

Googles besöksfrekvens & uppdatering av cache

**Besöker och cachar Google sidor oftare om sidan
uppdateras/förändras oftare?**

av Ulf Liljankoski

<http://www.liljankoski.se>

2007-03-16

INTRODUKTION

Vad går undersökningen ut på?

Denna undersökning utgår från teorin att Google oftare besöker och uppdaterar cachan för sidor som förändras/uppdateras.

För att göra testet har jag använt en redan etablerad sida där jag med hjälp av slumpen försökt simulera en förstasida som uppdateras ofta.

Undersökningen har delats upp i fem faser som sammanlagt sträcker sig över 103 dagar.

Att tänka på när man läser denna rapport är att undersökningen endast gäller Googles besök, indexering och uppdatering av cachan för förstasidan.

Huvudfråga: Besöker och cachar Google sidor oftare om sidan uppdateras/förändras oftare?

Delfråga: Har tidstämpeln i en webbplatskarta någon inverkan på hur ofta Google besöker och cachar förstasida.

Hur fungerar Googles cachefunktion?

För att kontrollera när en sida senast besöktes av Google samt när Googles cache av en sida uppdaterades kan man göra en *cache*:-sökning på Google.

Exempel:

Sök på frasen *cache:www.domännamn.com* för att kontrollera förstasidan.

För att göra en sökning på en specifik undersida kan du göra en sökning på t ex

cache:www.domännamn.com/sida1.htm

Då får du se den cachade versionen av sidan samt det datum och tid då sidan cachades, t ex 23 dec 2006 02:10:42 GMT.

Senaste besök kan även kontrolleras i Googles Verktyg för webbansvariga (inklusive webbplatskartor) (<http://www.google.com/webmasters/sitemaps/>).

Enligt Vanessa Fox på Google, cachas en sida varje gång Google varit på besök, och om sidan om sidan förändrats sedan den tidigare cachan. Dock uppdateras datumet för när sidan senast besöktes varje gång Google varit på besök oavsett om sidan förändrats eller ej (läs mer på <http://googlewebmastercentral.blogspot.com/2006/09/better-details-about-when-googlebot.html>).

För att Google ska bedöma att en sida inte uppdaterats ska servern svara med 304 (Not modified) <http://www.google.com/support/webmasters/bin/answer.py?answer=40203>. Detta är något som framförallt används för att spara bandbredd. I denna undersökning använder jag mig *inte* av 304 (Not modified).

Vad har betydelse för hur ofta Google kommer på besök?

Det pågår diskussioner kring vad det är som får Google att ofta komma på besök till en sida. Inga diskussioner ger några tydliga och säkra svar, men det som diskuteras och som i dessa diskussioner anses vara en viktig faktor för att få Google att komma på besök samt cacha en sida ofta är bl a:

1. Antal inlänkar till en sida.
2. Inkommande länkar med, i Googles ögon, högt förtroende (Trust).
3. PageRank på sidan.
4. Att sidan förändras och uppdateras. I denna punkt ingår även diskussion om huruvida stor en förändring ska vara samt hur ofta man bör uppdatera en sida.

TILLVÄGAGÅNGSSÄTT

Testet utfördes på förstasida på en etablerad sajt (funnits i ca 1 år). Innan testet påbörjades besökte Google förstasidan samt uppdaterades Googles cache av förstasidan med 5-7 dagars mellanrum. Själv uppdaterade jag förstasidan *en* gång i månaden. Ser man på sajten i sin helhet uppdateras undersidor och nya sidor läggs till så gott som varje dag, samt besökte Google diverse undersidor varje dag.

Förstasidan har i starskedet PageRank 0 (hade tidigare PageRank 3, men har ännu inte helt återhämtat sig efter ett domänbyte med 301-redirect) och i slutskedet PageRank 4. Antalet inkommande länkar och de inkommande länkarnas förtroende (Trust) under den tid undersökningen gjorts har inte förändrats nämnvärt.

