

## Forskare vid Stockholms Universitet ger råd

# Hur man hjälper besökare hitta på en webbplats

Av: Jacob Palme

Filnamn:URL: <http://dsv.su.se/jpalme/web-structure/hitta-webben.pdf>

Senast ändrad: 04-02-19 11.43

Publicerad i Computer Sweden, 16 februari 2004

Hur ofta hör man inte klagomål över hur svårt det är att hitta på olika webbplatser? Varför är det så, och hur kan man göra för att besökare lättare skall kunna hitta på en webbplats? Jacob Palme, professor i Data och systemvetenskap vid Stockholms Universitet, ger råd.

Den grundläggande orsaken till att det ofta är svårt att hitta på en webbplats är att de som konstruerar webbplatsen har svårt att leva sig in i besökarens behov. De som konstruerar webbplatsen är experter på sin egen webbplats, därför har de inga egna problem att hitta information. De är också ofta experter på det faktaområde som webbplatsen täcker.

Utmärkande för experter är att de oftast har ett eget språk. De vet det rätta ordet för varje begrepp, och kan inte förstå att människor som inte är experter kan vilja söka information med andra ord än de experten använder.

För att det skall vara lätt att hitta i en webbplats måste man göra en noggrann manuell organisation av informationen, och kolla denna mot många användares sätt att leta information. Denna organisation av informationen måste kunna se problemen ur användarens synpunkt. Experter på sakinnehållet i webbsajten har ofta svårt att se problemen ur användarens synpunkt. Expertens aktiva ordförråd är ofta mera begränsat än icke-experten. Experten vet det rätta ordet, och använder inte andra ord än just det rätta ordet. De som söker på en webbplats, och som inte är experter, söker ofta med andra ord än expertens. Men webbplatsen är konstruerad, så att man bara hittar information om man söker med det rätta ordet.

Exempel: En expert vet att det rätta ordet för anpassning av pensioner till inflationen är ”basbelopp”. Den som inte är expert, söker kanske med ord som ”inflation” eller ”stigande priser”, och hittar då inte svaret på sin fråga, därför att de som konstruerat webbplatsen inte förstått vilka ord en icke-expert kan tänkas använda för att hitta denna information.

## Kolla hur besökare söker information

Det allra viktigaste rådet, om man vill hjälpa användare att hitta på en webbplats, är att observera ett flertal verkliga användare, personer som inte har varit med och konstruera webbplatsen, och som inte är experter på webbplatsen innehåll. Logga hur de gör, analysera loggarna för att se hur de verkliga användarna gör, vilka problem de får. Gör om webbplatsen, så att den fungerar bättre när andra gör likadant som de observerade användarna. När man har analyserat hur ca hundra olika personer söker en viss information, och anpassat webbplatsen så att andra som gör likadant som dessa hundra personer får bra svar, då kan man vara ganska

säker på att webbsajten för de allra flesta i framtiden kommer att göra det lätt att hitta denna information på webbsajten.

## **Två metoder att söka – bläddring och att skriva sökfrågor**

Det finns i huvudsak två metoder att söka information på en webbplats. Den ena metoden är att bläddra och klicka sig fram, den andra är att skriva en sökfråga i ett sökfönster. Den ena metoden utesluter inte den andra, i den mycket framgångsrika webbkatalogen Yahoo, börjar man ofta med att skriva en sökfråga, hittar då en katalogsida ur Yahoos katalog, och bläddrar vidare därifrån för att hitta det man söker.

## **Hantering av sökfrågor från användare**

En av de allra viktigaste metoderna att söka information är att användaren skriver en sökfråga i ett sökfönster. Det är t.ex. den metod användare söker med i världens idag mest framgångsrika verktyg för att söka information på Internet, sökmotorn Google.

Många webbsajter erbjuder en söktjänst baserad på s.k. fritextsökning. Det innebär att den letar reda på alla sidor, där orden i sökfrågan förekommer. Om flera sidor innehåller orden i sökfrågan, prioriteras svaren, så att de sidor där orden förekommer i rubriken, i början av texten, eller många gånger, visas först.

