

Informatievisualisatie

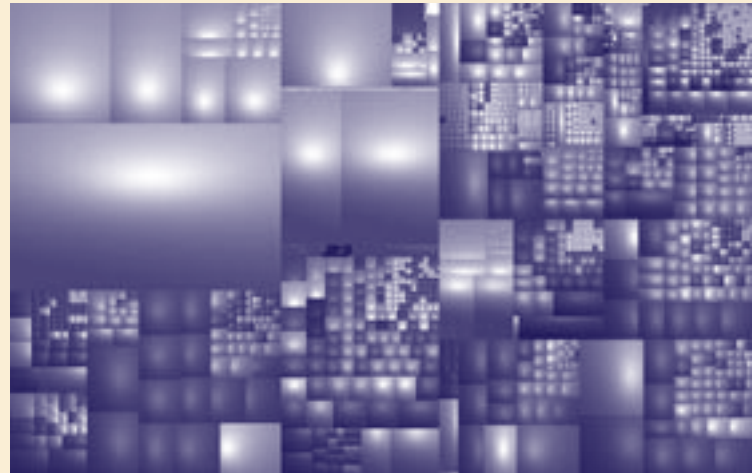
Jan Kraak j.kraak@rc.rug.nl

Informatievisualisatie was een van de onderwerpen van de Nationale Visualisatiedagen die eind november 2002 werden gehouden in de Zernikeborg in Groningen. Jack van Wijk, hoogleraar Visualisatie aan de TU Eindhoven, toonde daar de grafische afbeelding van een filesysteem als een treffend voorbeeld. Met deze afbeelding is in één oogopslag te zien welke files de meeste ruimte in beslag nemen. De afbeelding is daarom een ideaal hulpmiddel om grote, overbodige files, op te sporen en te verwijderen als uw harde schijf of uw ruimte op een fileserver vol dreigt te raken.

De afbeelding van het filesysteem is een zogenaamde 'squarified cushion treemap', die is te maken met het freeware Windows-programma SequoiaView.

Ruimtegebrek op schijf?

Harde schijven met meerdere gigabytes (Gb) fileruimte worden steeds goedkoper. Maar omdat onze behoeften vaak meegroeien met de mogelijkheden, raken deze grote schijven meestal even snel weer vol als vorige generatie schijven. Daarom blijven goede tools voor filebeheer nodig. Dit geldt vooral voor het beheer van fileservers, waarvan er momenteel ongeveer 130 zijn binnen de Rijksuniversiteit Groningen. Hierop worden files van medewerkers en studenten bewaard. Geregeld



Figuur 1. Zwart-wit afbeelding van een hiërarchisch filesysteem als een squarified cushion treemap, gemaakt met SequoiaView. De filetypes kunnen met kleuren worden getoond.

wordt er een backup van de data gemaakt op tape, zodat er bij computerstoringen zo weinig mogelijk waardevolle gegevens verloren gaan. Grootschalige backup-faciliteiten zijn tamelijk prijzig en vaak minder gemakkelijk uit te breiden dan schijfruimte. Ondanks het feit dat veel fileservers ruim zijn bemeten, komt het geregeld voor dat een systeembeheerder een verzoek moet doen aan zijn gebruikers om het gebruik van de fileserver te beperken en files die niet strikt nodig zijn te verwijderen.

Bijna iedere computergebruiker is bekend met Windows Explorer die de directory-structuur toont door middel van inspringen. Explorer toont verder een lijst van files in een bepaalde directory. Via de rechter muisknop krijgt men informatie over de aangewezen file, zoals de grootte en de aanmaak-

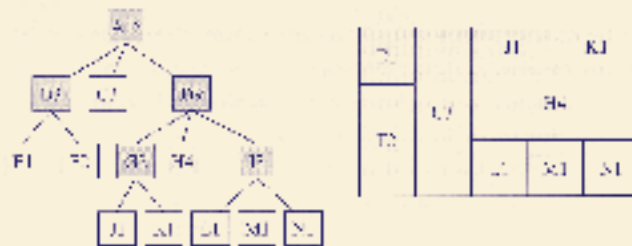
datum. Aldus kan men met Explorer stap voor stap een filesysteem exploreren, om knelpunten te vinden.

