

de databank van de new york times

dony van hoof

1. Traditionele Problemen en Behoefte van de Krant.

Een grote krant zoals de New York Times heeft voor haar dagelijkse berichtgeving behoefte aan een enorme hoeveelheid informatie. Die wordt haar verstrekt door haar reporters en de traditionele bronnen zoals de persagentschappen. Dit is lang niet voldoende. Voor haar historische gegevens en achtergrondinformatie heeft de New York Times (NYT) de beschikking over een uitgebreid archief, in het jargon van haar staf de «morgue» (dodenhuisje) genoemd. Dit archief gaat terug in de tijd tot 1900. De traditionele problemen van zulk een archief zijn plaatsruimte en gebrekkige en zeer dure retrieval. De archiefstukken zijn ofwel knipsels of het hele blad, zowel het eigen blad als andere belangrijke publikaties. Voor de NYT neemt het archief reeds enkele etages van een torengebouw in beslag, en hoe uitgebreider het wordt des te moeilijker en duurder wordt het opzoeken. Bovendien zit er geen systematiek in het klasseren: de onderwerpen waarnaar het archief is opgebouwd, werden volgens een volkomen arbitrair selectiepatroon gekozen. Iedere week komen ongeveer 15.000 archiefstukken de morgue aandikken. Reeds meer dan 27 miljoen stuks zaten er een paar jaar geleden in het archief: dit neemt enorm veel plaats in beslag. Buiten de reeds vermelde hoofdproblemen zijn er nog een paar secundaire (echter niet minder belangrijk), nl. de beslissing wat te behouden en wat weg te gooien, en ook de fysische overleving van de knipsels na enkele jaren gebruik. Bovendien is die informatiebank alleen toe-

gankelijk voor de staf van de NYT (400 man) en kan, omwille van de te hoge exploitatiekosten, niet economisch rendabel gemaakt worden. Een heleboel geld ligt aldus te rusten.

Het klasseren van de stukken is een probleem op zichzelf: de knipsels worden gestockeerd per onderwerp en per naam, wat meebrengt dat de meeste in meer dan één file terecht komen, sommige zelfs in negen verschillende bestanden. Daar is natuurlijk op de duur geen orde meer in te houden. Daarom besloot de directie van de NYT in mei 1965 over te gaan tot de studie van de mogelijkheid om het archief om te bouwen tot een computergestuurde databank. Het zou een on-line databank dienen te zijn, d.w.z. de gebruiker zou ten allen tijde moeten in staat zijn de bank te raadplegen.

De directie zag deze beslissing om drie redenen gerechtvaardigd, hoewel men vooraf wist dat dit een bom geld zou kosten:

- de bank zou betere zoekmogelijkheden bieden dan de morgue;
- ze zou bovendien een besparing betekenen;
- ze zou ook van buitenuit toegankelijk kunnen gemaakt worden en aldus zou ze economisch rendabel worden.

In juni 1966 werd er, na een heleboel werkzaamheden zoals bv. het kiezen van de consultant — in dit geval IBM — gestart met het zogenoemde toepasbaarheidsonderzoek (feasibility study). Faze 1 werd afgesloten in januari 1968; de software- en hardwarecontracten werden afgesloten in maart 1969. De basis-software of programma's waren beëindigd in juni 1971 en de eerste klant (de CIA!) werd ge-

boekt in februari 1973. Wat de mogelijke gebruikers van de databank betreft, zijn er geen beperkingen. Ook kranten zijn kandidaten. Immers, de informatie die in de bank opgestapeld is, is publiek.

Zo zijn de volgende kranten gebruikers: de Detroit Free Press en de Louisville Courier Journal; de persagentschappen Reuter en Associated Press; het nieuwsmagazine Time; het TV-station NBC. De eerste Europese klant is Der Stern.

Wat zit er zoal in de bank?

In principe gaat er in de bank wat er vroeger in ging, behalve dat de gecomputeriseerde databank veel meer mogelijkheden biedt op het stuk van het aantal bronnen. Het eigen blad is nog steeds de hoofdbron. Er worden nu nieuwe bronnen toegevoegd.

Wat het teruggaan in de tijd betreft, worden er in principe geen beperkingen gesteld, en zal gaandeweg de hele oude morgue worden toegevoegd. Maar eerst komen de recentere gegevens, want het is gebleken dat de klanten meer belang hechten aan de nieuwste informatie. Men gaat ervan uit dat men een databank in de breedte wil uitbouwen, met het accent op de recentste gegevens. Wat het archief van de niet-NYT bronnen betreft, heeft men vooralsnog 1971 als eindgrens gesteld.

