

## Connecties met JKL

Peter Grünwald

### 1. Inleiding: een ander Liber Amicorum:

Het was eind juni 1989. Ik was eerstejaars student informatica aan de VU Amsterdam en ik bereidde me voor op mijn tentamen kansrekening. In diezelfde maand verscheen een boekje aan de TU Eindhoven met de titel "Op de Snede tussen Theorie en Praktijk". Op de achterflap lezen wij:

*In deze bundel opstellen over de toepassing van wiskunde in de praktijk, probeert een uitverkoren groepje (hoog)geleerde heren zichzelf en elkaar te overtreffen in originaliteit en diepte van visie...*

Het betreft hier een Liber Amicorum voor Prof. dr. J.F. Benders, hoogleraar aan de TU Eindhoven. Benders ging juli 1989 met pensioen, en tot zijn opvolger was benoemd: Jan Karel Lenstra, afkomstig van het CWI in Amsterdam! Het was daarom niet verbazingwekkend dat JKL ook als co-auteur in het Liber voorkwam. Wellicht iets verbazingwekkender is dat er ook een bijdrage in staat van mijn vader, Herbert Grünwald (destijds werkzaam bij Philips, inmiddels 80 jaar oud en nog steeds zeer actief als milieu-activist in het lieflijke Oostenrijkse dorpje Pitten). Ziehier een eerste connectie met JKL. Met enig hangen en wurgen (Bending Benders, zie Figuur 1 in Sectie 5) is het mij gelukt alle in het oog vallende connecties tussen JKL en mij aan het Benders Liber op te hangen:

### 2. Organisatie van het CWI-Onderzoek

Naast het feit dat behalve JKL en mijn vader bijna iedere auteur een grote bril en snor droeg, valt vooral de naam op van Jan Karels bijdrage aan het Benders Liber (co-auteurs K. Anthonisse en M. Savelsbergh):

*De Bochtige Weg van Idee naar Implementatie*

Dit lijkt wel een profetische verwijzing, twintig jaar vooruit in de tijd, naar Spinoza: JKL's plan om de organisatie van het onderzoek op het CWI eens heel flink op de schop te nemen. Ik bedoel dit geenszins sarcastisch: naar mijn mening is de verdeling van zowel onderzoekers als onderzoeksthema's over de onderzoeksgroepen onder JKL's leiding een heel stuk verbeterd: er zijn minder groepen, er wordt aan minder onderwerpen tegelijkertijd gewerkt, en er wordt beduidend meer samengewerkt. In feite zijn delen

van het Spinozaplan uiteindelijk wel degelijk gerealiseerd, en is het CWI een coherenter geheel geworden. Dit is mij pas echt duidelijk geworden toen ik, in de herfst van 2010, meeschreef aan de periodieke zelfevaluatie van het CWI. Door het meeschrijven en discussies met medeschrijvers Sander Bohte, Jason Frank, Frank Roos en vooral JKL zelf, kreeg ik een overzicht van wie op het CWI nu eigenlijk wat doet, en in welke groep. Het was overduidelijk dat er sinds het aantreden van Jan Karel een duidelijke trend is naar minder groepen en meer bundeling van onderzoek, en dat er, gezien de inderdaad immer bochtige weg van idee naar implementatie eigenlijk heel veel is gebeurd.

Desalniettemin is een van de hoofdmoten van het Spinozaplan niet uitgevoerd: de onderzoeksgroepen aan het CWI zijn zelf weer gegroepeerd als *clusters*, en Jan Karel wilde die clusters heel graag afschaffen. Dat is niet gebeurd. Maar, over clusters gesproken:

### 3. Clustering

Ook aan het begin van zijn carrière was JKL heel erg bezig met clusters, zij het van een ander soort. Eén van zijn allereerste publicaties, uit 1974, was:

*Clustering a Data Array and the Traveling-Salesman Problem*  
J.K. Lenstra, Operations Research 22(2), pp. 413-414.

Het onderwerp van dit artikel is inderdaad het clusteren van data, samengevat in een matrix, en dit is iets waar onze onderzoeksgroep zeer actief mee bezig was toen JKL begon aan het CWI directeurschap. Mijn promotor Paul Vitányi en zijn oio Rudi Cilibrasi boekten grote successen met *Complearn*. Deze software implementeerde de zgn. *normalized compression distance*. Het idee is als volgt: neem twee bestanden *A* en *B* (bijvoorbeeld twee DNA-sequenties, of twee teksten, of twee muziekbestanden in MIDI-format). We comprimeren eerst beide bestanden apart met een standard data compressor (bijvoorbeeld *bzip* of *gzip*). Dit geeft twee nieuwe bestanden *A.zip* en *B.zip*. Vervolgens maken we een nieuw bestand door *A* en *B* aan elkaar te plakken, en comprimeren we het resulterende bestand. Dit geeft een nieuw bestand, *AB.zip*. Als *A* en *B* op elkaar lijken, dan zal de lengte van *AB.zip* in het algemeen veel kleiner zijn dan de som der lengtes van *A.zip* en *B.zip*. Als *A* en *B* volstrekt ongerelateerd zijn, dan zal de lengte van *AB.zip* ongeveer gelijk zijn aan de som der lengtes van *A.zip* en *B.zip*. Bijvoorbeeld, als *A* het volledige DNA van JKL is en *B* het volledige DNA van ondergetekende, dan is het verschil tussen *lengte(AB.zip)* en *lengte(A.zip)+lengte(B.zip)* veel groter dan als *A* het DNA van JKL is en *B* het