För att simulera att förstasidan på sajten uppdaterades ofta, implementerade jag ett PHP-script som varje gång sidan besöktes slumpade fram en HTML-sida utifrån *tio* olika förstasidemallar. Detta innebär att varje gång förstasidan besöktes av människor såväl som sökmotorer, slumpades en HTML-sida fram utifrån dessa tio mallar.

Tanken med detta var att varje gång Googles sökrobot kom på besök till förstasidan skulle sökroboten möta en förändrad sida (med en risk på 10 % att det ändå skulle bli samma sida

som tidigare som slumpades fram – dock bedömde jag den risk som i längden irrelevant för Google syn på sidan).

Nämnas bör också att förändringen som gjorts (implementeringen av det PHP-scrip som slumpar fram en ny förstasida) inte gjorts för sökmotorernas skull, utan för besökarnas skull. Tidigare har fyra utvalda produkter/kategorier marknadsförts på förstasidan, men med den nu framslumpade förstasidan marknadsförs olika produkter/kategorier beroende på vilken sida som dyker upp. Förändringen har alltså framförallt en uppgift att förändra upplevelsen av förstasidan för besökaren.

På sajten använder jag sedan tidigare en webbplatskarta (Google sitemaps) som visar en tidstämpel och talar om för Google när sidan senast uppdaterades. Vikten av en uppdaterad tidstämpel för förstasidan testades i samband med denna undersökning.

Faserna

I *Fas 1* fortsatte jag använda webbplatskartan, tillsammans med den framslumpade förstasidan. *Webbplatskartans tidstämpel uppdaterades en gång per dag och förstasidan förändrades vid varje besök.* Både förstasidan och tidstämpeln talade om för Google att sidan var uppdaterad.

Fas 1 sträcker sig från dag 1-15.

I *Fas 2* använde jag webbplatskartan, tillsammans med den framslumpade förstasidan, men denna gång lät jag tidstämpeln i webbplatskartan stå oförändrad. *Webbplatskartans tidstämpel uppdaterades inte, medan förstasidan fortfarande förändrades.*

Förstasidan talade om för Google att sidan förändrats, men webbplatskartan talade om för Google att förstasidan inte förändrats.

Fas 2 sträcker sig från dag 16-30.

I *Fas 3* tog jag helt bort webbplatskartan och använde mig enbart av den slumpade förstasidan.

Det enda som talade om för Google att förstasidan förändrats var sidans ”utseende”.

Förstasidan förändras vid varje besök.

Fas 3 sträcker sig från dag 31-45.

I *Fas 4* tog jag bort den framslumpade förstasidan och ersatte den med en statisk sida.

Ingen webbplatskarta använd.

Google möter vid varje besök en oförändrad sida.

Fas 4 sträcker sig från dag 46-88.

I *Fas 5* återinför jag den framslumpade förstasidan.

Ingen webbplatskarta använd.

Det enda som talade om för Google att förstasidan förändrats var sidans "utseende".

Förstasidan förändras vid varje besök.

Fas 5 sträcker sig från dag 89-103.

Fas 0

Dessa fem faser jämför jag även med vad jag kallar *Fas 0*. *Fas 0* är utgångspunkten och tiden *innan* PHP-scriptet implementerades och testerna påbörjades.

RESULTAT

Fas 0 (Oförändrad förstasida + Webbplatskarta)

Datum: -4 december 2006

Förstasidan genomgår en förändring *en* gång i månaden och tidstämpeln i webbplatskartan uppdateras 1-2 gånger i månaden.

Under *Fas 0* besöker Google och uppdaterar cachen för förstasidan på sajten med 5-7 dagars mellanrum.

Fas 1 (Förändrad förstasida + Förändrad webbplatskarta)

Datum: 3 - 17 december 2006

Förstasidan genomgår en förändring *varje* gång sidan besöks och webbplatskartan uppdateras *en* gång per dag.

Under *Fas 1* besöker Google och uppdaterar cachen för förstasidan med 3 dagars mellanrum.

Tiden mellan Googles besök och uppdatering av cachen av förstasidan har alltså krympt.

