

De "andre" søkemaskinene

I forrige Bok og Bibliotek skrev Even Flood en del om de store søkemaskinene. Spesielt om Google og om Yahoo som via oppkjøpt av Inktomi, AltaVista og Alltheweb seiler opp som en av de store konkurrentene til Google. I denne artikkelen følger han opp med å omtale en del av de andre søkemaskinene som prøver å komme inn på markedet. Felles for dem er at de er en del mindre enn de to store og dekker ikke så mange vevsider, eller at kvaliteten på databasen er ikke så god når det gjelder oppdateringer. Men mange av disse representerer en fornyelse av søkemaskinene når det gjelder enten søkegrensesnittet eller hvordan resultatene velges ut og presenteres. Begge deler er viktig, men det er i oppsummering og presentering av resultatene at det mest spennende skjer.

Av Even Flood, *førstebibliotekar*

Informasjonsmengden på veven er (det er ikke gjentatt ofte nok!) enormt. Noe tall har jeg ikke, men når man opererer med mengder på flere milliarder vevsider er tallene allikevel meningsløse. Og det blir stadig vanskeligere i en mengde av millioner av svar å finne de mest relevante. Her er det spennende og kreative algoritmer kan hjelpe. Disse "andre" søkemaskinene har i en del tilfelle laget sine egne databaser, i andre bruker de eksisterende (Google, Inktomi) eller de er metasøkemaskiner som søke i flere baser og leverer tilleggstjenester i listen over resultater.

Felles er at det ikke har skjedd noe på med søkeprofilene. Med AltaVista mistet vi den siste søkemaskinen som hadde trunkering og den eneste som kunne skille mellom små og store bokstaver. For oss gamle søkere som er bortskjemt med å finpusse og variere søkeprofilene for mest mulig presiste resultat er dette et stort tap, men vi får leve med det. Jeg vil derfor si lite om input, det er liten forskjell.

Så til de forskjellige maskinene.

Teoma og Ask Jeeves:

Teoma, <http://www.teoma.com/> kom på luften for vel tre år siden i årsskiftet 2001/2002 og var et friskt pust. Den skåret høyt på Greg Notess oversikt over store søkemaskiner fra 31/12 2002 med litt over en milliard vevsider. Siden falt brukbarheten en del, det skyldes blant annet at de var meget sene med å oppdatere databasen. Det tok for eksempel mer enn et halvt år etter 1. januar 2003 før sidene fra vevstedene til ABM-Utvikling kom inn i basen, det var ikke tillitsvekkende. Det virker imidlertid som de har forbedret seg noe der, og Teoma er et interessant alternativ i mange søkinger, spesielt hvis man har litt store og omfattende emner.

Det som gjør Teoma spennende er et par ekstra muligheter på siden over søkeresultater. Den ene er "Refine: Suggestions to narrow your search" hvor Teoma foreslår ekstra søketermer. Klikk på disse og søket gjøres om igjen med denne søketermen lagt til. Dette er imidlertid mange søkemaskiner nevnt her som tilbyr. Den andre er "Resources: Link collections from Experts and Enthusiasts". Dette er en enkel, men meget spennende mulighet. Det de (antagelig) har gjort er å telle lenker

på en side og samle dem i en egen kategori "lenkesamling". Dermed blir vi (ofte) presentert med kataloger over vevsider i emnet vi er ute etter, med andre ord oversikter gjort av fagfolk og entusiaster og som ofte hjelper i arbeidet med å lage våre egne og å finne frem til gode vevsider. Det fritar ikke en søker fra ansvaret om å evaluere og sortere resultatene etter beste evne, men det er en stor hjelp og inspirasjon å finne andre som har forsøkt å gjøre det samme.

Teoma eies av AskJeeves som også presenterer en egen søkemaskin, <http://www.ask.com/>. Denne databasen er så vidt jeg kan se nesten identisk med den Teoma bruker, basert på resultatene av noen søkinger.