DNA van, zeg, een eekhoorn. Op basis van dit idee kan een afstandsmaat tussen bestanden gedefinieerd worden - hoe meer extra compressie bereikt kan worden door de bestanden aan elkaar te plakken, hoe dichter ze bij elkaar liggen in *compression space*. Die afstanden kunnen weer gebruikt worden om allerlei soorten data te clusteren - we bepalen eerst de afstanden tussen de bestanden, stoppen die in een matrix en gebruiken vervolgens (bijvoorbeeld) JKL's algoritme voor het clusteren. Dit onderzoek was 'hot' rond de tijd van JKL's aantreden als directeur van het CWI, en een van mijn eerste interacties met JKL was op de Open Dag van 2005, toen ik het algoritme uitlegde aan JKL en zijn zoontje. JKL begreep precies hoe het werkte en waarom wij het zo leuk vonden, en dat vond ik heel prettig.

#### 4. **CAR** of: *het plezier van een directeur die begrijpt waar je mee bezig bent.*

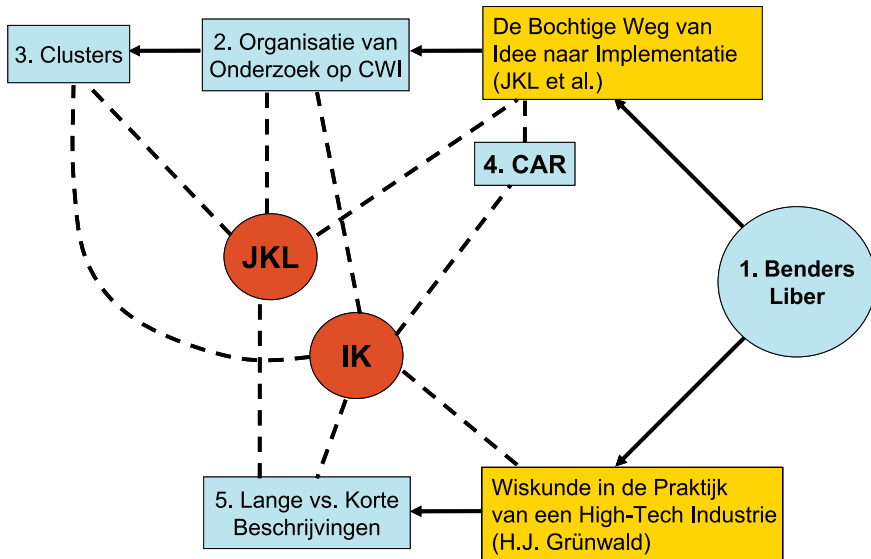
Ook later bleek vaak dat JKL affiniteit had met mijn onderzoek. In 2008 gaf ik een mini-voordracht in Leiden ter viering van de samenwerkings-overeenkomst die het CWI met Leiden had gesloten (en die leidde tot mijn snelle aanstelling als deeltijdhoogleraar aldaar). Ik probeerde in tien minuten eerst het statistische begrip *CAR* uit te leggen, en besprak vervolgens de algoritmische karakterisatie die Richard Gill en ik eraan hadden gegeven. Uit JKL's opmerkingen achteraf kon ik opmaken dat hij het verhaal van A tot Z had begrepen. Overigens betekent *CAR* "*Coarsening at Random*", en is er ook hier weer een verbinding met het Benders Liber: het onderwerp van JKL's 'Bochtige Weg' artikel in dat Liber is...*CAR!!!* Het woord *CAR* komt acht keer voor op de eerste pagina. Hier betekent *CAR* echter "*Computer Aided Routing*"....

Het komt denk ik niet zo vaak voor dat een wetenschapper die bestuurder is geworden zijn interesse in wetenschappelijke details blijft behouden. Jan Karel is dit gelukt. Tegelijkertijd is hij heel goed in het ontwaren van de grote lijnen in iemand's onderzoek. Tijdens het schrijven van de interne evaluatie viel me op dat hij een adequaat high-level idee heeft van het onderzoek van vrijwel iedere senior onderzoeker op het CWI, waarvan akte. In mijn geval bleek dit eens te meer toen ik, samen met Harry van Zanten, de Van Dantzigprijs uit zijn handen ontving: met zijn beschrijving van ons onderzoek en het belang ervan sloeg JKL de spijker geheel op de kop. Mijn ervaring is dat dit meestal niet het geval is bij prijsuitreikingen...

