



Fremtidens Søgemaskine

Denne artikel er IKKE et must for alle. Den er ikke skrevet til masserne og sproget er ikke altid enkelt. Hvis du interessere dig for søgemaskiner og gerne vil kunne optimere dit website nu og i fremtiden, bør du læse den. Artiklen vil kunne give dig en f

Skrevet den **05. Feb 2009** af **bufferzone** | kategorien **Generelt IT / Søgemaskiner** | ★★★★★

Fremtidens Søgemaskine.

Hvordan fremtidens søgemaskine kommer til at virke kan alle gætte på, og ingen har i princippet mere ret end andre. Hvis man læser lidt om emnet på den mere teoretiske del at nette, hvis man kikker lidt på algoritmer og den slags, er der dog nogle tendenser der er begyndt at tegne sig i det (ikke så) fjerne. Det buzz word der tales og skrives en del om i denne tid er **Det Semantiske Web.** og **Latent Semantisk Indeksering.** Disse begreber vil have stor betydning for den måde vi søger på, og også for hvordan vi som web mastere designer vores websites så vi får den nødvendige trafik fra søgemaskinerne.

Fremtiden er her nu. Prøv at søge i google med ~inkjet. Du vil her se at google præsenterer resultater med andre søgeord også, f.eks. epson og HP. Disse to ord er altså semantisk forbundne med inkjet. Prøv også at søge med ~mike ~tyson og se at ordet boksning er semantisk forbundet.

Hvis man skal kunne finde hoved og hale i fremtiden er man nødt til at forstå fortiden og nutiden. Lad os derfor begynde med at kikke på hvordan søgemaskiner virker i dag og hvilke principper der er styrende for de algoritmer der er i anvendelse i øjeblikket.

Når man skal designe et søgesystem i dag er der to grundlæggende forskellige metoder man kan vælge. Enten vælger man tekst indeksering eller META indeksering, eller også kombinere man dem, men det ser vi på til sidst.

Tekst indeksering.

Tekst indeksering virker på den måde at søgemaskinen henter alt sitets tekst ned og behandler det for at kunne lave en liste med relevante indholdsord for hver side. Denne behandling kan selvfølgelig gøres på mange måder, men konsensus "i branchen" er at opskriften er som følger

1. Smid alle kendeord, forholdsord, og bindeord væk.
2. Smid alle Almindelige verber (ved, se, gøre, være, have) væk.
3. Smid alle stedord væk.
4. Smid alle almindelige tillægsord (stor, sen, høj) væk.
5. Smid alle fylde ord (derfor, alligevel, imidlertid) væk.
6. Smid alle ord der forekommer I alle (de fleste) dokumenter væk.
7. Smid alle ord der kun forekommer en gang eller/og kun I et dokument væk.

De ord der så er tilbage er indholdsordene

META Indesering.

META indeksering virker ved at META data indskrives i hvert dokument og at der er META data der søges på. Fordelen er, at søgningerne kan gøres meget hurtige og effektive, hvis META standardiseres og indsættes intelligent. Desværre er dette ikke tilfældet på internettet, hvor skumle web mastere hurtigt fandt ud af at indsætte populære, men ikke for deres site relevante, søgeord i META for at få trafik. Der er således INGEN rene META søgemaskiner tilbage og alle de seriøse søgemaskiner bruger META meget sparsomt.

Den virkelige verden.