Greg Notess har skrevet en omfattende omtale om Teoma som finnes på hans side om søkemaskiner, <http://www.searchengineshowdown.com/features/teoma/review.html>

KartOO

KartOO <http://www.kartoo.com/> er en meget spennende nykommer. Det er en fransk metasøkemaskin, og bruker fjorten av de store søkemaskinene som basis, blant annet Yahoo-basene (Yahoo, AltaVista med mer) og Teoma. Derimot er Google ikke med. Det er mulig å velge ut enkelte av basene og utelukke andre.

Det som er nytt her er den grafiske presentasjon av resultatene. De presenteres grafisk som et kart med dokumenter som faste punkter og linjene mellom dokumentene med ekstra emneord. Fører man musen over et dokument dukker det opp en kort oversikt over nettstedet bak med adresse og også hvordan emneorden man har søkt etter, et slags "keyword in context". Klikker man på dokumentet kommer det opp i nettleseren. Fører man musen over emneordene imellom dokumentene kommer emneordet opp som "veier" frem mellom dokumentene, klikker man på ett av ordene så søker man på det i tillegg til det opprinnelige emnet. I tillegg til kartet kommer alle dokumentene (vevstedene) som er funnet i en liste til høyre for kartet og emnene i en liste til venstre.

Det virkelig verdiøkende er inndelingen i emner og delemner. Søker man for eksempel på en sykdom kommer behandling opp som delemne. Dette er gjort som en del av den kunstige intelligensen bak kartOO, og er det som gjør søkemaskinen meget spennende,

Vær obs på knappen som heter Kapitalyser. KartOO husker søkingene som har vært gjort, klikker man på knappen Kapitalyser kommer tidligere søkeord opp. Kjekt å ha mange ganger, andre ganger derimot... Denne kan slettes, det bør man ha i bakhodet når man deler maskinen med andre!

Oppsummering: Det originale med kartOO er inndeling i emner og bruken av disse til å presentere resultatet i den grafiske presentasjonen av søkeresultatene. KartOO er meget effektiv og det er lett å bli fascinert av denne måten å presentere av søkinger. Og jeg har funnet mange meget verdifulle kilder på en lettvinnt måte. Så dette er absolutt en peker på i hvilken retning søkemaskinene går.

WiseNut

WiseNut <http://www.wisenut.com/> er sammen med Teoma en base som kom som en løve i 2002, men minnet mer om en skinnfell ut i 2003. Sen oppdatering gjorde at maskinen ikke imponerte. Det er imidlertid en interessant søkemaskin. I Greg Notess

oversikt fra 31/1 2001 var den fjerde størst, med 1,4 milliarder vevsider.

Den originale med den er at resultatene ved hjelp av Algoritmen til WiseGuide samles i grupper, clusters, under overskriften "Result Clustering. Den er i dokumentasjonen beskrevet som at WiseGuide automatisk lager kategorier som ut fra semantikk er koblet til ordene i søkspørsmålet. Det fungerer brukbart, jeg vil ikke si mer en det, men det er ledd i utviklingen om hvordan kunstig intelligens brukes til å sortere søkeresultatene. Den mest vellykkede søkemaskinen som gjorde dette var den, nå dessverre nedlagte, Northern Light. Greg Notess har omtale av Wisenut på: <http://www.searchengineshowdown.com/features/wisenut/review.html>

Vivisimo

Vivisimo, <http://vivisimo.com/> er en metasøkemaskin som dekker flere databaser, både på web (MSN, Wisenut og andre), nyhetstjenester, medisinske baser (PubMed), kataloger og mye annet. Et av valgene er auksjonsstedet eBay. Styrken til Vivisimo er, som for WiseNut, å samle resultatene i grupper og søkemaskinen er egentlig laget for å demonstrere hvordan denne inndelingen fungerer, målet er å selge produktet til de som vil organisere sin egen informasjon. Vivisimo er med rette populær, enkelte har det nesten opp mot Google i effektivitet. En omtale av Vivisimo finnes på http://www.arnoldit.com/articles/iwr_vivisMay2003.html.

Gigablast

Gigablast <http://www.gigablast.com/> er en liten søkemaskin i forhold til de andre. Nå er liten et relativt begrep, den har over 400 millioner vevsider i basen. Det er et meget respektabelt tall, men konkurrentene er langt oppe i milliardene, og da er Gigablast ikke så imponerende.

