

# Sökmotorer På Internet – Google

Tomas Franzen, Fredrik Fröjd, Johannes Hassmund,  
Ylva Hecktor, Gabriel Jägenstedt, Christopher Karlsson

Linköpings universitet Ht 2006

## Sammanfattning

Vi har i denna rapport undersökt Google. Hur de blev det storföretag de är idag. Vi har studerat de tre områdena: Företaget, tekniken och missbruk av google. I det första området, företaget, undersökte vi googles policy och anställningsvillkor. På detta sätt har vi bildat oss en uppfattning om hur dessa områden påverkat företagens inriktning och utveckling. I det andra området har vi inriktat oss på teknologin bakom google. Vi har bland annat studerat spindeln som söker upp information och indexerar denna inför sökningar. Vi har studerat hur Pagerank fungerar, när den jämför länkar mellan olika sidor. Vi har undersökt optimeringen som sker för att motorn ska kunna arbeta så snabbt. Vi har också frågat oss om det verkligen är möjligt att ge så snabba svar som google ger med bara hårddisk-cachning? Samt att vi har undersökt det sista området som handlar om några av de sätt som Google kan missbrukas på. Under rapportens gång kom vi fram till att det är svårt att hitta mycket information om hur Google fungerar då detta räknas som affärshemligheter. Mycket av det vi kommit fram till är teorier som grundar sig på patent och rena gissningar. Vi tror inte att Google använder sig av vanliga superservrar utan snarare många små datorer som tillsammans använder sitt minne för att ge snabba svar. Vi vet att Google har flera hundra tusen servrar stående på olika platser så det är inte helt orimligt att anta att det fungerar så.

# Innehåll

<b>1</b>	<b>Inledning</b>	<b>2</b>
1.1	Bakgrund . . . . .	2
1.2	Syfte . . . . .	2
1.3	Metod . . . . .	2
1.4	Diskussion kring källor . . . . .	2
1.5	Typografiska konventioner . . . . .	3
<b>2</b>	<b>Företaget</b>	<b>4</b>
2.1	Historia . . . . .	4
2.2	Policy . . . . .	4
2.3	Annonsering . . . . .	6
2.4	Anställda . . . . .	6
<b>3</b>	<b>Teknik i Googles sökmotor</b>	<b>7</b>
3.1	Optimering . . . . .	7
3.2	Spindeln . . . . .	8
3.2.1	Sökresultat . . . . .	8
3.2.2	Pagerank . . . . .	8
3.3	Exempel på länkning . . . . .	9
3.4	Länkar till andra sidor . . . . .	10
3.5	Ålder och uppdateringsfrekvens . . . . .	10
3.6	Google Toolbar . . . . .	10
<b>4</b>	<b>Missbruk av Google</b>	<b>11</b>
4.1	Spamdexing . . . . .	11
4.2	Spamdexing-tekniker . . . . .	12
4.3	Cloaking . . . . .	12
4.4	Vidarebefordring . . . . .	12
4.5	Länk-spam . . . . .	13
4.6	Google-bomber . . . . .	13
<b>5</b>	<b>Avslutande diskussion</b>	<b>14</b>

# Kapitel 1

## Inledning

Det här är en rapport om Google. Ursprungligen skulle vi undersöka Sökmotorer överlag men bestämde oss för att begränsa urvalet lite.

### 1.1 Bakgrund

Bakgrunden till projektet är att vi tilldelades uppgiften som en del i kursen TDDC10 på Linköpings universitet.

### 1.2 Syfte

Vi har bestämt oss för att fokusera på Google för enkelhets skull eftersom sökmotorer är ett stort område. Google är den mest populära sökmotorn och speglar hela området på ett bra sätt. Vi tänker ta reda på hur Google lyckades bli så dominerande och bra, och dessutom undersöka de mer tekniska bitarna. En intressant frågeställning är hur de lyckas presentera sökresultaten så snabbt och precist.

### 1.3 Metod

Mycket av den information som är intressant för oss hålls hemlig av Google eftersom det är affärshemligheter. Därför är en stor del av all information spekulationer. Informationen är dessutom en färskvara och det gör att det inte finns mycket i böcker som går in på djupet i ämnet.

