, en daarna een aantal jaren naar Twente. Toen naar de Vrije Universiteit in Brussel. Daarna ben ik hier terechtgekomen.

"Wat ik eigenlijk altijd gedaan heb, zijn in de eerste plaats dingen die te maken hebben met formele taaltheorie. Grammatica's, automaten en zo. In eerste instantie is dat gericht geweest op de compilerbouw wereld, programmeertalen, attribootgrammatica's, et cetera. Langzamerhand is daarin een verschuiving gekomen naar natuurlijke-taalverwerking, grammaticaformalismen en ontledingsmethoden voor natuurlijke-taalverwerkingstaken.

"Een tweede interesse die altijd zichtbaar is geweest is de belangstelling voor de geschiedenis van de informatica, de geschiedenis van de technologie en de sociaal-maatschappelijke aspecten van technologie, in het bijzonder van de informatica. Ik ben altijd wel betrokken geweest bij activiteiten die niet puur informatica waren, maar die te maken hadden met informatica en maatschappij.

Het Parlevink Project

"Het project parlevink was in eerste instantie een zogenaamd 'studentenproject', bestaande uit een aantal afstudeerders en mijzelf. Zo langzamerhand is het gegroeid tot één groot project, dat allerhande aspecten van computerlinguïstiek en natuurlijke-taalverwerking bestrijkt, waaronder ook de maatschappelijke en sociale context. Het enthousiasme van studenten die binnen Parlevink wilden afstuderen straalde ook af op andere mensen, waaronder de stafmedewerkers. Die zeiden: "Ja, dat is toch wel de moeite waard, dat is toch wel leuk, en je kunt er op een leuke manier onderzoek aan doen". Veel stafmedewerkers hadden sluimerende interesses in de richting van de linguïstiek, computerlinguïstiek, of taalfilosofie. Die zijn door die activiteiten naar boven gebracht, en nu is er een redelijk enthousiaste club mensen onder de staf die zich met taaltechnologie bezighoudt. Dat is altijd leuk, om enthousiasme mee te maken. Op het ogenblik zijn er tussen 16 en 20 studenten die officieel een D-opdracht doen bij Parlevink, of te kennen hebben gegeven dat ze dat voorhebben. Het aantal is stijgend, zou je voorzichtig kunnen zeggen. Dit jaar zijn er negen of tien mensen afgestudeerd. In juni een stuk of vier, in augustus vier of vijf.

"Parlevink is een koepel waarbinnen verschillende deelprojecten hun plaats hebben. Een van de projecten is Schisma. Schisma is praktisch georiënteerd: er moet zo snel mogelijk een prototype gemaakt worden van een schouwburginformatiesysteem. Verder is er een deelproject dat zich bezighoudt met taalspecificatie en ontleden, waarbinnen het syntactische gedeelte inclusief unificatie is ondergebracht.

"Een derde project betreft semantiek en pragmatiek. Hierin wordt fundamenteel onderzoek verricht op het gebied van dialoogmodellering en het functioneel gebruik van taal. Jan Schaaake doet pragmatiek; Toine Andernach werkt aan semantiek. Zowel vanuit het syntax-deelproject als het semantiek/pragmatiek-deelproject vloeien er resultaten richting het Schismaproject.

Daarnaast is er een deelproject dat zich bezighoudt met societal and philosophical aspects in een maatschappelijke context van taal. Ook is er een project non-traditional approaches to language and information processing. Daarin wordt onderzoek gedaan naar neurale netwerken, genetische algoritmen, fuzzy logic en cellulaire automaten. Op dit moment is er geen stroom van resultaten uit dat project richting het taalverwerkende-systemengedeelte. Maar dat kun je beschouwen als exploratief onderzoek. We weten niet of er iets uit komt, maar we vinden het voldoende leuk en voldoende interessant om er tijd en mankracht aan te besteden.

"Dat zijn vijf deelprojecten, waarbij één deelproject echt product- of systeemgeoriënteerd is. Een aantal mensen ziet toe op de coördinatie. Ook zijn veel mensen ingedeeld bij verscheidene deelprojecten en zo weet men grotendeels van elkaar wat men doet. Daarnaast zijn er wekelijkse bijeenkomsten en ook de ParleBode fungeert als bindmiddel. Bovendien is er sinds kort ook een nieuwsgroep op het net, inf.seti.parlevink, waar mensen allerhande dingen kwijt kunnen die met het project te maken hebben. Hoewel ik moet zeggen dat er de laatste tijd meer dingen op staan die niet direct te maken hebben met het project, discussies over allerhande zaken, grapjes, "Wie gaat er mee naar een concert?" en zo. En we zitten op dezelfde gang, dus het is moeilijk elkaar te ontlopen.

