

Meesterlijk complex

De toepassingen van zijn theorie over complexiteit groeien explosief. Onder collega's genoot hij al veel langer aanzien. Informaticus Paul Vitányi is meester in zijn vak. En kleurrijk bovendien.

Bennie Mols

EEN ECHTE theoreticus: zo noemt theoretisch informaticus prof. dr. ir. Paul Vitányi (63) zichzelf. “Maar in tweede instantie vind ik het geweldig als je een theorie kunt toepassen.” En daarover heeft Vitányi geen klagen. Het aantal toepassingen van zijn complexiteitstheorie is sinds het jaar 2000 wereldwijd explosief toegenomen. Ter ere van zijn werk organiseerde het cwi gisteren een middag waarop drie toonaangevende collega-informatici lezingen hielden. Onder hen de Amerikaanse Turingprijswinnaar (een soort Nobelprijs voor de informatica) Andrew Yao. Tevens ontving Vitányi tijdens het symposium een koninklijke onderscheiding: hij werd benoemd tot Ridder in de Orde van de Nederlandse Leeuw. Het begon met een computerprogramma dat soorten op grond van hun DNA classificeert in een evolutionaire stamboom. Daarna volgden programma's om talen te rangschikken in taalbomen, om muziekstukken in te delen naar componist en literatuur naar schrijver. En dat allemaal zonder dat zo'n programma een flauw benul heeft van muziek, literatuur of biologie. De crux zit in een wiskundige vergelijking van de complexiteit van digitale bestanden. Daarmee kan de computer fragiele Chopin-sonates niet alleen van jazzy Miles Davis-beats onderscheiden, maar ook van meer verwante Debussy-composities, zonder verder iets te weten van ritmes, toonhoogten of melodie.

KOLMOGOROV Vitányi verdient zich aan het Amsterdamse Centrum voor Wiskunde en Informatica (cwi) al drie decennia in de vraag hoe je complexe informatie en ingewikkelde berekeningen zo precies mogelijk wiskundig beschrijft. En vervolgens in de vraag hoe computers slim kunnen redeneren met die complexe informatie. Zijn werk borduurde onder meer voort op dat van de Rus Andrei Kolmogorov (1903-1987). Samen met de Canadees Ming Li beschreef Vitányi dat in 1993 in het boek *Introduction to Kolmogorov complexity and its applications*. Als je beeld, geluid of tekst codeert in enen en nulen, dan is de Kolmogorovcomplexiteit de lengte van het kortste programma dat dat stuk informatie beschrijft. Stel dat je een bestand hebt bestaande uit honderdduizend nulen. Dan is dat te beschrijven met een programma dat veel korter is. Het kortste programma hoeft alleen maar te zeggen: 'print honderdduizend nulen'. Dit bestand heeft dus een lage Kolmogorovcomplexiteit. Maar als je honderdduizend maal een muntstuk opgooit, en achter elkaar noerteert of je 'kop' (1) of 'munt' (0) krijgt, dan kun je het resultaat vrijwel nooit comprimeren. Het gaat om zuivere wil-

keur, en de kortste beschrijving is precies de oorspronkelijke reeks van nulen en enen – geen bit korter. De Kolmogorovcomplexiteit is nu dus exact de lengte van de oorspronkelijke rij. Kolmogorov zelf dacht dat zijn wiskundige definitie onbruikbaar zou zijn, maar Vitányi, Li en anderen lieten zien dat het aantal toepassingen legio is. Vitányi: “De Kolmogorovcomplexiteit is in de praktijk onberekenbaar. Maar het bijzondere is dat er toepassingen mogelijk zijn waarvoor goede compressieprogramma's een resultaat leveren dat waarschijnlijk nauwelijks afwijkt van de theoretische Kolmogorovwaarde.”

LUI Paul Vitányi werd in 1944 geboren in Hongarije – “en dus ben ik automatisch niet-links” – uit een Hongaarse vader en een Nederlandse moeder. Op zijn tweede verhuisde hij naar Nederland. In de jaren vijftig en zestig waren het boeken als *Giant brains* (1949) van Edmund Callis Berkeley, het vertaalde *Menselijk en machinaal denken* (1964) van de Duitse ingenieur Karl Steinbuch en *Cybernetics – Or the control and communication in the animal and the machine* (1948) van de Amerikaanse wiskundige Norbert Wiener die hem intrigeerden. “Cybernetica – dat wou ik studeren. Het dichtste wat daar in die tijd bij in de buurt kwam, was een studie elektrotechniek in Delft, met als specialisatie computers. Maar ik was lui. Ik lag liever in bed dan 's ochtends naar college te gaan. En verplichte practica lopen vond ik ook erg vermoeiend. Na een tijdje dacht ik: ik kan beter wiskunde gaan doen, dan hoef ik helemaal geen colleges en practica meer te lopen. Dan kan ik gewoon thuis nadenken.” Hij zou cum laude als wiskundige afgestudeerd zijn, ware het niet dat Delft – op de golf van de universitaire democratiseringsbeweging van de jaren zestig – het cum laude niet had afgeschaft. Te elitair.