Wat gaat de bank de gebruiker kosten?

De klant moet beschikken over een video-terminal: een schrijfmachineachtig eindstation dat zowel een klavier als een videoscherm heeft (zie illustratie).

Via een gewone telefoonlijn krijgt de klant verbinding met de computer en kan dan zijn vragen aan de bank

stellen. De klant heeft de keuze tussen twee soorten terminals. Men kan via de terminal slechts output krijgen. Dus de klant kan de databank niet wijzigen. Het systeem werd ontworpen door IBM Federal Systems Division en wordt genoemd 'Digital TV Display System'. Op dit moment heeft de NYT 64 terminals in gebruik: 21 in de redactie, 28 in de indexeringsdienst en de rest verspreid. De NYT terminals kunnen zowel in- als output geven. Normaal krijgt de klant zijn informatie op het video-scherm, maar hij kan er ook een hard copy printer verkrijgen. Hij kan ook over een microfiche beschikken, waarvan eveneens een hard copy kan gemaakt worden.

Toegang tot de bank wordt betaald op meter-basis: de eerste 10 uren per maand kosten \$45/uur; meer dan 50 uren \$30. De gemiddelde kost per maand aan 20 u./maand is \$8.000 per jaar, aan meer dan 40 u. per maand \$15.000. Interessant om weten is dat de doorsnee-gebruiker ongeveer 6-8 minuten nodig heeft per opzoeking. De telefoonkosten komen daar natuurlijk nog bij.

2. Hoe is de NYT Databank opgebouwd?

De indexeringsafdeling werkt voortdurend de databank bij. Het NYT-materiaal wordt dagelijks up to date gebracht, en wordt rechtstreeks via de terminals ingegeven. Het niet-NYT materiaal wordt via ponskaarten ingegeven. Hier geschiedt de updating 2-3 maal per week.

De indexeersers krijgen fotocopies van de knipsels; na lectuur redigeren ze een abstract rechtstreeks op het scherm. Zij voegen er alle noodzakelijke indexcodes aan toe en de bibliografische gegevens, en geven dan het geheel in het systeem. Het materiaal komt in een tijdelijk opslag terecht. Er volgt een print-out voor correcties en dan pas wordt het definitief in de bank ingevoerd.

Wat wordt zoal als index-woorden gebruikt?

Meestal zijn het sleutelwoorden die in zeer beknopte vorm het essentiële van de tekst weergeven: topics, eingenamen, namen van organisaties, geografische namen...

Nemen we een eenvoudig voorbeeld hoe de gebruiker via zulke keywords aan de informatie kan komen.

De opzoeker geeft in: BPN (GNP). Eerst wordt deze term getest tegen de thesaurus, d.i. alle woorden die als sleutelwoorden werden weerhouden en die cumulatief in een lijst terechtkomen. Als hij aanwezig is, wordt op het scherm geprojecteerd: «Is this the term you mean; if so accept it». Meestal zal gelijk welk sleutelwoord meer dan 1-maal in de lijst zitten. In dat geval worden alle termen die met dat woord beginnen gegeven. De gebruiker selecteert dan die term die hem het best lijkt. Bv. GNP/France. Dan wordt hem gevraagd: «How do you wish 2 terms combined logically?». Beschikbaar zijn de Booleaanse operators and, or, or not. Dus zegt hij: GNP and France. De computer geeft dan bv. 73 items en vraagt «How do you wish them arranged? Latest date first, or earliest date first?» Dan krijgt hij een abstract van het stuk dat hij als eerste heeft geklasseerd. Als de nodige informatie er is, hoeft hij niet verder te gaan. Men kan via de microfichecode de microfiche met de volledige tekst krijgen en ook een hard copy daarvan.

In de bibliografische gegevens die aan iedere abstract worden gehecht vindt men: bron, datum, pagina, kolom. Welk soort artikel het is: een editoriaal, een overzicht, een brief... En ook het microfichenummer.

Niet zomaar iedereen kan toegang tot de bank krijgen: ieder legitiem gebruiker heeft een nummer en een 'password' (wachtwoord). Dat kan naar believen gewijzigd worden.