Vi har hittat det mesta av all fakta på Internet, med hjälp av just Google. För att vara säkrare på att den fakta vi använder oss av stämmer har vi försökt hitta flera källor. Förutom Internet har vi använt oss av en hel del resonemang och personlig erfarenhet i ämnet.

### 1.4 Diskussion kring källor

Många av de uppgifter som cirkulerar om hur Google fungerar är högst spekulativa. Nästan all information om de inre delarna av Google är väl skyddad. De källor som finns är Google själva samt olika webbplatser. Vi hade gärna haft

någon av de böcker som skrivits om Google men misstänker att det troligtvis inte hade gjort någon större skillnad.

## 1.5 Typografiska konventioner

Matematiska formler visas med kursiv stil. Utöver det bör det mesta tala för sig själv.

## Kapitel 2

# Företaget

Under de senaste åren har Google utvecklat sitt företag mycket och utforskar ständigt nya marknader. Vad är det som ger dem sådan framgång? Det går bra för Google, nyligen redovisade de sina resultat under det tredje kvartalet. Siffrorna visar att Google har ökat sin försäljning med 70 %. De fördubblade också nästan sin vinst, en ökning med 85 %.

[Ahlbom, 2006]

### 2.1 Historia

Lawrence Page och Sergey Brin, två studerande på Stanford, Kalifornien påbörjade 1996 arbetet på ett forskningsprojekt som senare kom att bli grunden till Google. Ursprungligen var projektet inte meningen att bli en sökmotor utan skulle bara ranka webbplatser på ett bra sätt. Steget till sökmotor föll sig sedan ganska naturligt. Det första namnet på projektet var Backrub. Ett koncept som byggde på att undersöka länkar mellan webbplatser för att se hur de förhöll sig till varandra. Page och Brin antog att de sidor med flest länkar till sig från mest relevanta sidor borde få hög rank. Den första revisionen av Google låg på Stanfords universitetets webbplats. Domänen [www.google.com](http://www.google.com) registrerades 14 September 1997. Sökmotorn lockade till sig många lojala användare, mycket på grund av sin enkelhet. Den första versionen hade ingen reklam vilket troligtvis många fastnade för. Google har idag ca 50 % av andelen sökningar på den amerikanska marknaden. [Seoblogg, 2006]

### 2.2 Policy

Google själva har slagit fast en policy som publiceras på deras webbplats. Här kommer vi kortfattat att gå igenom de 10 punkter som finns där. Punkterna är fritt översatta till svenska. [Google.com, 2005]

1. Fokusera på användaren så kommer allt annat att följa med.

Teorin är att om man bryr sig om användaren och vad användaren behöver kommer användaren bli nöjd och fortsätta använda produkten och även sprida produktnamnet vidare. Detta är tämligen enkelt att se bl.a. på IRC-kanaler där Google allt som oftast ges som svar på olika frågor.

- Gränssnittet är enkelt och tydligt.
  - Sidor ska laddas på ett ögonblick.
  - Det går inte att köpa en placering bland sökresultat.
  - Reklam på sidan måste vara relevant och inte störa användaren.
2. Det är bäst att göra en sak riktigt bra.

Google hanterar sökning av information och detta ligger alltid i fokus. Genom att ständigt utveckla sökprocesser har Google även kommit in på närliggande områden och har utvecklat nya produkter där expertisen som de fått kunnat utnyttjas.
  3. Snabbt är bättre än sakta.

Google tror på att ge användaren svar omedelbart. Google skulle kunna vara det enda företaget i världen som har som mål att användaren ska kunna lämna deras sida så snabbt som möjligt.
  4. Demokrati på webben fungerar

Eftersom grunden för sökningar med Google är att mäta hur många som länkar till en viss sida fungerar sökning bättre ju fler sidor som ansluts till servicen. Pagerank är en demokratisk process. Ju fler som gillar en sida desto bättre Pagerank kommer sidan få.
  5. Du måste inte vara vid ditt skrivbord för att behöva ett svar.

Världen blir mer och mer mobil. Användare kan behöva få svar oavsett om det är via en PDA eller mobil-telefon. Google tillhandahåller tjänster som gör det enkelt att använda mobil eller dylikt till att surfa sökresultat.
  6. Det går att tjäna pengar utan att vara ond.

