

Google Scholar – det globale e-bibliotek?

Af Peter Søndergaard

For de biblioteksansatte kan fornøjelsen over forsøget med et kraftfuldt enhedssøgeværktøj, som en udviklet Scholar vil være, nemt blive blandet med frygten for fremtiden, hvis den bibliotekariske ekspertise ikke mere efterspørges af brugere, der mener, at Scholar gør dem selvhjulpne.

Peter Søndergaard
Publikumschef RUB
pso@ruc.dk



Googlemanien synes ingen ende at ville tage. Den 17. november 2004 søsatte Google beta-versionen af en ny specialdatabase ved navn Scholar, altså den videnskabelige Google, og derfor en database som vil have stor interesse for forskningsbibliotekerne og deres brugere. Databasen indeholder indeks over videnskabelige tidsskrifter i fuldtæst, videnskabelige papirer og rapporter i fuldtæst, afhandlinger og preprints ligeledes i fuldtæst; i foråret 2005 angiveligt 50 millioner indekserede dokumenter og henvisninger til dokumenter. Disse henvisninger består som oftest af metadata, citationer og bogkatalogposter, såkaldte metadata, over dokumenter, som ikke findes i fuldtæst i Scholar, men som citeres i de indekserede fuldtæstdokumenter.

Det er et nyt træk ved Googles forskellige indeksdatabaser og spørgsmålet er, om det i første omgang er interessant for Googles traditionelle brugere. Men for forskningsbibliotekerne er det velkendt og velkomment. Vi er på hjemmebane og det præger den modtagelse Google Scholar har fået i den internationale bibliotekspresse, men herom lidt senere.

Googlemanien er så gennemtrængende, at man dårligt kan huske, hvad man gjorde før Googletidsregningens begyndelse. I dag er det et faktum, at kun 3 % af de danske studerende på videregående uddannelser aldrig bruger Google, 40 % bruger søgemaskinen dagligt, 80 % mindst en gang om ugen; det viser tallene fra en sprintny undersøgelse¹.

Den første web-baserede søgemaskine var Lycos fra 1994, udviklet på Carnegie Mellon University. Google var et resultat af et forskningsprojekt på University of Stanford 1995 med deltagelse af de studerende Sergey Brin og Larry Page, som i 1998 stiftede et it-firma og introducerede den ny søgemaskine. Til forskel

fra det fremherskende princip med opbygning af directories/vejviserindeks, som det f.eks. kendes fra Yahoo, indekserede Googles net-crawlers netdokumenter og web-sites ved hjælp af computergenerede ordindeks. Ordindeksene dannes ved at søgerobotter, også kaldet crawlers eller spiders, systematisk og repetitivt gennem søger og kopierer sider fra nettet, mens indeksprogrammer registrerer ord, navne, tidsangivelser, andre stamdata og deres placering. Disse indeks lagres i databaser. Når man som bruger benytter Google, har man en oplevelse af at søge direkte på nettet; det er ikke korrekt, man søger i Googles indeksdatabaser.

Visningsrækkefølgen af de fund man gør i Googles databaser baseres på PageRanking, opkaldt efter den ene af udviklerne. Den algoritme, som afgør visningsrækkefølgen, er patenteret og hemmelig, men man ved at det bærende princip i rankingen, visningsrækkefølgen afgøres af antallet af nethenvisninger, altså netcitationer fra andre web-sites. Principielt kan man altså ikke manipulere med rækkefølgen og f.eks. betale sig til at få sit eget site op blandt de første 10 visninger. Det er relevansen for andre sites, som afgør, om et site ender med at komme højt op i visningshierarkiet.

Det er selvfølgelig et problem, at den præcise funktion af Google-algoritmen er en af flere patenterede firmahemmeligheder, men man kan mene, at det er et endnu større problem, at brugeren ikke kan bestemme en anden visningsrækkefølge, hvis man måtte ønske det. Visningsrækkefølgen ligger fast.

Googles ambitiøse mission er "to organize the world's information and make it universally accessible and useful". Google har medtænkt, Google er populær og skaber løfterige historier eller tager initiativer, hvor bibliotekerne nøjes med overvejelserne.

