

Search Engine Optimization

Profesor coordonator

Conf. Dr. Ing. Stefan Stancescu

Masterand

Florescu Iulian

Master IISC

Cuprins

1. Istoric

a. Relatia cu Google

2. Metode de optimizare

a. Indexare

b. Crawling

c. Cresterea importantei

3. Arhitectura paginii web si factori de success pentru SEO

4. Concluzii

Optimizare pentru motoare de căutare sau **Search Engine Optimization** (SEO), este un proces de perfecționare (favorizare) a vizibilității site-urilor web sau paginilor web în cadrul ordonării rezultatelor căutării în lista făcută de motorul de căutare. SEO a apărut în anul 1990, odată cu apariția primelor site-uri pe Internet, și care reprezintă totalitatea tehnicilor prin care un site web este adus la o formă în care este propulsat mai sus în lista de rezultate date de un motor de căutare pentru diverse cuvinte-cheie. Cu timpul, optimizarea unei pagini web a unui site a devenit un serviciu oferit de unele companii și/sau corporații.

Webmaster-ii și furnizorii de conținut au început să optimizeze situri pentru motoarele de căutare începând de la mijlocul anilor 1990, când primele motoare de căutare catalogau rețeaua Web timpurie.

Inițial tot ce trebuia făcut de webmasteri era să adauge pagina, sau URL, la diferitele motoare de căutare, care trimiteau un "paianjen" (en: "spider") pentru a "se târî" (en: "crawl") prin pagină, extrage legături către alte pagini și să returneze informațiile de pe o pagină pentru a fi indexate. Procesul presupune copierea paginii de către paianjeni și stocarea pe propriile servere, unde un al doilea program, numit "indexator " (en: "indexer"), extrage diferite informații despre pagină, precum cuvintele conținute și locul lor în pagină, precum și orice ponderi pentru cuvinte specifice, dar și legăturile către alte pagini, care sunt adăugate printr-o manieră de programare care le face adecvate pentru târâre la o dată ulterioară. [1]

Proprietarii de situri au început să recunoască valoarea clasificării superioare și a vizibilității, creând o oportunitate pentru practicanții SEO "white hat" (pălărie albă) și "black hat" (pălărie neagră).

Printre "White hat" se numără acele companii care folosesc tehnici considerate "legale" pentru promovarea site-urilor în cadrul motoarelor de căutare. În mare parte, acestea se bazează pe optimizări aduse structurii paginilor conținute în cadrul site-ului pentru a fi corect interpretate și evaluate de algoritmi de atribuire a rangurilor folosite de motoarele de căutare.

Un număr mare de bloguri și de forumuri de discuții au apărut, inclusiv în limba română, care oferă creatorilor de site-uri îndrumări și sfaturi referitoare la cum trebuie construite paginile web pentru a fi optimizate din punct de vedere al motoarelor de căutare.

Cea mai devreme folosire a frazei "search engine optimization" a fost într-un mesaj spam postat pe Usenet pe data de 26 Iunie 1997 [2]

Relatia cu Google

Larry Page si Sergey Brin fondat Google în 1998. Google a atras un grup loial în rândul numărului tot mai mare de utilizatori de Internet, carora le-au placut un design simplu. Atât factorii off-page (cum ar fi PageRank și analiza hyperlink-urilor) au fost considerate cât și factori on-page (cum ar fi frecvența de cuvinte cheie, meta tag-uri, titluri, link-uri și structura site-ului) pentru a permite Google să evite un fel de manipulare văzută în motoarele de căutare, care numai considerau doar factorii on-page în sistemul lor de clasament. Deși PageRank a fost mai dificil de realizat, webmasterii au dezvoltat deja unelte de link building și scheme de a influența motorul de căutare Inktomi, și cu condiția aplicabilă în mod similar cu PageRank jocuri de noroc aceste metode. Multe site-uri axat pe schimbul de, cumpararea, și vânzarea link-uri, de multe ori pe o scară masivă. Unele dintre aceste sisteme, sau ferme link-ul, implicat crearea a mii de site-uri pentru unicul scop de a link-spam. [3]

Până în 2004, motoarele de căutare au încorporat o gamă largă de factori care nu sunt divulgați în algoritmiile lor de clasare pentru a reduce impactul manipularii link-urilor. În iunie 2007, The New York Times, Saul Hansell a declarat: Google clasează site-uri, folosind mai mult de 200 de semnale diferite. Cele mai importante motoare de căutare, Google, Bing și Yahoo, nu dezvăluie algoritmiile pe care le folosesc pentru a clasifica paginile.

