

# Tore på sporet: Hvordan søker man best på Internett?

av Tore Brattli, universitetsbibliotekar ved Universitetsbiblioteket i Tromsø

Utviklingen på Internett har endret på fremgangsmåten for å søke etter informasjon. I tillegg til vår eksisterende kunnskap om søking, må vi nå også kunne bruke nye søkeverktøy som fungerer etter andre prinsipper enn tidligere. Det er også kommet nye søketeknikker, og vi må forholde oss til informasjon som kan være ganske forskjellig fra den vi er vant til.

Det skrives mye om søking på Internett for tiden, men mye av dette er bare lettvinne gjennomganger av søkemaskinenes muligheter og finesser, eller oppramsing av mer eller mindre gode websider. Dette er for så vidt interessant, men kanskje viktigere er hvordan man bruker de nye søkemaskinene til å finne informasjon, eller hvordan informasjons-søking generelt har endret seg med utviklingen på Internett. Websider og søkemaskiner kommer og går, mens generell kjennskap til søkeprosessen er mer tidløs.

## Biblioteket i forandring

Biblioteket har tradisjonelt ikke vært en plass der det skjer mye. På biblioteket finner du bøker og tidsskrifter som kan kopieres, leses på lesesalen eller lånes med hjem. Slik har det vært siden de første bibliotekene ble opprettet for flere tusen år siden, og slik vil det nok fortsette å være fremover vil mange hevde. Vesentlige hendelser i biblioteket opp gjennom historien har det ikke vært mange av. Tar vi med trykkekunsten og utviklingen av bibliotekskatalogen, burde vel det meste være dekt. Biblioteket er likevel en av de mest spennende arbeidsplasser. Vi er inne i en tid der det sannsynligvis vil skje mer enn gjennom alle tidligere tider til sammen.

## Informasjonsrevolusjonen

De siste 20 årene har vi vært igjennom en revolusjon på referansesiden. Referanser er blitt digitale og tilgjeng-

elige via databaser og nettverk. Fra kortkataloger, referatorganer, onlinesøk og CD-ROM er det nå Internett og WWW som har overtatt. Dette har revolusjonert både våre og våre brukeres muligheter til å holde seg orientert om hva som eksisterer av informasjon.

De neste 20 årene vil selve informasjonen stå for tur. Dokumenter med muligheter for multimedia og hypertekst vil bli tilgjengelige i digital form via nettet. Tidsskrifter og oppslagsverk er allerede godt i gang, mens det nok vil ta litt lengre tid før bøker blir vanlige. Dette vil revolusjonere tilgjengeligheten og bruken av informasjon.

## Internett i biblioteket

Frem til nå har Internett på mange måter ikke passet helt inn i bibliotekets verden. Mange har oppfattet Internett som ei samling ustrukturert, ustabil og upålitelig informasjon som er vanskelig å finne frem i.

I dag har den seriøse og kommersielle informasjonen inntatt nettet og for referanser har Internett blitt den viktigste kilden. Av klassisk bibliotekslitteratur er tidsskrifter godt på vei, mens bøker i liten grad har begynt å komme. I tillegg til dette kommer all den andre informasjonen som etter hvert finnes i store mengder, og der mye holder høy kvalitet.

Informasjon er ikke lengre noe man bare finner i hyllene på biblioteket, og våre klassiske referansedatabaser dekker langt fra alt som er relevant for brukerne. Dette er en situasjon som vi må forholde oss til og ta konsekvensen av, ellers vil vi snart bli forbikjørt av den raskt voksende informasjonsindustrien. Utfordringen nå er å ta med all den generelle kunnskapen om søking over i den nye verden, og i tillegg lære seg det lille ekstra som er spesielt der. Gode gamle regler for søking gjelder i stor

grad fortsatt, men innholdet har endret seg litt.



## Grunnleggende ferdigheter i informasjonssøking

Vi skal se litt på hvordan utviklingen av Internett har endret innholdet til det vi kan kalle for overordnede prinsipper for informasjonssøking og hvilke konsekvenser har det for oss fagreferenter. Prinsippene er delt inn i 5 deler:

- Kjennskap til hva det letes etter
- Kjennskap til faget
- Kjennskap til søkeområdet
- Kjennskap til lokaliseringverktøy
- Kjennskap til søketeori

## Kjennskap til hva det letes etter

Kjennskap til hva brukeren skal ha gjør det enklere å finne relevant informasjon. Dette vil stort sett være som vi er vant med fra tidligere. Det nye med Internett er at det har blitt tilgjengelig mye mer ikkefaglig informasjon, i tillegg til andre former for informasjon som bilder, film, lyd, dataprogrammer, osv. Dette må vi være forberedt på å få spørsmål om.