Denna metod, som den används av sökmotorer på myndigheters webbplatser, ger nästan alltid usla resultat. Det är för det mesta mycket svårt även för en mycket intelligent användare att hitta svar på en fråga med hjälp av denna typ av automatisk fritextsökning. Ofta hittar man många sidor som är föga relevanta, och när frågan matchar många svar, kommer ofta det relevanta svaret långt ner i listan över hittade svar.

Google har ju nått stor framgång genom att de kan ge bra svar. Deras metod är att kvalitén på varje webbsida bedöms. Denna kvalitet har sedan stor betydelse vid sorteringen, så att sidor med hög kvalitet kommer högst på listan. Googles metod att bedöma kvalitén på en webbsida är att se på hur många andra webbsidor har länkar till denna webbsida. Vid denna bedömning värderas länkar från webbsajter med gott rykte bättre än länkar från webbsajter med sämre rykte. Googles betyg på webbsajter sätts i en skala från 1 till 10, och bara länkar från webbsajter med betyg högre än 5 räknas med när man bedömer kvalitén på en webbsajt.

Googles metod bygger alltså på att många olika människor visar sin kvalitetsbedömning av andras webbsajter genom att lägga länkar till dem. Det är alltså i grunden ett mänskligt jobb att sätta betyg på webbsajter som Google utnyttjar. (Många andra sökmotorer än Google har börjat ta över liknande metoder.) En nackdel med Googles metod är att svaren ibland är mera bra än relevanta. Man hittar intressant information, men inte alltid det man letar efter.

Orsaken till att fritextsökning på myndigheters webbplatser ger så dåliga resultat är att det helt enkelt inte går att hitta det bästa svaret på en fråga med denna typ av automatiska metoder. Enligt min erfarenhet får man oftast bättre resultat med Google än med myndigheternas egna fritextsökningssystem, även när man vill söka i en viss myndighets webbsajt.

Man brukar mäta kvalitén på sökresultat med två kriterier relevans (eng. recall) och precision. Med täckning menas hur många av alla relevanta dokument som hittas, med precision hur många av de hittade dokumenten som är relevanta. En perfekt sökmotor skall alltså ge högsta betyget (1,0) för både täckning och precision.

Dessa mätetal är omdiskuterade. Det finns det som hävdar att det är bättre om man mäter hur nöjda användarna är med de svar de får. Hur nöjda användarna är beror inte bara på täckning och precision, utan också på att det finns bra och begripliga svar på det användarna söker. Även en sökmotor med perfekt täckning och precision, kan vara dålig om de används på en dålig dokumentsamling, där det inte finns några bra och begripliga svar på det användarna vill veta. Och om det omvänt finns många relevanta svar, är användaren kanske mer betjänt av att få det eller de bästa än att få en lista över alla.

## Klassificering (Indexering)

En metod att höja kvalitén på sökresultaten är klassificering (även kallat indexering). Med klassificering menas att man kompletterar varje dokument med en beskrivning av vilka frågor som bör leda mot detta dokument. Ett dokument om hur pensioner skyddas mot inflation kan t.ex. ha en klassificering som säger att detta dokument skall hittas av alla frågor som dels innehåller ordet "pension\*", dels något av orden "inflation\*", "penningvärde\*", "pris\*" där "\*" anger alla ord som börjar på det som står före "\*". penningvärde\* hittar alltså även när användaren använder ord som t.ex. "penningvärdesförsämring" eller "penningvärdesanpassning". En bra klassificering måste göras manuellt, av människor med kompetens att göra en bra klassificering. En automatisk klassificering som görs av ett datorprogram blir aldrig särskilt bra.

Om man vet att de som söker är experter, kan man göra klassificering med bara de rätta orden, de som experterna använder. Men om man skall tillhandahålla en sökmöjlighet för allmänheten, så måste man vid klassificeringen inte bara ta med experternas rätta ord, utan också alla varianter som icke-experter kan tänkas använda. Man kan kalla det för synonymer, men det är ofta synonym i en vidare betydelse än detta ord vanligen används för. T.ex. är ju inte orden "inflation" och "basbelopp" och "värdesäkring" synonymer, men för en användare som söker information om värdesäkring av pensioner, bör svaret hittas vilket som helst av dessa tre ord som användaren använder.