"Iets wat hier - voornamelijk ook in het begin - speelde, zijn parallelle (ontleed-)methoden: Hoe kun je dingen parallel aanpakken? Daarin hebben we ons ook vanaf het begin onderscheiden van onderzoek op andere plaatsen. Dat heeft te maken met de informaticaomgeving waarin je toch vrij gemakkelijk afstapt van linguïstische en psycholinguïstische ideeën en modellen van taalverwerking. Je kijkt hier naar taal als een informaticus, dus je ziet het als iets dat geformaliseerd moet worden: je moet uiteindelijk een werkend staalproduct krijgen. In hoeverre dat nou psycholinguïstisch of cognitief plausibel is, dat is een tweede - soms zelfs een derde. Ik wil niet zeggen dat dat niet interessant is en dat we daar ook niet naar kijken - want je kunt daar ontzettend veel van leren, zonder meer. Alleen is het niet het uitgangspunt van het onderzoek hier. We zijn op een heel abstracte manier bezig, gericht op modellen, efficiëntie en dat soort zaken, volledig abstraherend van allerlei zaken die in de praktijk spelen: je hebt geen uitzonderingen, je hebt enkel een elegant model. Dat maakt ook dat je niet al te veel bagage meesleept die je belemmert om dingen op een nieuwe manier aan te pakken.

Computerlinguïstische Vakken in Twente

"In eerste instantie waren er in Twente geen vakken die ook maar iets te maken hadden met linguïstiek of computerlinguïstiek. Enkel bij de faculteit Wijsbegeerte & Maatschappijwetenschappen was een vak Taalwetenschap. Nu zijn er bij Informatica in ieder geval een drietal. De basis wordt gevormd door Taal & Informatica, dit jaar voor het eerst gegeven. Het is de opvolger van het vak Computerlinguïstiek. Je zou het een algemene inleiding in de linguïstiek kunnen noemen, waarbij de computationele invalshoek voortdurend aanwezig is. Daarnaast is er nu het vak Taaltechnologie. Het behandelt allerhande bestaande natuurlijke-taalverwerkings-systemen, waarbij ook de bijbehorende literatuur wordt aangeboden. Waarschijnlijk zullen er ook een aantal van de systemen aan de orde komen die op de workshop TWLT5 over natural language interfaces (Twente Workshop on Language Technology 5, thema Natural Language

Interfaces, gehouden op 3 en 4 juni 1993) gepresenteerd werden. Het is de bedoeling dat studenten zelf ook met die systemen gaan werken. Het derde vak, Formele Analyse van Natuurlijke Taal, is historisch gezien het eerste. Meteen nadat ik hier kwam, ben ik daarmee begonnen. Het behandelt op dit moment onder andere allerhande ontleedmethoden, unificatie en ook een stuk Montague-semantiek. Daarnaast zijn er vakken als Neurale Netwerken & Neurocomputing, en het vervolgvak Seminarium Theoretische Informatica, waar je ook theorie aangereikt krijgt die je kunt gebruiken bij natuurlijke-taalverwerking. Ongeveer hetzelfde geldt voor de colleges Theoretische Informatica en Compilerbouw. Voor Neurale Netwerken & Neurocomputing is dat wat duidelijker, omdat er veel studenten zijn die belangstelling hebben voor de combinatie neurale netwerken en natuurlijke taal.

Kritiek op Informatica

"Mijn kritiek op de Informatica is dat men voortdurend bezig is geweest zich af te schermen: Dit is wel informatica en dat niet. Misschien dat men zich altijd heeft moeten afzetten, dat zou kunnen. Maar je kunt het ook niet vasthouden. Iedereen gaat computers gebruiken en gaat specifiek voor zijn vakgebied met informatica bezig. Ik denk dat het heel goed is voor de faculteit dat er zo nu en dan echt nieuwe dingen ontstaan die niet zo passen binnen het traditionele informaticaonderzoek. Het geeft informatici ook de gelegenheid breder bezig te zijn met haar eigen puur informatica-aspecten van onderzoek. Bijvoorbeeld als je kijkt naar informatiesystemen, kun je zeggen: "We kijken puur naar de databases". Je kunt ook wat ruimer gaan kijken en ergonomische aspecten erbij betrekken, bijvoorbeeld human computer interaction. Dan kun je wel zeggen: "Dit stukje hoort bij ergonomie en dat stukje hoort daarbij", maar je kunt ook expertise binnen je eigen club op dat gebied hebben. Dat zie je nu wat natuurlijke taal betreft binnen de informatica gebeuren. Ik zie ook niet dat wij in de toekomst allerhande dingen gaan dupliceren wat betreft het onderzoek van taal, maar ik denk dat ons onderzoek meer geïntegreerd zal worden in zaken als informatiesystemen, computer supported cooperative work, en multimediasystemen.

