

MOTOARE DE CĂUTARE A INFORMAȚIEI

Pavel GHERCIU^{1*}, Vlaicu MAȚARIN¹, Mihail GHIMP¹,
Dorian MUNTEAN¹, Daniela ISTRATI²

¹Universitatea Tehnică a Moldovei, Facultatea Calculatoare, Informatică și Microelectronică, Departamentul Informatică și Ingineria Sistemelor, Grupa IA-201, Chișinău, Republica Moldova

²Universitatea Tehnică a Moldovei, Facultatea Calculatoare, Informatică și Microelectronică, Doctorand specializat în 05.04 „Modelare, matematică, metode matematice, produse de program”

*Autorul corespondent: Gherciu Pavel, email: pavel.gherciu@iis.utm.md

Rezumat: În acest articol va fi prezentată importanța motoarelor de căutare, precum și modul în care acestea funcționează. Motoarele de căutare sunt sisteme software care sunt concepute pentru a efectua căutări pe web într-un mod sistematic pentru anumite informații specificate într-o interogare de căutare textuală. Prin intermediul roboșilor de căutare utilizatorului îi sunt oferite o serie de răspunsuri, care i-ar putea fi utile în dependentă de ceea ce a căutat acesta pe net.

Cuvinte cheie: Motor de căutare, internet, web crawler, SEO, PageRank, Google, algoritmi, marketing.

Introducere

Un motor de căutare este un program de rețea care colectează și organizează conținut pe baza informațiilor pe care utilizatorii le caută. De exemplu, o persoană care folosește un motor de căutare pentru a căuta „restaurante în apropierea mea” va vedea o listă de site-uri web pentru restaurantul în care se află acea persoană. Deși motoarele de căutare sunt programe extrem de complexe, nu este dificil să înțelegem conceptele de bază care le conduc. Motoarele de căutare sunt piatra de temelie a internetului. Pentru mulți oameni, motoarele de căutare sunt punctul lor de plecare ori de câte ori deschid un browser web. Un motor de căutare este practic un program de computer bazat pe Internet. Programul are trei funcții principale: colectează o cantitate mare de informații despre conținutul de pe Internet, clasifică informațiile și ajută utilizatorii să caute informațiile clasificate [1].

Cum funcționează un motor de căutare

Fiecare motor de căutare are trei elemente principale: În primul rând, motoarele de căutare colectează informații printr-un proces numit „accesarea cu crawlere”. Un „crawler web” proiectat de motoarele de căutare verifică metodic adresele URL. Aceste motoare de căutare primesc informații de la sute de miliarde de site-uri web și evaluează diferite aspecte ale site-ului, cum ar fi codul back-end și copia pe care vizitatorul o citește atunci când vizitează pagina. Informațiile sunt indexate pe baza cuvintelor cheie, data când site-ul a fost publicat și a altor factori. După indexarea informațiilor, utilizatorii pot căuta în index tastând în bara de căutare. Aceste căutări declanșează algoritmi complecși care utilizează un număr mare de indici pentru a găsi cele mai relevante rezultate [1].

Popularitatea motoarelor de căutare

Există mai multe tipuri de motoare de căutare contemporane, în funcție atât de piața de consum, cât și de preferințele persoanelor care le folosesc. Prin urmare, sunt distinse câteva motoare de căutare majore, cum ar fi: Google, Yandex, Yahoo, Bing, Shenma, Mail.ru, Haosou și Baidu. Popularitatea lor depinde de statutul economic global. Prin urmare, Google se bucură de cea mai mare vizibilitate din America de Nord, Australia (peste 88,14%) și cea mai mare parte a

Europei (peste 93,25%). Yandex (39,07%) și Mail.ru sunt utilizate pe scară largă în Federația Rusă și sunt competitori direcți ai Google (58,37%). China este singura țară mare Baidu (74,95%) care folosește propriul motor de căutare, care este de fapt o companie cheie pe piața chineză. Utilizarea fiecărui motor:

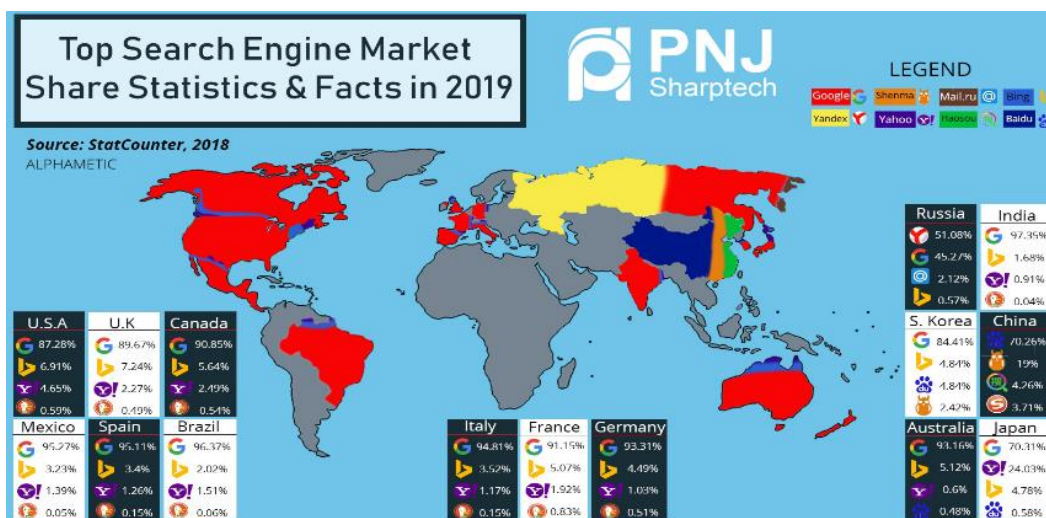


Figura 1. Statistica popularității motoarelor de căutare [2]

La nivel mondial cel mai folosit browser este Google-93.25%, care este urmat de Bing-3.18% și Yandex-1.49%, celelalte utilizând până la 1% de utilizatori de pe piață. De aici deducem faptul că Google practic este un hegemon mondial, a cărui capitalizare pe piață este aproximativ: \$1.088 tril. Și cu un venit de peste \$166 miliarde anual [2].

PageRank

Motorul de căutare creat de Google conține două caracteristici foarte importante care îl fac să fie atât de eficient în comparație cu alte motoare de căutare. Prima caracteristică este numită **PageRank**, aceasta folosește structura legăturilor de pe web pentru a calcula o clasare a calității pentru fiecare pagină web. PageRank poate să fie descris ca un model sau mai bine zis un algoritm al unui utilizator. Page Rank descrie popularitatea a paginii web sau a site-ului web. Acest algoritm Page Rank depinde de link-ul "Analiză" în care clasamentul paginii web este decis pe baza legăturilor de ieșire și a legăturilor de intrare. Asta înseamnă că este complet pe link-ul WWW și Google folosește acest algoritm pentru căutarea în paginile web bazate pe numărul de hyperlinkuri precum **Inbound** (Link-uri de intrare) și **Outbound** (Link-uri de ieșire). Link-urile de intrare sunt acele linkuri care provin de pe alt site către site-ul web, este cunoscut și sub denumirea de „**backlink**”. Google consideră că numai linkurile relevante indică site-ul, dar nu e posibil de controlat ce site-uri indică site-ul. Link-urile de ieșire sunt acele linkuri care trimit către alt site de pe site-ul web [3].

O pagină are un rang înalt dacă celelalte pagini cu un rang înalt sunt conectate la aceasta. Calculul acestui lucru este făcut cu ajutorul:

$$PR(A) = (1-d) + d(PR(T_i)/C(T_i) + \dots + PR(T_n)/C(T_n))$$

Fie A pagina și al cărei rang de pagină este PR(A).

PR(T_i) este Pagerank-ul paginilor T_i care leagă pagina A.

C(T_i) este numărul de legături de ieșire care ies din pagina T_i și d este un factor de amortizare presupus a fi între 0 și 1 de obicei 0,85. Uneori nu faceți clic pe orice legături și sare la alte pagini la întâmplare. Urmează legăturile directe.

(1-d) este probabilitatea să sară la unele pagini aleatorii; fiecare pagină are un minim rangul paginii de (1-d). Urmează legăturile non-directe [4].

Pentru a calcula Page Rank-ul oricărei pagini, trebuie cunoscut Page Rank-ul fiecărei pagini care indică spre acesta și numărul de link-uri de ieșire din fiecare dintre aceste pagini.

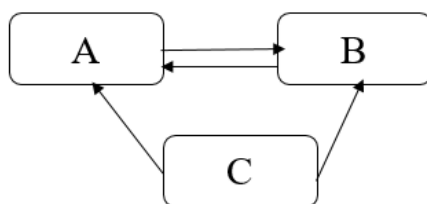


Figura 2. Legătura dintre trei pagini

Un exemplu simplu de trei pagini web A, B și C prezentate în figura 2:

1. Pagina A conține 1 link de ieșire care indică pagina B.
2. Pagina B conține 2 linkuri de ieșire care indică pagina A și pagina C.
3. Și Pagina conține 1 link de ieșire care indică pagina A
4. Clasamentul inițial al fiecărei pagini este considerat a fi 1.

Page Rank-ul fiecărei pagini este calculat prin următoarea ecuație

$$PR(A) = 0,2 + 0,4 PR(B) + 0,8 PR(C)$$

$$PR(B) = 0,2 + 0,8 PR(A)$$

$$PR(C) = 0,2 + 0,4 PR(B)$$

Rezultatul ecuației de mai sus este dat

$$PR(A) = 1,2$$

$$PR(B) = 1,0$$

$$PR(C) = 0,66$$

Evoluție pentru rangul de pagini de 4 pagini cu factor de amortizare diferit. Aici legătura de ieșire este constantă de fiecare dată când se schimbă factorul de amortizare [5].