Googles intäkter kommer från att erbjuda sin sökteknologi till företag och från försäljning av reklam som visas på söksidan. Google tillåter inte reklam att visas på en sökning om det inte är relevant. Därför är det inte säkert att alla sökresultat kommer visa reklam. Den enda reklam som tillåts är textbaserad reklam. Detta har visat sig ge fler klick än bilder och pop-ups. De som önskar göra reklam kan dra nytta av det.
  7. Det finns alltid mer information.

När Google hade indexerat mer HTML-sidor än någon annan sök-service började de skapa produkter för att söka efter annat såsom bilder och inlägg i Usenetgrupper. Eller göra det lätt att titta på pdf:er.
  8. Kravet på information är gränsöverskridande.

Googles högkvarter må ligga i Amerika men det finns kontor över hela världen. Det är möjligt för användare att specificera vilka språk de vill hitta resultat på.
  9. Du kan vara seriös utan kostym.

Företaget är bara seriöst när det gäller sökning. Det bygger på att arbete ska vara utmanande och att utmaningen ska vara rolig. På kontoret är det de anställda som kommer först. Grupparbete ges en viktig position på

företaget och anställda känner sig stolta över att vara del i en grupp. Även individuella prestationer är givetvis viktiga. Möten som vanligtvis skulle ha tagit flera timmar kan lätt genomföras i lunchkön.

10. Jättebra är inte bra nog.

Leverera alltid mer än förväntat. Google accepterar inte att vara bäst i slutändan utan vill vara det från början. Att aldrig vara nöjda är det som driver företaget framåt och håller dem framför konkurrenterna.

## 2.3 Annonsering

När man annonserar på Google gör man detta via AdWords. En annonsör bestämmer själv sin dagliga budget och hur mycket den kan tänkas betala per klick. Antal visningar av reklam kommer sedan anpassas av Google. På webbplatsen finns det specialiserade verktyg för att beräkna hur mycket vissa sökord och vissa placeringar på reklamsidan skulle kosta. Det är även möjligt att tjäna pengar på sin webbplats genom att tillhandahålla en plats där reklam kan visas på en sida. Detta kallas AdSense och har varit ganska omdebatterat i relation till förekomsten av click-fraud. Click-fraud kallas det fenomen där någon tros förfälska klick på en länk för att kunna tjäna mer pengar. Detta är givetvis inte särskilt omtyckt av de företag som annonserar och därför har Google ett antal gånger anlagats för att inte ta tillräckligt allvarligt på problemet. Resultatet blir hårdare regler från Googles sida och det resulterar i att även de som legitimt får många klick ibland kan stängas av från tjänsten.

[Google.com, 2006]

## 2.4 Anställda

Ännu en anledning till att Google blivit den gigant de är, är troligtvis att de behandlar sina medarbetare på ett vettigt sätt. Google själva anser att det de kallar 20 % tid är det som har gett upphov till de flesta innovativa ideerna och bland annat Gmail är ett resultat av denna policy. 20 % tid innebär att alla medarbetare har rätt att använda 20 % av arbetstiden till att jobba med de projekt som intresserar dem. Google är ett företag som bildades mitt under dot.com-boomen och ett av de ytterst få som lyckades överleva.

## Kapitel 3

# Teknik i Googles sökmotor

I det här kapitlet undersöker vi sökmotorn. Vi tittar på optimeringen och hur sidor värderas.

### 3.1 Optimering

För att du ska få dina sökresultat så snabbt som möjligt har Google olika knep för att optimera allt som behövs för att leverera låga responstider.

Spindeln är den teknik som söker efter texter på internet och skickar tillbaka dem till Google. När spindeln letar upp sidor ger den hela texten till Google Indexer som i sin tur ska indexera allt. För att optimera indexeringen tar indexeraren bort så kallade stoppord. Det är vanliga ord som förekommer i texter som man kan ta bort utan att förstöra sammanhanget i texten, till exempel 'the', 'is', 'or', 'how' och 'why'.

Indexeraren tar även bort dubbla mellanrum, viss interpunktering och konverterar alla bokstäver till gemener för att kunna söka igenom allt mycket snabbare. Något den även gör är att rättstava texterna den tar in för att kunna matcha mer exakt när en användare söker.