Søgemaskinerne i dag bruger faktisk en kombination af META og tekst indeksering. De bruger tekst indeksering til at finde indholdsordene og META++ til at værdisætte de enkelte indholdsord. Når websidens indholdsord er fundet, kigger robotten sidens koder igennem, for at kunne give de enkelte søgeord en værdi, som den senere skal anvende, til at rangere de forskellige websites, når søgemaskinen skal præsentere resultatet på en søgning. Robotterne kigger efter indholdsord i Title taggen, i overskrifter (H1 tag), i title attributter og ALT tag, i Link teksten (den tekst der er den aktive del af et link) i fil og biblioteks navne (URL) og i kommentarer. Derefter laves en analyse af sidetekstens indholdsordtæthed (densitet), frekvens og nærhed (proximitet, dvs. hvor tæt på starten af teksten, den enkelte sætnings og afsnits start indholdsordet står). Dette, sammen med forskellige site parametre, f.eks. Googles PageRank der primært bruger inbound (dvs. links fra andre til dig) bruges som grundlag for beregning. Alle disse parametre samles i databasen og bruges når et søgeord indtastes. Den side der bedst passer algoritmens grænseværdi vil stå som nummer 1 på resultatsiden.

Vær opmærksom på at enkelte søgemaskiner faktisk er begyndt at straffe sider på baggrund af META tag. Hvis du har søgeord i din META tag, der ikke genfindes blandt indholdsordene på siden, vil flere søgemaskiner betragte det som et forsøg på spam fra din side, og som minimum give dine sider dårlig placering eller måske helt slette siden fra sin database.

Problemet.

Hvad er problemet så med nutidens søgemaskiner?. Mangel på intelligens hvis det skal siges kort. Søgemaskinerne kan kun finde sider der har søgeordet (det samme som indholdsordet, men nu set fra søgerens side og ikke søgemaskinen) stående i teksten. Hvis du f.eks. skal bruge informationer om "Fransk impressionisme" vil søgemaskinerne kunne finde de fleste af de sider hvor disse ord står på. Men hvad med sider der handler om Claude Monet, sider om Renoir udstillinger eller sider fra museet i Giverny, eller Salon des Refusés. Hvis du ikke selv ved lidt om fransk impressionisme og kan foretage separate søgninger på disse ord og sætninger finder du aldrig disse sider. Til gengæld finder du måske ikke relevante sider hvor ordene forekommer.

Den ideelle søgemaskine.

Den ideelle søgemaskine findes ikke, og kommer aldrig til det, men vi er nødt til at beskrive den, for at kunne tale om en søgemaskine der er bedre end den vi har i dag.

Igen kan dette stilles op på flere måder, men konsensus "i branchen" gør det således:

Scope: Den ideelle søgemaskine er i stand til at indekserer alle sider på nettet
Speed: Resultatet kommer med det samme
Currency: Alle informationer er altid up-to-date
Recall: Vi er i stand til at finde alle relevante dokumenter. Ingen falske
negativer
Precision: Der er ingen ikke relevante dokumenter i resultater. Ingen falske
positiver
Ranking: De mest relevante resultater præsenteres først, de mindst relevante til
sidst

Fremtidens Søgemaskine.

Mit bud på fremtidens søgemaskine er den Semantiske Søgemaskine, der anvender latent semantisk indeksering som yderligere et element til de elementer der allerede er i brug i dag. Hvad er det så for en størrelse? Latent Semantisk Indeksering er et velbeskrevet matematisk begreb, der anvender ren matematik til at opstille semantiske sammenhænge mellem dokumenter og samlinger af dokumenter, men den matematiske beskrivelse ligger nok noget uden for denne artikels område. Jeg vil derfor forsøge at give en lægmandsbeskrivelse af princippet. Personer der er interesseret i matematikken bag ved, kan følge de opgivne links i slutningen af artiklen.

Den semantiske søgemaskine uddrager indholdsord fra websiderne på samme måde som nutidens søgemaskine, og den vægter disse ord på næsten samme måde, men hvor nutidens søgemaskine stopper behandlingen, fortsætter den semantiske et par trind endnu. Den ser nemlig på samlingen af indholdsord, deres relative vægt og sammenhængen mellem indholdsordenes indbyrdes vægt og finder andre websider og samlinger af websider med samme semantiske profil eller med en semantisk profil der ligger inden for de fastsatte grænseværdier.