În anul 2005, Google a început personalizarea rezultatelor căutării pentru fiecare utilizator. În funcție de istoricul lor de căutările anterioare, Google reține rezultate pentru utilizatorii logați. În 2008, Bruce Clay a spus că "ranking is dead" din cauza căutărilor personalizate. El a apreciat că aceasta ar deveni lipsit de sens, pentru a discuta modul în care un site web clasat, deoarece rangul lui ar fi diferit pentru fiecare utilizator și fiecare căutare.

În 2007, Google a anunțat o campanie împotriva link-uri plătite prin care se transferă PageRank. La 15 iunie 2009, Google a dezvăluit faptul că au luat măsuri pentru a atenua efectele PageRank prin utilizarea atributului no-follow pe link-uri. În decembrie 2009, Google a anunțat că va folosi istoricul căutărilor web a tuturor utilizatorilor săi în scopul de a popula rezultatele de căutare.

În data de 8 iunie 2010 a fost anunțat un nou sistem web indexare numit Google Caffeine. Conceput pentru a permite utilizatorilor să găsească rezultate șitiri, mesaje și alte tipuri de conținut mult mai devreme după publicarea decât înainte, Google Caffeine a fost o schimbare a modului Google actualizat indexul său, în scopul de a face lucrurile apărea mai repede pe Google decât înainte. Potrivit

lui Carrie Grimes, inginerul software care a anunțat Caffeine pentru Google, " Caffeine oferă rezultate cu 50 la sută mai recente pentru căutările pe web decât indexul nostru ..." [4]

În februarie 2011, Google a anunțat actualizarea Panda, care penalizează site-uri cu conținut duplicat de pe alte site-uri și surse. Punct de vedere istoric site-uri au copiat conținut de pe unul de celălalt și au beneficiat în clasamentul motor de căutare prin angajarea în această practică, cu toate acestea Google a implementat un nou sistem care pedepsește site-uri al căror conținut nu este unic.

In 2012 Google Penguin a încercat să penaliza site-uri care au utilizat tehnici de manipulare pentru a îmbunătăți locul lor în clasament pe motorul de căutare, iar actualizarea 2013 Google Hummingbird prezentate o schimbare algoritm conceput pentru a îmbunătăți procesarea limbajului natural Google și înțelegerea semantică a paginilor web.

Metode de optimizare – Indexarea

Cele mai importante motoare de căutare, precum Google, Bing și Yahoo !, utilizează crawlere pentru a găsi pagini pentru rezultatele lor în algoritmi de căutare. Paginile care sunt legate de paginile indexate ale altor motoare de căutare, nu trebuie să fie încrcate, deoarece acestea sunt găsite în mod automat. Două directoare majore, directorul Yahoo și DMOZ, ambele necesită prezentarea manuală și revizuire editorială umană. Google oferă Google Search Console, pentru care un XML Sitemap poate fi creat și prezentat pentru a se asigura că toate paginile sunt găsite. [8]

Crawlerele motorului de căutare se pot uita la o serie de factori diferiți, atunci când fac crawling unui site. Nu orice pagina este indexat de motoarele de căutare. Distanța de pagini din directorul rădăcină al unui site poate fi, de asemenea, un factor dacă pagina este sau nu indexata.

Indexarea se face în două etape: întâi este analizat textul care urmează a fi indexat a afla termenii din interiorul lui, în a doua etapă se construiește indexul ce va fi căutat pentru a genera răspunsul numit index invers

Prevenirea crawling

Pentru a evita un conținut nedorit în indici de căutare, webmasterii pot instrui păianjeni să nu se adune anumite fișiere sau directoare prin intermediul fișierului robots.txt standard, în directorul rădăcină al domeniului. În plus, o

pagină poate fi exclusă în mod explicit din baza de date unui motor de căutare cu ajutorul unui meta tag specific. Atunci când un motor de căutare vizitează un site, fisierul robots.txt localizat în directorul rădăcină este primul fișier citit. Fișierul robots.txt este apoi analizat și va instrui robotul cu privire la care paginile respective, care nu vor fi luate de crawler. Crawler-ul motorului de căutare poate să păstreze o copie cache a acestui fișier. Paginile de obicei care nu se doresc a fi adunate includ pagini cum ar fi cărucioarele de cumpărături și conținutul specific utilizatorului, detalii de securitate din paginile de login, cum ar fi rezultatele de căutare din căutările interne. În martie 2007, Google a avertizat webmasteri că acestea ar trebui să împiedice indexarea rezultatelor căutării interne, deoarece aceste pagini sunt considerate ‘spam.

