

This is an author produced version of a paper published in Tidskrift för Dokumentation (Nordic Journal of Documentation).

This paper does not include the final publisher journal pagination.

Citation for the published paper:

Våge, L., "Söktjänster för bloggar",

Tidskrift för Dokumentation, 2004, volume 59, issue 2, pp. 45-50.

Published with permission from: Svensk förening för Informationspecialister

Söktjänster för bloggar

L. VÅGE
Mid Sweden University

Abstract

More and more people nowadays are turning to so called weblogs to be able to cope with the ever increasing flow of news that is available on the Internet. This article describes the emergence of specialised search engines designed for those who want to search for weblogs to read and to search through the contents of weblogs. The most important of these new blog search engines and their functions are reviewed.

Keywords

Blogs, search engines.

Behovet av bloggsöktjänster

Bloggar (eng. blogs eller weblogs) är ett slags webbsidor som uppdateras ofta och där nya daterade texter placeras överst. Eftersom allt fler använder sig av nyhetsbloggar som förstahandsval för att följa utvecklingen inom olika intresseområden har det uppstått ett behov av att kunna söka efter intressanta bloggar och att söka i bloggars textinnehåll. Det finns inget bra sätt att söka i bloggarna med hjälp av de vanliga söktjänsterna så därför har det under de senaste åren vuxit fram söktjänster som specialiserat sig på bloggar. Medan utvecklingen går mot allt större konsolidering och mindre mångfald när det gäller de generella sökmotorerna pågår det bland bloggsöktjänsterna fortfarande ett intensivt experimenterande på gräsrotsnivå. Ofta är det en eller två personer som ensamma driver bloggsöktjänster och ibland helt och hållet på sin fritid. Detta möjliggörs av att det än så länge inte finns ett så stort antal bloggar att en sökmotor måste ha mängder av parallellt arbetande datorer för att klara av sin uppgift. Bloggosfären växer dock hela tiden i storlek och betydelse och under våren har mjukvarujätten Microsoft aviserat att de arbetar med att utveckla en bloggsökmotor (1).

Sökmotorer och webbkataloger för bloggar

De vanligaste formerna av bloggsöktjänster är sökmotorer och webbkataloger. Utmärkande för bloggsökmotorerna är att de redan från början använt sig av länkanalys på ett medvetet sätt. Flera av bloggsökmotorerna har genom att analysera länkar genererat olika former av topplistor för t.ex. mest länkade webbsida under dagen. I själva verket var det just sådant som de första bloggsökmotorerna gjorde snarare än att erbjuda fulltextsökning i bloggar. Typiskt för bloggsöktjänster är att de ofta integreras med nyhetssöktjänster. Detta beror främst på det gemensamma syndikeringsprotokollet RSS (Real simple syndication) som bloggar och nyhetswebbplatser i hög utsträckning använder sig av för att sprida sina artikelrubriker. Nu är det å ena sidan inte så att alla vare sig bloggar eller nyhetswebbplatser har en RSS-ström (eng. RSS feed). Å andra sidan finns det både bloggsökmotorer och bloggwebbkataloger som bara indexerar eller katalogiserar

bloggar (och ibland nyhetswebbplatser) som använder RSS eller det nyare Atom-protokollet.

De första bloggsöktjänsterna

Den första webbkatalogen för bloggar verkar ha varit EatonWeb Portal som skapades av Brigitte Eaton redan i slutet av 1999. Inledningsvis hade The Original Blog Directory som man idag kallar sig länkar till ett femtiotal bloggar. Ett år senare var siffran 1300 och numera finns över 17000 bloggar ordnade i olika typer av kategorier som ämnen, länder och språk. Under 2000 skedde förmodligen de tidigaste försöken med spindlande sökmotorer för bloggar. Michael Stilwell hade det året en webbplats kallad Beebo.org med Metalog som var en lista över de webbsidor som hade flest inlänkar från 400 webblogger. Dessutom fanns Metalog Weblog Ratings som visade vilka bloggar som det fanns flest inlänkar till. Stilwell tappade dock intresset för projektet och Metalog lades ner i december 2000. I The Web Log Scoop Index som också tycks ha levt och dött under år 2000 använde sig Jim Aspnes av länkanalys av innehållet i över 800 bloggar. Liksom i Metalog fanns där en topplista över mest länkade webbsidor men förutom det fanns också ett Scoop Index som mätte vilka bloggar som var snabbast med att länka till en webbsida som sedan fler bloggar länkade till. Slutligen kunde man i hans Scooped Index se vilka bloggar som var de som oftast blev länkade till av de andra.

