

Google og bibliotekene

Mange vil dele utviklingen på Internet inn i tiden før og etter Google.

Google presterte det utrolige å sette en, i utgangspunktet enkel, søkemaskin øverst på listen over populære internettsteder og har på få år vokst fra å være en av mange, og ikke alltid den beste, søkemaskinen på veven til å bli en av verdens mest kjente varemerker og brukte vevsteder.

Av Even Flood

Før Google var det også mange søkemaskiner, men ingen så noen fremtid i å bare satse på en tjeneste hover det eneste man kunne gjøre var å søke etter vevsteder. Tvert imot, et vevsted hvor hovedoppgaven var å holde på brukerne kortest mulig og sende dem videre til andre steder, hvorfor skulle noen satse penger og innsats på å lage noe slikt? Portaler med totaltilbud hvor man skapte lojalitet og holdt på kundene, ga dem tid til å lese reklamen (som man tjente penger på), det var tidens løsning!

Mest brukt

Google satte for alvor en søkemaskin i sentrum og har i få år vokst fra å være en av mange, og ikke alltid den beste, søkemaskinen på veven til å bli en av verdens mest kjente varemerker og brukte vevsteder. Det var mange om beinet da Google ble lansert og lenge var det alvorlige konkurrenter i markedet, som Alltheweb, AltaVista og NorthernLight, for å nevne noen. Men etter hvert har Google med sitt enkle søkegrensesnitt og, viktigst, en suveren algoritme for å finne frem til det som virket som de mest relevante resultatene gjort at den er blitt markedsleder og kan oppleve det verste av alle dominerende merker: At navnet blir et vanlig ord, løsrevet fra opprinnelsen. Kanskje kommer en dag da folk vil snakke om å google Yahoo?

Populær

Det de hadde gjort var å bruke antall lenker til et sted som et relevanskriterium for å rangere utskrifter. Et enkelt prinsipp som ga overraskende gode resultater. Selv om ikke alltid de aller beste stedene kom øverst på svarlisten. Popularitet er ikke alltid et godt kvalitetskriterium. Dette prinsippet for rangering er for øvrig meget likt det ISI lanserte i Science Citation Index, og tanken bak Googles rangering og siteringsanalyser er mye det samme. Googles historie er beskrevet i flere bøker som er kommet ut de siste månedene og jeg skal ikke gjennomgå den her. Men det er en meget interessant historie, spesielt hvordan de greide å gjøre prosjektet lønnsomt og eierne til noen av de rikeste i dag og spesielt boken til John Batelle anbefales.

Google ga fra starten biblioteket mange utfordringer. Mange spurte, hva er vitsen med bibliotek når alt finnes på nettet og vi har Google til å finne det? Svaret for oss som hadde søkt en del fra før var enkelt. Google finner ikke alt. Det finner noe og ikke alltid de beste vevstedene sett fra et kvalitetssynspunkt. Steder som ser interessante ut kan være fulle av desinformasjon, og enda verre, den gode kvalitetssikrede informasjonen er ofte i den skjulte del av veven, inne i databaser, bøker, tidsskriftartikler og oppslagsverk som man må søke direkte i for å finne det som er mest pålitelig.

