

DOCUMENTARE ONLINE

I. CĂUTARE INFORMAȚIEI PE INTERNET - NOȚIUNI INTRODUCATIVE

Căutarea informației pe Web, în contextul actual al evoluției Internetului (miliarde de pagini existente în cadrul a milioane de situri) nu este o sarcină tocmai ușoară, din cauza faptului că aproape de fiecare dată serviciul de căutare ales ne returnează mii sau zeci de mii de legături, cele mai multe cu informație sterilă (mai corect spus nerelevantă din punctul nostru de vedere).

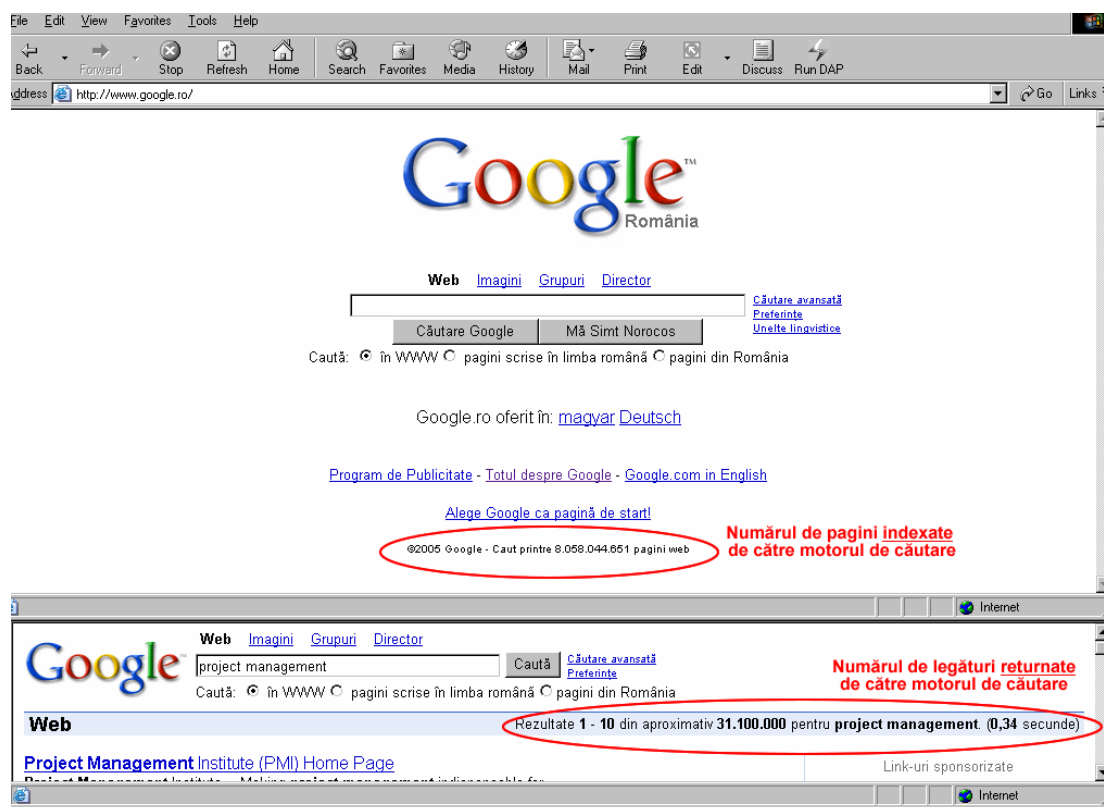


Fig. 1. Problematika căutării informației pe Web

Acest material pleacă de la premiza că ați experimentat deja o situație de acest gen și încercați să vă ajutați trasându-vă niște jaloane pentru a vă putea optimiza căutarea de informație pe Web.