Search Engine Optimization

Termenul SEO este o abreviere a conceptului „**Search Engine Optimization**”, care se traduce prin „Optimizare motor de căutare”. Procesul de optimizare și promovare a rețelei are ca scop plasarea site-ului web pe prima pagină a rezultatelor afișate de motorul de căutare - acesta este cazul căutării pe baza anumitor cuvinte cheie. Motoarele de căutare precum Google și Bing folosesc roboți pentru a accesa cu crawlere paginile web de la un site la altul, colectează informații despre aceste pagini și le plasează în index. Apoi, algoritmul ia în considerare sute de factori de clasificare sau semnale pentru a analiza paginile din index pentru a determina ce pagini de comandă ar trebui să apară în rezultatele căutării unei anumite interogări.

Factorii de clasificare a căutării pot fi văzuți ca un proxy pentru experiența utilizatorului. Tabelul periodic al factorilor SEO împarte acești factori în șase categorii principale și îi ponderează în funcție de importanța lor generală pentru SEO (Figura 3). De exemplu, calitatea conținutului și cercetarea cuvintelor cheie sunt factori cheie pentru optimizarea conținutului, iar accesarea cu crawlere și compatibilitatea cu dispozitivele mobile sunt factori importanți în arhitectura site-ului web.

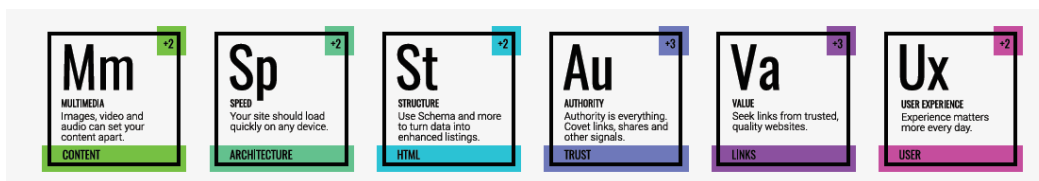


Figura 3. Categoriile și factorii de SEO [7]

Optimizarea site-ului și a conținutului pentru acești factori vă poate ajuta paginile să se claseze mai bine în rezultatele căutării. Spre deosebire de publicitatea cu căutare plătită, nu puteți plăti pentru motoarele de căutare pentru a obține clasamente organice de căutare superioare. SEO este o parte fundamentală a marketingului digital, deoarece oamenii efectuează miliarde de căutări în fiecare an și găsesc adesea informații despre produse și servicii în scopuri comerciale. Obținerea unei vizibilități și a unui clasament mai mare în rezultatele căutării decât concurenții poate avea un impact semnificativ asupra propriilor rezultate. Cu toate acestea, în ultimii ani, rezultatele căutării au continuat să evolueze, oferind utilizatorilor răspunsuri și informații mai directe, care sunt mai susceptibile să țină utilizatorii pe pagina de rezultate în loc să îi conducă către alte site-uri web [6].

Concluzie

În concluzie poate fi afirmat că motoarele de căutare au schimbat complet lumea și oferă mai mult acces la informații decât oricând, ceea ce oferă nenumărate posibilități. Motoarele moderne de căutare știu să clasifice conținutul cerut de algoritmi complecși, fără a cere utilizatorilor să depună prea mult efort. Dezvoltarea tehnologiei oferă modalități și soluții umanizate pentru diferite nevoi, iar cererea de informații este plasată într-o poziție mai înaltă. Utilizatorii au nevoie de multe informații, conținut de înaltă calitate și ușor accesibil, motoarele de căutare au devenit cea mai bună soluție.

Referințe:

1. Google. "How Search Organizes Information". [online]. [accesat 10.11.2020]. Disponibil: <https://www.google.com/search/howsearchworks/crawling-indexing/>
2. Statcounter GlobalStats. Search Engine Market Share Worldwide. [online]. [accesat 12.11.2020]. Disponibil: <https://gs.statcounter.com/search-engine-market-share>
3. Parveen Rani, Er. Sukhpreet Singh: *An Offline SEO (Search Engine Optimization) Based Algorithm to Calculate Web Page Rank According to Different Parameters*, INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY Vol 9, No 1, July 15, 2013.
4. Tamanna Bhatia, "Link Analysis Algorithms For Web Mining", IJCST Vol. 2, Issue 2, June 2011
5. R.Cooley, B.Mobasher and J.Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web". In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), 1997.
6. Mary Cushman. Search engine optimization: What is it and why should we care? [online]. [accesat 10.11.2020]. Disponibil: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6055554/>