En parisare tappar tålamodet

Michael Krus heter en fransman bosatt i Paris som under 2000 tillsammans med en schweizisk kollega tröttnade på att behöva surfa till mängder av webbplatser för att hitta nyheter (2). Därför skapade de NewsIsFree som var ett av de första projekten på webben där man intresserade sig för spridningen av syndikerade nyheter på Internet. En av idéerna var att man skulle kunna använda webbplatsen som en personlig nyhetsportal. Man fick registrera sig och kunde därefter välja bland de tillgängliga nyhetskällorna och skapa sin egen anpassade version av NewsIsFree. Ända från början har här funnits en särskild kategori för bloggar även om denna inte varit underindelad annat än alfabetiskt. I augusti 2000 fanns redan närmare 800 nyhetskanaler att välja bland varav en del men långt ifrån alla använde RSS. På startsidan visas i dag utvalda större nyhetsrubriker inom några olika områden men man kan också bläddra bland rubriker indelade efter länder. Newsisfree har dessutom ett söksystem där man kan söka i innehållet av de över 9000 nyhetskanalerna eller i databasen med beskrivningar av dessa. NewsIsFree är alltså både en sökmotor och en webbkatalog med en överskådlig katalogstruktur i vilken man kan bläddra bland både bloggar och nyhetswebbplatser.

Cameron och Internetmemetiken

Cameron Marlowe, den då 24-årige studenten vid MITs Media Laboratory, skapade 2001 som doktorandprojekt ett "weblog diffusion index" som han gav namnet Blogdex. Tanken var att man skulle kunna spåra vilken information som sprider sig snabbast på nätet. Han använde termen meme för informationsenheterna och kallade det han gjorde ett sätt att studera Internets memetik (3). Detta gör han genom att titta på utlänkarna från webblogger och den webbsida som har flest inlänkar kommer högst på topplistan i Blogdex som alltså är en slags sökmotor. Man kan också titta på vilka webbsidor som länkats till en sida på topplistan (track this site). Projektets

webbplats blev tillgänglig för allmänheten den 18 juli 2001 och spindlade då ca 9 000 bloggar. Det finns ännu inga arkiv för att se äldre topplistor och är man nyfiken på dessa får man nöja sig med de kopior som finns lagrade i Internet Archive. Sökning i Blogdex inskränker sig till möjligheten att söka bland namnen och beskrivningarna av de bloggar som Blogdex indexerar för att skapa sin topplista.

