OVER HET VERWERKEN VAN GRUIS

H.J. Hopman

In het informatieblad no. 1, dat ingesloten was bij CB van april 1973, gaf de heer A.W. Lacourt een handleiding voor gruisonderzoek. Hierin maakte hij een opmerking die me aan het denken zette en die ik hieronder nader wil uitwerken. De heer Lacourt schreef nl. dat men bij gruisonderzoek een paar schaaltjes bij de hand moet hebben om alle uitgezochte exemplaren in te deponeren en men kan dan in eerste instantie een scheiding maken op grove kenmerken als schelp en slak, groot of klein, lang of kort etc. Ik vroeg me toen af wat de snelste manier zou zijn om zo'n schaaltje, waarin nog vele verschillende soorten bij elkaar liggen, verder uit te werken.

De snelste werkwijze is het splitsen van het gehele materiaal in een paar grote groepen; vervolgens elke groep te splitsen in kleinere groepen en zo door te gaan tot de soorten overblijven. Een andere werkwijze zou die zijn waarbij telkens één soort wordt afgescheiden uit het gehele materiaal. Dit kost veel meer tijd maar minder inspanning. Achteraf blijkt nu dat ik zelf altijd de langzame en dus de verkeerde methode gebruikt heb. Voor vele lezers is het bovenstaande misschien overbekend maar in de professionele literatuur ben ik nooit een behandeling van dit probleem tegengekomen. Daarom heb ik het opgeschreven.

- Eerst -

Eerst een voorbeeld om de twee wijzen van werken toe te lichten. Stel er is een bakje met Tricolia pulla, Peringia ulvae, Putilla semistriata, Cingula cingillus, Folinia costata, Alvania montagui, A. reticulata, A. lactea en A. punctura. De langzame methode werkt nu zo dat er eerst één soort, b.v. Peringia, uit het bakje bij elkaar gezocht wordt. Vervolgens een tweede soort, b.v. Folinia, en dan de derde soort etc. Telkens blijft er een restant achter waar alle overige soorten inzitten. De snellere methode werkt zo dat het hele materiaal in twee groepen wordt gescheiden b.v. gladde en geribde soorten. De gladde groep wordt weer in twee groepen gescheiden b.v. gelige exemplaren, Putilla en Cingula, en anders gekleurde exemplaren, Tricolia en Peringia. De laatste splitsing geeft de soorten. Op dezelfde wijze worden de geribde exemplaren weer in groepen gesplitst tot de soorten overblijven.

Een eenvoudige geïdealiseerde berekening geeft het volgende resultaat. Stel het materiaal bestaat uit n exemplaren die behoren tot m soorten. Het splitsen in soorten vergt dan volgens de langzame methode ongeveer 1/4 nm handelingen tegen 0,3 n log m handelingen met de snelle methode. Met een voorbeeld komen deze formules tot leven. Stel dat het materiaal bestaat uit n=500 exemplaren die in totaal tot m=50 soorten behoren. Iedere soort is dan gemiddeld met 10 exemplaren vertegenwoordigd. Als voor elke handeling één seconde nodig is, dan kost de langzame methode 2 uur arbeid, terwijl voor de snelle methode maar drie kwartier nodig is! De tijden in dit voorbeeld zijn natuurlijk niet reëel, maar het gaat mij er om het verschil in tijd te illustreren, dat bestaat tussen beide manieren van sorteren.

In het bovenstaande wordt met handeling bedoeld: het beslissen tot welke groep een bepaald exemplaar behoort en het er vervolgens aan toevoegen. Eén seconde voor een handeling is vrij kort en geldt zeker niet voor het scheiden van moeilijke soorten. Dit laatste valt dan ook geheel buiten het bestek van dit verhaaltje.

Tot slot de afleiding van beide formules. Te verdelen n voorwerpen in m groepen van elk n/m exemplaren middels r bakjes.

Snelle methode. Splits de n voorwerpen op in r groepen. Hiervoor zijn n handelingen nodig. Splits de inhoud van elk bakje met n/r voorwerpen weer op in r groepen. Aantal benodigde handelingen n. Aantal voorwerpen in elke groep nu n/r^2. Ga zo door. Na p keer de opsplitsing gemaakt te hebben zijn er pn handelingen verricht en bevinden zich ongeveer n/r^p voorwerpen in elke groep. We zijn klaar als $n/r^p = n/m$, want dan is alles in soorten opgesplitst. We vinden $p \approx {}^r\log m = \log m/\log r$. Totaal aantal handelingen $n\ {}^r\log m$. Met 2 schaaltjes wordt dit $n\ {}^2\log m = 0{,}3\ n \log m$.

