

”Nya” hemliga Google sökmotor en revolution bland sökmotorer där miljoner siter kommer att sorteras bort

Google ännu ”hemliga” nya sökmotor kommer att förändra hela världen där inte minst alla svenska glädjemätare som lurar kunderna ”hemsida ägarna” genom länkfarmar som driver upp besökarantalet som är en ren skär bluff. VisitSweden, Stockholm Visitors Board och andra kända svenska turistrelaterade siter tillhör dom som uppger högre besökaretal än vad som är sant. Jag har otaliga gånger påpekat ”cyber” generalen Thomas Brühl VisitSweden´s luftnummer med branschen där okunniga åhörare faller för osanningarna som presenteras på löpande band.

Google TrustRank generell sammanfattning.

Är nästa generation av algoritmer från Google. Google har ett patent som läckt ut till branschen. Vad står det för?, Hur berör det er som hemsidesägarna?, Hur påverkar detta resultaten. Troligtvis bara positivt eftersom TrustRank är utvecklat för att bekämpa cloaking (fusk) och spam mot sökmotorerna.

Vad är det då som Google kommer att titta på?.

Hemsidors historik:

När startades hemsida. Har den en historik på över två år så betecknas den trovärdig.

Whois data:

Vem äger hemsidan?. Är den registrerad med företagsnamnet så bör det vara trovärdigt.

Eget IP nummer:

Ligger hemsidan på ett eget IP nummer eller delas det med 1000-tals andra. Eget IP är en trovärdighets-faktor.

Adress och telefonnummer:

Finns det adress och telefonnummer på hemsidan så är det troligt att den är trovärdig.

Ligger hemsidan i rätt land:

Ligger den på en .dk domän i Danmark eller i Sverige, Dansk webbhotell minskar förtroendet för en svensk baserat site eller vice versa.

Som ni förstår så skulle allt ovanstående (och lite till) ställa till det ordentligt för merparten av hemsidor idag. Men, om och när Google kommer att implementera detta patent är en annan sak.

Patentet Google TrustRank (Google Patent Analys)!

Vi pratar här om ett helt nytt sätt att värdera hemsidor på, kallat TrustRank. Google har ju sedan länge sitt PageRank system. Detta är helt nytt och vissa delar är redan implementerade.

Referering till patentet som ni hittar här:

<http://appft1.uspto.gov/netacgi/nph-Parser...&RS=20050071741>

Någon enstaka kanske redan har sett patentet och avfärda det som nonsens och enbart till för att förvilla, kanske för att det är svårläst, otroligt komplext och besvärligt att förstå.

Jag tror patentet går ut på 3 saker:

1. Faktorer som Google (G) tycker är viktiga och finns i dagens algoritm redan.
2. Faktorer G tycker är viktiga och kommer att implementeras i algoritmen under dom kommande 3-5 åren.
3. Faktorer G vill sätta ett tidigt patent på för att befästa att ingen konkurrent tar det i beaktande.

Om ni läser igenom patentet kommer ni att se en bred och ofta motsägelsefulla redogörelser vilket får folk att avfärda den som skräp. Men jag tror att dom isåfall missar poängen med patentet. Vad G vill säga är att händelser och egenskaper på optimerade hemsidor efterliknar "riktiga" hemsidor men skiljer sig i skala, syfte och i förhållande till andra faktorer. Som exempel om en hemsida helt plötsligt får 500 nya länkar till sig på en vecka, är det bra eller dåligt?. Svaret är, det beror på.

Om länkarna är för ett genombrott, trendigt, eller hett sökord så är det troligtvis inte dåligt. I annat fall så är det säkert dåligt eller onaturligt. Så om en hemsida får en onaturlig hög mängd med inkommande länkar men det inte finns någon likvärdig ökning av sökningarna för sökordet så känns det naturligt att anta att dom inkommande länkarna kan anses som spam. Ur en algoritmisk synvinkel så är det vad som står i länkarna och relevansen i det (samma överallt) som kan vara en indikation över att det är sökmotorsspam. När man ser på en hemsida eller ett dokument i sin helhet och det finns flera faktorer som indikerar spam, finner dom det troligtvis onaturligt och hemsidan filtreras (sandboxad).

Se det så här att du kör din bil nedför gatan. Det skulle kanske inte uppmärksammas speciellt mycket. Om du kör 10km/timma över hastighetsbegränsningen så är det fortfarande ingen som lyfter på ögonbrynen. Om du dessutom har ett trasigt bakljus så börjar det se misstänksamt ut. Om du dessutom cabbar ner och musiken flödar, din passagerare står upp och skriker och dansar så blir du garanterat stoppad. Förutom din passagerares agerande så var det inget som skulle vara en anledning till att du blev stoppad av sig självt. Ju fler kombinationer av "överträdelser" dessutom större chans att åka dit.

Här är en lista över några av dom faktorer som omnämns I patentet. Det känns som om det finns en normal anledning till att dessa faktorer förändras. Min poäng med att peka ut just dessa är att jag vill lyfta en varningsflagga mot möjligheter för sökmotorsspam. Jag har inkluderat sektionerna från vart jag dragit slutsatserna I patentet.

Domän faktorer:

Domänregistrationens längd.