Google Scholar

I november 2004 kom Google med et nytt produkt som overrasket alle: Google Scholar. Det var Googles svar på alle kommentarer om at Google ikke dekket vitenskapelige kvalitetssikrede arbeider. Google hadde av flere store forlag fått tillatelse til å indeksere hele teksten i vitenskapelige arbeider, inklusive litteraturlisten, og gjøre dem søkbar i en egen database over vitenskapelige arbeider. I tillegg tok Google med alle de offisielle åpne arkivene for vitenskapelige publikasjoner som finnes og la dem inn i basen. Mange forlag var ikke med i prosjektet, spesielt Elsevier holdt sine utenfor siden de hadde sin egen base som gjorde noe tilsvarende, Scirus. Google Scholar ble meget hyppig kommentert mange steder. Basen har mange gode og interessante muligheter. Den er multifaglig og egner seg godt innen de aller fleste fag. Rangering av svarlister er primært antall siteringer av et arbeid. Dermed opplever man ofte at verkene som kommer øverst ikke er artikler som ligger inne i basen i seg selv, men de som forekommer oftest i litteraturlistene i basen. Den kan ofte gi overraskende og meget gode resultater. Når Google Scholar allikevel er blitt mye kritisert er det fordi det er usikkert og uklart hva som egentlig foregår. Når det gjelder søkning i noe så stort og ustrukturert som verdensveven er ikke det så farlig, det er urimelig å kreve høy presisjon der. Men her dreier det seg om en mindre mengde data av høy faglig karakter, og da vil gjerne brukerne, spesielt bibliotekarer som er vant til søk og oversikter over hva man jobber med ha mer informasjon. Som: Hvilke kilder er egentlig inne i basen som primærdokumenter? Hvor ofte oppdateres den (meget viktig!).

Brukerne får i første omgang bare tilgang til de bibliografiske opplysninger. Men hvis det dreier seg om en

artikkel i et åpent arkiv så ligger det en lenke til hele teksten. Det kan ligge lenker til flere versjoner av samme arbeid, både den som er publisert i et tidsskrift og til forfatterens egne arkiver. De siste er ofte gratis. Google Scholar har også mulighet for å koble direkte til bestandsopplysninger for de enkelte bibliotek. Dermed kan brukere som får bibliografiske opplysninger også få direkte tilgang til den hele teksten hvis han har lokal adgang. Hvordan dette fungerer er litt uklart, Google gir denne beskrivelsen: "Once you tell us what library you're a member of, we'll keep an eye out for that library's subscription materials and provide special links to them in your search results." Jeg er ikke sikker på om jeg synes den setningen er betryggende.

Google og Worldcat

Når man søker i Google Scholar er ofte en god del av trefflisten bøker som er mye sitert, og der kan det dukke opp et valg som heter "Library Search". Går man til dette kommer man til en innførsel i WorldCat fra OCLC. OCLC har inngått et samarbeid med Google og Yahoo om å legge hele katalogen til WorldCat inn i databasene og gjøre den søkbar der, prosjektet heter Open Worldcat.. Foreløpig er bar en liten del av WorldCat lagt inn. Planene er å gjøre hele katalogen søkbar, men det ser ut til at det er et godt stykke igjen der. Tross alt er WorldCat en av verdens største katalogdatabaser med godt over 50 millioner poster. Og selv når den er komplett er dette på ingen måte ment å være en erstatning for FirstSearch, det er alt for få søkemuligheter. Når man har funnet en bok kan man også finne bibliotek som har den. Ingen norske bibliotek er med i WorldCat, så brukere som vil finne boken det må søke i norske databaser etter den for å lokalisere verket. Mer om dette er på www.oclc.org/worldcat/. Det er mulig å søke direkte i postene fra Open Worldcat i Google og Yahoo ved å avgrense søkingen til domenet worldcatlibraries.org. Firefoxbrukere kan legge inn søking direkte i Open Worldcat som alternativer i søkeboksen.