Man kan säga att klassificeringen handlar om att skapa mallar som täcker alla olika varianter av formuleringar av frågor som användarna kan tänkas ställa.

Som exempel följer här den klassificering vi gjort för webbsajten <http://web4health.info/sv/> på en sida med information om hur barn reagerar på traumatiska upplevelser:

**Kräv:** \$barn ; trauma grym\* ohyttig\* illdåd\* ondska\* vidrig\* sjukhus\* krig\* våld\* slag\* misshand\* smäll\* dask\* smisk\* aggress\* pisk\* ilsk\* övergrepp\* olyck\* sorg\* hot\* risk\* fara\* slåss\* kämp\* kamp\* boxn\* spark\* [ när om vid # dör död\* dog avlid\* ] / PTSD [post ; trauma\* ; stress\* ]

**Prioritet:** reager\* uppför\* påverk\* symptom\* effekt\*

**Alternativ:** \$brådåligt \$sjukdom

Ovanstående klassificering hittar detta svar på så olika frågor som ”Hur reagerar barn på krig” eller ”Vad har sjukhusvistelse för inverkan på barn” eller ”Vad är effekten på barn när någon dör” eller ”PTSD”.

Förklaring: \$barn \$brådåligt och \$sjukdom hänvisar till termer i en separat synonymlista med synonymer som ofta behövs i klassificeringen för denna webbsajt. T.ex. är \$barn ekvivalent med ”baby\* barn\* bebi\* dotter\* flicka\* grabb\* jänta\* pojke\* son\* sonen söne\* spädbarn\* tjej\* tös\* ung\* ettåring\* tvååring\* treåring\* fyraåring\* femåring\* sexåring\* sjuåring\* åttaåring\* nioåring\* tioåring\* elvaåring\* tolvåring\* trettonåring\* fjortonåring\* femåring\* tonåring”. Mer information om hur den sökmotor vi utvecklat fungerar finns på <http://web4health.info/KOM2002/natural-language-answering.htm>.

Det tar ca 10-15 minuter att göra en första sådan noggrann klassificering av ett dokument. Sedan måste man logga användarnas frågor, speciellt frågor där inget bra svar hittades, och ändra klassificeringen, så när en webbsajt är färdigklassificerat har man lagt ner ca 30 minuter på klassificeringen av varje dokument. Klassificeringen bör göras av en person som är duktig på att göra sådana klassificeringar. Experter är ofta olämpliga, därför att de gör inskränkta klassificeringar med bara de ”rätta” termerna. Företag som säljer söksystem erbjuder ofta paket i vars pris ingår både söksystemet och klassificering av en webbsajt.

Givetvis behöver man inte lägga ner detta arbete på alla dokument, utan man kan välja ut de dokument som är viktigast för att ge den information allmänheten behöver.

Många myndigheter drar sig antagligen för att lägga ner så mycket arbete på sin webbsajt. Men man skall då tänka på det arbete man lagt ner på att producera dokumenten, jämfört med det är ju klassificeringen ändå en liten marginell merkostnad. Och man måste tänka på den ökade nyttan av att besökare på webbsajten hittar svar på sina frågor, och den ökade goodwill för myndigheten som detta skapar hos allmänheten.

## Länkar och länklister

Utöver sökmotor hittar folk information på en webbsajt via länkar och länklister. Vad man där bör tänka på är att det inte behöver vara en rent hierarkisk struktur, alltså att man bara kan hitta ett dokument via en väg genom trädstrukturen. Utan ofta behöver samma dokument kunna hittas på flera ställen i olika länklister.

Referenser:

Dalianis, Hercules: Hur do ordnar så att besökarna hittar på din webbplats. Kunskapsbrevet (<http://www.popkom.se/dio.html>) nr 6,2003.

Bales, Marcia: Are there any ways to evaluate how good our Web-site search is for answering questions? CRM Magazine Weekly,  
<http://www.destinationcrm.com/articles/default.asp?ArticleID=3029> (note: For some reason this link does not work with the Mozilla/Netscape web browser).