După parcurgerea acestui capitol ar trebui să fiți familiarizați cu următoarele aspecte:

- cum indexează serviciile de căutare informația publicată pe Web
- diferențele existente între diferitele tehnici de căutare, respectiv diferențele dintre diferitele „baze de date Web”
- cum să selectați termenii cheie potriviți pentru căutare
- cum să folosiți diferitele „unelte” puse la dispoziție de diversele motoare de căutare, combinându-le în mod optim

Ce este o „bază de date Web” (Web database) ?

O „bază de date Web” este o listă organizată de pagini web. O astfel de listă poate fi imaginată ca un gigantic repertor ce conține un “surogat” al fiecărei pagini înregistrate în listă (practic părți mici din acea pagină, cum ar fi titlul, antetul etc).

Crearea acestor liste cu surrogate ale paginilor poartă numele de **indexare** și fiecare bază de date web o realizează în stilul său caracteristic. În general, bazele de date web conțin astfel de astfel de “surrogate” pentru un număr imens de pagini web (de la câteva milioane la câteva miliarde de pagini).

Pentru utilizatorul final, baza de date web furnizează o interfață ce are ca și caracteristică fie un “câmp” special în care utilizatorul tastează cuvintele după care va efectua căutarea (vezi interfața cu utilizatorul de la Google), fie o listă cu “directoare” din care utilizatorul poate alege legătura dorită (vezi Yahoo directories). Trebuie menționat însă faptul că (aproape) fiecare bază de date web folosește metode de indexare diferite, respectiv interfețe de căutare diferite.

Metode de indexare

Principial, există două modalități mai importante de indexare a informației în timpul generării bazelor de date web: *indexarea full-text*, respectiv *indexarea “manuală” (human indexing)*.

Indexarea full-text

Așa cum îi spune și numele, indexarea full-text se caracterizează prin includerea tuturor cuvintelor dintr-o pagină în baza de date pentru căutare, cu ajutorul unor programe speciale de calculator numite “păianjeni sau roboți” (spiders sau robots). Altavista și Google folosesc pe scară largă această tehnică de indexare.

Indexarea full-text ne permite astfel să regăsim toate referirile la un anumit termen din documentul indexat. Pe de altă parte însă, căutarea după termeni generici în baze de date web ce folosesc indexarea de tip full-text va duce la returnarea a numeroase pagini cu informație „sterilă”, cu alte cuvinte nerelevantă pentru autorul căutării.

Indexare “manuală”

Spre deosebire de tehnica de indexare de mai sus, în cazul indexării manuale (human indexing) o persoană examinează paaginile ce urmează a fi indexate și decide asupra câtorva cuvinte (fraze) cheie ce descriu cel mai bine informația conținută în respectiva pagină. Aceasta permite utilizatorului să regăsească mai multe legături utile în urma căutării, tocmai pentru că un om și o mașină a ales cuvintele cheie ce au fost incluse în indexul bazei de date.

În mod obișnuit această tehnică de indexare este folosită cu precădere în cazul serviciilor de directoare de pe web, ce vor fi tratate mai târziu în lucrare (vezi Yahoo directories sau Magellan).

Păianjeni, roboți sau oameni ?

Cum selectează bazele de date web ce pagini vor fi indexate ?

Deoarece, așa cum știm, nu există un Sistem Central de Comandă al Internetului, nu există nici un loc anume unde aceste servicii de căutare să „învețe” despre paginile web nou apărute. Din această cauză foarte multe servicii de căutare folosesc programe automate, numite "spiders" or "robots" (păianjeni sau roboți), ce călătoresc de la site la site (= a se citi colecții de pagini web) căutând pagini WWW nou apărute sau modificate substanțial.

Unii roboți caută pagini noi numai în secțiunile sitului intitulate "What's New" sau "What's Hot" și le folosesc numai pe acestea pentru a indexa respectivele situri, în timp ce alți roboți examinează metodic fiecare legătură ce duc la o anumită pagină, respectiv ce pleacă de la acea pagină. De la caz la caz, în funcție și de serviciul de căutare, paginile furnizate de acești roboți pot fi excluse manual de la indexare dacă nu îndeplinesc anumite criterii considerate importante de deținătorii serviciului de căutare.