En amerikan i Hongkong ledsnar på att surfa efter nyheter

Den första större sökmotor med vilken man kunde söka i bloggarnas innehåll hette Daypop. Sökmotorn är skapad av och sköts av den amerikanske programmeraren Dan Chan ensam. Skälet för tillkomsten var enligt Chan själv att han hade svårt att följa med i det kaotiska amerikanska presidentvalet år 2000 från Hongkong där han då bodde (4). Efter återkomsten till Los Angeles ägnade han ett halvår åt att programmera och gjorde därefter sökmotorn tillgänglig i augusti 2001. Daypop indexerar alltså inte bara bloggar utan också traditionella nyhetswebbplatser. Man kan välja att söka i bägge eller söka separat i endera. Sedan tillkomsten har Chan efter hand lagt till en hel rad topplistor baserade på länkanalys. Daypop Top 40 är t.ex. de mest populära länkarna bland bloggare, Top Posts är de bloggartiklar som har flest inlänkar och Top News Stories de populäraste nyhetsartiklarna. Word bursts är lite mer specialiserad och försöker fånga ökat användande av vissa ord bland bloggare under de senaste dagarna. Idén är att identifiera vad bloggare skriver om snarare än vilka webbsidor de länkar till. News bursts listar på liknande sätt ord som använts i ovanligt hög utsträckning under de senaste dagarna på nyhetswebbplatsernas hemsidor. De populäraste webbloggarna rankas i listor dels efter antal inlänkar, dels efter inlänkarnas kvalitet enligt vad Chan kallar Daypop Score. Om man utgår från en viss blogg i Daypop kan man också få tips om bloggar som har liknande innehåll, s.k. similar blogs. Till detta finns också möjligheten att hitta s.k. related blogs som är bloggar som i sin tur länkar till de som kallas related blogs i relation till webbloggen man utgick från. Antalet indexerade webbplatser hos Daypop närmar sig nu 60 000.

Syndic8 sprider syndikeringsmodellen

Syndic8 var det väl valda namnet på ett projekt som Jeff Barr vid Vertex development i Sannamish, USA började jobba med ungefär samtidigt som Daypop dök upp. Barr designade webbplatsen själv och är fortfarande den som är ansvarig för driften. Enligt ett pressmeddelande lanserades Syndic8 i april 2002 men redan i oktober 2001 verkar den ha varit igång på nätet. Meningen var att webbplatsen skulle vara en resurs för alla som var intresserade av syndikering av nyhetsrubriker på nätet via RSS-protokollet och senare även Atom-protokollet. Syndic8 har en stor användargrupp som försöker fånga in och bedöma olika kvaliteten hos olika strömmar (feeds). Man har också hela tiden uppmuntrat webbplatsägare som har ett bra nyhetsinnehåll men saknar RSS-ström att överväga att etablera en sådan (5). I deras webbkatalog som har flera kategoriseringssystem finner man både nyhetsströmmar och rena bloggar och det totala antalet är idag uppe i över 40 000. Man kan söka efter strömmar i Syndic8-databasen men det finns ingen möjlighet att söka i själva innehållet. Överlag är Syndic8 en mycket välskött webbresurs och har en samling med dokument runt syndikering, intressant statistik, bra beskrivningar och mycket annat.

Open Source-entusiasten Shantis Popdex

En bloggsöktjänst liknande Blogdex utvecklades på en månad under 2002 av amerikanen Shanti Braford vid Washington University i St. Louis. Braford gjorde söktjänsten med hjälp av enbart Open Source-programvaror på sin fritid (6). Hans skapelse som kallas Popdex använder sig av en sökmotorspindel som varje dag besöker ca 14 000 bloggar och skapar en topplista över de webbsidor som dessa länkar till. Popdex har tillnamnet "The website popularity index" men vid genererandet av listan sker rankingen i två steg. Det är inte bara antalet inlänkar som räknas utan även vikten av inlänkarna i Brafords s.k. Popscore-system. Således räknas en inlänk från en webbplats som själv har fler inlänkar för mer värd än en inlänk från en webbplats som har färre inlänkar ungefär som i Google. För varje webbsida på listan kan man se "citations", d.v.s. vilka det är som länkar till sidan ifråga. Den här typen av tjänst är utmärkt för att ta reda på vad som rör sig på Internet och vad bloggarna skriver om och hur. Dessutom finns en lista över Top News Stories och en lista som visar de 100 populäraste webbsidorna över hela Popdex historia. Det finns också särskilda news feeds för olika ämnesområden som baseras på nyhetswebbplatsers artiklar. För varje dag från december 2002 finns Popdex-listan arkiverad så att det går att gå in och titta på individuella datum i efterhand. Att söka i Popdex fungerar kanske inte riktigt som skulle kunna vänta sig för istället för en vanlig träfflista får man träffar på sökordet redovisade per dag.