Google Scholar

Også modtagelsen af Google Scholar var positiv, velvillig, forventningsfuld, i bibliotekskredse måske lidt afventende, tvivlende. Sådan er vi jo. Den 14. december, en lille måned efter lanceringen af Google Scholar, lynede Google igen, denne gang med Google Print, et initiativ som biblioteksvæsenet selvfølgelig selv skulle have taget. Ikke bare noget pjat med en digitaliseret kulturarv, men for fuld skrue. Finansieret af Google og i samarbejde med universitetsbibliotekerne i Harvard, Stanford, Oxford og Michigan – tre amerikanske og et engelsk forskningsbibliotek – samt New York Public Library indledte man ved årsskiftet et digitaliseringsprojekt, som på få år skal nå målet at digitalisere 15 millioner bøger. Om der bliver adgang via Scholar, ønsker Google endnu ikke at oplyse. Suppleret med 2004-aftalen med OCLC om at gøre WorldCats knap 60 millioner katalogposter søgbare i Google er man tilsyneladende godt på vej til at opfylde missionen. Også bibliotek.dk har i al beskedenhed taget kontakt til Google.

Craig Silverstein, it-leder på Google, udtalte i marts 2004 at "the ultimate goal is to have a computer, that has the kind of semantic knowledge, that a reference librarian has". Ved lanceringen af Scholar kunne referencebibliotekarerne så konstatere, hvad deres ny kollega var i stand til og lige siden har den internationale bibliotekspresse sprudlet med diskussion og analyser af Scholar. På kreditsiden tæller, at Google Scholar har en fri og brugervenlig grænseflade, og at dens ambition om at være den eksklusive fælles søgemaskine til alle frie og klausulerede videnskabelige fuldtæstresourcer på nettet befordres af at bibliotekerne, også de danske, frit kan søge deres beholdninger af bl.a. licenstagsskrifter direkte i Scholar.

Google har medietække, Google er populær og skaber løfterige historier eller tager initiativer, hvor bibliotekerne nøjes med overvejelserne.

scholar.google.com

Googlemania

På creditsiden tæller også databasens mange millioner metadata og gennemstilling til bibliotekskataloger. Rimeligt forekommer det også at spørge: Hvor perfekt skal en betaversion være?

Kritikken

Men der er ingen nåde: Den skal være perfekt fra starten, synes kritikerne at mene, og kritikerne er fortrinsvis de professionelle biblioteksansatte, referencebibliotekarerne. De ønsker en avanceret søgefacilitet. Det får de så i løbet af foråret 2005, men den er selvfølgelig ikke tilstrækkelig avanceret, mener man. Vejet og fundet for let sammenlignet med søge- og visningsformaterne i eksisterende, velafprøvede professionelle referenceværktøjer som ISI's Web of Science og den nye Elseviers Scopus.

Kritikernes tests viser, at der er noget galt med limiteringsfaciliteterne i den avancerede søgefacilitet, når man får flere hits i en limiteret søgning end i en ulimiteret søgning. Det kan mine tests dog ikke bekræfte. Derimod kan jeg bekræfte kritikken af, at man ikke kan browse i Google Scholar, en mangel som gør f.eks. forfatter-søgning og søgning på tidsskrifttitel usikker, besværlig og undertiden uoverkommelig.

Man kritiserer ligeledes, at indholdet i Google Scholar er mangelfuldt deklareret, at fordelingen på fagområder er uoplyst og at Scholar-bagmændene aldrig har defineret, hvad der afgrænser videnskabelig litteratur og web-sites fra anden litteratur og andre web-sites. Man kritiserer, at man ikke selv kan bestemme visningsformat eller visningsrækkefølge af resultater i en Googlesøgning, men er overladt til Googles egen, uigennemsigtige, hemmelige PageRanking. Man kritiserer, at det tilsyneladende kun er de første 100 - 120 K-bytes i et fuldtekstdokument, der indekseres og at det ofte kun er 10 - 30 % af de relevante fuldtekstdokumenter i et fuldtekstarkiv, som indekseres. Det sidste kan mine test ikke bekræfte.

Endelig kritiserer man med rette, at opdateringshyppighed er ukendt, at der er dubleringer, og at Scholars indhold er domineret af indeks over amerikanske netdokumenter. Den gode, måske benovede presse, Google havde før man trådte ind på bibliotekernes domæne, er afløst af en stærk kritisk bibliotekspræse.

