

Informationssökning med sökmotorer på gott och ont - vad behöver vi veta?

Nytt och nyttigt på nätet

Lars Våge

Mittuniversitetets bibliotek

Sogndal 2007-09-21

Del 1

Vad är bra med sökmotorerna?

Vad erbjuder sökmotorerna för tjänster idag?

- One-stop shops
- Webbsökning
- Bildsökning
- Nyhetssökning
- Videosökning
- Boksökning
- Hårddisksökning
- Gratis e-post
- Annonsering

Så fungerar en sökmotor i korthet

- Uppbyggda av tre väl avgränsade delar
- spindelprogrammet – hämtar webbsidor och söker efter länkar till fler webbsidor och besöker dessa i sin tur. Hur ofta sker detta?
- indexeringsprogrammet – analyserar webbsidor och extraherar information som indexeras för att göras sökbar. Hur fort görs det?
- sökmodulen – är sökmotorns ansikte utåt på webben och möjliggör sökningen mot indexet. Vilka sökfunktioner finns? Hur väl fungerar relevansrankingen?

Vilka större sökmotorer finns det egentligen?

- Yahoo Search (=Inktomi+Altavista+Alltheweb), ca 20 miljarder?
- Google, ca 10-20 miljarder webbsidor?
- MSN Live Search, ca 5 miljarder?
- Exalead, fransk uppstickare – 4 miljarder?
- Ask.com (tidigare AskJeeves+Teoma) – några miljarder
- Gigablast – minst bland de stora – 2 miljarder

Varför har Google blivit så dominant?

- I slutet av 90-talet drabbades Internet av portalitis, alla webbplatser skulle vara portaler
- Google dök upp 1998-99 när sökningen kommit i skymundan och hade ett gränssnitt som var rent och avskalat
- Googles träffar upplevdes som relevanta, de var först med att använda länkanalys
- Idag är Google vare sig störst, "bäst" eller ger relevantast träffar men vanans makt är enorm
- Google dominerar Internetannonseringen och lever främst på denna och är en global maktfaktor i stil med Microsoft

Specialiserade sökmotorer och metasökning

- Den generella webbsökningen räcker ibland inte
- Det finns mängder av specialiserade söktjänster som är utformade för en viss typ av sökning och är bättre på detta än de generella
- Nyhetssöktjänster är ett bra exempel
- Google och Yahoo har många specialiserade söktjänster
- Metasöktjänster söker i flera sökmotorer samtidigt och presenterar ett samlat resultat
- Ger mindre träffar än att söka separat i flera sökmotorer

Vilka akademiska sökmöjligheter ger sökmotorerna?

- Scirus har funnits sedan 2001 – sökning i gratis artikelarkiv på nätet
- Google Scholar – söker även i vetenskapliga artikeldatabaser som inte är gratis
- Windows Live Academic – söker i samma typer av källor som Google Scholar
- Windows Live Academic har ett bättre gränssnitt för akademisk sökning, ger fler funktioner, bättre träffhantering, lista över källor...
- ...men Google Scholar är mycket mer använd

Del 2 –

Vad är inte lika bra med sökmotorerna?

Sökning i sökmotorer vs forskningsdatabaser

- Att söka i t.ex. Google Scholar är ett alternativ till att söka direkt i forskningsdatabaserna
- Gränssnittet är enklare men i längden handikappande
- En forskningsdatabas är en ordnad, välindexerad datamängd medan ...
- ... ett sökmotorindex är en representation av en kaotisk textmassa utan pålitliga metadata
- I Sverige finns Samsök byggd på MetaLib som är en metasökning i forskningsdatabaserna – men med komplext gränssnitt

Bristen på precisa sökverktyg i sökmotorerna

- Du kan inte söka på författare
- Det finns ingen pålitlig uppgift om utgivningstid
- Det finns inga ämnesord
- En webbsidas titel är inte detsamma som en artikels titel
- Du kan inte följa citeringsmönster bara länkmönster
- Det finns inga abstract i t.ex. Google Scholar