Men denna lilla optimering räcker inte för att kunna ge så pass snabba sökresultat som Google ger. Utför vilken sökning som helst på Google och du får vanligtvis ett resultat inom 4/10 sekund. Vanligtvis brukar sökmotorer lagra allt på hårddiskar, och när en sökning ska utföras läser man från dem, eller om man ska indexera något, då skriver man till hårddiskarna.

En vanlig hårddisk har ungefär 3/10 sekunders åtkomsttid. Det vill säga, tiden det tar att läsa från hårddisken och skicka tillbaka resultatet. Då är det 1/10 sekund över till sortering för att få fram de mest relevanta resultaten, Pageranking eller din egna sortering som man kan ange under advanced search. Men hur är det med de sökresultaten som tar mindre än 4/10 sekund? Hur kan man läsa snabbare än vad hårddiskar klarar av? [Sullivan, 2004] Det kan man helt enkelt inte. Det de måste göra är att använda sig av RAM istället för hårddiskar. Google måste ha ettill exempelremt stort kluster med massor av RAM för att kunna hantera alla sökningar och kunna ge så pass snabba resultat som de gör. Detta betyder att hela indexet (i alla fall den viktiga delen) ligger i minnet och inte på hårddiskar.

Genom att använda RAM möjliggör man jämföring per kategori, sortering



per Pagerank och att kunna leverera resultaten extremt snabbt. Det är ungefär 100 olika faktorer som inträffar på ett sökresultat. Även om man skulle göra simultana läsningar och skrivningar till hårddiskar skulle det fortfarande vara slött. Därför bör allt ske i minnet för att kunna utföra allt snabbt. Dock ligger inte allt i minnet, vissa delar cachas till minnet och cachade sidor sparas till hårddisken, men i majoritet utförs största delen i minnet.

[Hunter, 2006]

## 3.2 Spindeln

### 3.2.1 Sökresultat

Det finns många faktorer som avgör vilka sidor som visas först när man gör en Google-sökning på ett visst ord. Naturligtvis spelar de ord man söker på in, men hur vet Google vilka sidor som är mest relevanta för dig som surfar? Varje webbsida som Google indexerat ges en viss ranking, sidor med hög ranking får en placering högt upp bland sökresultaten. Själva rankingen bestäms av flera faktorer. Detaljerna kring hur rankingen beräknas är en väl bevarad hemlighet eftersom Google inte vill att man ska kunna manipulera sökmotorn så att irrelevanta sidor får en hög placering. Nedan följer dock ett antal faktorer bekräftade genom bland annat Googles patent.

[Acharya et al., 2005]

### 3.2.2 Pagerank

Den mest avgörande faktorn för rankingen av en sida är tekniken Pagerank. Pagerank bygger på följande enkla princip: Ju fler som länkar till din sida, desto högre Pagerank ges den. Man kan betrakta varje länk till en viss sida som en röst på att sidan är intressant. Länkar från sidor som själva har en hög Pagerank är mer värda än länkar från sidor med låg Pagerank. Matematiskt kan Pagerank beskrivas enligt nedan.

Vi har en webbsida,  $A$ . Pagerank för denna sida,  $PR(A)$ , beräknas genom följande formel:  $PR(A) = (1 - d) + d(PR(T_1)/C(T_1) + (PR(T_2)/C(T_2) + \dots + (PR(T_n)/C(T_n))))$

$T_1, T_2, \dots, T_n$  är sidor som länkar till sidan.