Een belangrijk bezwaar van Explorer is dat men geen globaal overzicht krijgt van de gehele inhoud van het filesysteem, omdat het aantal files en directories dat gelijktijdig kan worden getoond beperkt is, vanwege de beschikbare ruimte op het scherm. Zonder een globaal overzicht, zijn vragen als 'Welke directory neemt de meeste ruimte in?' of 'Waarom is mijn harde schijf of mijn beschikbare ruimte op de file server vol?' niet snel te beantwoorden. Vooral als het om duizenden files gaat, hetgeen tegenwoordig niet ongebruikelijk is.

SequoiaView voor Windows

SequoiaView biedt een elegante oplossing voor dit probleem door de beschikbare ruimte op een

met treemaps



Figuur 2. Links een boomstructuur en rechts de daarmee corresponderende treemap.

computerscherm optimaal te gebruiken. In plaats van het tonen van het filesystem door middel van een lange lijst van files en directories, worden files afgebeeld door middel van rechthoekjes, waarvan de oppervlakte evenredig is met de file-grootte. Vrijwel het gehele computerscherm wordt benut en in principe kan elke pixel een file weergeven. Er kunnen dus zeer veel files tegelijk worden afgebeeld.

Vanwege de evenredigheid van de oppervlakte met de filegrootte, zien we onmiddellijk welke files de meeste ruimte in beslag nemen. Lang vergeten grote files komen zó aan het licht. Zelf vond ik een aantal 'hidden' files van Word, die standaard niet door Explorer worden getoond. Een collega, die last had van een volle schijf, ontdekte dat hij een installatiefile van 3 Gb had laten staan.

Files in dezelfde directory staan gerangschikt binnen een rechthoek. Door aanwijzen met de muis wordt de omtrek rood. Na het aanklikken van een file in een bepaalde directory wordt ingezoomd op de betreffende directory. Van een file waar de muis op staat worden de eigenschappen onder aan het scherm getoond.

SequoiaView werkt onder Windows 95/98/2000/XP alsmede onder Windows NT, op een scherm van tenminste 640*480. Downloaden (zie onder voor adres) gaat snel, omdat het programma slechts ongeveer 0.5 Mb groot is. Installatie op de harde schijf kost

ongeveer 2 Mb. Het programma heeft een fraaie gebruikersinterface en een goede Help-faciliteit. Daarom kost het weinig tijd om er mee te leren omgaan. Om een overzicht van een filesystem te krijgen, hoeft u alleen de naam op te geven, bijvoorbeeld 'c:\' voor een harde schijf of 'g:\' voor een Novell fileserver.

SequoiaView is genoemd naar de mammoetbomen van Californië. Ze kunnen zo hoog worden als de Martinitoren; de laagste takken beginnen vanaf ongeveer 45 meter. Een Linux-versie van de treemap heet 'qtreesmap'. Dit is een library waarmee cushion treemaps in eigen programma's zijn te maken. Het Linux-programma 'kdirstat', mits dat gebruik maakt van qtreesmap, biedt ongeveer dezelfde mogelijkheden als SequoiaView voor Windows.

Toon me het plaatje van uw files

SequoiaView kent behalve de afbeelding doormiddel van een oppervlakte, nog veel meer mogelijkheden. De kleur van een rechthoek toont bijvoorbeeld het type file. Een Word-document is blauw, een JPEG-file is rood en een programma groen etc. Iedereen krijgt zó een eigen karakteristieke afbeelding met SequoiaView ("Toon me het plaatje van uw files, en ik zal zeggen wie gij zijt").

SequoiaView kent ook filters. Zo kan men bijvoorbeeld kiezen om alleen files groter dan 1 Mb en ouder dan één jaar af te beelden. Dit

is slechts één van de mogelijkheden van het programma.