Relevansranking och sortering i sökmotorer

- I sökmotorer används länkanalys som främsta kriterium vid relevansranking
- Relevansrankingen styr sorteringen
- Forskning är en "kumulativ process"
- I forskningsdatabaser presenteras sökresultat traditionellt kronologiskt
- Detta är inte möjligt i generella sökmotorer och inte i Google Scholar
- Beror alltså på att webbsidor som indexeras av sökmotorer inte kan datumbestämmas

Källkritiska problemställningar och kvalitetssäkring

- På Internet måste du alltid bedriva din egen källkritik
- Forskningsdatabaserna är kvalitetssäkrade genom den akademiska publiceringsprocessen
- På Internet kan alltid namn förfalskas, datum förvanskas, innehåll "lånas"
- Utan egen ämneskunskap är du utlämnad
- Kolla vem som länkar till en webbsida som du vill bedöma kritiskt – det är ofta avslöjande

Fixeringen vid en sökmotor

- Googlefixeringen är idag närmast total
- Syns du inte i Google finns du inte...
- ...men Google har inte störst index
- Google brukar slås av andra vid studier av relevans
- Google har inte så avancerade sökfunktioner som vissa andra sökmotorer
- Överlappningen mellan de största sökmotorerna är överskattad
- Övertron på ett sökverktyg är fördummande och utnyttjas

Svårigheter med länkning till vetenskapliga artiklar

- Hur citera vetenskapliga artiklar när det inte är tryckta versioner (i fulltextdatabaser och på nätet)
- Kan man använda webbadressen som den är och ska man ange när man besökt webbadressen?
- Bäst är att ange DOI-nummer om det finns
- DOI=Digital Object Identifier
- Standard för att identifiera digitala representationer av vetenskapliga artiklar hos utgivande förlag
- Windows Live Academic ger DOI-nummer

Del 3

Vad är nytt och på gång?

Trender i sökmotorernas utveckling

- Allt fler specialiserade söktjänster som inte är så lätta att hitta till
- Många har börjat med videosökning
- Att inte spara användarnas sökdata för länge har blivit politiskt korrekt
- Att erbjuda gratis e-post med massor av hårddiskutrymme slåss man om
- Att inte utveckla sökfunktionaliteten i sig...
- ...men att istället erbjuda fler möjligheter att jobba med träffarna

Nyhetsökning och betydelsen av RSS-prenumerationer

- Att kunna hålla koll på nyheter på nätet är av oerhörd strategisk betydelse
- RSS-prenumerationer förenklar detta kraftigt
- Den viktigaste utvecklingen de senaste tio åren
- RSS-läsare finns nu inbyggda i webbläsare, som webbtjänster, som separata program, plugins...
- Många nyhetswebbplatser har ett antal RSS-kanaler för olika typer av nyheter
- Bloggar har RSS, Flickr har RSS, massor av olika typer av webbplatser använder tekniken inkl. sökmotorer

Utvecklingen av sökningen efter bilder och video

- Bildsökning har alltid varit populärt men det är svårt att ge bra sökmöjligheter
- Den omåttligt populära Flickr använder sökord som användarna själva hittat på (tags)
- Ansiktsigenkänning har blivit på modet
- Videosnuttar på YouTube har blivit enormt stort – mest piratkopierat material och hemvideo
- Många använder YouTube som en marknadsföringskanal idag
- Google äger YouTube efter att ha betalat ett fantasibelopp

Podcasting och automatisk taligenkänning i sökning

- Ny publiceringsform på nätet - Podcasting, ljudbloggar, asynkron amatörradio
- Nu kommer en ny generation sökmotorer som har stöd för automatisk taligenkänning
- De skapar transkriberingar av det "igenkända talet" som kan sökas med vanliga sökord
- Tre som klarar av engelskt tal – Everyzing, Podscope/TVEyes och Blinkx, av dessa klarar Everyzing också spanska
- De stora sökmotorerna har ännu ingen taligenkänning (men AOL har)