Van bescheidenheid heeft de Amsterdamse informaticus weinig last. Collega Ming Li zei in 2004 over Vitányi: “Paul heeft overal een sterke mening over, en met goede redenen.” Ronald de Wolf promoveerde 2001 cum laude bij Vitányi. In een e-mail vanuit Japan schrijft hij: “Als mens is hij kleurrijk, erudiet, overal op de wereld geweest, kent iedereen, zit vol anekdotes en goede adviezen, sterke meningen, en (is) dus niet altijd politiek correct.” Anekdoten vertelt de flamboyante Vitányi inderdaad graag. “In 1969 studeerde ik nog in Delft. In mei van dat jaar zat ik een keer in Amsterdam buiten het Bier te drinken bij café Hoppe aan het Spui, toen ik ineens een boel studenten het Maagdenhuis zag binnengaan. Het begin van de Maagdenhuisbezetting. Het leek me leuk daar bij te zijn, en dus ben ik achter ze aan gegaan. Ik heb er twee dagen en nachten doorgebracht, maar toen was de bezetting nog bezig. Het probleem was alleen dat ik in Delft voor mijn afstudeerwerk rekentijd had gereserveerd op de computer die mijn afstudeerhoogleraar Van der Poel voor zijn studenten had neergezet. Dus zei ik tegen de sociologen met wie ik optrok: ‘sorry, ik moet gaan, want ik heb een afspraak met een computer.’” Het geheim van de programma's, die muziek, talen of planten rangschikken in clusters, zit in een door Vitányi en Ming Li bedachte ‘universele maat van

gelijkenis'. Deze geeft aan hoezeer digitale bestanden op elkaar lijken. Een computerprogramma comprimeert eerst alle bestanden. Vervolgens vergelijkt het paarsgewijs de gecomprimeerde bestanden met elkaar. Hoe meer twee bestanden op elkaar lijken, hoe minder ruimte nodig is om de combinatie van beide op te slaan. Tenslotte drukt het programma in een getal tussen 0 en 1 uit hoezeer het ene op het andere bestand lijkt. Met die getallen kan dan een boomstructuur worden gemaakt waarin bestanden die meer op elkaar lijken dichter bij elkaar zitten. In recent onderzoek zijn de resultaten van deze methode vergeleken met die van de best beschikbare *datamining* technieken, die wel a-priori kennis op een vakgebied gebruiken. Vitányi: “En dan blijkt dat onze methode in het alg-

Paul Vitányi

‘De vlucht van de toepassingen geeft een kick’

meen minstens even goed werkt, en dat ze zelfs veel beter presteert als er gekke afwijkingen in de gegevens zitten. Andere programma's zijn slecht in het herkennen daarvan, omdat ze van te voren gedefinieerde eigenschappen gebruiken. Onze methode heeft geen voorkennis nodig. Dat is de crux. Of het nu gaat om talen, muziek, DNA of wat dan ook. Het maakt niet uit. En dat is vooral handig in gevallen waarbij je niet precies weet naar welke regelmaat je zoekt.” Ten tijde van de SARS-uitbraak concludeerde het clusterprogramma dat SARS sterk verwant is aan het Corona-229-virus, een conclusie die de biologen met in jarenlang opgebouwde kennis van virussen ook stelden.

“Maar”, aldus Vitányi, “het is altijd maar afwachten hoe snel andere vakgebieden zulk theoretisch werk oppikken. Dat is de malke van abstracte wiskunde: hoe eenvoudig je zelf ook denkt dat het is, hoe helder je ook denkt het te hebben opgeschreven, voor niet-vakgenoten is het vaak heel moeilijk te geloven waarom het zou moeten werken. Tijdens de vogelgrieppepidemie stuurden we enkele van onze resultaten over mutaties in het virus naar de groep van viroloog Ab Osterhaus in Rotterdam. ‘Hoe komen jullie daar aan?’ vroegen ze. ‘Tja, dat doet ons programma’, zeiden we. ‘Ja hallo, hoe weten we dat we daar op kunnen vertrouwen?’ was toen het antwoord.”