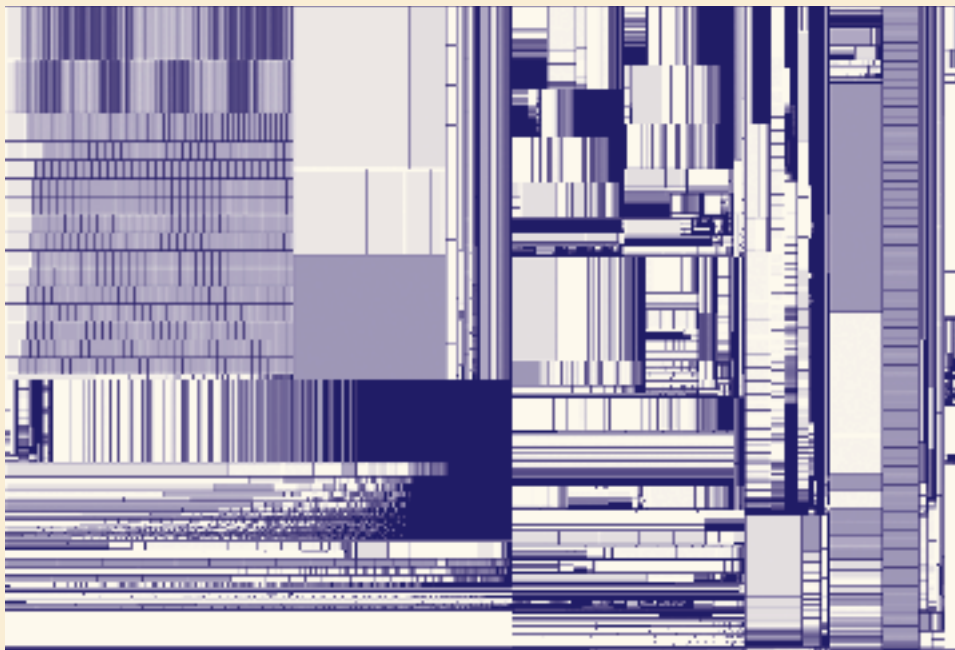
Met SequoiaView zelf kan men geen files verwijderen of andere file-bewerkingen uitvoeren. Door echter een file met de rechter muisknop aan te klikken, verschijnt er een menu met onder meer als opties het openen van de file en het activeren van Explorer, waarna men bijvoorbeeld de aangewezen file kan verwijderen.

Treemap

Ik ga nu wat dieper in op de visualisatie van hiërarchische gegevens. Als u SequoiaView alleen voor file-beheer wilt gebruiken, dan is het niet strikt noodzakelijk om verder te lezen.

Vele soorten gegevens zijn te ordenen in een zich vertakkende boomstructuur. Niet alleen files op een harde schijf, maar ook personen in een organisatie, aandelen, de voorraad van een grootwinkelbedrijf etc. Ben Shneiderman maakte de eerste treemap van een boomstructuur in 1992.

Figuur 2 toont links een boomstructuur met knooppunten en rechts de overeenkomstige treemap. Elk knooppunt is afgebeeld door een rechthoek, waarvan de oppervlakte evenredig is met een attribuut van het knooppunt, zoals de (file)grootte. Elk knooppunt heeft een label X_n , waar 'X' de naam is en 'n' de (attribuut)grootte. De grootte van een zich vertakkend knooppunt is gelijk aan de som van de groottes van de kinderen.



Figuur 3. Prototype van de treemap.

➤ Een treemap is als volgt geconstrueerd. We beginnen met een gegeven rechthoek die het begin van de boom A, de *root*, voorstelt. Vervolgens verdelen we de rechthoek van links naar rechts in drie kleinere rechthoeken die de kinderen van A, B, C en D voorstellen. Elke rechthoek heeft een oppervlakte evenredig aan de som van de groottes van zijn kinderen. In de volgende verdelingen herhalen we dit proces tot de onderverdeling stopt, ofwel als we op een eindknooppunt (*leaf*) zijn aangekomen. De onderverdeling vindt afwisselend horizontaal en verticaal plaats. Bij een filesysteem stelt elke binnenste rechthoek een file voor. Uit de wijze waarop de binnenste rechthoeken 'samenklonteren' tot grotere rechthoeken, kan men de directory-structuur aflezen.