Et interessant spørsmål i denne sammenhengen er hvor grensene går for hva vi skal ha ansvar for å hjelpe brukerne med å finne. Tidligere var dette ofte begrenset til informasjon som var

tilgjengelig i bibliotekene. Med Internett i biblioteket er grensene mer diffuse. Hva gjør vi for eksempel med en bruker som vil ha oversikt over priser på datamaskiner som skal brukes i et forskningsprosjekt? Det finnes mye om det på Internett.

## Kjennskap til faget

Kjennskap til emnet som brukeren skal ha informasjon om er forhåpentligvis også viktig. Spesielt siden det er dette som skiller oss fagreferenter fra bibliotekarene. Kjennskap til faget er viktig for å kunne kommunisere med brukeren, formulere søk og evaluere søkeresultatet. Å holde seg faglig oppdatert vil alltid være ei utfordring og dette har nok ikke endret seg vesentlig fra tidligere.

## Kjennskap til søkeområdet

»Søkeområdet« er der informasjonen finnes, dvs. i bøker, artikler, oppslagsverk, websider, osv. Med kjennskap til søkeområdet menes kjennskap til hva som finnes av informasjon og hvor, uten bruk av lokaliseringsverktøy. Dette er viktig siden det ikke er alt som er like lett å søke etter, og siden det finnes informasjon som ikke lar seg søke frem i det hele tatt. I tillegg vil slik kjennskap både forenkle valg av lokaliseringsverktøy og bruken av dem. En god fagreferent vet hvor den relevante informasjonen finnes og går direkte dit. Internett endrer ikke på det. Det nye er at søkeområdet har utvidet seg kraftig, det flytter på seg og både innhold og terreng er ikke helt som vi er vant til.

## Utvidet søkeområde

Med Internett har vi fått tilgang til en stor mengde informasjon som vi tidligere ikke hadde. Enten fordi den ikke eksisterte, eller fordi den var vanskelig tilgjengelig. Mye av den ikke-kommersielle informasjonen som vi i dag finner på Internett eksisterte også tidligere, men den var som regel svært utilgjengelig. Mesteparten av informasjon som vi for eksempel har på Universitetsbiblioteket i Tromsø sine websider eksisterte også tidligere som informasjonsbrosjyrer, men da uten at det var mulig å søke dem frem. Det er derfor litt absurd å klage på at det er så vanskelig å søke etter informasjon på

nettet, når det tidligere i praksis var umulig å finne den samme informasjonen.

## Søkeområdet flytter seg

I tillegg til at søkeområdet utvides, har innholdet begynt å flytte på seg. Spesielt gjelder dette en del av innholdet i vårt tradisjonelle søkeområde, dvs. tidsskrifter, oppslagsverk og CD-ROM som nå begynner å flytte seg over til vårt nye område Internett.

## Uvant terreng og innhold

På Internett er innhold og terreng i tillegg svært ulikt det vi er vant til fra før med bøker, tidsskrifter, oppslagsverk og referansedatabaser. Dette er med på å sette nye krav til fagreferentene.

- Informasjonen er mer variert og finnes i mange former (tekst, bilder, film, osv) og formater (HTML, PDF, tekst, osv).
- Informasjonen er dårlig merket (mangler metadata) og er derfor vanskeligere å oppdage.
- Informasjonen er dynamisk i innhold og eksistens. Den kan når som helst oppstå, endre seg, flytte på seg eller opphører å eksistere.
- Informasjonskvaliteten varierer og det kan være vanskelig å stole 100% på nettinformasjon. Likevel er det her som tidligere at forfatter eller organisasjon som tilbyr informasjonen er viktige indikatorer på kvaliteten. For mange typer informasjon kan kvaliteten på det som finnes på nettet være bedre enn på papir, spesielt for informasjon som endrer seg ofte.
- Informasjonsmengden er ustrukturert og mange er av den oppfatning av at Internett er et eneste stort kaos. Kanskje er det tilfelle, men da må man igjen sammenligne med tidligere. Har det noensinne tidligere vært orden på denne typen informasjonen?
- Hvis man vet hvor noe er, er det mye enklere og raskere å hente det enn før.

## Hva kan man nå finne og hvor?

På nettet finnes store mengder ikke-kommersiell informasjon om alt mulig og informasjon om kommersiell informasjon. Den kommersielle informasjonen er på vei.