Cresterea importantei

O varietate de metode pot crește proeminența unei pagini web în cadrul rezultatelor căutării. Cross-link-urile între paginile ale aceluiași site ofera mai multe legături către pagini și pot îmbunătăți vizibilitatea în cadrul cautarilor. Includerea în conținutul paginii a unui cuvânt cheie sau a unei expresii des cautate sporește vizibilitatea. Actualizarea conținutului, astfel încât să mențină motoarele de căutare crawling înapoi frecvent poate da o greutate suplimentară la un site. Adăugarea cuvinte cheie relevante pentru meta-date la o pagină Web, inclusiv tag-ul titlu și descrierea meta, vor avea tendința de a îmbunătăți relevanța listării de căutare pe un site, sporind astfel de traficul. [8]

Tehnici white hat vs tehnici black hat

Tehnicile SEO pot fi clasificate în două mari categorii: tehnici pe care motoarele de căutare recomanda ca parte a unui design bun, și acele tehnici de care motoarele de căutare nu sunt de acord. Motoarele de căutare încearcă să minimizeze efectul acestora, printre care spamdexing. Comentatori din industrie au clasificat aceste metode, și practicanții care îi angajează, fie ca White-Hat SEO, sau Black-Hat SEO. [5] White-Hat tind să producă rezultate care sa dureze o lungă perioadă de timp, în timp ce Black-Hat anticipeaza că site-urile lor pot fi în cele din urmă interzis temporar sau permanent o dată ce motoarele de căutare descopera ceea ce fac.

O tehnica de SEO este considerată white hat, dacă aceasta este conformă cu liniile directoare motoarele de căutare și nu implică înșelăciune. Pentru ca

regilile motoarelor de căutare, nu sunt scrise ca o serie de reguli sau porunci, aceasta este o distincție importantă de reținut. White-Hat SEO nu este doar despre următoarele reguli, dar se referă la asigurarea că un conținut al unui site va fi indexat de motorul de căutare, în general rezumată ca crearea de conținut pentru utilizatori, nu pentru motoarele de căutare, și apoi făcând acel conținut ușor accesibil pentru păianjeni, nu mai degrabă decât a încerca să trucheze algoritmul de la scopul propus. White-Hat SEO este în multe privințe similar cu dezvoltarea web care promovează accesibilitatea, deși cele două nu sunt identice.

Black-Hat SEO încearcă îmbunătățirea ranking-ului prin moduri care sunt respinse de către motoarele de căutare, sau implică înșelăciune. Black-Hat utilizează text care este ascuns, fie ca text de culoare asemănătoare cu fundalul, într-un div invizibil, sau poziționat în afara ecranului. O altă metodă arată o pagină diferită în funcție fata de pagina solicitata de către un vizitator uman sau un motor de căutare, o tehnică cunoscută sub numele de camuflare (cloaking).

O altă categorie folosită uneori este de Gray-Hat SEO. Acest lucru este o abordare între Black-Hat SEO și White-Hat SEO, caz în care metodele utilizate evita site-ul să fie penalizat cu toate acestea nu acționează pentru producerea de cel mai bun conținut pentru utilizatori, mai degrabă s-au concentrat în întregime pe îmbunătățirea ranking-ului motorului de căutare.[7]

Motoarele de căutare pot penaliza site-uri pe care le descoperă folosind metode Black-Hat, fie prin reducerea locului lor în clasament sau eliminarea cataloagele lor sau din bazele lor de date cu totul. Astfel de sancțiuni pot fi aplicate fie în mod automat prin algoritmi de motoarele de căutare sau printr-o examinare manuală a site-ului. Un exemplu a fost îndepărtarea în februarie 2006 Google atât a BMW Germania și Ricoh Germania pentru utilizarea practicilor înșelătoare. Ambele companii, cu toate acestea, a cerut scuze rapid, reparand paginile supărătoare, și au fost restaurate la lista de Google. [8]

Arhitectura paginii web si factori de succes

Crawbilitatea site-ului

Motoarele de căutare "adună" website-uri, mergând de la o pagină la alta incredibil de rapid. Ele fac copii ale paginilor care sunt stocate în ceea ce se numește un "indice", care este ca o carte masivă a web-ului.

Când un utilizator caută, motorul de căutare rasfoiește prin această carte mare, găsește toate paginile relevante și apoi alege ceea ce crede el ca este cel mai

bun pentru a arăta mai întâi. Pentru a fi găsit, va trebui să fie în carte. Pentru a fi în carte, va trebui să fie “adunat”.

Cele mai multe site-uri, în general, nu au probleme de crawling, dar sunt lucruri care pot cauza probleme. De exemplu, JavaScript sau Flash poate ascunde potențiale legături; ceea ce face ca paginile acestor link-uri să conducă ascundera de motoarele de căutare.

