

## Nytt om søkemaskiner

# Gamle venner forsvinner, nye kommer

**Siste året har det vært en skikkelig omkalfatring i markedet for søkemaskiner, og det brygger opp til stor kamp om brukerne. Allianser er inngått, oppkjøp er gjort og røken har ennå ikke lagt seg, utfallet er uklart. Kjempen er fortsatt Google, men Yahoo! er blitt den andre virkelig store aktøren og sterke rykter går om at Microsoft samler krefter til en krig. Andre mindre aktører er også inne med spesialvarianter som har sine goder.**

Av Even Flood, førstebibliotekar

Det er ikke lenger nok bare å ha en søkemaskin, her brukt i betydningen database over internettsteder som er laget av en robot og gjort søkbar. Det må være noe mer. Dette "noe mer" kan være en indeksert internettkatalog, skapt av mennesker over utvalgte vevsteder. Det var starten på Yahoo!s storhet. Eller det kan være mange gode søkemuligheter. Det man forsker mest på er å organisere søkeresultatene på en oversiktlig måte ved å skape en "relevans" basert på forskjellige antagelser om hva som er viktig for brukeren. Her kan Google av og til være nifst god! Andre muligheter er å automatisk sortere svarene i forskjellige rubrikker. Søkemaskinen "Northern Light" var meget gode på dette, den er fortsatt savnet! Eller det kan være å inkludere ekstra baser i sin søkemaskin som andre ikke kan, for eksempel Amazons nye søkemaskin A9.

Yahoo! har vært meget aktiv på markedet med mange oppkjøp. Etter at alt støvet har lagt seg står de som eiere av Inktomi, Overture og AltaVista. Overture eier igjen AlltheWeb, søkemaskinen som ble laget av det norske selskapet FAST, Inktomi er primært kjent som databaseprodusent. De laget en meget stor robotbasert database over veven som de så leide ut til forskjellige verter som gjorde den eller deler av den søkbar på sin tjeneste, som for eksempel HotBot og Microsoft. Dette medførte at Yahoo! nå hadde tre firma som laget databaser over veven og flere søkemaskiner med forskjellig funksjonalitet. Spekulasjonene var mange om hva de ville gjøre med alt dette. Optimistene håpet at de ville ta den største databasen, som Alltheweb hadde, og legge den under den kraftigste søkemaskinen med de fleste søkemulighetene, som var AltaVista. I april ble det klart at dette definitivt ikke hadde skjedd. Gamle AltaVista ble lagt ned, samme med gamle AlltheWeb, og begge har blitt Inktomibaserte søkemaskiner. Inktomi er dermed blitt grunnlaget for Yahoo! søkemaskin. Spesielt tapet av AltaVista er følbart og mange er skuffet over hva Yahoo! har gjort her. AltaVista hadde søkemuligheter de andre manglet, som bruk av nærhetsoperator, forskjell på små og store bokstaver og, aller viktigst, trunkering. I AltaVista kunne man også lage store søkesetninger med parenteser og boolske operatører. Kort sagt, de hadde tatt mye fra de tradisjonelle klassiske søkegrensesnittene da de laget sitt. Alt dette er nå borte. Det virker rart på utenforstående at når Yahoo! forbereder seg til å starte kampen mot Google om brukerne begynner de med å kaste fra seg det som burde være de skarpeste våpnene, men jeg skjønner meg ikke så mye på business.

Dermed er det nå to store produsenter av vevdatabaser, Yahoo! ved Inktomi og Google. De leier ut databasene sine eller deler av dem til egne verter, slik at nå er både Alltheweb, AltaVista og Microsofts søkemaskin Inktomibaserte databaser. Google har sin egen, men bidrar også til Amazons nye søkemaskin, A9 og til American Online og Netscape sine søkemaskiner.