Dette lyder måske som en umulig opgave, men de forsøg der allerede er foretaget på forskellige amerikanske universiteter (og her er det værd at huske af f.eks. Google faktisk oprindeligt er et universitets projekt og at Google stadig har meget tætte forbindelser til Stanford University <http://www.google.com/corporate/history.html>) viser at det faktisk lader sig gøre og at søgemaskinerne kan levere relevante dokumenter der ikke indeholder de primære søgeord overhoved.

Fremtiden set fra søgerens perspektiv.

Når søgemaskinen indeholder semantisk intelligens betyder det at søgningerne bør gennemføres semantisk. Dette vil sige at man i stedet for at søge på "Fransk Impresionisme" i stedet bør søge mere præcist, f.eks. "Den franske impressionistiske malerkunst, i særlig grad Claude Monet og Renoir; ikke udstillinger". Om søgningen skal formuleres på denne måde er svært at sige noget afgørende om, hvilke operatører der bliver stillet til rådighed, hvilke indsnævring og wildcards der gives vil fremtiden vise, "enkelt ords" søgningernes dage er nok talte.

Fremtiden set fra web masterens perspektiv.

Den web master der ønsker at have en mærkbar tilstedeværelse i morgendagens semantiske søgemaskiner, bør allerede starte overvejelserne og forberedelserne i dag. Hvor nutidens søgemaskiner er relativt lette at gennemskue og relativt lette at opnå rimelige placeringer, i hvis man ikke er i et meget kompetitivt markedssegment, vil den semantiske søgemaskine være noget mere uoverskuelig. Alene det at søgerne oftere vil bruge sætninger i stedet for enkelt ord, giver jo uendeligt mange flere muligheder. Her er mit bud på hvordan det kan se ud:

Indhold.

Indhold bliver endnu mere vigtigt end det er nu, og måden hvorpå vi skriver indhold vil ændre sig. Hvor det i dag er vigtig at kende sine relevante og relationelle søgeord, og så sørge for at disse forekommer i de rigtige mængder, med den rigtige placering, tæthed og frekvens i teksten, vil fremtidens indhold skulle skrives meget mere varieret. Det vil dreje sig om at have mange forskellige variationer af det indholdsord (jeg skriver med vilje indholdsord her og ikke søgeord. Når søgeren ikke bruger enkelt ord, men sætninger, har det ikke mere den samme mening at se på dette fra søgerens perspektiv, du kan alligevel ikke gætte alle de kombinationer af sætninger der er mulige), det drejes sig om at have synonymer, alternative begreber og variation (men husk at alle indholdsord, skal forekomme to gange i hver tekst, se oven for). Indholds bredden i hele sitet får også større betydning. Hvor nutidens søgemaskine primært kikker på hver side for sig (tillagt site parametre som f.eks. PageRank) vil de semantiske søgemaskiner kikke mere på den semantiske sammenhæng de enkelte indholdssider indgår i på sitet som helhed. Det er derfor vigtigt at have meget større dækning af sitets indholdsrum.

Med andre ord, vil du gøre dig i fremtiden, skal dit indhold skrives om og der skal skrives meget mere.

Interne Links.

Da søgemaskinerne bevæger sig af links og da links binder nettet sammen, er det naturligt at de interne links i et site vil sige en masse om den semantiske sammenhæng i selve sitet. Det er min vurdering af f.eks. Google vil bruge links i deres semantiske indeksering, selvom links strengt taget ikke er en del af den matematik der ligger bag ved Latent Semantisk Indeksering. Som webmaster vil du skulle se på din linkstruktur som en semantisk vejviser, der forstærker det semantiske indhold i dit site. Analyser dine links og lav dem så de understøtter det indholdsrum du ønsker at arbejde med. Prøv dig frem og kik på linkstrukturen når du analyserer dine konkurrenter.

Eksterne Links.