Feedster + rssSearch = bästa bloggsökmotorn?

Feedster är en av de största och mest mångsidiga sökmotorerna för nyhetsströmmar och bloggar som använder RSS och Atom. Ursprungliga Feedster slogs ihop med konkurrerande rssSearch under sommaren 2003 (7) och resultatet har blivit en utomordentligt kraftfull bloggsökmotor. Underligt nog var både skaparen av Feedster, Scott Johnson, och mannen bakom rssSearch, François Schiettecatte, baserade i närbelägna städer i delstaten Massachusetts, USA. Den sammanslagna bloggsökmotorn indexerar nu över en halv miljon webbplatser som spindlas mycket ofta och hela indexet innehåller närmare 38 miljoner artiklar. Feedster hör också till de sökmotorer som erbjuder en cachad kopia av de artiklar man indexerar. Man kan gratis skaffa ett Feedster-konto och erhålla sin egen MyFeedster-tjänst där man kan lägga upp de strömmar som mest intresserar en och lätt hålla koll på dessa i ett webbgränssnitt och därigenom slippa att installera en separat RSS-läsare. På så sätt kan MyFeedster användas som en webb-baserad news aggregator med den utökade möjligheten att enkelt söka i de bevakade strömmarna. Feedster stöder även OPML-filer vilket gör det lätt att importera sitt urval strömmar från någon annan news aggregator som man använt. En annan utmärkt funktion hos Feedster är att man kan skapa e-postbaserade News Alerts för vilka sökuttryck som helst. Dessa alerts går att få varje dag eller t.o.m. varje timme om det dyker upp något nytt som svarar mot sökningen.

Technorati – största bloggsöktjänsten hittills

Technorati är den senaste större bloggsöktjänsten som har dykt upp och är den som idag indexerar allra flesta bloggar, nämligen över 2 miljoner om man får tro Technoratis hemsida. Sökmotorn konstruerades av David Sifry i samband med att denne tillsammans med redaktören för den tryckta tidskriften Linux Journal, Doc Searls, förberedde en artikel om bloggar (8). När Technorati dök upp i slutet av 2002

kunde man bara söka på vilka bloggar som länkade till en given webbsida (dess s.k. LinkCosmos) men efter en omfattande omarbetning i år kan man göra vanliga fulltextsökningar som i vilken sökmotor som helst. På samma sätt som i flertalet andra bloggsöktjänster kan man studera topplistor över vad bloggarna länkar till som NewsTalk, Current Events och i BookTalk även vilka böcker som diskuteras av bloggarna. I Top100 finns de bloggar som har flest inlänkar av alla i det gigantiska indexet. Det sammanlagda antalet länkar som indexet innehåller anges till ofattbara 286 miljoner vilket i så fall är jämförbart med den generella webbsökmotorn Gigablast. Technorati hävdar att de besöker alla sina källor med en mediantid på 9 minuter och kallar att söka i Technorati att söka på The World Live Web snarare än The World Wide Web. Technorati inbjuder villiga programmerare att delta i utvecklingen av sökmotorn och annonserar om att de också anställer personer.

Ännu fler bloggsöktjänster på väg

Det finns ytterligare ett antal mindre kända bloggsökmotorer som är värda att nämna. BlogStreet har funnits sedan sommaren 2002 och sedan dess förbättrats en hel del och har idag en mängd funktioner och indexerar över 140 000 bloggar samt har en egen bloggwebbkatalog. Bland de möjligheter som BlogStreet erbjuder finns att se en Javabaserad grafisk bild av en bloggs virtuella grannskap (visual neighborhood). Waypath är en annan bloggsökmotor som tillåter fulltextsökning, sökning efter relaterade bloggartiklar, sökning efter inlänkar under den senaste månaden och en del annat. The Waypath Project kom igång under 2002 och erbjöd från november 2002 vanlig fulltextsökning. Bakom Waypath som enligt uppgift spindlar 1,6 miljoner bloggar står företaget ThinkTank23 i Seattle. Bloglines är ytterligare en bloggsökmotor men kan också fungera som en RSS-läsare/aggregator. Här finns en snabb, bra sökning i 124 000 bloggar och en topplista för mest länkad webbsida (Top Links). Bloogz är en bloggsökmotor som spindlar RSS-baserade bloggar flera gånger per dag. Sökmotorn förefaller ha italienskt ursprung att döma av den knapphändiga informationen på webbplatsen. Speciellt för Bloogz är att man kan söka bland bloggar med begränsning till olika språk.