3. Iets over het Systeem.

De machine is een IBM 370/145 (384 K geheugen) met een 3330 schijvengeheugen waarop de databank zit. Het werd door de bovengenoemde groep van IBM ontwikkeld, maar strikt volgens de specificaties van de NYT. Aan het ontwerp hebben 8 mensen 6 maanden gewerkt (een totaal van 21 manjaren). IBM stelde 4-8 mensen gedurende 26 kalendermaanden ter beschikking. Het probleem was sterk onderschat: IBM meende dat 4 personen gedurende 18 maanden zouden volstaan. Het was wat men noemt een vaste-prijs contract. Het is wellicht het laatste dat IBM verder nog zal afsluiten. De NYT heeft nu zelf een DP-team dat het systeem onderhoudt en bijwerkt. De software

is eigendom van de NYT. Het systeem wordt op dit moment als bevredigend aangezien.

Nochtans zijn er nog problemen.

Er is namelijk altijd een verschil tussen wat het systeem geeft en wat de gebruiker wil. Er bestaat voor een zodanig systeem geen model waarvan men kan afkijken. Zelfs de volgende bank voor een andere krant zal verschillend zijn. Een van de grootste problemen was, te voorzien hoe groot de bank uiteindelijk wel zou worden: het is ondertussen klaar geworden dat de bank jaarlijks zal groeien met ongeveer 250.000 abstracts. De antwoordtijden van het systeem gaan van virtueel niets tot een halve minuut.

4. De Rentabiliteit van het Systeem.

Een zeer heet hangijzer niet alleen bij de NYT maar in de hele computergemeenschap. Sommige mensen beweren dat de NYT nooit de kosten zal kunnen recupereren, laat staan winst erop te maken. Nochtans heeft de directie van in het begin gezegd dat dit een project op lange termijn was dat pas na jaren een dividend zou opbrengen. Volgens deze mensen zal informatie in de toekomst één van de duurste betaalde goederen worden omdat slechts weinig instellingen genoeg financiële middelen zullen hebben om een bevredigend apparaat op te bouwen.

De databank heeft over de periode 1966-73 de NYT de ronde som van \$5 miljoen gekost. Nu wordt jaarlijks een budget van \$2 miljoen voorzien om ze te onderhouden en uit te breiden. Het is het grootste project dat de NYT ooit heeft ondernomen. Op dit ogenblik moeten er ongeveer 200 gebruikers meer bijkomen om het systeem rendabel te maken. (N.B. De oude morgue loopt nog steeds parallel!)

Wie zijn nu de potentiële klanten? In de eerste plaats de nieuwsmedia. In feite zal dit altijd in de optiek van de sponsors een kleine groep blijven. Men gokt op de markt van de big business: meestal wordt een onderneming abonnee langs één afdeling, bv. de public relations. Spoedig gaat ook de juridische afdeling gebruik maken van de bank voor problemen in verband met anti-trustwetgeving of copyrightproblemen. De voorzitter moet een belangrijke speech houden en wil

graag de juiste en recentste gegevens. Voor vestigingspolitiek wil men exacte data over milieu, energie of andere problemen. Bij vestigingen in het buitenland wil men de VSA-ambassadeur kennen enz. En de firma zit hals over kop in de databank en kan ze niet meer missen.

Welke bronnen zitten er zoal in de bank ?

- In de eerste plaats de NYT zelf ;
- de Washington Post, de Los Angeles

Times, de Wall Street Journal, de London Times ;

- de nieuwsmagazines zoals Time ;
- de economische bladen zoals Fortune ;
- gespecialiseerde bladen in verband met wetenschappen, buitenlandse zaken, economie...

Men wil de bronnen nog verder uitbreiden, ook tot Europese.

Hoe zien de sponsors van deze DB de toekomst op dit gebied ?

Een toekomstige databank moet in staat zijn niet alleen de geschreven documenten op te bergen en gemakkelijk toegankelijk te maken, ook foto's, films, videotapes moeten in de toekomst digitaal gestockeerd worden en even gemakkelijk toegankelijk worden als de beschreven bronnen. De mensen van de NYT geloven rotsvast in de toekomst van de databank die zij in het leven geroepen hebben, en die door de computer wordt beheerd en gestuurd. □