Hvor Google i dag bruger eksterne links til at udregne dine sideres PageRank, tror jeg at eksterne links, BÅDE inbound og ourbound i fremtiden vil få mere betydning. Hvem du linker til, og hvem der linker til dig, siger noget om hvilken semantisk sammenhæng du tydeligst indgår i og dermed hvilken semantisk værdi dine sider skal tillægges. PageRank vis sikkert stadig have betydning for vægtningen af de forskellige indholdsord, men du vil også være nødt til at se på din linksammenhæng som et middel til at definere dit indhold. Outbound tror jeg især vil få betydning her.

Misc.

Der sker et skift i den måde vi som web mastere bør kikke på vores sites på. Hvor vi i dag tager udgangspunkt i de vigtige søgeord, når vi optimere vores sites til søgemaskinerne, vil sammenhængen i fremtiden få mindst lige så stor betydning. Vi vil stadig kunne bruge vores søgeordsanalyse som et udgangspunkt, men vi er nødt til at gå et skridt videre end vi gør i dag, idet vi i fremtiden både er nødt til at se på sammenhængen mellem de enkelte søgeord og på den semantiske sammenhæng for det enkelte søgeord.

Sagt på en anden måde. Hvor vi i dag f.eks. har en liste med 25 søgeord, som vi optimere til enkeltvis, vil vi i fremtiden have 25 sider, hver med et ord som overskrift, og så alle synonyme, akronymer, beslægtede ord, modsætninger, sammenhænge stående neden under. Disse 25 sider vil munde ud i en samling af indhold, hvor ordet **samling** er mindst lige så vigtigt som indhold.

Afsluningt.

Du har nu læst mit bud på fremtiden, hvordan det faktisk går, må tiden vise. Det kan være denne artikel kommer til at udvikle sig meget i tiden frem over eller det kan være den forsvinder helt fordi jeg simpelthen tager helt fejl.

Jeg vil gerne have at du gør følgende. Hvis du har kommentarer, ønsker, hvis der er ting du ikke forstår eller hvis du mener/tror noget andet end mig, så mail mig på kim@bufferzone.dk, husk at de kommentarer du lægger her på artiklen ikke er til mig, men til de brugere der overvejer om denne artikel er de 5 point værd. Anmeld artiklen til de andre læser med kommentaren og send de ting du har til mig pr. e-mali.

Hvis du gerne vil læse noget om dette, kan du søge i google på Semantic Web eller Latent Semantic Indexing. Se disse links

http://javelina.cet.middlebury.edu/lisa/out/lisa_definition.htm

<http://www.w3.org/2001/sw/>

<http://infomesh.net/2001/swintro/>

God informativ artikel

Kommentar af steven d. 14. Apr 2004 | 2

Kommentar af steen_hansen d. 19. Jul 2005 | 3

Thumbs up herfra. Imponerende, og meget interessant læsning :)

Kommentar af eric-pedersen d. 13. Oct 2005 | 4

Kommentar af ellert d. 14. Apr 2004 | 5

Rigtig god artikel! Måske lidt lang, men dette blir man fornuftigt advaret om, inden man gir nogen point! Bestemt noget enhver (seriøs) webdesigner kan bruge..

Kommentar af lenk d. 11. Oct 2004 | 6

Meget spændende artikel på ganske højt niveau

Kommentar af phoenixv (nedlagt brugerprofil) d. 16. Apr 2004 | 7

Virkelig spændende artikel om et vigtigt emne. Læs den ikke så meget for rådene, men snarere for den interessante teori bag fremtidens søgemaskiner.

Kommentar af phunkypopcorn d. 17. Jan 2005 | 8

puha en lang smøre - men bestemt interresant læsning.
Gem den til en dag hvor du ikke har noget at give dig til (som jeg ikke har idag hehe) så går tiden med det og du bliver lidt klogere :)

Kommentar af schaefer d. 17. Apr 2004 | 9

Rigtig god og interessant artikel

Kommentar af kgp43 d. 18. Apr 2004 | 10

Hmmm.
"6. Smid alle ord der forekommer I alle (de fleste) dokumenter væk"
Burde det ikke være:
"6. Smid alle ord der IKKE forekommer I alle (de fleste) dokumenter væk"

Kommentar af al1407 d. 24. Jun 2004 | 11

Rigtig god artikel!