Google Book Search

Google Scholar skapte mange diskusjoner, men det er ikke noe mot den oppstandelsen det ble da Google i desember 2004 kunngjorde og i mai 2005 lanserte sitt mest kontroversielle produkt: Google Print, ganske nylig omdøpt til Google Book Search, og populært kalt Google Books. Google hadde inngått en avtale med en rekke meget store bibliotek: Bibliotekene til Harvard, Stanford, University of Michigan, University of Oxford, og The New York Public Library om å digitalisere samlingene og gjøre dem søkbare. Det var et enormt program og spørsmålene sto i kø. For det første: Var det teknisk mulig å gjennomføre det? Digitalisering er tid- og resurskrevende, hver enkelt side skal leses inn og gjøres lesbar digitalt. Det andre spørsmålet var: Har de lov til å gjøre det, det er tross alt rettighetsbeskyttet materiale mye av det. Det var ikke helt nytt det Google gjorde. Amazons søkemaskin A9 hadde allerede latt brukere få søke i bøker. Men det var med forlagenes samtykke. Google hadde ikke tenkt å spørre om lov. De påberopte seg "fair use" siden de ikke la ut hele verket til brukerne, de bare gjorde det søkbart og så lar brukeren få lese de få (?) sidene hvor de relevante søkeordene forekommer. Dermed blir det ekvivalent med å lage en indeks til verkene. Dette har ikke beroliget rettighetshaverne og flere har gått til søksmål mot Google for å få stoppet digitaliseringen. Saken har skapt reaksjoner også fra Europa, ikke minst fra Frankrike hvor de er bekymret for at ikke-engelsk litteratur vil bli liggende i bakleksa når det gjelder å bli digitalisert. EU har svart med prosjektet: i2010 for å lage et digitalt bibliotek, mer om det finnes på europa.eu.int/information_society/activities/digital_libraries/index_en.htm og www.i2010.org.uk/. Et annet initiativ for å konkurrere med Google Book Search og støttet av erkekonkurrenten Yahoo er Open Content Alliance, www.opencontentalliance.org. Statusen otill Google Books er usikker akkurat nå på grunn av søksmål fra flere amerikanske rettighetsorganisasjoner. Googles visjon og forklaring om hva de gjør og deres synspunkt på rettigheter finnes på <http://books.google.com/googleprint/library.html>.

Utfordringene for bibliotekene

Denne utviklingen har ført til nye utfordringer for bibliotekene. For det første, den forslitte uttalelsen "Hvorfor skal vi bruke biblioteket, alt finnes på Internett" har kommet litt nærmere en realitet. Alt finnes ikke, men det blir stadig mer, også av informasjon som tidligere var skjult inne i artikler, bøker og oppslagsverker og ikke direkte tilgjengelig via de store søkemaskinene. Det "skjulte" del av veven er blitt mer synlig. Og når utstyret for å lese elektroniske bøker blir bedre og mer tilgjengelig, har vi en del problemer. Men fortsatt er det slik at kunnskap om litteratur og oppslagsverker er uvurderlig. Google har ingen kildekritikk når det gjelder kvalitet, og det vil heller ikke digitalisering én masse gi. Det er fortsatt mye tvilsomt som publiseres også i bøker, og det er en oppgave å formidle kildekritikk og hjelpe brukere til å evaluere materialet de finner. Og mye av den "skjulte" veven (populært kalt "invisible web" eller "deep web" på engelsk) er fortsatt skjult. Skjult i trykte oppslagsverk, i betalingstjenester, inne i databaser som er på kommersielle informasjonstjenester. Selv om man finner mye mer via Googles forskjellige tjenester enn man gjorde før så er det fortsatt mye igjen der ute.

En ny utfordring er tendensen til overforenkling av søkegrensesnitt. Brukerne er blitt så vant til Google at de venter et like enkelt grensesnitt for alt, og tjenester forsøker å etterkomme deres ønsker: Dette fenomenet har til og med fått et navn, googlifisering. Vi får et søkebilde som er en enkel linje hvor man kan skrive inn noen få søkeord og få ut svarene. De gamle kompliserte grensesnittene er ikke brukt, hvor gjøre livet så vanskelig? Booleske operatører, trunkering, nummererte søkesett – hvem trenger det? Det er utrolig irriterende å få som kommentar når man påpeker mangler i et søkegrensesnitt: "Dette er ikke for eksperter, men for sluttbrukere!", som om det å være ekspert gjør at man ikke er meningsberettiget.