#### 5. **Kort en Helder (be-) Schrijven!**

JKL is een man van korte en heldere teksten. Maar een plaatje zegt meer dan duizend woorden, en JKL is dan ook de meester van de *visuals*: tijdens

het schrijven van de zelfevaluatie was hij vaak tot in de kleine uurtjes bezig met het bedenken en vereenvoudigen van figuren en diagrammen (ik denk eigenlijk dat hij veel meer tijd aan de zelfevaluatie heeft besteed dan Sander, Jason en ik...). De zelfevaluatie van het CWI staat vol met dit soort visuals waarin een grote hoeveelheid informatie op verrassend duidelijke wijze wordt samengevat. Een voorbeeld van een visual is te vinden in Figuur 1. Wie deze visual naast de vele visuals van de zelfevaluatie legt ziet meteen dat JKL er veel beter in is dan ik.



*Figuur 1: een "visual" die de hier besproken connecties tussen JKL en mij weergeeft, en tevens de superioriteit van JKL op dit gebied illustreert.*

Ikzelf ben de auteur van het boek *The Minimum Description Length Principle*, een methode binnen de statistiek en machine learning gebaseerd op het idee dat het zoeken van patronen in data gezien kan worden als het pogen de data samen te vatten, dat wil zeggen op zo kort mogelijke wijze te beschrijven: *hoe korter het kan, hoe beter het is* - weer een connectie met JKL. Helaas ben ik niet consistent: mijn eigen boek is, ondanks de titel, 570 pagina's dik - als ik er een jaar langer aan had gewerkt had ik het waarschijnlijk tot 300 pagina's kunnen indikken. Maar daar had ik na vijf jaar werk geen zin meer in. Wellicht zit deze eigenschap in mijn genen: in het stuk van mijn vader in het Benders Liber staan af en toe wat lange, omslachtige zinnen. Mijn vader verontschuldigt zich hiervoor in een voetnoot, waarin hij schrijft:

*Gezien "Die Knappheit der Zeit und die Vordringlichkeit des Befristeten" (Goethe) was het niet te vermijden dat de formuleringen bij het eerste lezen van dit stuk waarschijnlijk een cryptische indruk maken. Ik houd mij aan een uitspraak van mijn wiskundeprof toen hij bezig was een boek te schrijven en men hem alsmaar kwelde duidelijker te formuleren: "ik moest ook altijd moeite doen om wiskundeboekjes van anderen te ontcijferen, dus waarom zou ik het zelf duidelijk doen?")*

JKL kan dit beter dan de Familie Grünwald. Toen ik mijn boek aan JKL presenteerde en hij opmerkte dat het wel erg lang was, zeker gezien de titel, verontschuldigde ik me op mijn beurt met het citeren van Pascal, die naar ik meende ooit had geschreven "Ik stuur u een lange brief, omdat ik niet voldoende tijd had om een korte te schrijven". JKL kende een soortgelijk citaat (van Benjamin Franklin, meen ik). Dit leidde tot een langer gesprek waarin wij ons afvroegen wie nu precies wat hierover had gezegd. Ik besloot het uit te gaan zoeken. Met behulp van Teddy Seidenfeld maakte ik een lijst met citaten van (inderdaad) Pascal, maar ook Thoreau, Cicero, Gauss, Nietzsche, Felelon (mij onbekend), Voltaire, Woodrow Wilson en Mark Twain. De mooiste vier repliceer ik even:

*Je n'ai fait cette lettre-ci plus longue parce que je n'ai pas eu le loisir de la faire plus courte*

Blaise Pascal, Lettres provinciales (Lettre No. XIV), Pierre le Petit, Paris 1656

*If I had more time, I would have written a shorter letter.*

Marcus Tullius Cicero

*If you want me to give you a two-hour presentation, I am ready today. If you want only a five-minute speech, it will take me two weeks to prepare.*

Mark Twain

*You know that I write slowly. This is chiefly because I am never satisfied until I have said as much as possible in a few words, and writing briefly takes far more time than writing at length.*

Carl-Friedrich Gauss.

Toen ik de gehele lijst aan JKL mailde, kreeg ik onmiddellijk het volgende korte maar krachtige antwoord:

*Peter -- Dank! Twain is prachtig. Maar Cicero is het kortst dus hij wint.*

*--- Jan Karel*

Naar aanleiding van deze citaten is het trouwens de hoogste tijd om te stoppen met dit stukje!

## **6. Conclusie**

Laat ik het verder inderdaad maar kort houden en afsluiten zoals JKL het zelf zou doen:

*Jan Karel -- Dank!*

*--- Peter*