Kommentar af dripz d. 18. Mar 2005 | 12

Meget interessant artikel. Tak!

Kommentar af per-olof d. 19. May 2004 | 13

Meget godt Driller min filosofiske sans

Kommentar af htmkongen d. 15. Apr 2004 | 14

Meget langt! Meget godt! Meget overskueligt! Meget flot! Meget god! Og alle pointne værd !! :)
/Htmkongen

Kommentar af hcichosz d. 25. May 2004 | 15

Kommentar af skarvenneverdies d. 28. Oct 2004 | 16

Lækkerier

Kommentar af wicez (nedlagt brugerprofil) d. 23. Nov 2004 | 17

meget god artikel.. du skriver imidlertid to gange :D

Kommentar af jankramer d. 15. Apr 2004 | 18

Kommentar af learner d. 12. Jan 2005 | 19

Men jeg tror nu at fremtidens søgemaskiner bliver lidt anderledes end han skriver. Jeg tror at fremtidens søgemaskine stiller spørgsmål til personen der søger. Således at hvis du søger efter 'billige mobiltelefoner', så kommer den og spørger dig om du leder efter markedspladser, vil sammenligne priser, leder efter artikler/forums eller bare vil se alle resultaterne på en liste. Måske ikke det bedste eksempel, men jeg tror at søgemaskinerne kommer til at sortere informationerne i grupper på en måde, der gør det lettere at overskue. Altså ikke bare en lang liste som idag... men jeg er bare newbie og vil ikke spille smart her eller noget :o)

Kommentar af sorensbs d. 30. Aug 2004 | 20

kanon du :)

Kommentar af optical d. 19. Jul 2006 | 21

!

Kommentar af kelfe d. 19. Apr 2005 | 22

Rigtig spændende læsning. Sætter en masse tanker igang i ens hoved om hvordan man skal lave sin hjemmeside indholdsmæssigt. Helt klart det hele værd. Læs den

Kommentar af ingenproblem d. 13. Sep 2005 | 23

her er artiklens indhold: Fremtidens Søgmaskine.

Hvordan fremtidens søgemaskine kommer til at virke kan alle gætte på, og ingen har i princippet mere ret end andre. Hvis man læser lidt om emnet på den mere teoretiske del at nette, hvis man kikker lidt på algoritmer og den slags, er der dog nogle tendenser der er begyndt at tegne sig i det (ikke så) fjerne. Det buzz word der tales og skrives en del om i denne tid er Det Semantiske Web. og Latent Semantisk Indeksering. Disse begreber vil have stor betydning for den måde vi søger på, og også for hvordan vi som web mastere designer vores websites så vi får den nødvendige trafik fra søgemaskinerne.

Hvis man skal kunne finde hoved og hale i fremtiden er man nødt til at forstå fortiden og nutiden. Lad os

derfor begynde med at kikke på hvordan søgemaskiner virker i dag og hvilke principper der er styrende for de algoritmer der er i anvendelse i øjeblikket.

Når man skal designe et søgesystem i dag er der to grundlæggende forskellige metoder man kan vælge. Enten vælger man tekst indeksering eller META indeksering, eller også kombinere man dem, men det ser vi på til sidst.

Tekst indeksering.

Tekst indeksering virker på den måde at søgemaskinen henter alt sitets tekst ned og behandler det for at kunne lave en liste med relevante indholdsord for hver side. Denne behandling kan selvfølgelig gøres på mange måder, men konsensus "i branchen" er at opskriften er som følger

1. Smid alle kendeord, forholdsord, og bindeord væk. 2. Smid alle Almindelige verber (ved, se, gøre, være, have) væk. 3. Smid alle stedord væk. 4. Smid alle almindelige tillægsord (stor, sen, høj) væk. 5. Smid alle fylde ord (derfor, alligevel, imidlertid) væk. 6. Smid alle ord der forekommer i alle (de fleste) dokumenter væk. 7. Smid alle ord der kun forekommer en gang eller/og kun i et dokument væk.