Är domänen köpt på ett eller 10 år I förskott (sektion 0099)

Domäner övervakas

För förändringar när dom går ut. (sektion 38,39)

Namnserver, och Who is data

Övervakas utifrån förändringar och giltig fysisk adress (samma teknologi som till Google Maps)

Namnserver och troligtvis klass C IP-serier

Bör ha olika whois data, registraturer och sökord resp inga sökord på domäner eftersom många hemsidor delat samma IP nummer i en C-klass serie. (sektion 0101)

Dokument/websiter ges ett datum för upptäckt
När G hittade den genom någon av nedanstående sätt:

- 1) extern länk
 - 2) registrerad
- (sektion 1,2,3,4, 38)

Förändringar av en domäns "teman"

Som inte matchar innehållet som är i fokus och gentemot existerande länkars innehåll övervakas. (sektion 0084)

Websiter måste bestå av fler än ett dokument (sektion 5)

Förändringar av söktermer för en domän övervakas. (sektion 50)

Dokument och sidor

Dokument jämförs för förändringar av:

- 1) frekvens (tidsfaktor)
- 2) mängd av förändringar (sektion 6,7,8, 9, 11, 12)

Antal nya dokument (interna ?) länkade till dokumentet övervakas (sektioner 9,13)

Förändringar av sökordsdensiteten på ett dokument övervakas (sektion 10, 14)

Dokument som stagnerat (inte förändrade) övervakas. (sektion 19)

Hur innehållet på en sida förändras hur texten på externa länkar förändras, registreras. (sektion 31, 33)

Utgående länkar till dåliga, eller spammande hemsidor kan vara en indikation på låg kvalitet. (sektion 0089)

Förändra inte fokus på för många dokument samtidigt. (sektion 0128)

Länkar

En länks (extern till hemsidan) upptäckts datum registreras (sektionerna 54, 55, 56, 57, 58)

Länkar ges en upptäckts och försvinnande datum. (sektion 22,26, 58)

Länkar och länktext övervakas utifrån volym. (sektion 48)

Länkar övervakas utifrån förändringar i länktextern över en given tidsperiod. (sektioner 27, 30, 54, 55, 56, 57, 58)

Länkar viktas utifrån trovärdighet eller auktoritet av det länkande dokumentet, liksom för åldern på länken. (sektion 28, 58, 0074)

Länkars tillväxt oberoende av dokumenten (olika C-klass nätverk, IP-series) övervakas.

I vilken takt nya länkar till ett dokument uppkommer övervakas. (sektionerna 23, 24)

Det fastställs om ett dokument har en trend till ökning eller minskning av inkommande länkar. (sektion 25)

Ett distributions faktorsvärde för länkars ålder övervakas. (sektion 29)

Länkar som har en lång historik har högre värde än länkar med kort livslängd. (sektion 59)

Länkar från O-uppdaterade hemsidor nedvärderas.

Länkar från frekvent uppdaterade hemsidor premieras och ges en boost. (sektion 60)

Nya hemsidor förväntas inte få stora mängder länkar direkt. (sektion 0038)
Länkars tillkommande bör vara konstant och långsamt växande. (sektion 0069, 0077)

En för snabb tillväxt av länkar kan vara indikation på sökmotorspam. (sektion 0077)
Om ett dokument förblir orört länge men kontinuerligt får nya länkar kommer sidan att betraktas som fräsch (sektion 0075)

Om ett dokument förblir orört och utan tillväxt eller minskning av länkar kommer dom utgående länkarna att inte få något värde. (sektion 0080)

En stor mängd länkar kan accepteras om dokumentet har någon eller några länkar från auktoritära hemsidor (sektion 0110)

Sökresultaten

Volymen på sökningar utifrån sökord övervakas över tid utifrån ökningar.(sections 17, 18)
Information beträffande dokumentets placeringar övervakas och noteras utifrån förändringar. (setioner 41, 42, 43)

Click through rates (CTR, antal klick) övervakas utifrån förändringar per sessioner eller andra faktorer (sektion 43, 44)

Antal klick övervakas utifrån ökande eller minskande trender (sektion 51, 52, 53)

Antal klick övervakas för att se om ouppdaterade eller uppdaterade dokument är mer efterfrågade vid sökresultaten. y (sektioner 20, 21)

Antal klick för ett dokumentets sökord övervakas (sektionerna 15, 16, 37, 43)

Användardata

Trafiken till ett dokument övervakas utifrån förändringar (troligtvis via Google Toolbar eller via sökning på datorns cache och historis) (sektion 34, 35)

Användarnas beteende på hemsidor övervakas och förändringar registreras. (klick på bakåt knapp etc)(sektion 36, 37)

Användares beteende övervakas genom bokmärken, cache, favoriter och temporära filer (sektion 46)

Bokmärken och favoriter övervakas utifrån tillägg eller borttagningar. (sektion 0114, 0115)

Användaregenskaper för dokument övervakas utifrån förändringar. (sektion 47)

Tiden användare spenderar på en hemsida kan användas för indikering av en hemsidas kvalitet (sektion 0094)

Övrigt

Dokument som förändras konstant i placeringarna kan antas vara icke trovärdiga. (sektion 0104)

Sökord med liten eller ingen förändring i resultaten bör matcha domänen med stabil placering (sektion 0105, 106, 107)

Sökord med hög livlighet (ombytlighet) bör ha domäner med mer livlighet (sektion 0105, 106, 107)