Ca urmare a utilizării acestor tehnici sunt create în principal trei clase de baze de date web:

- *baze de date ce monitorizează TOATE categoriile de pagini WWW*
- *baze de date ce monitorizează NUMAI paginile WWW considerate populare (în principiu cele cu număr mare de vizitatori)*
- *baze de date ce monitorizează NUMAI paginile WWW ce îndeplinesc anumite criterii (legate fie de calitatea informației furnizate, fie de tipul de informație urmărit – ex. medicală, științifică, știri etc)*

II. UNELTE PENTRU CĂUTAREA INFORMAȚIEI PE INTERNET

Există o multitudine de unelte disponibile pentru căutarea informației pe WEB: *search engines* (motoare de căutare), *subject directories* (servicii de directoare) / *virtual libraries* (biblioteci virtuale), *invisible (deep) web databases* (baze de date web invizibile), *metasearch engines* (motoare de meta-căutare), etc.

Dintre acestea, cel mai mare număr de utilizatori le întrunesc primele două unelte: motoarele de căutare, respectiv serviciile de directoare.

i. Motoare de căutare

Un **motor de căutare** este, așa cum am văzut, o bază de date conținând pagini Web ce pot fi regăsite pe baza unor cuvinte cheie și care continuă să scaneze Internetul, cu ajutorul unor programe automate (spiders, robots) în căutare de pagini noi. Informația rezultată în urma activității roboților este apoi indexată și stocată în baza de date.

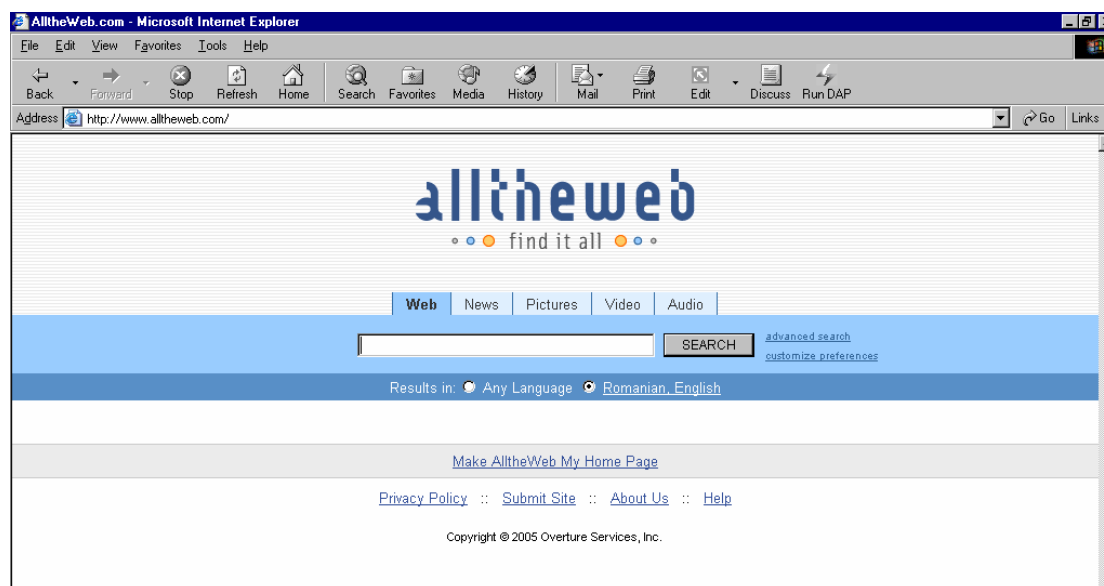


Fig. 2. Motorul de căutare AlltheWeb

Motoare de căutare renumite:

Google (www.google.com)
AlltheWeb (www.alltheweb.com)
MSN (www.msn.com)
Teoma (www.teoma.com)
AltaVista (www.altavista.com)
WiseNut (www.wisenut.com)

ii. Directoare web (anuale online, repertoare tematice)

Un **serviciu de directoare web** (numit uneori și subject directory - director de subiecte) este o colecție de pagini Web selecționate și organizate ierarhic în categorii de subiecte de către un editor uman. Mai mult, un concept și mai selectiv este cel de bibliotecă virtuală, care este un director web ce include legături spre pagini cu informație de înaltă specializare, pe domenii alese de editor (bibliotecarul virtual).