Hvor finnes informasjonen som man er på jakt etter? Utfordringen er å tenke seg til hvor det er sannsynlig at den kan finnes. Typiske steder kan være informasjonsprodusenter som organisasjoner, forlag, bedrifter, forskningsmiljøer, forskere, privatpersoner, osv. alt etter hvilken informasjon det er snakk om. Produktinformasjon finnes hos bedrifter. Artikler eller forskningsresultater hos forlag, organisasjoner eller forskere osv.

## Kjennskap til lokaliseringsverktøy

Lokaliseringsverktøy er hjelpemidler for å finne frem til relevant informasjon som man ikke vet hvor er, uten å måtte sjekke alt som finnes. Fra tidligere er vi kjent med å søke etter referanser i databaser som direkte eller indirekte tilbyr boolsk søkespråk. Med Internett har vi plutselig fått nye søkeverktøy som søker etter fulltekstdokumenter i stedet for referanser, bruker andre søkespråk og fungerer etter andre prinsipper enn det vi er vant med. Hver av disse krever nye ferdigheter som vi snarest må lære oss.

Store informasjonsmengder, skiftende informasjon, varierende kvalitet på informasjonen og manglende interesse fra bibliotekshold har gjort at mesteparten av informasjonen på Internett indekseres automatisk. Problemet er at automatisk indeksering i hovedsak bare fungerer for tekstdokumenter, og selv da er resultatet langt fra fullgodt. For bilder, kodet tekst, osv. blir gjenfinning nesten umulig. Årsaken til dette er i stor grad mangelfull merking (metadata).

## Søkemaskiner

Søkemaskinene på Internett/WWW har utviklet seg fra World Wide Web Worm, som kom i september 1994, til den mer moderne generasjonen søkemaskiner som AltaVista, som kom i desember 1995.

## Vektorbaserte modeller

De fleste søkemaskinene benytter seg av eller fungerer som vektorbaserte modeller for gjenfinning. Ved å representere både dokumenter og søk som vektorer av sine ulike ord, blir det mulig å ta i bruk avanserte matematiske verktøy for å beregne »avstanden«

mellom vektorene. Fordelen med en slik fremgangsmåte er at dokumentene kan sorteres etter avstand, dvs. relevans, slik at de »beste« dokumentene kommer øverst på trefflista.

Vektorbaserte modeller har utviklet seg gjennom de siste 20-30 år og retter seg i hovedsak mot fulltekstdokumenter. De har derfor ikke fått det helt store gjennombruddet før utviklingen på WWW kom i gang, men der har de til gjengjeld hatt stor suksess. Søkemaskinene kan raskt og billig indeksere og følge med endringer i 100 millioner dokumenter, og daglig ta imot millioner søk fra brukere over hele verden. Søkemaskinene er av de viktigste verktøyene vi har i dag for å finne dokumenter på Internett.

Ved vektorbasert søking er det mye vanskeligere å forstå kriteriene for treff enn ved boolsk søking. For det første vil ikke søkemaskinene avsløre nøyaktig hva de egentlig gjør, og for det andre er matematikken som ligger bak mer komplisert enn ved boolsk søking. Bakgrunnen for modellen er resultater fra forskning innen informasjons-gjenfinning som bl.a. sier at den relative frekvensen til et ord og plassering i et dokument (tekst, tittel, overskrifter, metadata) sier noe om hvor mye dokumentet omhandler dette begrepet. I tillegg er ord som opptrer i få dokumenter gunstige for gjenfinning.

Dette utnyttes ved at brukeren kan søke ved å skrive inn alle relevante søkeord innen aktuelt emne uten å bruke et komplisert eller formelt søkespråk. Dokumentene vektet og sorteres etter en kombinasjon av hvor mange av søkeordene de inneholder, hvor søkeordene er plassert og frekvens. Resultatet med denne fremgangsmåten blir jevnt over bedre og mer brukervennlig enn ved boolsk søking.

- Vi får de beste dokumentene øverst på trefflista.
- Brukerne slipper å lære seg et formelt søkespråk.
- Det er uproblematisk å bruke mange søkeord.
- Antall treff spiller ingen rolle!
- Søkresultatet blir jevnt over betydelig bedre!

## Videreutvikling av vektormodellen

En videreutvikling av vektormodellen er latent semantic indexing (LSI), som i tillegg ser på korrelasjonen mellom ordene i dokumentene. Ord som ofte opptrer i lag har ofte noe med hverandre å gjøre. Ved å bruke LSI vil vi kunne få relevante treff på dokumenter som ikke inneholder noen av søkeordene, hvis det i dokumentet er tilstrekkelig med ord som vanligvis opptrer sammen med søkeordene. LSI gir jevnt over bedre resultater enn klassiske vektorsystemer. Søkemaskinen Excite baserer seg på et lignende prinsipp.