Fas 1 säger alltså oss att Google gillar att besöka och cacha den numera till synes väluppdaterade förstasida bättre än den mer sällan uppdaterade förstasidan (*Fas 0*).

I Fas 1 kan jag dock inte avgöra om det är förändringen av förstasidan (den framslumpade sidan) som gett upphov till att Google kommer på besök och cachar sidan oftare, eller om det är den förändrade tidstämpeln som ger upphov till detta.

Fas 2 (Förändrad förstasida + Oförändrad webbplatskarta)

Datum: 18 december 2006 - 1 januari 2007

Förstasidan genomgår en förändring *varje* gång sidan besöks medan webbplatskartan *aldrig* uppdateras. Webbplatskartans tidstämpel står still på 17 december 2006.

Under Fas 2 besöker Google och uppdaterar cachen för förstasidan på sajten med 1-3 dagars mellanrum. Tiden mellan uppdateringen har alltså sjunkit ytterligare efter Fas 1, trots att tidstämpeln i webbplatskartan talar om för Google att sidan inte förändrats sedan 17 december 2006..

Fas 2 säger oss att Google uppskattar den numera till synes väluppdaterade förstasida och att obetydlig eller ingen hänsyn tas till tidstämpeln i webbplatskartan (sidan räknas som uppdaterad trots att webbplatskartan hävdar motsatsen).

I Fas 2 drar jag slutsatsen att Google ser förändringarna på förstasidan som en uppdatering och därmed anser det vara av värde att på nytt cacha och uppdatera sitt index för förstasidan på sajten. Den framslumpade förstasidan verkar alltså vara av värde inte bara för besökarna, utan även för den som vill få sin sida besökt och cachad av Google oftare.

Utifrån Fas 2 kan man även anta att om man tar bort webbplatskartan helt men låter den framslumpade förstasidan vara kvar, bör detta göra att Google antingen fortsätter att besöka och cacha förstasidan i samma takt eller till och med besöker och cachar sidan oftare (om nu tidstämpeln i webbplatskartan gett någon effekt under Fas 2 så är det en negativ effekt). Detta kommer vi att få svar på i Fas 3.

Fas 3 (Förändrad förstasida + Ingen webbplatskarta)

Datum: 1 - 15 januari 2007

Förstasidan genomgår en förändring *varje* gång sidan besöks. Någon webbplatskarta används ej.

Under Fas 3 besöker Google och uppdaterar cachen för förstasidan med 1-3 dagars mellanrum. Tiden mellan uppdateringarna har alltså inte förändrats sedan Fas 2. Jag tolkar detta resultat som att *webbplatskartan och tidstämpeln inte haft någon större inverkan på hur*

ofta Google uppdaterar sin cache av sidan, och att Google uppskattar den numera till synes väluppdaterade förstasida.

Fas 4 (Oförändrad förstasida + Ingen webbplatskarta)

Datum: 16 januari - 28 februari 2007

Den framslumpade förstasidan är nu borttagen och sajten har nu åter en *statisk förstasida*.

Fas 4 - Del 1

Under första delen av Fas 4, 16 januari - 6 februari 2007 fortsätter Google att uppdatera cachen/besöka förstasidan med 1-3 dagars mellanrum. Ingen förändring sedan Fas 3 alltså. För att se om det finns någon form av fördröjning innan Google inser att förstasidan inte förändras längre förlänger jag Fas 4 till 28 februari.

Fas 4 - Del 2

Under andra delen av Fas 4, 7 februari – 28 februari, börjar Google att mindre frekvent besöka/uppdatera cachen för första sidan. Nu besöks sidan av Google med 4-6 dagars mellanrum.

Jag tolkar detta resultat som att *Google numera ser sidan som mindre frekvent uppdaterad, men att det finns en viss fördröjning innan Google bestämmer sig för att besöka sidan mer sällan. Google visar en förståelse för att sidor ibland kan stanna i uppdateringen.*

Fas 5 (Förändrad förstasida + Ingen webbplatskarta)

Datum: 1 mars - 15 mars 2007

Fas 5 liknar Fas 3, förstasidan genomgår en förändring *varje* gång sidan besöks. Någon webbplatskarta används ej.