Problemet er at verden er ikke alltid så enkel. Søkegrensesnittene i de klassiske tjenestene ble så kompliserte fordi vi brukere ba om det for å få renere og mer presise resultater. Når man søker i en så ustrukturert

informasjonsmengde som verdensveven, er det ikke noe behov for sofistikerte søkemuligheter. Men når man arbeider med kvalitetssikrede og høyt strukturerte data og en mye mindre database er det stor gevinst i å ha muligheter for å presisere søket.

Ta for eksempel en så aktuell sak som spørsmålet om bruk av mobiltelefoner gir øket risiko for skade, spesielt kreft, i hjernen. Mobiltelefon på engelsk kan være cellular phone, cellular telephone, cell phone, mobile telephone, mobile phone og en del andre alternativer, og både entall og flertallsformer må være med.. Hjerne, brain, er enkelt. Men kreft kan være cancer, tumour, tumor. Og carcinogenic eller carcinogens er også aktuelle søkeord. Her er det nødvendig, eller i alle fall en stor hjelp hvis man kan bruke trunkering og boolske operatører og lagre resultater i søkesett.

De store kommersielle tjenestene: Dialog, Web of Science, FirstSeach, Scopus og mange flere har alle disse mulighetene. Googlifisering kan hjelpe mange å komme i gang med å søke og finne informasjon. Men hvis man vil være en seriøs bruker vil man føre eller senere komme til en oppgave som krever mer avansert enn bare et par søkeord. Og da er det vår jobb å vise brukere at det finnes bedre muligheter og få dem til å benytte disse mulighetene og at de er viktige for svært mange søk. Hvis de ikke tar poenget, vis dem en mobiltelefon.

Mere litteratur:

Dette er bare noen få aktuelle kilder. Det skjer så mye her, både i trykt materiale og på web at det er umulig å gi en god bibliografi. Et råd jeg kan gi er ø bruke søkemaskinene for å finne omtaler, men ikke bare Google. Både Yahoo og Clusty gir interessante svarlister.

Om Google: John Battelle :The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture, Nicholas Brealey Publishing 2005, ISBN: 1857883616
David A. Vise: The Google Story, Macmillan 2005, ISBN: 1405053712

Google Scholar: Denne har jeg skrevet om i mer detalj med en del henvisninger i Bibliotekforum nr 3, 2005. Den er også online: <http://www.norskbibliotekforening/article.php?id=911>
Dessuten anbefales Lars Iselid: Kritisk granskning av Scopus, Web of Science och Google Scholar, powerpointpresentasjon som finnes på <http://www.ub.umu.se/tjanste/lais/>

Google Books: Umulig å gi en skikkelig bibliografi siden diskusjonen går hett og kontinuerlig, men søkemaskinene gir mange meget relevante treff på ordene Google og print, eller Google og books, selv fikk jeg et godt resultat med Clusty, www.clusty.com. Det er også omfattende diskusjoner om Google Books og Google Scholar på listen Web4lib: <http://lists.webjunction.org/web4lib/>, og der kan man også lese eldre poster, eller man kan søke i Google og legge på web4lib som søkeord.

Når det gjelder Gorman og hans synspunkter og diskusjonen rund dem anbefales å søke etter ordene michael gorman google. Det blir mange interessante treff. Et referat fra seminaret hvor Gorman innledet om dette på Onlinemøtet i London i november i år er på <http://www.iwr.co.uk/information-world-review/news/2148333/gorman-slams-digitalisation>.

Usynlige web: En god artikkel med lenker finnes på Wikipedia: http://en.wikipedia.org/wiki/Deep_web

Googlifisering: Lars Iselid: Google vs. biblioteket: bibliotekets roll i en googlifierad värld, <http://iselid.info/ppt/googlifiering.ppt>. Foredrag på NFF seminar i Trondheim 3/11/2005, andre foredrag fra dette seminaret er på http://biblioteknett.no/alias/nff/nfftrondheim/seminar_host05/program.htm.