Voor bezitters van aandelen is een treemap van de aandelenmarkt van belang. Per marktsector zijn beursgenoteerde firma's weergegeven door een rechthoek, waarvan de oppervlakte evenredig is met de totale waarde van de aandelen. Zoals te verwachten heeft Microsoft de grootste rechthoek

in de technologiesector, IBM is tweede etc. Kleur codeert het koersverloop, groen betekent winst en rood verlies.

Squarified cushion treemap

Informatievisualisaties worden dikwijls ontworpen en geïmplementeerd door onderzoekers met een primaire belangstelling voor bijvoorbeeld cognitieve psychologie, informatica of grafisch ontwerp, maar zonder een grondige kennis van computer graphics. Daarom is de toegevoegde waarde van die visualisaties soms beperkt, en moet men enige tijd studeren voordat men het nut ervan ontdekt, terwijl het kenmerk van een goede grafische weergave juist is dat de bedoeling onmiddellijk duidelijk is.

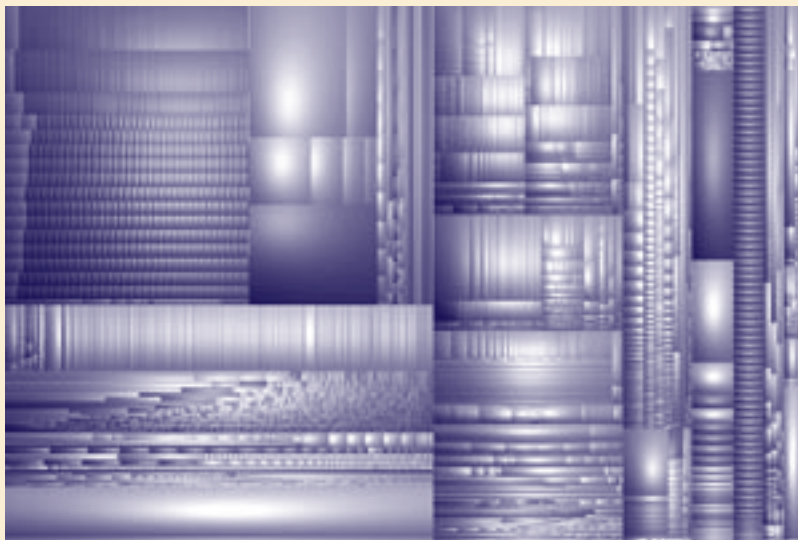
Voordat Jarke J. (Jack) van Wijk zich met informatievisualisatie ging bemoeien, was hij al lang werkzaam geweest op visualisatiegebied, ondermeer op het ECN in Petten. Arthur Veldman, hoogleraar Toegepaste Mechanica, liet tijdens een lezing op de Visualisatiedagen een stroming zien, die was gevisualiseerd met de *spot noise*-methode van Van Wijk. Zijn lange ervaring met

computer graphics kwam Van Wijk goed van pas in zijn nieuwe onderzoeksgebied. Samen met zijn groep paste hij de treemap van Shneiderman zodanig aan, dat het een nuttig gereedschap werd in de vorm van het programma SequoiaView.

Figuur 3 toont het prototype van de treemap. Het vergt enige studie om de directory-structuur te halen uit de 'samenklonteringen' van vlakken. Daarom beeldt SequoiaView een rechthoek af als een opbollend kussen. Bovenop dit kussen worden onderverdelende rechthoeken afgebeeld door kleinere kussens. Dit proces wordt herhaald. Aldus wordt een hiërarchie afgebeeld als een stapeling van kussens, een zogenaamde *cushion treemap*, (zie figuur 4).

De cushion treemap wordt beschreven door een lamp, waardoor schaduwen ontstaan. Waar kussens op hetzelfde niveau elkaar raken, is de kloof niet diep en is er weinig schaduwwerking. De kloof tussen kussens dicht bij de root is diep en de schaduwwerking sterk. Daarom kan men uit een cushion treemap beter de hiërarchie aflezen dan uit het prototype. Maar afgezien daarvan, ziet de stapel kussens er erg mooi en 'aaibaar' uit, vooral als de kussens zijn gekleurd. Op de SequoiaView-internetsite staat een plaatje van de huiskamer van Birgit en Marcel Rieder, met boven de (IKEA?) zitbank de ingelijste cushion treemap van hun harde schijf.