Under Fas 5 besöker Google och uppdaterar cachen för förstasidan med 1-2 dagars mellanrum och under de sista 7 dagarna av Fas 5 besöks sidan med 1 dags mellanrum. Google besöker åter sidan oftare än vad som gjordes under Fas 4 (oförändrad sida).

Jag tolkar detta resultat som att *Google åter uppskattar den numera till synes väluppdaterade förstasida och därför kommer på besök allt oftare.*

SAMMANSTÄLLNING

Undersökningen säger att:

1. Google besöker och uppdaterar cachen för en sida mer frekvent om sidan uppdateras/förändras.
2. Webbplatskartans tidstämpel har obetydlig eller ingen inverkan på hur ofta Google besöker förstasidan.
3. Det finns en viss fördröjning innan Google slutar besöka en sida ofta när en sida slutar att förändras/uppdateras.
4. Det går att simulera en förändring på en sida för att Google ska besöka och uppdatera cachen av sidan oftare.

Observera att webbplatskartans låga relevansnivå i denna undersökning inte säger att webbplatskartan inte har en positiv eller negativ effekt för att hjälpa Google hitta sidor/undersidor på en sajt. Den låga relevansnivån säger oss enbart att webbplatskartans tidstämpel för senaste uppdatering har föga (eller ingen) inverkan på Googles tolkning om huruvida en sida uppdaterats sedan tidigare besök. Tidstämpeln i webbplatskartan kan också tänkas ha en viktig uppgift att informera Google om att sidor som mycket sällan besöks, har uppdaterats.

DISKUSSION

Det finns alltid faktorer som jag kan ha förbisett. Exempel på faktorer som kunnat inverka är t ex:

- Nya inkommande länkar till sajten med förtroende (Trust) kan ha påverkat utan att de kunnat hittas med olika mätinstrument.
- Liknande undersökningar borde ha gjorts på flera sajter för att kunna utesluta att resultaten uppkommit av en ren slump.
- Tidsrymden för varje fas hade kunnat utökas för att utesluta naturliga variationer i Googles besöksfrekvens.

Ytterligare undersökningar skulle kunna göras gällande:

- Hur stor (t ex procentuellt eller i tecken) en förändring bör vara för att påverka Googles besöksfrekvens.
- Kommer Google med tiden inse att det är samma sidor som återkommer (slumpas fram) och därmed inte besöka sidan lika ofta?

Tankar i slutskedet:

- Att Google på slutet besökte sidan med 1 dags mellanrum innebär det att Google just håller på att kartlägga förändringen?

LÄS MER

Om Googla och Bloggar

<http://www.seo-forum.se/google/1057-om-googla-och-bloggar.html>

Hur ofta går Google igenom min sida.

<http://www.seo-forum.se/google/238-hur-ofta-gar-google-igenom.html>

Uppdateringsfrekvens

<http://www.seo-forum.se/google/261-uppdateringsfrekvens.html>

Crawl dates importance

<http://forums.digitalpoint.com/showthread.php?t=137905>

Better details about when Googlebot last visited a page

<http://googlewebmastercentral.blogspot.com/2006/09/better-details-about-when-googlebot.html>

Crawl dates in the Google cache

<http://www.mattcutts.com/blog/video-crawl-dates-in-the-google-cache/>

Crawl dates in the Google cache (video)

<http://video.google.com/videoplay?docid=3227639395500097587>

Google likes updated content!

<http://forums.digitalpoint.com/showthread.php?t=5021>

What is considered updated content?

<http://forums.digitalpoint.com/showthread.php?t=8096>

How frequent google update cache pages?

<http://forums.digitalpoint.com/showthread.php?t=143226>

304 (Not modified)

<http://www.google.com/support/webmasters/bin/answer.py?answer=40203>