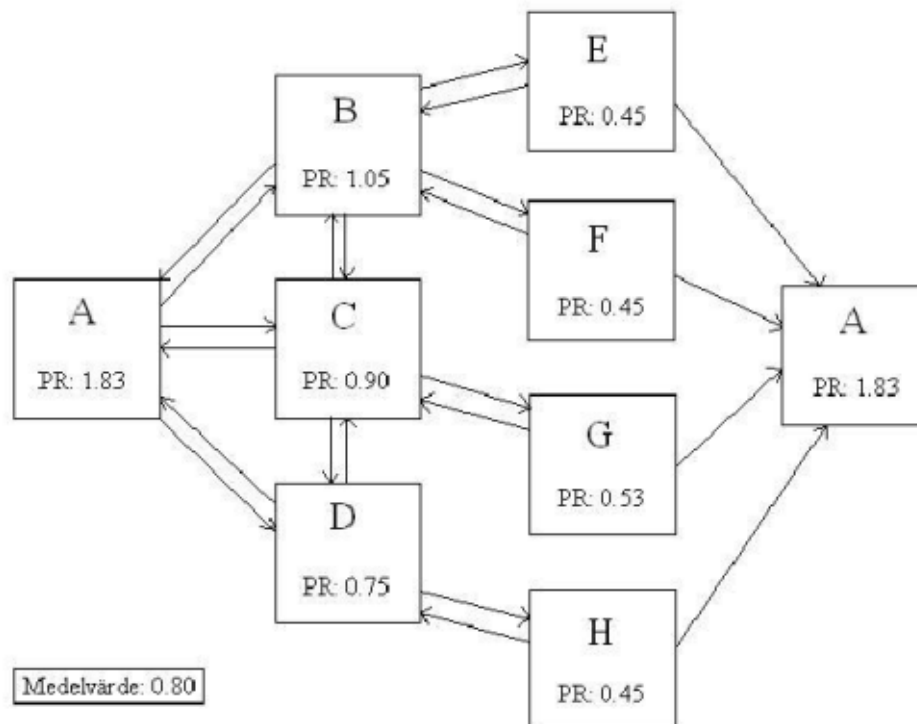
$PR(T_x)$  är Pagerank för sidan.

$C(T_x)$  är antalet utgående länkar från sidan.

$d$  är en dämpningsfaktor,  $0 < d < 1$ .

Vi kan konstatera:

- Ju fler sidor som länkar till sidan  $A$ , desto högre Pagerank får sidan  $A$ .
- Ju högre Pagerank de sidor som länkar till  $A$  har, desto högre Pagerank får sidan  $A$ .
- Ju fler länkar sidor som länkar till  $A$  innehåller, desto lägre Pagerank får sidan  $A$ .



Figur 3.1: Exempel på Pagerank.

[Brin and Page, 1998]

### 3.3 Exempel på länkning

För att visa lite hur Pageranking fungerar så kommer här ett till exempel på hur länkning bland webbplatser påverkar rankingen.

I figur 3.1 så ser vi en webbplats A som länkar till tre undersidor B, C och D som i sin tur länkar vidare till E, F, G och H, varandra och tillbaka till A. De lägsta platserna i hierarkin länkar tillbaka till A och till sin föregående webbplats. För att räkna ut vad Pagerankvärdet är för alla dessa webbplatser så används formeln för Pagerank som kan hittas i 3.2.2. Det första man får göra är att gissa ett Pagerankvärde för alla webbplatser och sedan sätts dem in i formeln som sedan körs ett godtyckligt antal gånger och för varje gång den körs om så ser man att det värde man får för till exempel plats A beräknas igen med de nya värdena från de sidor som den länkar till. Alltså det blir iterativt för att värdet ändras från gång till gång. Till slut får man ett värde som skiljer sig tillräckligt lite från föregående att man kan stanna och få ett godtyckligt Pagerankvärde för varje webbplats.

Värdena kan ses i figur 3.1

[Rogers, 2003], [Webworkshops, 2003]

### 3.4 Länkar till andra sidor

Länkar till andra webbplatser med hög ranking kan höja rankingen på din egen sida. Trasiga länkar ger motsatt effekt. Om man eftersträvar hög ranking är det därför viktigt att hålla koll på att de länkar man har på sin sida fungerar.

### 3.5 Ålder och uppdateringsfrekvens

Att en sida har funnits länge på nätet tyder enligt Google på att sidan är stabil och etablerad och ger en högre ranking. Regelbundna uppdateringar av sidan höjer också rankingen, men att uppdatera siden för ofta kan tyda på att du försöker manipulera din ranking och kan därför bidra till att sänka densamma. Lagom är bäst.

