

De vele kanten van information retrieval

Information retrieval is een van de belangrijkste gereedschappen dat de informatie professional bij zijn werk ten dienste staat. Een goede reden om een heel themanummer van *Informatie Professional* aan dit onderwerp te wijden.

VANAF DE JAREN zestig tot nog maar enkele jaren geleden, werd dit gereedschap vrijwel uitsluitend door informatie professionals gebruikt. Pas sinds de opkomst van het world wide web werkt vrijwel elke computergebruiker er mee. Voor de oorspronkelijke gebruikersgroep is het, ondanks wat kleine vernieuwingen, twintig jaar lang een tamelijk stabiel gereedschap geweest, met weinig revolutionaire veranderingen. Wie in 1975 voor het laatst bij een on line host gezocht zou hebben, had in 1995 zonder al te veel problemen de draad weer kunnen oppakken. Hoewel ook in die jaren al veel fundamenteel en toegepast wetenschappelijk onderzoek naar information retrieval werd gedaan, worden nieuw ontwikkelde technieken pas op veel grotere schaal in praktische producten toegepast sinds de enorme www-gebruikersgroep is ontstaan. En dat heeft het beeld wel heel grondig gewijzigd.

Aan information retrieval zitten veel verschillende facetten en het kent veel uiteenlopende toepassingen. Het wordt gebruikt om grote gestructureerde bibliografische databases bij on line hosts of op cd-rom te doorzoeken. Het wordt gebruikt voor het doorzoeken van ongestructureerde full-text documenten. Die documenten kunnen zowel wetenschappelijke artikelen zijn, als de door een afdeling van een bedrijf verstuurd brieven die ergens bij een secretaresse op de harde schijf staan. En retrieval wordt al evenzeer gebruikt voor het zoeken naar informatie in een volstrekt ongeordende, ongestructureerde en onbeperkte vergaarbak als het world wide web.

Er kan ook vanuit allerlei verschillende invalshoeken naar information retrieval worden gekeken. In de eerste plaats kan dat vanuit het gezichtspunt van de informatie zelf, zoals in de vorige alinea werd gedaan. Het kan ook vanuit het gezichtspunt van de ontsluiting van informatie. Welke voorbewerking moet informatie ondergaan om het met retrievalssystemen beter terugvindbaar te maken en is zulke voorbewerking bij de nieuw ontwikkelde technieken eigenlijk nog wel nodig? Dit aspect komt zelfs in verschillende bijdragen terug, waarbij de nadruk ligt op mogelijkheden tot automatische voorbewerking.



FOTO: EGON VIEBRE

Een andere invalshoek gaat uit van de verschillende technieken die in moderne software voor information retrieval worden toegepast. Zeker niet de minst belangrijke invalshoeken, tenslotte, zijn die van de gebruikers van de terug te vinden informatie: de informatiebehoefte van die gebruikers en de door hen uitgevoerde zoekprocessen. Ook al deze invalshoeken komen op een of andere manier in de artikelen in dit themanummer aan de orde – zelfs in de *Opinie*, *Column* en de eerste aflevering van de nieuwe serie ‘Herlezen’.

En er komt nog meer!

Er was namelijk zo veel materiaal voor dit nummer – en naar aanleiding van de aan dit nummer gekoppelde *IP-Lezing* verwachten we nog meer materiaal –, dat de lezer ook in het volgende nummer van *Informatie Professional* nog enige bijdragen kan verwachten die bij dit thema aansluiten. Bovendien kan ik nu al opzienbarende resultaten aankondigen van een onderzoek naar irreproducibiliteit en wispelturigheid van veel bekende zoekmachines op het world wide web, resultaten die net niet meer op tijd zodanig konden worden verwerkt, dat ze op verantwoorde wijze in dit nummer konden worden opgenomen.

Mede namens de andere leden van de deelredactie voor dit themanummer – drs. Hans van der Laan, dr. Paul Nieuwenhuysen en drs. Marten Hofstede – wens ik u veel nuttig leesplezier.

Eric Sieverts