Este evident faptul că serviciile de directoare acoperă și indexează o porțiune mult mai mică din paginile WEB existente, comparativ cu motoarele de căutare. Dar folosirea lor poate duce la regăsirea unor rezultate ale căutării mult mai relevante pentru utilizator.

Cele mai extinse servicii de directoare web indexează cel mult câteva milioane de pagini, comparativ cu cele câteva miliarde indexate de către motoarele de căutare mai importante. Serviciile de directoare NU interoghează direct paginile WEB, ci caută mai degrabă în interiorul bazei lor de date. Din acest motiv, rezultatul căutării poate duce la returnarea unor rezultate expirate, care uneori nu mai au relevanță, pagina originală suferind între timp schimbări majore de conținut sau chiar putând să dispară.

Trebuie menționat însă că o serie de motoare de căutare sunt de fapt unele hibride, fiind în același timp atât motoare de căutare cât și servicii de directoare (Google™, de exemplu, unul dintre cele mai cunoscute motoare de căutare, are și un serviciu de directoare, bazat pe soluția Open Directory Project).

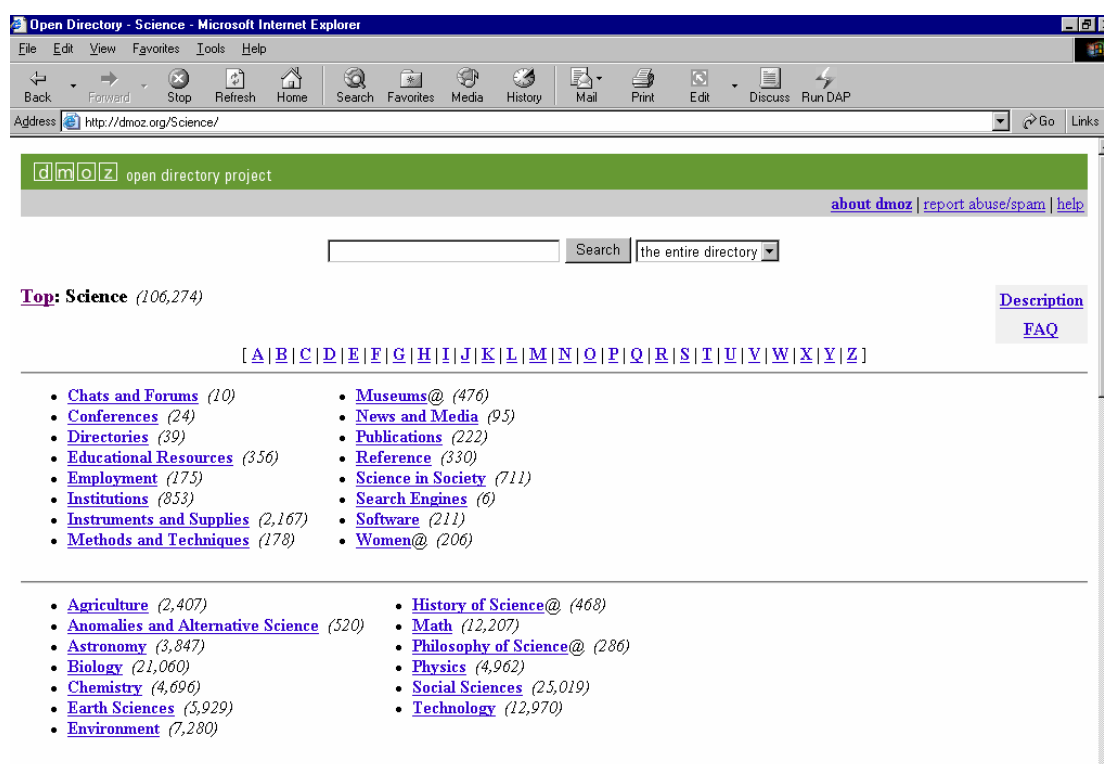


Fig. 3. Serviciul de directoare Open Directory Project

Dintre **serviciile de directoare web**, mai cunoscute sunt cele de mai jos:

Yahoo! (www.yahoo.com)

Google™ Directory (www.google.com)

Open Directory Project (ODP) (http://dmoz.org)

Zeal (www.zeal.com)

JoeAnt (www.joeant.com)

Gimpsy (www.gimpsy.com)

MOTOARELE DE CĂUTARE VERSUS SERVICII DE DIRECTOARE

Până acum am văzut cum indexează bazele de date web paginile. În următoarele rânduri vom vedea cum accesează utilizatorul aceste pagini.

Cele mai uzuale modalități de accesare a acestor pagini se regăsesc în așa numitele motoare de căutare, respectiv în așa numitele directoare web (numite și anuare în cazul anumitor traduceri).

Modalităților de căutare a informației -prezentare comparativă

Motoarele de căutare permit utilizatorului să folosească orice termeni de căutare (combinație de termeni) dorește, motorul căutând acești termeni în baza de date și returnând paginile care conțin aceste combinații de termeni.

Spre deosebire de motoarele de căutare, *serviciile de directoare web* au paginile organizate în baza de date pe subiecte (o bună comparație ar fi cea cu structura anuarelor de tip Pagini Aurii). Paginile cu informația dorită pot fi selectate navigând prin acest arbore de directoare. Navigarea prin structura directoarelor web poate fi un foarte bun start pentru căutarea informației dorite, cu condiția ca topicul ales să fie ales ca subiect pentru o arborescență de directoare (de exemplu vei găsi cu ușurință un arbore de directoare dedicat sectorului educational în orice serviciu de directoare web, însă nu vei găsi prea ușor un arbore de directoare dedicat să zicem consumului de droguri)

ATENȚIE – uneori se pot produce confuzii cu privire la modul de căutare folosit de un anumit serviciu. Astfel renumitul site Yahoo include atât un motor de căutare cât și un arbore de directoare web. Diferența față de un motor de căutare obișnuit este dată de faptul că motorul de căutare al Yahoo caută cuvintele cheie doar în interiorul directoarelor web din baza de date a Yahoo.

Din acest motiv motorul de căutare al sitului poate fi folosit cel mult pentru a identifica zona din arborele de directoare ce conține informația dorită.

Prezentarea rezultatelor

În ambele cazuri, odată ce ai introdus termenii de căutare, serverul îi caută în baza de date web indexată și vă returnează lista cu rezultate. Din cauza multitudinii de legături cu informație sterilă (nerelevantă pentru utilizatorul ce a efectuat căutarea), trebuie subliniate două concepte deosebit de importante ce servesc la evaluarea listei cu rezultate ale căutării: *rata de relevanță (relevancy ranking)*, respectiv *rezumatul (abstract)* ce însoțește legătura către pagină.

Rata de relevanță (Relevancy Ranking)

Paginile găsite în urma căutării sunt aproape întotdeauna prezentate în ordinea relevanței lor, în conformitate cu termenii după care s-a efectuat căutarea. Algoritmii după care este cuantificată rata de relevanță diferă de la sit la sit și, în general, nu sunt făcuți publici. Pentru a vă face însă o idee imaginați-vă un exemplu simplu: paginile vor fi considerate cu atât mai relevante (și de vor fi mai sus în lista cu rezultate) cu cât termenul(ii) de căutare se regăsește de mai multe ori în acea pagină.