### 3.6 Google Toolbar

Många användare installerar Google Toolbar som ett tillägsprogram till sin webbläsare. Google Toolbar hjälper dig att till exempel markera dina sökord på den sida du besöker. Google Toolbar kan också ge en indikation om Pagerank för den sida du besöker. Det spekuleras i huruvida Google använder data från användandet av Google Toolbar för att förbättra rankingen av de sidor du besöker. Det är till exempel möjligt att Google mäter:

- Antal användare som besöker en viss sida
- Hur ofta sidan besöks
- Om användare tillbringar lång tid på en viss sida
- Om användare lägger till bokmärke till sidan

Dessa uppgifter är inte helt bekräftade men är helt klart i enlighet med det finstilla i Google Toolbars privacy policy:

*“Except for information sent through Toolbar for use with a separate Account-based service such as Gmail, we do not associate any of the information that Toolbar sends with other personal information about you. However, it is possible that a URL or other page information sent to Google may itself contain personal information.”*

[Acharya et al., 2005]

## Kapitel 4

# Missbruk av Google

I takt med att sökmotorer som Google växer och blir allt mer populära ökar också deras makt. Att komma högt upp på resultatlistan är viktigt för att en webbplats ska bli populär och välbesökt. Därför försöker allt fler manipulera sökmotorernas indexering för att fuska sig till bättre placeringar. De algoritmer som Google använder sig av för att prioritera sökträffar kan utnyttjas på många sätt.

På senare år har det också blivit vanligare att starta webbplatser som bara innehåller reklam. Webbplatsernas ägare får betalt för varje gång reklambannern visas, och lurar användare till webbplatsen för att tjäna pengar. Eftersom webbplatserna saknar användbart innehåll länkar ingen till dem, och det blir viktigt att få träffar från sökmotorer.

### 4.1 Spamdexing

En teknik kallas Spamdexing (orden 'spam' och 'indexing' sammansatta). Spamdexing går ut på att få Google att associera sidan med nyckelord som egentligen inte är särskilt relevanta. Nyckelorden indexeras av Google och ger sidan bättre placeringar bland sökresultat, eller placeras under en kategori som sidan egentligen inte hör hemma i.

Tekniken började bli populär i samband med att sökmotorerna blev allt viktigare i mitten av 90-talet. Spamdexing skiljer sig från vanlig sökmotoroptimering på det sätt att uppsåtet är att lura användaren, istället för att använda relevanta nyckelord. [Wikipedia.org, 2006e]

Det finns ett antal sätt att lura indexeringsalgoritmerna att associera en sida med nyckelord. Att på vanligt sätt lägga text som innehåller dessa söktermer direkt i sidans innehåll skulle ta mycket plats och förstöra sidans riktiga innehåll, i de fall där sidan har ett annat innehåll som man vill att användaren ska se. Därför kan det vara fördelaktigt att nyckelorden placeras på andra ställen där de inte syns för webbplatsens besökare, men som Google ändå tar med när sidan indexeras.

## 4.2 Spamdexing-tekniker

Ett mycket enkelt sätt att uppnå detta är att helt enkelt färga texten i samma kulör som sidans bakgrund. Andra liknande tekniker går till exempel ut på att göra texten så liten så den inte syns med blotta ögat. Eftersom Google bara bryr sig om sidornas innehåll och inte deras utseende betraktar indexeringsmotorn gömd text på samma sätt som all annan text.

I HTML finns det ett sätt att ange nyckelord för att hjälpa sökmotorer att hitta relevanta sökord som sidan har att göra med. Under sökmotorernas barndom blev det populärt att missbruka denna funktion för spamdexing. Man fyller helt enkelt på med populära sökord så att sidan dyker upp oftare. Nyare sökmotorer som Google har dock aldrig använt sig av dessa så kallade meta-taggar för sin indexering, av just den anledningen. [Wikipedia.org, 2006c]

En teknik som har blivit populär på senare år är att sätta upp s.k. Doorway pages. Det är en sida som enbart innehåller stora mängder sökord, utan något viktigt innehåll. Användare vidarebefordras sedan till den riktiga webbplatsen automatiskt. Ofta försöker sidor dölja vidarebefordringarna eftersom sökmotorer annars kan känna av dem, och värdera sidorna lägre. [Wikipedia.org, 2006b]

Google försöker skydda sig så väl det går mot missbruk, och anpassar algoritmerna för att upptäcka falska sidor. Oftast ges sidorna bara lägre Pagerank, men tas inte bort från databasen. Eftersom det hela tiden upptäcks nya sätt att lura indexeringen måste Google hela tiden vara snabba att rätta till brister i systemet.