Een van de meest recente clustertoepassingen is het gebruik van zoekmachine Google als woordenboek om computers automatisch woorden te laten begrijpen. Samen met zijn promovendus Rudi Cilibrasi ontwikkelde Vitányi daarvoor een ‘afstandmaat’ die woorden aan elkaar relateert op grond van hun betekenis. Zo levert het woord ‘horse’ 155 miljoen Google-hits op; het woord ‘rider’ 60 miljoen. Zoeken op pagina's waarin beide woorden tegelijk voorkomen, levert 2,8 miljoen treffers. Uit die getallen kan worden berekend hoezeer ‘horse’ en ‘rider’ qua betekenis met elkaar in verband staan. Door een web van meer en minder verwante woorden te

creëren, kunnen computers zo automatisch woorden duiden. Het lijkt typisch een methode voor Google zelf. “Ik weet niet of ze het al gebruiken”, zegt Vitányi. “Wel weet ik dat ze vijftien van mijn boeken over Kolmogorovcomplexiteit hebben besteld.”

NEUS Vitányi's interesse – zowel binnen als buiten de theoretische informatica – reikt veel verder dan clustering. Ongeveer elke zeven jaar begon hij aan een nieuw terrein. Paralleel rekenen, netwerkalgorithmes, de fysica van computing, machinelere, informatietheorie en quantumrekenen – hij heeft zich er overal in verdiept. Een goede neus voor baanbrekende nieuwe ontwikkelingen heeft hij ook. Toen de Amerikaan Peter Shor in 1994 op een congres zijn Shor-algoritme presenteerde als een mogelijke toepassing van de nog volstrekt hypothetische quantumcomputer, realiseerde Vitányi zich dat dit de geboorte was van een nieuwe tak van wetenschap: het quantumrekenen. Terug op het cwi heeft hij zich meteen ingespannen om een onderzoeksgroep op dit nieuwe terrein op



• Prof. dr. ir. Paul Vitányi
FOTO FREDDY RIKKEN

te richten. “Als je als een van de eersten een nieuw terrein betreedt, dan kun je de grootste vruchten plukken.” Hoogleraar Harry Buhman is de huidige leider van de onderzoeksgroep Quantum Computing bij het cwi. Hij werkte als postdoc met Vitányi samen. “Paul is een van de weinige wetenschappers die ook op latere leeftijd nog altijd volop publiceert. Iemand die nog steeds gaat voor het onderzoek en niet voor het management. Hij is echt gericht op onderzoek en niet op politiek. Daardoor heeft hij wel vaak bonje met de bureaucratie.” Vitányi, naast onderzoeker aan het cwi ook hoogleraar aan de Universiteit van Amsterdam (UvA), is de eerste om zijn problemen met de regels-omde-regels te beamen.

Hij vertelt over de Amerikaanse Rudi Cilibrasi, die promovendus werd zonder een mastertitel. “Toen Rudi nog student was aan het California Institute of Technology, stuurde hij een e-mail over een artikel van mij. Hij had het bewijs daaruit geprogrammeerd om te kijken of het klopte. Hij schreef ook dat hij erover dacht om, na zijn bachelor van vier jaar aan CalTech, een master in de logica in Amsterdam te doen. Maar ik dacht: iemand die dat moeilijke artikel zo goed snapt dat hij het kan programmeren, kan wel meteen gaan promoveren.”

GESTEGGEL Cilibrasi deed “prachtig werk”, maar vlak voor de officiële promotie maakte de UvA bezwaren omdat je zonder een masterdiploma niet kunt promoveren. Vitányi: “Wat een onzin, gaat voor het onderzoek en niet voor het management. Hij is echt gericht op onderzoek en niet op politiek. Daardoor heeft hij wel vaak bonje met de bureaucratie.” Vitányi, naast onderzoeker aan het cwi ook hoogleraar aan de Universiteit van Amsterdam (UvA), is de eerste om zijn problemen met de regels-omde-regels te beamen.

zijn 65e wil hij niet weten, zelfs niet nadat hij in 2006 door een herseninfarct werd getroffen. “De schade is puur motorisch, aan de rechterkant van mijn lichaam. Het heeft gelukkig geen invloed op mijn denken. En dus wil ik doorwerken. Ik werk aan de derde druk van mijn boek, en er moet nog een aantal artikelen de deur uit.” Hij kan de toepassingen van zijn werk niet meer allemaal bijhouden. “Maar ik weet bijvoorbeeld dat MIT-onderzoekers het recent zijn gaan gebruiken voor onderzoek naar de voorspelling van orkanen, Stanford-onderzoekers voor het voorspellen van aardbevingen en bio-informatici gebruiken het voor het vergelijken van DNA-strengen.” Behalve voor het vergelijken van bestanden, wordt Kolmogorovcomplexiteit ook steeds vaker gebruikt voor het comprimeren van individuele bestanden van CalTech is toch zeker meer waard dan een masterdiploma van de UvA! CalTech heeft tot nu toe 32 Nobelprijswinnaars gehad, het dubbele van heel Nederland!” En na veel gesteggel kon Cilibrasi toch promoveren. Vitányi heeft nog twee jaar te gaan tot zijn pensioen. Maar van ophouden na