Langzame methode. Afsplitsen eerste r soorten vergt n handelingen. Restant is n - rw/m. Afsplitsen in tweede ronde van r soorten neemt nu n - rn/m handelingen. Het subtotaal is 2n - rn/m. Ga zo door. Na p rondes is het restant n - prn/m. Bestaat dit net uit r soorten dan zijn er rn/m exemplaren en vinden we p = -1 + m/r. Totaal aantal verrichte handelingen is $(p + 1)n - (1 + 2 + 3 + \ldots\ldots p)\ rn/m = (mn/2r)\ (1 + r/m) \approx mn/2r$. Voor r = 2 volgen mn/4 handelingen voor het uitsorteren.

Geschreven is deze afleiding vermoedelijk onleesbaar. Wie het wil nagaan doet 't het handigste door een tabel te maken met de afgescheiden soorten, het restant en de in de laatste ronde ver-

--richte -

richte handelingen. Optellen van de tabel geeft het antwoord. De gevolgde methode is correct maar het antwoord is een benadering omdat werd aangenomen dat alles klopte en er net niets overbleef. In de praktijk gaat dat natuurlijk nooit op.

Naschrift van de redaktie

Voor het gemak van de lezers hebben wij voor U de belangrijkste formules getabelleerd:

Snelle methode.
Aantal handelingen is:
$n\ {}^{r}\log m = n \frac{\log m}{\log r}$

	m=50	m=100	m=150	m=200
r=2	5,7n	6,7n	7,3n	7,7n
r=4	2,8n	3,3n	3,6n	3,8n
r=8	1,9n	2,2n	2,4n	2,6n
r=16	1,4n	1,7n	1,8n	1,9n

Langzame methode.
Aantal handelingen is:
$n\ \frac{m}{2\ r}$

r=2	12,5n	25n	37,5n	50n
r=4	6,2n	12,5n	18,7n	25n
r=8	3,1n	6,2n	9,4n	12,5n
r=16	1,6n	3,1n	4,7n	6,3n

n= aantal individuen in het monster
m= aantal soorten in het monster
r= aantal schaaltjes, waarop wordt gesorteerd.

Uit deze tabellen blijkt het voordeel van het werken met een groot aantal schaaltjes, al is de tijdwinst bij de snelle methode bij meer dan 8 schaaltjes niet groot meer. Voorts blijkt de langzame methode inderdaad heel wat langzamer dan de snelle, en wel meer naarmate het aantal soorten groter is en het aantal schaaltjes kleiner. Bij m=200 en r=2 vraagt de langzame methode 50/7,7=6,5 maal meer tijd! Alleen bij een klein aantal soorten en een groot aantal schaaltjes komen beide methoden in elkaars buurt.

In de praktijk zal men veelal een geheel andere samenstelling van het monster ontmoeten, waarbij enkele soorten zeer goed vertegenwoordigd zijn, andere in slechts enkele exemplaren en de overige ergens daar tussen in. Eén van Uw redakteurs heeft er zijn duim eens op aangekeken, die hem vertelde dat in dat geval de langzame methode aanmerkelijk versneld kan worden door eerst de goed vertegenwoordigde soorten er uit te halen, doch dat de snelle methode er niet veel vlugger door zou kunnen. Het tijdsverschil tussen beide methoden zou er dus belangrijk kleiner door worden.

Ook vertelde zijn duim hem dat de invloed van het aantal schaaltjes bij de langzame methode er nog wat sprekender door zou worden.

Of die duim de waarheid kent weten Uw redakteuren echter niet, duimen mogen nu eenmaal spreken, ook al kennen ze de waarheid niet. Misschien ook sprak hij de waarheid en luisterde Uw redakteur niet zo goed. Hoe dit alles ook zij, misschien zoekt U het zelf eens uit aan de hand van een paar voorbeeldjes. Laat U ons het resultaat dan weten?

A.V.