Verder heeft het prototype van de treemap in figuur 3 als nadeel dat kleine files dicht bij de root aanleiding geven tot lange zeer smalle rechthoeken, die vrijwel niet van elkaar zijn te onderscheiden. Als remedie hiertegen kent SequoiaView de 'squarify'-optie die rechthoeken zo goed als mogelijk in vierkanten omzet (figuur 1). Lan-



Figuur 4. Cushion treemap, corresponderende met het prototype uit figuur 3.

ge zeer smalle rechthoeken worden nu vierkantjes die beter zijn te onderscheiden. Met SequoiaView zijn de drie varianten van de treemap te maken.

Natuur als inspiratie

Met de uitvinding van de cushion treemap is nog lang niet het laatste artikel geschreven over de visualisatie van hiërarchische datastructuren. Ook al geven de opbollende kussens de hiërarchie redelijk weer, toch gaat het 'in één oogopslag'-criterium hier niet helemaal op. Enige studie blijft nog steeds gewenst. Daarom zet de groep van Van Wijk zijn werk voort aan hiërarchische structuren.

Van Wijk toonde tijdens zijn lezing twee experimentele methoden. In de eerste plaats de zogenaamde *beam tree*, waarbij de derde dimensie goed is benut. Interessanter is de andere methode, de *botanical tree*, waarvoor een boom uit de natuur model staat (figuur 5).

Mensen zijn gewend aan bomen en kunnen daarin snel bepaalde aspecten ontdekken. Zo wordt bij een botanical tree van een filestelsel de afzonderlijke files afgebeeld als een zwerm vruchten in een bol, waarvan de inhoud even-

redig is met de filegrootte en de kleur het filetype aangeeft.

De reacties op de botanical tree methode zijn tot nu toe gemengd, aldus Van Wijk. Iedereen vindt de plaatjes mooi, maar voor praktische toepassingen geven velen de voorkeur aan cushion treemaps. Belangrijk is evenwel de constatering dat de natuur, en onze bekendheid daarmee, een rijke inspiratiebron vormt voor de vertaling van abstracte gegevens in geometrische vormen.

Te zijner tijd zullen er, net als het programma SequoiaView voor de treemap, implementaties komen voor de beam tree en de botanical tree, die vrij zijn te downloaden. De groep van Van Wijk heeft veel geleerd van het commentaar van gebruikers bij de ontwikkeling van SequoiaView. Tevens was het erg stimulerend. Ze verwachten hetzelfde effect van de twee toekomstige implementaties.

Informatievisualisatie aan de RUG

Een hiërarchische structuur is één van de soorten ordeningen van gegevens, die zich goed leent voor visualisatie. Grote collecties documenten, bijvoorbeeld alle webdocumenten over een be-

paald onderwerp, kan men met bepaalde methoden visualiseren, zie hiervoor mijn artikel 'Voor de dolende webspeurder - visualisatie van documenten'.

In het Informaticacollege 'Wetenschappelijke Visualisatie', gegeven door prof.dr. Jos Roerdink, wordt een overzicht gegeven van de voornaamste methoden van informatievisualisatie.



Figuur 5. Botanical tree.

Links:

- Het freeware Windows-programma SequoiaView: www.win.tue.nl/sequoiaview
- Een Linux-versie van de treemap, 'qtreamap': qtreamap.sourceforge.net/
- De eerste treemap van een boomstructuur gemaakt in 1992 door Ben Shneiderman: www.cs.umd.edu/hcil/treemaps/
- Een treemap van de aandelenmarkt kunt u vinden op www.smartmoney.com/marketmap
- Jarke J. (Jack) van Wijk: www.win.tue.nl/~vanwijk
- Het artikel 'Voor de dolende webspeurder - visualisatie van documenten': www.rug.nl/hpc/projects/docvis
- Prof.dr. Jos Roerdink: www.cs.rug.nl/Onderwijs/Roerdink/scivis.html