"Binnen het project Parlevink willen we absoluut niet alleen productgeoriënteerd bezig zijn. We schermen wel met de kreet language engineering, linguistic engineering of taaltechnologie, maar zodra je sociaal-maatschappelijke aspecten erbij gaat halen, ben je al veel minder produktgeoriënteerd bezig. Ook wordt hier fundamenteel onderzoek verricht op het gebied van formele talen en dialoogmodellering. En dat heb je ook wel nodig als input voor de mensen die wèl meer productgeoriënteerd bezig willen zijn, die echt systemen willen bouwen die werken.

Breed Georiënteerd Onderzoek

"Het is een bewuste keus om breed bezig te zijn met het onderzoek. Het merendeel van de mensen die erbij betrokken zijn, heeft geen vrede met het losstaand bestuderen van een wiskundig modelletje. We willen dingen in hun context zien. Dit komt voort uit de brede belangstelling van een aantal mensen binnen het project. Bovendien sluit het aan bij de formulering van het beleid ten aanzien van het informaticaonderzoek door de faculteit.

"Die brede oriëntatie is binnen Parlevink heel uitdrukkelijk aanwezig. Meer dan bij alle andere projecten binnen de faculteit. Meestal belijdt men in woorden wel dat men dat wil, maar in de praktijk blijft men toch wel dicht bij de kerninformatica. En voor de rest zegt men: "Nou ja, dat moet maar door andere mensen gebeuren." Misschien dat natuurlijke-taalverwerking zich ook meer voor een bredere oriëntatie leent. Natuurlijke taal is zo typisch gekoppeld aan mensen, dat het eigenlijk wel moet.

Niet Conventioneel

"Ik denk dat het clubje mensen hier wat betreft belangstelling duidelijk afwijkt van wat standaard is voor een informaticaomgeving. Als er iets nieuws is en men daar sceptisch tegenaan kijkt, zijn wij

al gauw bereid om er wel iets aan te doen en er juist wel naar te kijken. Dat zijn de niet-traditionele benaderingen. Je merkt bijvoorbeeld dat men eerst heel vreemd tegen zo iets als taal aankijkt. Dan werd er gezegd: "Dat doen wij informatici toch niet, dat hoort er toch niet bij. Dat doet de zachte sector." Dat verandert langzamerhand. Voor een informaticaomgeving zijn we niet conventioneel, kun je zeggen. Maar als je naar andere faculteiten of universiteiten kijkt, dan valt het nog wel mee. Kijk bijvoorbeeld naar Remko Scha in Amsterdam. Die houdt zich bezig met alfa-informatica en koppelt dat aan onderzoek naar muziek. Dat is toch veel meer afstand nemen van je concrete onderzoeksgebied dan hier. Zo zijn er wel meer mensen aan te wijzen in Nederland die er andere dingen bij halen dan gewoon is. Jammer genoeg vind je die niet of nauwelijks binnen een informaticafaculteit.

Centre for Telecommunications & Information Technology

"Net opgericht is het Centre for Telecommunications and Information Technology (CTIT). Binnen dat instituut gaan groepen samenwerken die op dit moment los van elkaar staan binnen dezelfde faculteit, of binnen dezelfde universiteit. Er gaan mensen van Elektrotechniek, Toegepaste Wiskunde en Wijsbegeerte & Maatschappijwetenschappen meewerken. Ook Technische bedrijfskunde is een kandidaat. Op dit moment zijn er onder andere de projecten computer supported cooperative work en multimedia gedefinieerd. In de loop der tijd zullen er meer projecten bij komen. Dit zijn typische integratieprojecten: verschillende aspecten van de informatica worden bij elkaar gebracht en het onderzoek wordt op elkaar afgestemd. Prof. Niemegeers van de vakgroep TIOS (Tele-Informatica en Open Systemen) is daar directeur.

"Parlevink is op dit moment ook één van de aanloopprojecten binnen het CTIT. Dit zal meer gelegenheid bieden om aan te sluiten bij ander onderzoek binnen de faculteit of universiteit, waar natuurlijke-taalaspecten aan verbonden zijn. Het kan alleen maar goed zijn en stimulerend werken.

Telematica Onderzoeksschool

"Er is wel sprake van een onderzoeksschool in de richting van de telematica, die eventueel gekoppeld zou kunnen worden aan een instituut als het CTIT. Daar zouden dan eventueel plaatsen voor een tweedefase kunnen komen. Maar dat is absoluut nog niet zeker. Eigenlijk wil men overall onderzoeksscholen hebben gekoppeld aan universiteiten. Dat is een hele concurrentiestrijd, en wie er op een gegeven moment wint, dat is politiek.

