

Căutarea și regăsirea informației . Tehnici de căutare

GOOGLE

- 1995: *Sergey Brin* și *Larry Page*
- Propria tehnologie patentată
- PageRank* → cele mai importante rezultate sunt afișate primele
- Instrumente puse la dispoziție: <http://www.google.com/options/>
- Listă de termeni → pagini incluzând toți termenii
- "Lista de termeni": paginile care conțin termenii respectivi consecutiv, ca și cum ar alcătui o expresie.
- Cuvinte ignorate în căutare (cuvinte stop): *http*, *.com*, cifre sau litere singulare. Dacă îi precedăm de caracterul + → vor fi căutați
- Google nu caută după familia de cuvinte, și nici nu suportă căutare parțială, ci caută exact cuvântul introdus
- Google nu face diferență între literele mici și majuscule (*HTML* ↔ *html*)
- Căutarile Google sunt insensibile și la accente și caractere diacritice (*Iași* → *Iasi*, *Muenchen* → *München*)

Tehnici de căutare pe Web

- Internetul → una dintre multiplele surse de informații disponibile
- Preconizăm să găsim informația dorită pe siturile unor:
 - organizații (biblioteci, universități, colegii, școli, instituții guvernamentale, companii, magazine, muzee etc.),
 - publicații (cărți, ziare, reviste, rapoarte tehnice),
 - persoane (profesori, experți, colegi, prieteni, familie)
- În ce format am dori să obținem această informație (multimedia, audio, video, text).
- Clarificați-vă foarte bine ceea ce doriți să căutați și rămâneți focalizat asupra acestui scop
- Gândiți-vă la posibilele cuvinte cheie
- Când ajungeți la situri de interes, utilizați facilitățile de navigare și instrumentele de căutare (indexul sau harta sitului, documentele de FAQ - Frequently Asked Questions, paginile de ajutor).
- Dacă o adresă Web nu funcționează corect sau produce mesaje de eroare, încercați să eliminați o parte de la sfârșitul acestei adrese

Instrumente de căutare:

- Motoare de căutare* - folositoare atunci când doriți să găsiți o informație exactă
- Cataloage Web* - colecții de adrese Web grupate tematic, întreținute de o echipă de specialiști. Utile atunci când doriți să parcurgeți situri populare sau de specialitate: *Yahoo!*, *DMOZ Open Directory*, *Lycos*, *Kappa* etc.
- Porți către subiecte (Gateways)* - sunt site-uri informative într-un anumit domeniu, cu un grad mare de control al calității. Grupează tematic adrese Web adăugate, descrise, indexate și clasificate de către experți. Folositoare când aveți nevoie de informații obținute din surse de încredere. Ex: *The Resource Discovery Network* (RDN - <http://www.rdn.ac.uk/>), sau Education Virtual Library menținut de Charles Stuart University din Australia: www.csu.edu.au/education/library.html
- Portaluri specializate* – precum:
 - Portalul destinat profesorilor și elevilor (<http://portal.edu.ro/>)
 - portalul mobilității cercetătorilor din Europa (<http://europa.eu.int/eracareers/>)
- Blog-uri și Wiki-uri* – site-uri tematice specializate în publicarea de noutăți, având un pronunțat caracter colaborativ

Rafinarea cautării

Sfaturi Generale

Din moment ce Google întoarce pagini ce contin **toate** cuvintele din cererea dumneavoastră, restrângerea sau rearanjarea cautării este la fel de simplă precum adăugarea mai multor cuvinte în termenii cautării pe care i-ati adăugat deja. Prin adăugarea mai multor cuvinte, rezultatele vor conține un subset specific al paginilor întoarse de către cererea originală "prea largă".

Excluderea cuvintelor

Puteti exclude un cuvânt din cautare prin scrierea semnului minus ("-") imediat inaintea termenului pe care vreti sa il evitati. (Fiti sigur sa includeti un spatiu inainte de semnul minus.)

Cautarea frazelor

Cu Google poti cauta fraze prin adăugarea ghilimelelor. Cuvintele incluse între ghilimele ("asa") vor apărea împreună în toate documentele întoarse (doar dacă nu sunt cuvinte de sfârșit, cuvinte comune speciale ce necesită un semn "+" -- [vezi aici pentru detalii](#)). Cautarea frazelor folosind ghilimelele sunt folositoare cautarilor de fraze celebre sau anumite nume.

Anumite caractere au rol de conectare a frazelor. Google recunoaste cratimele, barele, punctele, egalurile și apostroafele ca elemente de conectare a frazelor.