## Utvidet boolsk (fuzzy boolsk)

Utvidet boolsk søking eller fuzzy boolsk er boolsk søking med rangerte trefflister og har etterhvert blitt vanlig i flere søkemaskiner. Utvidet boolsk kombinerer vektormodellens muligheter for rangerte trefflister og det boolske søkespråkets muligheter for presisering. Gjenfinningskvaliteten er likevel ikke påvist å være bedre enn for klassiske vektorsystemer. Vær oppmerksom på at operatoren NOT i de fleste søkemaskiner ikke er implementert fuzzy-boolsk, og bør derfor brukes med stor forsiktighet på fulltekst.

## Multisøkemaskiner

Selv de beste søkemaskinene dekker bare en liten del av den totale informasjonen på Internett og ulike søkemaskiner dekker ikke nødvendigvis det samme. Det kan derfor være lurt å bruke flere søkemaskiner eller såkalte multisøkemaskiner. Dette er søkemaskiner som ikke selv har oversikt over informasjon, men som i stedet sender søket videre til en mengde andre søkemaskiner samtidig. Etter å ha sendt ut søket, venter den på svar, sorterer trefflistene fra de andre søkemaskinene til ei treffliste og sender den til brukeren. Denne typen verktøy er grei for søk der man ikke forventer så mange relevante treff og der søket enkelt lar seg spesifisere.

Søkemaskinene med sine enkle søkeskjermer har av en eller annen grunn ikke vært spesielt godt ansett i de profesjonelles rekke. Med innføringen av boolsk søking har holdningen endret

seg i positiv retning. Likevel viser altså forskningen at den mer eller mindre vektorbasert søkingen som brukes i de enkle søkeskjermene fungerer minst like bra som boolsk søking. Det betyr at vi må få et mer positivt og aktivt forhold til denne typen søking både for vår egen og for brukernes del.

## Emneindekser

Emneindekser, som for eksempel YAHOO eller Kvasir, er også svært nyttige verktøy for å finne informasjon, og kan i mange tilfeller være det beste startstedet. Informasjonen er hierarkisk ordnet i kategorier, noe som gjør det enkelt å finne frem til informasjon uten å måtte formulere søk. Både kategori-strukturen og innholdet i dem blir vanligvis manuelt oppdatert, slik at det er høy kvalitet på informasjonen. Det gjør emneindekser brukervennlige og kanskje spesielt godt egnet til brukere som ikke helt vet hva de skal ha. Emneindekser er ryddigere på den måten at de presenterer hjemmesidene til organisasjoner, bedrifter, institusjoner osv. i stedet for mer eller mindre tilfeldige sider som søkemaskinene ofte gjør.

## Nyhetsgrupper/postlister/Personer

En annen viktig måte å finne informasjon på er å spørre noen som vet svaret. Med Internett kan dette gjøres på flere og mer effektive måter enn tidligere. Hvis man ikke finner den informasjon som man er på jakt etter, finner man sikkert navnet på personer som kan tenkes å vite noen om det. Dette kan for eksempel være personer som har skrevet artikler, laget websider eller som jobber innen emnet. Hvis man sender dem e-post og spør på den riktige måten, får man ofte positive svar. E-post har fortsatt den litt uformelle karakteren som gjør at man kan sende spørsmål til nesten hvem som helst. En mer indirekte mulighet, som gir omtrent samme resultat, er å sende spørsmålet til e-postlister eller nyhetsgrupper innen emnet.

Disse mulighetene bør ikke misbrukes og det er viktig å ha gjort godt forarbeid før man forstyrrer ukjente med spørsmål.

## Intelligente agenter

Den automatiske »fagreferent« er på vei i form av intelligente agenter. Dette er

dataprogrammer som kan læres opp til å forstå hvilket informasjonsbehov brukeren har. En slik agent kan for eksempel kontinuerlig surfe eller søke på nettet, i referansedatabaser, bøker, artikler eller oppslagsverk etter informasjon som kan være av interesse. Den kan organisere informasjonen og for eksempel presentere det som er nytt eller det som har endret seg siden sist.

## Kjennskap til søkteori

Søkteori går på kjennskap til selve fremgangsmåten for å finne informasjon. Dvs. hvordan man bruker sin kjennskap til hva brukeren skal ha sammen med kjennskap til faget, søkeområdet og lokaliseringsverktøy på en best mulig måte for å finne informasjon. Mye er som før med unntak av det som vi har vært inne på tidligere. I tillegg er det med WWW kommet en del nye teknikker for å finne informasjon.