Aan de VPRO, met bewondering voor het geleverde werk

De VPRO, de begaafde benjamin onder de oude Nederlandse omroepen, bestaat (15 juli 1926) 50 jaar. Feest dus. Dat wil zeggen: een plechtige bijeenkomst, feestelijk geroezemoes in de ether, een gedenkboek en felicitatie-stukjes. Want evenmin als bij een overlijden wordt men geacht bij een jubileum erg kritisch te zijn. Daar is in dit geval ook niet zoveel reden toe. Natuurlijk hebben wij voldoende kabelcontact met Holland om te weten dat het ook bij de VPRO niet altijd koek en ei is, dat niet alles volmaakt is. Detailkritiek is er op 'n totaalpakket van 11 uren radio en 2,5 uren televisie genoeg te leveren. Maar het staat eigenlijk buiten kijf dat de VPRO best een en ander heeft gedaan en bereikt. In de prijzen voor de beste radio- en televisieprogramma's zijn de VPRO-pioniers bij herhaling de sterke winnaars. Zij hebben van de stoomradio een fascinerend communicatiemedium gemaakt en verzorgen op de televisie uitzendingen die technisch en inhoudelijk voorop lopen in Nederland. Dat Nederland op het gebied van omroepwerk in de wereld op veel plaatsen redelijk vooraan staat, is voor een deel mee te danken aan de gouden VPRO. Met de avant-garde en de Blokkeriaanse avonden, de prima ingekochte Tweede Oorlog propagandafilms, de Archie-serie, met Culemborg, de Vrijdagse radiogeluiden, met programmafavorieten Kees en Wim - dit lijstje is kort, maar kan veel langer - laat de VPRO zien hoe goed het is dat er in één land grote én kleine omroepen zijn.

Gelukkig verlost van de zee van sport, nieuws en amusement maakt de VPRO bij tijden van het huismeubel TV in plaats van een «treurbuis» een gezegend instrument van bevrijding, waarvoor veel individuele krampachtigheid zich durft te ontspannen.

In België weerspiegelt het kijk- en luisteronderzoek van de BRT duidelijk de populariteit van de Nederlandse omroepen. Uit enquête-gegevens blijkt dat de portie die van de totaal gerealiseerde kijktijd wordt besteed aan door Nederlandse zenders uitgezonden programma's varieert van 15% tot 25%. Het laat zich licht berekenen dat de moderne VPRO-helden Fred Haché, Barend Servet, Sjef

van Oekel en de heren Kooten en Bie in die ruimte een belangrijke plaats innemen. Zij worden gevolgd met een zekere gespannenheid, vol hooghartige minachting, streng vorsend, sceptisch blikkend, met mildheid, berusting, met natte ogen of met een knipoog. Ook de opinievormende critici schrijven in gelijke mate met verbazing, hard doorprikkend, vol wrevel, met knorrigheid, met liefdevol begrip of gemelijke weerzin over die vreemde, vieze, negatieve, pornografische, beledigende, sarcastische, opruiende etc. programma's en de daaraan medewerkende onfrisse, niezerige en misselijke jongens. Belangwekkende staaltjes daarvan leveren alle Vlaamse dag- en weekbladen. De VPRO boeit, verwart en beangstigt ook hier.

In verscheidene opzichten is dat gemis aan onbevooroordeeldheid ook erg geruststellend. Het zou jammer zijn als een Belg de VPRO beter begreep en taxeerde dan Nederlanders doen.

Die Vlaamse belangstelling voor de Nederlandse omroepprogramma's was voor de VPRO zelfs een aanleiding tot het instellen op 12 juli 1971 van een omroepdistrict België en het aandringen op de uitbouw van een gedifferentieerd Hilversums programma-aanbod over België waarbij ook de Nederlanders baat zouden vinden. Daar heb ik toenmaals in «vrije geluiden» en de VPRO-ledenvergadering nog voor gepleit. Met Radio Vrij België en de schrikbarende verhalen van Walter Slosse over zijn land, waar het maar niet pluis wil worden, zette de VPRO dit voornemen in daden om. Was daarmee mijn wens vervuld? Als ik over die ontwikkelingshulp nadenk, geregeld en steeds meer wauwelwoorden, vermeend onrecht, doldriftig idealisme in het toontje van «het beter weten» hoor in plaats van omroepjournalistiek, zeg ik neen. De VPRO is geen schouder om op uit te hullen. Desondanks zou het verkeerd zijn het hoofd te laten hangen. Indien érgens, dan is er bij de VPRO de mogelijkheid om fouten te verbeteren, zwakke plekken te herstellen, vernieuwingen tot stand te brengen.

Herman VAN PELT