Restrangere domeniu

Unele cuvinte, când sunt urmate de două puncte, au o anumită semnificație pentru Google. Un asemenea cuvânt pentru google este operatorul "site:". Pentru a cauta un anume domeniu sau sit, folosiți sintaxa "site:domeniuexemplu.com" în caseta de cautare Google.

Cautare pe Categori

Directorul Google (care se afla la directory.google.com) te ajuta sa redefinesti cautarea ta bazata pe o idee particulara. De exemplu, cand cauti "Saturn" care se afla in domeniul stiintei > Categoria Astronomie din directorul Google iti dezvaluie pagini numai despre planeta Saturn. Nu iti da pagini despre masina Saturn, jocul Saturn, si multi alti termeni

Cautand o categorie te poate ajuta sa descoperi paginile cele mai bune pentru tine

Cautare Avansata Deseori, pot adauga mai multe cuvinte in spatiul gol, care te poate ajuta pana cand gasesti ce iti trebuie. Dar Google iti poate oferi o varietate de cautari care te lasa:

- Restrictioneaza cautarea ta la paginile care ti-au fost date
- exclude paginile dintr-un site particular
- limiteaza cautarea dvs. numai la pagini intr-o anumita limba
- gaseste toate paginile care au un link la o pagina data
- gaseste pagini care sunt legate de o pagina data

Motorul avansat de cautare Google face mai usoara aplicarea acestor atribute.

Motoarele de cautare (search engines) sunt realizate cu ajutorul unor baze de date de volum mare, în care sunt stocate informații referitoare la paginile web.

Pentru a căuta o anumită informație într-un motor de căutare, trebuie formulată o întrebare alcătuită din cuvinte sau expresii având sensul cât mai apropiat de ceea ce se dorește să se afle (cuvinte cheie).

În urma interogării, motorul de căutare examinează baza lui de date și returnează o listă cu un număr mare de legături către site-uri web care conțin cuvântul cheie sau expresia folosită.

Din lipsa de spațiu această listă este însă fragmentată, astfel încât nu vor apărea mai mult de 10 sau 20 de legături pe fiecare pagină, așezate în ordinea relevanței, în funcție de un algoritm specific fiecărui motor de căutare. Restul legăturilor apar pe paginile următoare.

Orice motor de căutare are 3 elemente majore :

* Primul este **robotul**, care mai este numit un păianjen sau vierme. Robotul este un program care vizitează paginile web, le citește și apoi urmărește legăturile către alte pagini. Multe dintre paginile vizitate își pot modifica după un anumit timp conținutul, se pot adăuga noi legături sau unele pagini pot dispărea definitiv. De aceea roboții se întorc după un anumit interval de timp și vizitează din nou site-ul, căutând eventualele schimbări care apar. Aceste programe au o capacitate extraordinară de a citi și de a prelucra date.

Astfel, conform declarațiilor producătorilor de la *Altavista*, unul dintre cele mai folosite motoare de căutare, ele au capacitatea de a citi 1 GB de text pe secundă, respectiv 3 milioane de pagini pe zi.

* Toate informațiile preluate de robot, cuvinte cheie sau porțiuni de text din paginile vizitate sunt trecute în cel de-al doilea element al motorului de căutare, numit **index** sau **catalog**. Acesta este similar unei baze de date imense în care se află informații despre fiecare pagină web găsită de robot. Această pagină apare în catalog ca o nouă înregistrare, care va conține titlul și adresa paginii respective, cuvinte cheie folosite, legăturile dintre alte pagini, precum și porțiuni de text, care pot fi însă diferite de la un motor de căutare la altul.

Dacă robotul găsește schimbări într-o pagină web, atunci indexul este actualizat cu noile informații. Deci, o pagină web poate să fie vizitată de robot, dar este posibil ca ea să nu fie încă « indexată ». În acest caz pagina nu este disponibilă la o eventuală căutare.

* Al treilea element esențial este **software-ul de căutare**, un program care permite căutarea printre milioanele de pagini înregistrate în index precum și afișarea rezultatului sub forma unei liste de legături, în ordinea pe care el o consideră cea mai relevantă.

Tabelul următor prezintă câteva dintre cele mai cunoscute motoare de căutare :

Motoare de cautare internationale	Motoare de cautare romanesti
www.google.com	www.google.ro
www.yahoo.com	www.start.ro
www.lycos.com	www.cauta.ro
www.GoTo.com	www.click.ro
www.altavista.com	www.index.ro

APLICATIE

Sa se realizeze un document pe o tema la alegere din urmatoarele: masini, fotbal, flori, anotimpuri cu imagini de pe internet