Det er imidlertid en spennende søkemaskin. I oversikten over svar genereres emneord i "Giga Bits" med et prosenttall for hvor relevant det emneordet er. Nøyaktig hvordan de genereres og betydningen av prosentene er ikke gjort klart, men det virker som de er generert ut fra en analyse av svarmengden. Søker man på dem gjøres søket med de emneordene i tillegg til det opprinnelige. Gigablast arkiverer sidene, så hvis en side er forsvunnet finner man tilbake til den Gigablast har indeksert. Det samme gjør også Google. Men Gigablast går videre og legger inn pekere til Internettarkivet "Wayback machine", <http://web.archive.org> og der kan man følge historien og oppdateringene på en side mange år bakover.

De oppfordrer til å formulere et spørsmål i fritekst. Oppfordringer skal man følge. Spørsmålet: "Who is king of Norway" ga interessante resultater. I den vanlige svarlisten kom pekere til dokumenter om Olav Haraldsson, St. Olav, opp øverst. Selv om undertegnede som er både katolikk og trønder følte at dette var et usedvanlig godt svar, vil nok de fleste kanskje mene at det ikke er ikke helt korrekt. Pekere til dokumenter om kong Harald var også høyt oppe på svarlisten, så det blir godkjent. Mer interessant er listen fra Giga Bits hvor de høyest rangerte emnene ble "Sweden", "King of Sweden", "Denmark" før "King Harald" på fjerdeplass. Ikke helt godkjent, den rangeringen. Greg Notess har en anmeldelse av Gigablast på sin side om søkemaskiner, <http://www.searchengineshowdown.com/features/gigablast/review.html>.

Tre fra Infospace: Metacrawler, Webcrawler og Dogpile

Infospace, <http://www.infospaceinc.com/> eier og vedlikeholder tre søkemaskiner, MetaCrawler, <http://www.metacrawler.com>, WebCrawler, <http://www.webcrawler.com/> og Dogpile <http://www.dogpile.com/>

Dette er tre metasøkemotorer som søker i Google, Yahoo, AltaVista, Ask Jeeves, About, LookSmart, Overture, Teoma, og FindWhat, og er nesten identiske, samme grunnlag og samme algoritme for å presentere resultatene. Der blir også resultatene inndelt i underemner med forslag til å presisere resultatet, og i alle fall i noen av tilfellene jeg har prøvet det har det gitt gode resultater og fått frem artikler som andre søkemaskiner hadde begravet lenger nede i resultatlisten. Men det kan man vel si om alle søkemaskinene. Ellers er det ikke så mye å si om trioene. Men det må jo nevnes at WebCrawler er så vidt jeg vet den eldste søkemaskinen som fremdeles er i drift. Den kom først online i 1994, da med 4000 vevsider. Han som laget den, Brian Pinkerton har skrevet en kort historie om den, <http://www.thinkpink.com/bp/WebCrawler/History.html>. Nå må det vel sies at det bare er navnet som er igjen av den opprinnelige søkemaskinen.

Amazon A9

Amazon har kommet med en søkemaskin, A9, adressen er <http://www.a9.com>. Den ble nevnt i forrige Bok og Bibliotek, dette er bare som en påminnelse, siden det er sammen med disse basene den hører hjemme og fordi det er en interessant nykommer. I tillegg til å søke i vevdatabasen som blant andre Google leverer søker Amazon også i bøkene i katalogen til Amazon.com, så vi får full tekst søking i mye av faglitteraturen.

Og bakom synger Microsoft...

Microsofts søkemaskin, MSN search ble nevnt i forrige Bok og Bibliotek sammen med Yahoo og de andre Inktomibaserte søkemaskinene. Ikke noe har forandret seg der, men ryktene om at Microsoft har noe stort på gang blir stadig sterkere. Flere artikler skriver om det neste store skrittet i søkemaskinteknologi som er å gjøre dem mer tilpasset den enkelte bruker. Kombinasjon av å søke på nettet og på egen harddisk blir også stadig mer aktuelt. Google og Yahoo! varsler nyheter kommende år. Følg med, nå blir det spennende!