Förutom bloggsökmotorer har det under de senaste två åren dykt upp en rad nya webbkataloger för bloggar. Bland dessa har vi t.ex. den ämnesindelade Blogarama som startade i december 2002 och omfattar närmare 10 000 bloggar. Globe of Blogs som funnits sedan februari 2002 och som är ungefär lika stor som Blogarama har både en geografisk indelning och en ämnesindelning. En engelsman med det ovanliga namnet Sven Latham står bakom BlogWise som lanserades i oktober 2002. BlogWise har katalogiserat 17 000 bloggar och har indelning efter ursprungsland vilket gör den intressant att bläddra i. Till sist får vi självklart inte glömma den svenska Weblogs.se som bland annat har en lista över ett hundratal svenska bloggar. Vid sidan av de bloggsöktjänster som tagits upp i denna artikel finns det säkerligen andra som förmodligen förtjänat ett omnämnande men det blir allt svårare att hålla reda på de allt fler bloggsöktjänsterna. Detta känns dock egentligen bara trevligt i en tid då vi får allt färre generella söktjänster och dessa i allmänhet kontrolleras av ekonomiska intressen som inte så sällan står i bjärt kontrast till internetsökarnas behov.

Slutord

Är då bloggarna något att fästa så stor vikt vid egentligen? Ska vi bry oss om att använda alla dessa bloggsöktjänster när vi har sökmotorer som Google? Som svar

angående nyttan av bloggsoektjansterna skulle jag vilja citera en artikel i Discover Magazines novemhernummer 2003 dar forfattaren konstaterar att: "Professional writers and editors generate the stories and the Web's vast audience decides which ones deserve our attention" (8). Traffande kanns också det som Rebecca Blood skriver i den utmarkta boken "The Weblog Handbook" dar hon sammanfattar bloggarnas roll med orden: "Weblogs are not, as some people say, a new kind of journalism. Rather, they supplement traditional journalism by evaluating, augmenting, and above all filtering the information churned out by journalists and the rest of the media machine every day" (9).

Referenser:

1. Bazeley, Michael, Microsoft to create search site for Weblogs, 2004-03-27, Mercury News, www.mercurynews.com/mld/mercurynews/business/technology/8291524.htm
2. Lasica, J.D., News that comes to you, 2003-02-04, Online Journalism Review, www.ojr.org/ojr/lasica/1043362624.php
3. Kahney, Leander, Tracking Bloggers With Blogdex, 2001-07-30, Wired News, www.wired.com/news/culture/0.1284.45546.00.html
4. Glaser, Mark, Can Daypop Stay Out of Google's Headlights?, 2003-05-13, Online Journalism Review, www.ojr.org/ojr/glaser/1052801637.php
5. Bond, Julian, Content syndication: The Syndic8 Project, 2001-10-23, Content Wire, www.content-wire.com/Online/Syndication.cfm?ccs=111&cs=916
6. Shankar, Kriuba, Interview with Shanti Braford, creator of Popdex, 2002, www.kiruba.com/shanti.html
7. Feedster LLC, rssSearch Merges with Feedster, 2003-06-19, Pressmeddelande, www.feedster.com/press/rssmerger.php
8. Johnson, Steven, How the web edits news, 2003, Discover Magazine Vo. 24 No. 11, www.discover.com/issues/nov-03/departments/tech
9. Blood, Rebecca, The weblog handbook, Perseus Publishing: 2002, s. 23.