## Surfing

Surfing er å benytte seg av hypertekstens sammenveving av websider for å finne informasjon. Siden websider som omhandler samme emne ofte har lenker til hverandre, er det mulig å finne relevant informasjon ved å velge de mest relevante lenkene når man klikker seg fra side til side. Før søkemaskinene kom, var surfing den eneste måten å finne det lille som da fantes av informasjon. De fleste websider hadde lange lister med lenker til andre websider for å kompensere at de selv hadde lite informasjon å tilby. I dag inneholder sidene mer informasjon, og det er blitt viktigere å holde på brukerne enn å henvise dem videre. Selv om surfing ikke brukes like mye som før, er det fortsatt en viktig måte å finne informasjon, spesielt i kombinasjon med andre metoder.

## Gjetting av URL-er

En god del informasjon er det mulig å finne raskt og enkelt ved å tenke seg til hvor den kan være, og gjette URL-en (adressen) til tjenermaskinen som den

kan befinne seg på. Ofte er URL-en på en standardisert form som gjør dette relativt enkelt.

Et mulig sted for å finne informasjon om informasjonsvitenskap kan for eksempel være American Society for Information Science som vi kan gjette oss til har URL (www.asis.org). På jakt etter gamle nyheter eller programvare til PC-en? Prøv henholdsvis avisa Dagbladet (www.dagbladet.no) eller Microsoft (www.microsoft.com). Med litt kjennskap til hvor man finner informasjon, gjetter man raskt riktig URL.

## Klipping av URL-er

Gjennom surfing eller søk kan vi av og til komme til interessante sider som ikke har lenker videre eller mulige interessante sider som ikke eksisterer (døde lenker). Det kan da tenkes at det finnes andre tilsvarende sider eller at sidene som ikke eksisterer er flyttet til en annen plass, begge deler på samme tjenermaskin. Løsningen er å »klippe« av URL-en på høyresiden »slash« for »slash«, til det kommer opp ei overordnet side som man kan gå videre med. På denne måten kan man til og med finne informasjon som ikke »offisielt« er lagt ut

## Indirekte søk

Ved bruk av søkemaskiner eller emneindekser kan det noen ganger være vanskelig eller umulig å søke frem nøyaktig det man skal ha. Med den sammenvevde informasjonen på Internett er det i mange tilfeller heller ikke nødvendig. Det er nok å finne ei side som er lenket til ei side som er lenket til den relevante siden. Tidligere var databasen begrensningen, i dag fungerer den som utgangspunktet.

## Gjenfinning på Internett er en kombinasjon av mange teknikker

Det raskeste er å gå rett til websider med relevant informasjon hvis man vet hvor de er. Deretter kan emneindekser som YAHOO være en grei fortsettelse for å finne sentrale sider som kan gi svaret,

eller være et bra utgangspunkt for surfing. Hvis ikke det gir godt nok svar kan kanskje noen av søkemaskinene hjelpe. Søkemaskiner kan gi svaret eller være utgangspunkt for videre surfing og evt. klipping av URL-er. Det er viktig å være kreativ i forhold til søkeord og formulering av søkestreng. Til slutt kan det være aktuelt å spørre personer via e-post, nyhetsgrupper, osv. Det spesielle med Internett er at det ofte finnes ulike mange teknikker og verktøy for å komme frem til samme informasjon. Dette gjelder i hovedsak den åpne frie informasjonen. For klassisk biblioteks-litteratur vil selvsagt referansedatabaser og bibliotekskataloger være det naturlige stedet å starte.

## Oppsummering

Informasjonsbehov, informasjon, søkeverktøy og søketeknikker på Internett er ikke helt som vi er vant med fra tidligere. Mye er fundamentalt forskjellig og dette er det viktig at vi har oversikt over og lærer oss. Det finnes ingen snarvei for å bli ekspert i søking på Internett. Som vi er vant med fra tidligere er det også her snakk om hardt praktisk og teoretisk arbeid gjennom lang tid. Det er mange ferdigheter vi må tilegne oss:

- Vi må være god i å finne ut hva brukeren skal ha.
- Vi må kunne fagområdet som brukeren skal ha informasjon om.
- Vi må vite hva som finnes av informasjon og hvor
- Vi må vite hva slags lokaliseringsverktøy som finnes, hvordan de virker, hvordan man bruker dem og hva de er egnet til.
- Vi må ha god kjennskap til søkeprosessen og vite hvordan all denne kunnskapen kan brukes til å finne informasjon.

Vi må kunne dette best av alle. Det er vi som er ekspertene!