Det er så mye som skjer at det er vanskelig å holde seg oppdatert på feltet, selv for den som følger med nesten daglig. Men her kommer en liten oppsummering og vurdering av situasjonen som den er nå og hva noen av de forskjellige kan tilby. Det kan bare bli en liten oversikt over de største aktørene, en rekke mindre og spennende nykommere må vente til senere.

## Google

Google er fortsatt nummer en med en søkemaskin som stadig overrasker med å plukke ut gode steder og serverer dem langt opp. De introduserte lenking som et av kriteriene for relevans, jo flere lenker det var til en side, desto høyere kom siden opp på rangeringen. Dette er ensbetydende med å si at popularitet og kvalitet var det samme, det er ikke nødvendigvis korrekt. Men svært ofte fungerer det bra. Google er også raskt ute med fersk informasjon. I mai advarte Forbrukerådet mot reiseselskapet "One World Travel", en rask googling (Mye brukt ord. Google er allerede blitt et verb. Eller som Tommy i Tommy og Tigeren sier det, ordet er blitt verbet.) ga at de norske avisforsidene med advarselen var blitt fanget opp og lagt inn i Google dagen etter med høy rangering. Sidene var på norsk, så bare nordmenn kunne skjønne det, men poenget er at sidene kom inn med høy rangering på dagen. Det er interessant at det var vevdatabasen i Google som hadde denne informasjonen, ikke nyhetsdatabasen på <http://news.google.com/>.

Alle søkemaskinene gir problemer med oss ikke engelske som ikke kan stave skikkelig. De "gamle" tjenestene hadde gode hjelpemidler for dette som trunkering og ekspand kommandoen, som i BibSys lever som "søk i indeks". Ingen slik hjelp finnes. Men Google sjekker ordet og hvis det er noen lignende med høy frekvens spør den om det er det du mener. Søker du på ordet "teater" får du forslag om "theater" Da dette først bli introdusert søkte Google uten videre på det den trodde var korrekt, det ga store problemer, blant annet for kjemikeren som skulle vite om en gruppe kjemiske forbindelser som heter aminimal og ikke var interessert i animal. Nå kommer bare forslagene. Denne funksjonen var til meget stor nytte nylig da jeg skulle lete etter en katolsk gruppe som kalles Neokatekuminatene. Etter noen gjetninger fra min side foreslo Google Neocatechuminate, godt er det ellers hadde jeg vel sittet der fremdeles.

Et annet tilbud fra Google er arkivet til nyhetsgruppene i Usenet i <http://www.groups.google.com>. Dette er et enormt arkiv fra alle diskusjonsgruppene der, også de norske, og inkluderer vel en milliard meldinger som man ikke finner noe annet sted. Mye nyttig selv om det er mange ganger så mye søppel, så dette er en kilde som må brukes med stor varsomhet. Allikevel, det er meget bra tilbud.

Nyhetsjenersten, <http://www.news.google.com> er en kontinuerlig nyhetsovervåkning av en rekke store internasjonale aviser og nyhetsmedia. Meget bra kilde for å følge nyheter, men som nevnt tidligere er det også ferske nyheter i den vanlige vevdatabasen som ikke kommer inn i nyhetsbasen.

Google har også en indeksert katalog, <http://www.google.com/dirhp>. Det var tidligere

en peker til denne direkte fra Googles startside, den er av en eller annen grunn tatt bort. Man kommer til den, og mange andre av Googles ekstratjenester ved på klokke på "More" fra Googles startside. Googles katalog er basert på Open Directory Project, <http://dmoz.org>, et enormt frivillig prosjekt som nå dekker over fire millioner indekserte vevsteder i nesten seks hundre tusen kategorier og utført av over seksti tusen redaktører. Et så digert frivillig prosjekt må nødvendigvis ha svakheter og det er kanskje derfor Google nedprioriterer den. I tillegg er Google ofte sent oppdatert, så vil man bruke denne katalogen vil jeg anbefale å søke i den oprinnelige.