## 4.3 Cloaking

En vanlig teknik för att lura sökmotorer är att använda s.k. cloaking. Det innebär att webbplatsen presenterar en version av sidor till sökmotorn och en annan till användare. Detta är enkelt eftersom spindlar och webbläsare avslöjar sin identitet för varje anrop. Antingen baserar man detta på "User-Agent", ett fält i HTTP-anropens headers som identifierar programmet som hämtar sidorna, eller så kontrollerar man klientens IP-adress. Den sistnämnda tekniken är säkrare, men kräver att man vet spindelns möjliga IP-adresser från början.

Genom att använda Cloaking kan man till exempel fylla en sida med sökord som Google indexerar, så man får bra placeringar bland sökresultaten. När användaren besöker webbplatsen ser denne inte sökorden, utan istället något helt annat. [Wikipedia.org, 2006a]

## 4.4 Vidarebefordring

Ett effektivt sätt att lura Pageranksystemet är att använda vidarebefordring. Om en webbsida vidarebefordrar en förfrågan till en annan sida så ger Google den ursprungliga sidan samma rankning som sidan den pekar till. Detta kan tyckas vara ganska rättvist, eftersom till exempel flera domäner som pekar vidare till samma ställe bör värderas lika högt. Funktionen kan dock missbrukas genom att använda cloaking. Man låter Googles spindel bli videskickad, men låter vanliga användare se den riktiga sidan. Spindeln ger alltså sidan samma rankning som vilken annan sida man vill. [SEO Black Hat, 2005]

## 4.5 Länk-spam

Att länka till sin egen sida ger visserligen aningen högre ranking på Pagerank och andra sökmotorers motsvarande system, men det är mycket mer värdefullt att bli länkad från andras webbplatser. Därför har det utvecklats tekniker för att missbruka även den funktionen.

I början av 2000-talet började bloggar bli ett viktigt sätt att kommunicera, och antalet bloggar ökade drastiskt. I anslutning till bloggar finns ofta system som tillåter läsare att kommentera inläggen. Detta kan missbrukas med hjälp av automatiserade system som lägger ut länkar till webbplatser i bloggarnas kommentarer. Google indexerar bloggarna och dess kommentarer, och spammarnas webbplatser får höga placeringar på grund av alla länkar. [Wikipedia.org, 2006d]

På samma sätt missbrukas andra sorters system, som till exempel wikis, för länk-spam. Allt fler webbplatser kräver numera att man verifierar att man är en riktig människa genom att tolka en bild med bokstäver i.

När man besöker en webbsida innehåller HTTP-förfrågan adressen till den sida man kommer från, d.v.s. den som länkade dit. Ibland publicerar webbplatser statistik över vilka andra webbplatser som länkar mest till dem. Genom att skicka falska anrop med sin egen webbplats i det så kallade Referer-fältet kan man fuska till sig en plats i statistiken, och öka sin Pagerank. [Wired.com, 2002]

## 4.6 Google-bomber

Samma tekniker som används för att fuska till sig bra placeringar på sökmotorer kan också användas för att associera andras webbplatser med vissa nyckelord, mot webbplatsens ägares vilja. Detta kallas Google-bombning. Genom att många andra webbplatser lägger ut länkar som har nyckelordet som text och sidans adress som mål värderar Google sidan högt tillsammans med nyckelordet.

Google-bomber har använts för att till exempel smutskasta politiska motståndare. Mest känd av dessa kupper är den mot George W. Bush som påbörjades 2003. Söker man på frasen “miserable failure” kommer presidentens webbplats på första plats. Det har också förekommit motattacker som resulterar i att före detta presidenten Jimmy Carter och filmregissören Michael Moores respektive webbplatser hamnar högt upp på samma resultatsida. [BBC, 2005]

## Kapitel 5

# Avslutande diskussion

Resultatet av detta projekt är en avhandling om hur sökmotorn Google fungerar. Med tanke på den undanhållna informationen om detta så är det nog lite svårt att riktigt använda detta arbete till något konkret men det ger en insikt om hur en sökmotor fungerar och man skulle kunna säga att man i teorin själv skulle kunna bygga upp och marknadsföra en egen motor i princip som Google. I våran studie har vi bland annat tagit reda på hur Google blivit så stora genom marknadsföring och också hur sökmotorn kan fungera så snabbt och precist. Vi ser möjligheter att använda deras sökmotor inte enbart på webben utan också i interna stora nät och till exempel på sjukhus för att snabbt söka upp patienters sjukjournaler.