Web2.0 och tagging utförd av användarna

- Web2.0 – luddigt begrepp som bl.a. står för deltagande, vi skapar webben tillsammans
- Tagging – användarnas egna ämnesord
- Kända tjänster med tagging är Flickr (bilder) och del.icio.us (bokmärken)
- Men även bloggsökmotorer som Technorati använder dem – taggning av träffar
- Taggar är bara ett ord – en allvarlig begränsning
- Taggar har heller inga hierarkier
- Folksonomier vs taxonomier

Wikipedias allt större betydelse som källa och auktoritet

- Allt fler dagstidningar och tidskrifter citerar vanemässigt (engelskspråkiga) Wikipedia
- Ärevördiga Nature visade i en studie i december 2005 att kvaliteten var fullt jämförbar med Encyclopedia Britannica
- Deras undersökning kritiserades men Nature stod på sig och uppmanar forskare att skriva
- En av webbens mest besökta och lästa webbplatser vare sig man vill eller inte
- Ett fantastiskt socialt kunskapsprojekt som utmärks av social kontroll, global granskning, försök till manipulation (jfr WikiaScanner)

Del 4

Vad behöver vi veta?

Vilken typ av söktjänst ska jag använda när?

- Använd de specialiserade söktjänsterna
- Nyheter -> nyhetssöktjänster och bloggsöktjänster -> lär dig RSS!
- Nyheter om forskning -> vetenskapliga pressmeddelandetjänster t.ex. EurekAlert
- Nyheter om företag -> kommersiella pressmeddelandetjänster t.ex. PR Newswire
- Forskningsartiklar -> i första hand forskningsdatabaser i andra hand sökmotorer
- Fakta -> faktasöktjänster, webbversioner av uppslagsverk och Wikipedia med försiktighet

Hur ska jag kunna följa med i utvecklingen?

- Låt andra sköta bevakningen – läs de specialiserade bloggarna
- Pandia.com – ledande europeisk webbplats i bevakningen av söktjänster. Made in Norway!
- Internetbrus.com – svensk blogg
- Många amerikanska bloggar följer utvecklingen
- Search Engine Land – viktigast men publicerar nästan för mycket
- ResourceShelf och ResearchBuzz – mycket bra bloggar som funnits mycket länge
- John Battelle – problematiserar

Hur ska jag söka?

Vilka ord ska jag använda?

- Inget kan ersätta språklig fantasi
- Skriv ned din fråga, ta ut betydelsebärande ord och skriv in dem
- Använd alltid frastecken för flerordsuttryck
- Komplettera med negerade sökord för att ange aspekter och kontexter för mångtydiga ord
- Föreställ dig webbsidan som du vill ha som träff
- Vad för mening vill du läsa som svar på din fråga
- Skriv den meningen med frassökning och utelämnat "svar"

Bevakningar av sökuttryck i sökmotorer och forskningsdatabaser

- Många forskningsdatabaser erbjuder alerts i form av e-post
- Det kan vara nya träffar på sökord eller table of contents för nya tidskriftsnummer
- Vissa databaser erbjuder nu också RSS
- Vissa av nyhetssöktjänsterna erbjuder träfflistor i form av RSS-prenumerationer
- Så fort det dykt upp nya artiklar som svarar mot sökfrågan uppdateras listan med rubriker i RSS-läsaren

Vad kan bibliotekarierna hjälpa till med?

- Att ge vägledning i att på rätt sätt söka i bibliografiska forskningsdatabaser och fulltextdatabaser med forskningartiklar
- Att läsa och förstå referenser
- Att använda referenshanteringssystem som t.ex. EndNote
- Att tipsa om bra söktjänster och hur man får ut det mesta möjliga av sökmotorerna
- Att lägga upp sökstrategier och utforma fungerande sökuttryck som kan användas i automatiska bevakningar