## Yahoo! og AltaVista/AlltheWeb

Disse tre taes under ett siden de er under samme paraply. For det første er det selvfølgelig vevdatabasen de kan tilby, den er meget stor, spesielt virker det slik hvis man søker Yahoo! som gir de fleste treffene. Men det kan være en illusjon, det er meget vanskelig å bruke antall treff som eneste kriterium når man vurderer størrelsen på databasen. Søkemulighetene er dårligere enn i Google, spesielt kan man ikke søke på lenker, en slags siteringsøk. Det kunne man tidligere både i AltaVista og Alltheweb. På den annen side har AltaVista muligheten for å søke direkte etter lydfiler, det har ikke Google og heller ikke de andre.

En annen tjeneste som Yahoo! tilbyr er katalogen, som man kan gå direkte til via <http://dir.yahoo.com/>. Dette var det Yahoo! startet med og ble berømt på, så det er litt trist at den nå er gjemt bort på hovedsiden. Dette er vel den nest største internettkatalogen etter Open Directory Project og er en av Yahoo! sine sterke sider, deverre virker den noe nedprioritert når man ser på Yahoo! sin startside.

Diskusjonsgrupper er også et tilbud som er interessant. Google har hele Usenetarkivet så de har en enorm database her. Yahoo! har laget et mini Usenet i <http://groups.yahoo.com> hvor man også kan bli med i åpne diskusjonsgrupper. Det er mange grupper, men færre enn i Usenet, og langt lavere aktivitet. Det siste er ikke nødvendigvis et onde.

## Microsoft

Det har lenge vært skrevet om at Microsoft har planer om å kaste seg inn på markedet med søkemaskiner. Foreløpig har de ikke så meget, en database som er basert på Inktomi og en katalog som ikke er så innholdsrik som de andres. Man kan bruke den ved å gå direkte til <http://search.msn.com>. Der har også en vevkatalog med mye bra innhold, men ikke så innholdsrik som de andre. Det er interessant at kategoriene i katalogen er eldre enn innholdet, enkelte kategorier er tomme. Som en kuriositet, kategorien "Library > Sciences > Alternative Science > Flat Earth" er tom, selv om det vitterlig er vevsteder om dette. Det er skrevet mye om at Microsoft i løpet av året vil komme med sin egen database, og da får man se på hva som skjer. Foreløpig er de bak de to andre, men det er ikke lurt å undervurdere Bill Gates!

## Amazon

Amazon har kommet med en søkemaskin, A9, adressen er <http://www.a9.com>. Den er meget interessant. I likhet med Google kommer også de med hjelpsomme råd til dårlig søking, de hjalp meg også med Neokatekuminatene. Kanskje ikke så rart, siden Google står bak mye av databasen. I tillegg søker Amazon også i bøkene i katalogen til Amazon.com, så vi får full tekst søking i mye av faglitteraturen. Det er en personlig tjeneste, da jeg gikk inn siste fikk jeg hilsenen "Hello Even Flood" og den husker også søkingene jeg har gjort. Antagelig har de dataene mine fra cookies fra Amazon bokhandelen. Hvordan de bygger databasen er jeg ikke sikker på, nederst på siden står "Search results enhanced by Google. Results also provided by A9.com, Alexa, and Amazon.com". Google er altså også inne i bildet her, det virker som resultatet av

websøkingen kommer fra dem.

## Mer om dette?

Som nevnt skjer det utrolig mye og det er flere søkemaskiner som er på markedet. De må bli behandlet i en senere artikkel. Det er, selvfølgelig, mange nettsteder som overvåker og informerer om hva som skjer. Den viktigste er Danny Sullivans Searchenginewatch, <http://searchenginewatch.com/>. Men Greg Notess' <http://www.searchengineshowdown.com/> har også mange nyheter og gode oversikter over søkemaskinene. Han har også den mest siterte statistikken over størrelsen på søkemaskinene, men den ble siste oppdatert 1/1 2002 og må sies å være foreldet. Han har planer om å oppdatere den, men det er så mye som Google og Yahoo! gjør nå at det er vanskelig å få gjort noe skikkelig