Fiecare site este administrat un crawl budget, o cantitate de aproximativ de timp sau pagini pe care un motor de căutare îl va “aduna” în fiecare zi, în funcție de încrederea relativă și autoritatea unui site. Site-uri mai mari pot încerca să îmbunătățească eficiența lor crawl pentru a asigura faptul că paginile importante sunt adunate mai des. Utilizarea robots.txt, structura internă a link-urilor și anunțând special motoarele de căutare să nu adune pagini cu parametri URL special definiți pot îmbunătăți eficiența crawl. În plus, este bine să se folosească sitemap, HTML și XML, pentru a facilita motoarelor de căutare accesul cu crawlere la site. [6]

Dublarea / Canonizarea

Uneori, în acea carte mare, indexul de căutare, devine “dezordonat”. Rasfoind prin ea, un motor de căutare ar putea găsi pagină după pagină după pagină a ceea ce arată ca, practic, același conținut, ceea ce face mai dificil să dau seama care dintre aceste multe pagini ar trebui să se întoarcă pentru o anumită căutare. Acest lucru nu este bun.

Acest lucru devine chiar mai rău dacă oamenii leagă în mod activ la diferite versiuni ale aceleiași pagini. Aceste legături, un indicator de încredere și autoritate, sunt dintr-o dată împărțită între aceste versiuni. Rezultatul este o percepție distorsionată a utilizatorului asupra adevăratei care au fost alocate paginii. De aceea, canonizarea este atât de importantă: dacă doriți doar o versiune a unei pagini să fie disponibil pentru motoarele de căutare. [6]

Mobile Friendly

Mai multe căutări Google au loc pe dispozitive mobile decât pe desktop. Având în vedere acest lucru, nu este de mirare ca Google recompensează site-urile care sunt pentru mobil cu o șansă de o mai bună în clasamentele din căutările pe mobil în timp ce cele care nu sunt, au un punctaj mai mic în cadrul căutărilor. Bing, de asemenea, face același lucru. [6]

Viteza Site-ului

Google vrea să facă pe web un loc mai rapid și a declarat că site-urile rapide a obține un mic avantaj asupra site-urilor mai lente.

Cu toate acestea, ceea ce face site-ul dvs. vezicule rapid nu este o plimbare garantat în mod expres la partea de sus a rezultatelor căutării. Viteza este un factor minor care are impact doar 1 din 100 de interogări în funcție de Google, dar angrenare (și conversia) de pe un site poate da un timp de încărcare rapid.

Sunt URL-uri descriptive?

Avand cuvintele pe care doriți să fie găsite pentru domeniu sau adresele URL ale paginilor pot ajuta perspectivele de rang. Nu este un factor major, dar are sens sa fie cuvinte descriptive în adresele URL.

HTTPS / Secure Site

Google ar dori să vadă întregul web ca functioneaza pe servere HTTPS, în scopul de a oferi o mai bună securitate a web-ului. Pentru a ajuta la realizarea acestui lucru, se recompensează site-urile care folosesc HTTPS, cu un mic impuls in clasament.

Ca și în cazul crește viteza site-ului, acesta este doar unul dintre numeroșii factori pe care Google ii utilizează atunci când decide dacă o pagină web are sau nu un rang bun. Nu garantează în sine intrarea în rezultatele de top. Dar dacă te gândești care rulează un site securizat, oricum, atunci acest lucru ar putea ajuta contribui la succesul dvs. de căutare generală. [6]

Concluzii

In aceasta lucrare am prezentat ideile de baza ale SEO si cum poate fi imbunatatita structura unei pagini web pentru a fi pe placul unui motor de cautare. Se deosebesc doua tipuri de abordari in optimizarea paginilor web, cea orientate spre utilizator, in care continutul este optimizat pentru a conduce a un ranking mai bun si cea in care se

apleaza la tehnici nerecomandate pentru rezulte pe termen scurt, dar care duce de cele mai multe ori la scoaterea paginii web din baza de date a motorului de cautare.

Bibliografie

- [1] https://ro.wikipedia.org/wiki/Optimizare_pentru_motoare_de_c%C4%83utare accesat la 08.02.2016
- [2] *Beel, Jöran and Gipp, Bela and Wilde, Erik (2010). "Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar and Co." (PDF). Journal of Scholarly Publishing. pp. 176–190.*
- [3] Brin, Sergey and Page, Larry (1998). "The Anatomy of a Large-Scale Hypertextual Web Search Engine". Proceedings of the seventh international conference on World Wide Web. pp. 107–117.
- [4] "Our new search index: Caffeine". Google: Official Blog. Retrieved May 10, 2014.
- [5] Andrew Goodman. "Search Engine Showdown: Black hats vs. White hats at SES". SearchEngineWatch.
- [6] <http://searchengineland.com/guide/seo/site-architecture-search-engine-ranking> accesat la 08.02.2016
- [7] *Brian Pinkerton. "Finding What People Want: Experiences with the WebCrawler" (PDF). The Second International WWW Conference Chicago, USA, October 17–20, 1994.*
- [8] https://en.wikipedia.org/wiki/Search_engine_optimization accesat la 08.02.2016