# Litteraturförteckning

- Anurag Acharya, Matt Cutts, Jeffrey Dean, Paul Haahr, Monika Henzinger, Urs Hoelzle, Steve Lawrence, Karl Pflieger, Olcan Sercinoglu, and Simon Tong. Information retrieval based on historical data. [*www*] Hämtat 061014, 2005. URL [<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=1&f=G&l=50&co1=AND&d=PG01&s1=20050071741&OS=20050071741&RS=20050071741>].
- Helena Ahlbom. Därför vill alla stora leka med google. *Ny Teknik*, 25 Oktober, 2006.
- BBC. 'miserable failure' links to bush. [*www*] Hämtat 061027, 2005. URL [<http://news.bbc.co.uk/2/hi/americas/3298443.stm>].
- Sergey Brin and Lawrence Page. The anatomy of large-scale hypertextual web search engine. [*www*] Hämtat 061014, 1998. URL [<http://www-db.stanford.edu/~backrub/google.html>].
- Google.com. Marknadsför ditt företag med google. [*www*] Hämtat 061020, 2006. URL [<https://adwords.google.com/select/Login?sourceid=AW0&subid=SE-ET-ADS&hl=sv>].
- Google.com. 10 things google has found to be true. [*www*] Hämtat 061020, 2005. URL [<http://www.google.com/corporate/tenthings.html>].
- Michael Hunter. How do search engines differ from one other? [*www*] Hämtad 061020, 2006. URL [<http://www.searchengines.com/searchDiffer1.html>].
- Ian Rogers. The google pagerank algorithm and how it works. [*www*] Hämtat 061018, 2003. URL [<http://www.iprcom.com/papers/pagerank/>].
- SEO Black Hat. How to fake a google page rank 10... or any pr you want. [*www*] Hämtat 061027, 2005. URL [<http://seoblackhat.com/2005/07/26/how-to-fake-a-google-page-rank-10-or-any-pr-you-want/>].
- Seoblogg. Google stabila, yahoo växer och msn tappar trots nya live search. [*www*] Hämtat 061027, 2006. URL [<http://www.iprospect.se/seoblogg/google/google-vaxer-och-msn-tappar-trots-nya-live-search.html>].
- Rob Sullivan. Why is google so fast? [*www*] Hämtad 061020, 2004. URL [<http://www.searchengineposition.com/info/Articles/GoogleFast.asp>].



- Webworkshops. Pagerank calculator. *[www] Hämtat 061018*, 2003. URL [[http://www.webworkshop.net/pagerank\\_calculator.php?pgs=8](http://www.webworkshop.net/pagerank_calculator.php?pgs=8)].
- Wikipedia.org. Cloaking. *[www] Hämtat 061027*, 2006a. URL [<http://en.wikipedia.org/Cloaking>].
- Wikipedia.org. Doorway page. *[www] Hämtat 061027*, 2006b. URL [[http://en.wikipedia.org/Doorway\\_page](http://en.wikipedia.org/Doorway_page)].
- Wikipedia.org. Meta element. *[www] Hämtat 061027*, 2006c. URL [[http://en.wikipedia.org/Meta\\_tag](http://en.wikipedia.org/Meta_tag)].
- Wikipedia.org. Spam in blogs. *[www] Hämtat 061027*, 2006d. URL [[http://en.wikipedia.org/Spam\\_in\\_blogs](http://en.wikipedia.org/Spam_in_blogs)].
- Wikipedia.org. Spamdexing. *[www] Hämtat 061027*, 2006e. URL [<http://en.wikipedia.org/Spamdexing>].
- Wired.com. When the spam hits the blogs. *[www] Hämtat 061027*, 2002. URL [<http://www.wired.com/news/culture/0,1284,56017,00.html>].