Hvad siger Google så til kritikken af Scholar? Tja, det er ikke nemt at vide, for Google

har en adfærd man ikke er vant til. Google deltager ikke i diskussionen, tager sjældent til genmæle og det kan man undre sig over al den stund bibliotekskulturen, som er en diskussionskultur, der muligvis er fremmed for firmaet Google, udtrykker det synspunkt, at de dokumenter og dokumentkilder, man formidler, skal kunne deklarerer, hvad angår selektions-, søge- og visningskriterier. Adspurgt direkte om hvilke videnskabelige dokumenter, Scholar indekserer, svarer Google med et computer-genereret email-svar: "*While we're unable to send personal responses to emails, please be assured, that we do read all of the emails, we receive and your feedback will be used to improve Google Scholar. If you wrote to share encouragement for Google Scholar, thank you. If you didn't, know that we're working hard to improve Google Scholar, and we hope it'll be more useful to you soon*". Herefter henviser svaret til Scholars hjælpemenu, hvor man selvfølgelig ikke får svar på de store spørgsmål.

Hvorfor Google Scholar?

Hvad er ambitionen med netop denne store Google specialdatabase? Firmaet udtaler at: *The intent with Google Scholar is to build on what came before, exemplified by the Isaac Newton quote: "Stand on the shoulders of giants". Because Google has benefited greatly from scholarly research, Google Scholar is one way for Google to give back to the scholarly community for their help*". Taget for pålydende er Scholar altså en ubetinget gave, en gave til det videnskabelige samfund som tak for hjælpen på Stanford University. Men dette er indtil videre ikke tilstrækkeligt til at formilde kritikerne, som naturligvis også har ret i, at man kan stille krav til en gave. Kun Scholars faktiske ændringsstrategi i løbet af betaperioden i forhold til kritikken, kan afgøre om de biblioteksansattes traditionelle armslængde-princip og berøringsangst overfor Google vil vige. Brugerperspektivet synes klart det umiddelbart mest positive. For de biblioteksansatte kan fornøjelsen over forsøget med et kraftfuldt enhedssøgeværktøj, som en udviklet Scholar vil være, nemt blive blandet med frygten for fremtiden, hvis den bibliotekariske ekspertise ikke mere efterspørges af brugere, der mener, at Scholar gør dem selvhjulpne.

Selv om Google Scholar forsøget skulle slå fejl og databasen aldrig kommer ud over beta-stadiet, er der næppe tvivl om, at der stadig vil udvikles søgemaskiner, som bliver stadig mere effektive, også set med videnskabelige briller. Det særlige ved Scholar er, at den er gratis og derfor har store besparelspotentialer.

Men brugerne fortjener et modspil, som kunne være en omlægning af de traditionelle bibliotekskurser, hvis positive effekt kun vanskeligt lader sig påvise, til kurser, som tager udgangspunkt i de søgemaskiner, brugerne påviseligt benytter med stor flid. "*Grav hvor du står*", hed en god bog i begyndelsen af 80'erne om at tage udgangspunkt i det nære og nærværende. Måske kunne det være ledemotiv og hjælpe med at sortere i den megen tale om informationskompetence. Måske kunne man bryde dødvandet og nedbryde det opstyltede, hvis man tog praktisk afsæt i, hvad brugerne gjorde og ønskede og viste sig solidarisk med dem ved at give dem en kritisk kompetence og indsigt, der satte dem i stand til at vurdere resultaterne af nettets mageløse søgemaskiner. Jeg tror der ville være massiv tilstrømning til kurser, som kunne berige brugernes ukritiske glæde over Google og Google Scholar med evnen til kompetent vurdering af søgeresultater. Man kunne grave hvor brugerne står.

Danmarks Pædagogiske Bibliotek var det første danske forskningsbibliotek til at benytte sig af Google Scholars tilbud om at understøtte institutionel adgang til bibliotekets e-ressourcer og metadata, straks efter at muligheden i maj måned blev frigivet af Scholar. Kort efter fulgte også Aalborg Universitetsbibliotek. Nu kan disse bibliotekers brugere blive ført fra en søgning i Scholar direkte til de ressourcer, som universitetsbiblioteket har licens til plus alle de frie e-ressourcer, man har meddelt Scholar er relevante. Det er en spændende begyndelse.

Denne artikel er et sammendrag af et oplæg ved Forum for Fjernlans konference på Hindsgavl 19.-20 maj 2005. Præsentationen: Google Scholar - trussel eller alternativ til vore fuldtekst tidsskriftpakker? www.danbib.dk/docs/hindsgavl_2005_pres.htm Her findes også gennemgang af søgeeksempler.

¹ Niels Ole Pors (2005). Studerende, Google og bibliotekerne. Biblioteksstyrelsen 2005-06-12. www.bs.dk/publikationer/andre/studerende/index.htm