De ord der så er tilbage er indholdsordene

META Indesering.

META indeksering virker ved at META data indskrives i hvert dokument og at der er META data der søges på. Fordelen er, at søgningerne kan gøres meget hurtige og effektive, hvis META standardiseres og indsættes intelligent. Desværre er dette ikke tilfældet på internettet, hvor skumle web mastere hurtigt fandt ud af at indsætte populære, men ikke for deres site relevante, søgeord i META for at få trafik. Der er således INGEN rene META søgemaskiner tilbage og alle de seriøse søgemaskiner bruger META meget sparsomt.

Den virkelige verden.

Søgemaskinerne i dag bruger faktisk en kombination af META og tekst indeksering. De bruger tekst indeksering til at finde indholdsordene og META++ til at værdisætte de enkelte indholdsord. Når websidens indholdsord er fundet, kikker robotten sidens koder igennem, for at kunne give de enkelte søgeord en værdi, som den senere skal anvende, til at rangere de forskellige websites, når søgemaskinen skal præsentere resultatet på en søgning. Robotterne kikker efter indholdsord i Title taggen, i overskrifter (H1 tag), i title attributter og ALT tag, i Link teksten (den tekst der er den aktive del af et link) i fil og biblioteks navne (URL) og i kommentarer. Derefter laves en analyse af sidetekstens indholdsordtæthed (densitet), frekvens og nærhed (proximitet, dvs. hvor tæt på starten af teksten, den enkelte sætnings og afsnits start indholdsordet står). Dette, sammen med forskellige site parametre, f.eks. Googles PageRank der primært bruger inbound (dvs. links fra andre til dig) bruges som grundlag for beregning. Alle disse parametre samles i databasen og bruges når et søgeord indtastes. Den side der bedst passer algoritmens grænseværdi vil stå som nummer 1 på resultatsiden.

Vær opmærksom på at enkelte søgemaskiner faktisk er begyndt at straffe sider på baggrund af META tag. Hvis du har søgeord i din META tag, der ikke genfindes blandt indholdsordene på siden, vil flere søgemaskiner betragte det som et forsøg på spam fra din side, og som minimum give dine sider dårlig placering eller måske helt slette siden fra sin database.

Problemet.

Hvad er problemet så med nutidens søgemaskiner?. Mangel på intelligens hvis det skal siges kort. Søgemaskinerne kan kun finde sider der har søgeordet (det samme som indholdsordet, men nu set fra søgerens side og ikke søgemaskinen) stående i teksten. Hvis du f.eks. skal bruge informationer om "Fransk impressionisme" vil søgemaskinerne kunne finde de fleste af de sider hvor disse ord står på. Men hvad

med sider der handler om Claude Monet , sider om Renoir udstillinger eller sider fra museet i Giverny, eller Salon des Refusés. Hvis du ikke selv ved lidt om fransk impressionisme og kan foretage separate søgninger på disse ord og sætninger finder du aldrig disse sider. Til gengæld finder du måske ikke relevante sider hvor ordene forekommer.

Den ideelle søgemaskine.

Den ideelle søgemaskine findes ikke, og kommer aldrig til det, men vi er nødt til at beskrive den, for at kunne tale om en søgemaskine der er bedre end den vi har i dag.

Igen kan dette stilles op på flere måder, men konsensus "i branchen" gør det således:

Scope: Den ideelle søgemaskine er i stand til at indekser alle sider på nettet
Speed: Resultatet kommer med det samme
Currency: Alle informationer er altid up-to-date
Recall: Vi er i stand til at finde alle relevante dokumenter. Ingen falske negativer
Precision: Der er ingen ikke relevante dokumenter i resultater. Ingen falske positive
Ranking: De mest relevante resultater præsenteres først, de mindst relevante til sidst

Fremtidens Søgmaskine.