"Als het onderzoeksinstituut er echt komt, dan is het idee dat het onderzoek ook op een hoger plan gebracht wordt. En dat je meer aansluiting kunt krijgen bij befaamde onderzoeksinstituten in het buitenland. Zo kan er een netwerk van instituten ontstaan, waar D-studenten en Aio's gebruik van kunnen maken. Omdat zo'n onderzoeksinstituut zich niet puur met kerninformatica bezighoudt, kun je als student breder bezig zijn. Wellicht heb je zo meer kans dat bepaalde interesses ook werkelijk in je afstudeerpakket aanwezig kunnen zijn.

Neslia Paniculata

"Het eerste boek moest ik schrijven, want dat was mijn proefschrift (Context-free grammars: covers, normal forms and parsing, Anton Nijholt, 1980). Dat is vervolgens bij een uitgever aangeboden, en geaccepteerd. Het ging over contextvrije grammatica's, en het transformeren van grammatica's. Het tweede boek was 'Computers and Languages' (Computers and languages: theory and practice, Anton Nijholt, 1988). Dat heb ik geschreven in de periode dat ik afstand nam van formele taaltheorie en veel breder bezig wilde zijn. Typisch iets wat je na je proefschrift schrijft. Het boek gaat voornamelijk over de interactie tussen natuurlijke-taalverwerking, informatica en maatschappij: hoe die elkaar onderling beïnvloeden. Daarnaast zijn er een paar boekjes geweest waarbij ik als editor ben opgetreden. Eén over expertsystemen, en één ook weer

over parsing. Dit jaar komt er nog een boek uit van zo'n vierhonderd bladzijden over de geschiedenis van de informatica (De Geschiedenis van de Rekenkunst: Van kerfstok tot computer, Anton Nijholt en Jan van den Ende, Uitgeverij Academic Service, te verschijnen april 1994). Het boek betreft niet uitsluitend de technische geschiedenis, maar vooral ook de geschiedenis van de mensen die zich ermee beziggehouden hebben. Ook de mensen die ervoor zorgden dat dergelijk onderzoek plaatsvond: geldgevers, subsidiegevers, overheden, en hun ideeën. Ook zaken als Wat is informatica voor een soort wetenschap? komen in dat boek aan de orde.

"Een uitgeverij beginnen is nog één van mijn wensen, maar dat kan gecombineerd worden met mijn werk hier. De naam van de uitgeverij zal zijn 'Neslia Paniculata'. Om dingen te publiceren die ik leuk vind en die een mooie uitvoering te geven. Zoals de boeken die ik geschreven heb, of de proceedings van TWLT, dat zijn dingen die passen binnen mijn belangstellingsgebied. Misschien soms een aardig proefschrift, of een dichtbundel. Riex op den Akker heeft een poos geleden een boekje geschreven 'Over Informatica, Computers en Taal (Prolegomena voor een Antropologische Technologie, 1992). Dat is typisch een boekje dat zou passen binnen een uitgeverij die ik zou willen hebben. Samengevat: een beetje beschouwend over het onderzoek waar we mee bezig zijn, maar zo nu en dan ook iets over het onderzoek zelf.

De Toekomst

"Ik denk dat er in de toekomst meer sprake zal zijn van integratie van dit natuurlijke-taalonderzoek met andere aspecten die actueel zijn en worden binnen de informatica. Dus dat je niet naar natuurlijke taal sec gaat kijken, maar naar wat human factors wordt genoemd: mens-machine-communicatie.

"Verder zie ik ons de komende jaren werken aan de deelprojecten die geformuleerd zijn. Als de belangstelling van de mensen die hier werken verschuift, of als er opeens een Aio binnenkomt met wilde ideeën, is het niet bij voorbaat uitgesloten dat daar een poos de aandacht komt te liggen. Zo is neurale netwerken ook binnengekomen. Er was iemand die daar tijdens zijn studie aan gewerkt had en daar absoluut mee door wilde gaan. Ik vond het voldoende interessant om zo iemand hier te hebben. Maar dat betekent op een bepaald moment dat ook andere mensen naar neurale netwerken gaan kijken. Zoiets kan weer gebeuren. Door de D-opdracht van Willem-Olaf (W. Huijsen. Genetic grammatical inference: Induction of pushdown automata and context-free grammars from examples using genetic algorithms, master's thesis, 1993) zijn mensen bijvoorbeeld naar genetische algoritmen gaan kijken. Om nou alles vast te leggen in projectbeschrijvingen en daar niet van te willen afwijken, dat zal absoluut niet het geval zijn.

"Daarnaast zijn er plannen om hier in de regio Nederland-Duitsland een Forschungs- und Entwicklungszentrum in het leven te roepen op het gebied van neurale netwerken en fuzzy logic. Wij doen daaraan mee. Het is nog niet helemaal goedgekeurd, maar als het doorgaat, betekent dat dat er mankracht hiernaar toekomt."

(Willem-Olaf Huijsen)