Mit bud på fremtidens søgemaskine er den Semantiske Søgmaskine, der anvender latent semantisk indeksering som yderligere et element til de elementer der allerede er i brug i dag. Hvad er det så for en størrelse? Latent Semantisk Indeksering er et velbeskrevet matematisk begreb, der anvender ren matematik til at opstille semantiske sammenhænge mellem dokumenter og samlinger af dokumenter, men den matematiske beskrivelse ligger nok noget uden for denne artikels område. Jeg vil derfor forsøge at give en lægmandsbeskrivelse af princippet. Personer der er interesseret i matematikken bag ved, kan følge de opgivne links i slutningen af artiklen.

Den semantiske søgemaskine uddrager indholdsord fra websiderne på samme måde som nutidens søgemaskine, og den vægter disse ord på næsten samme måde, men hvor nutidens søgemaskine stopper behandlingen, fortsætter den semantiske et par trind endnu. Den ser nemlig på samlingen af indholdsord, deres relative vægt og sammenhængen mellem indholdsordenes indbyrdes vægt og finder andre websider og samlinger af websider med samme semantiske profil eller med en semantisk profil der ligger inden for de fastsatte grænseværdier.

Dette lyder måske som en umulig opgave, men de forsøg der allerede er foretaget på forskellige amerikanske universiteter (og her er det værd at huske af f.eks. Google faktisk oprindeligt er et universitets projekt og at Google stadig har meget tætte forbindelser til Stanford University <http://www.google.com/corporate/history.html>) viser at det faktisk lader sig gøre og at søgemaskinerne kan levere relevante dokumenter der ikke indeholder de primære søgeord overhoved.

Fremtiden set fra søgerens perspektiv.

Når søgemaskinen indeholder semantisk intelligens betyder det at søgningerne bør gennemføres semantisk. Dette vil sige at man i stedet for at søge på "Fransk Impresionisme" i stedet bør søge mere præcist, f.eks. "Den franske impressionistiske malerkunst, i særlig grad Claude Monet og Renoir; ikke udstillinger". Om søgningen skal formuleres på denne måde er svært at sige noget afgørende om, hvilke operatører der bliver stillet til rådighed, hvilke indsnævring og wildcards der gives vil fremtiden vise, "enkelt ords" søgningernes dage er nok talte.

Fremtiden set fra web masterens perspektiv.

Den web master der ønsker at have en mærkbar tilstedeværelse i morgendagens semantiske søgemaskiner, bør allerede starte overvejelserne og forberedelserne i dag. Hvor nutidens søgemaskiner er relativt lette at gennemskue og relativt lette at opnå rimelige placeringer, i hvis man ikke er i et meget kompetitivt markedssegment, vil den semantiske søgemaskine være noget mere uoverskuelig. Alene det at søgerne oftere vil bruge sætninger i stedet for enkelt ord, giver jo uendeligt mange flere muligheder.

Her er mit bud på hvordan det kan se ud:

Indhold.

Indhold bliver endnu mere vigtigt end det er nu, og måden hvorpå vi skriver indhold vil ændre sig. Hvor det i dag er vigtigt at kende sine relevante og relevante søgeord, og så sørge for at disse forekommer i de rigtige mængder, med den rigtige placering, tæthed og frekvens i teksten, vil fremtidens indhold skulle skrives meget mere varieret. Det vil dreje sig om at have mange forskellige variationer af det indholdsord (jeg skriver med vilje indholdsord her og ikke søgeord. Når søgeren ikke bruger enkelt ord, men sætninger, har det ikke mere den samme mening at se på dette fra søgerens perspektiv, du kan alligevel ikke gætte alle de kombinationer af sætninger der er mulige), det drejes sig om at have synonymer, alternative begreber og variation (men husk at alle indholdsord, skal forekomme to gange i hver tekst, se oven for). Indholds bredden i hele sitet får også større betydning. Hvor nutidens søgemaskine primært kigger på hver side for sig (tillagt site parametre som f.eks. PageRank) vil de semantiske søgemaskiner kikke mere på den semantiske sammenhæng de enkelte indholdssider indgår i på sitet som helhed. Det er derfor vigtigt at have meget større dækning af sitets indholdsrum.

Med andre ord, vil du gøre dig i fremtiden, skal dit indhold skrives om og der skal skrives meget mere.

Interne Links.

Da søgemaskinerne bevæger sig af links og da links binder nettet sammen, er det naturligt at de interne links i et site vil sige en masse om den semantiske sammenhæng i selve sitet. Det er min vurdering af f.eks. Google vil bruge links i deres semantiske indeksering, selvom links strengt taget ikke er en del af den matematik der ligger bag ved Latent Semantisk Indeksering. Som webmaster vil du skulle se på din linkstruktur som en semantisk vejviser, der forstærker det semantiske indhold i dit site. Analyser dine links og lav dem så de understøtter det indholdsrum du ønsker at arbejde med. Prøv dig frem og kik på linkstrukturen når du analyserer dine konkurrenter.

Eksterne Links.

Hvor Google i dag bruger eksterne links til at udregne dine sider PageRank, tror jeg at eksterne links, BÅDE inbound og outbound i fremtiden vil få mere betydning. Hvem du linker til, og hvem der linker til dig, siger noget om hvilken semantisk sammenhæng du tydeligst indgår i og dermed hvilken semantisk værdi dine sider skal tillægges. PageRank vis sikkert stadig have betydning for vægtningen af de forskellige indholdsord, men du vil også være nødt til at se på din linksammenhæng som et middel til at definere dit indhold. Outbound tror jeg især vil få betydning her.

Misc.

Der sker et skift i den måde vi som web mastere bør kikke på vores sites på. Hvor vi i dag tager udgangspunkt i de vigtige søgeord, når vi optimerer vores sites til søgemaskinerne, vil sammenhængen i fremtiden få mindst lige så stor betydning. Vi vil stadig kunne bruge vores søgeordsanalyse som et udgangspunkt, men vi er nødt til at gå et skridt videre end vi gør i dag, idet vi i fremtiden både er nødt til at se på sammenhængen mellem de enkelte søgeord og på den semantiske sammenhæng for det enkelte søgeord.

Sagt på en anden måde. Hvor vi i dag f.eks. har en liste med 25 søgeord, som vi optimerer til enkeltvis, vil vi i fremtiden have 25 sider, hver med et ord som overskrift, og så alle synonymer, akronymer, beslægtede ord, modsætninger, sammenhænge stående neden under. Disse 25 sider vil munde ud i en samling af indhold, hvor ordet samling er mindst lige så vigtigt som indhold.

Afsluningt.

Du har nu læst mit bud på fremtiden, hvordan det faktisk går, må tiden vise. Det kan være denne artikel kommer til at udvikle sig meget i tiden frem over eller det kan være den forsvinder helt fordi jeg

simpelthen tager helt fejl.

Jeg vil gerne have at du gør følgende. Hvis du har kommentarer, ønsker, hvis der er ting du ikke forstår eller hvis du mener/tror noget andet end mig, så mail mig på kim@bufferzone.dk, husk at de kommentarer du lægger her på artiklen ikke er til mig, men til de brugere der overvejer om denne artikel er de 5 point værd. Anmeld artiklen til de andre læser med kommentaren og send de ting du har til mig pr. e-mail.

Hvis du gerne vil læse noget om dette, kan du søge i google på Semantic Web eller Latent Semantic Indexing. Se disse links

http://javelina.cet.middlebury.edu/lisa/out/lisa_definition.htm

<http://www.w3.org/2001/sw/>

<http://infomesh.net/2001/swintro/>

Kommentar af hoppe_34 d. 04. Jan 2007 | 24