Pe baza ratei de relevanță computerul (serverul) care face căutarea va acorda fiecărui document ce conține termenii doriți o rată de relevanță și, în final, va lista primele acele pagini considerate a fi cu cea mai mare relevanță. Din acest motiv nu trebuie să ne facem griji dacă o anumită căutare returnează o listă cu 10 milioane de rezultate. Primele 20 - 40 de pagini sunt probabil cele mai relevante din punctul de vedere al termenilor folosiți pentru căutare. Totodată însă putem vedea importanța fantastică a alegerii celor mai potriviți termeni de căutare. Cu cât alegem cei mai potriviți termeni pentru a efectua căutarea, cu atât vom obține rezultate cu o relevanță mai mare.

Rezumatele paginilor (Abstracts)

În cazul celor mai multe servicii de căutare web puteți să observați că aproape fiecare legătură spre o pagină considerată relevantă este însoțită de un scurt rezumat al conținutului, multe servicii recurgând chiar la sublinierea (îngoșarea) termenilor folosiți în respectivul rezumat. Dacă acest scurt rezumat nu ar exista ne-am putea pune întrebarea ce caută legătura respectivă în lista cu rezultate, respectiv am fi obligați să pierdem mai mult timp urmând fiecare legătură pentru a vedea ce fel de informație conține.

Oricum trebuie menționat faptul că, chiar și atunci când există și este afișat, acest rezumat al paginii este mult mai scurt decât conținutul paginii aflat în baza de date web. De aceea de cele mai multe ori nu vom putea aprecia adevărata relevanță a informației conținute în acea pagină decât deschizând-o.

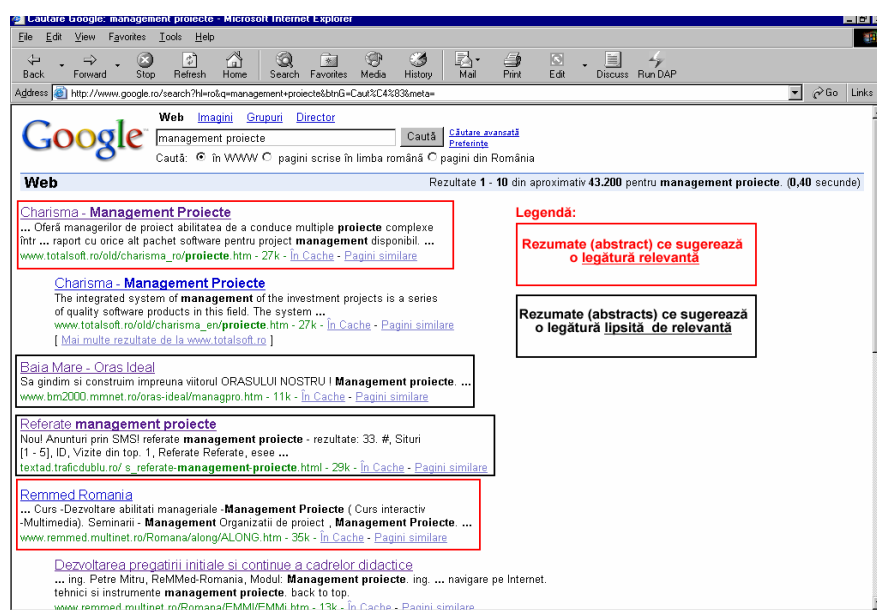


Fig. 4. Importanța rezumatelor (abstracts) pentru evaluarea rezultatelor căutării

ALTE UNELTE DE CĂUTARE A INFORMAȚIEI PE WEB

iii. Biblioteci virtuale:

Biblioteci virtuale de „uz general”:

Librarians' Index to the Internet (www.lii.org)

INFOMINE (www.infomine.com)

Internet Public Library (www.ipl.org)

The WWW Virtual Library (vlib.org)

Academic Info (www.academicinfo.net)

Internet Scout Project (scout.wisc.edu)

BUBL Link (academic resources) (bubl.ac.uk/link/)

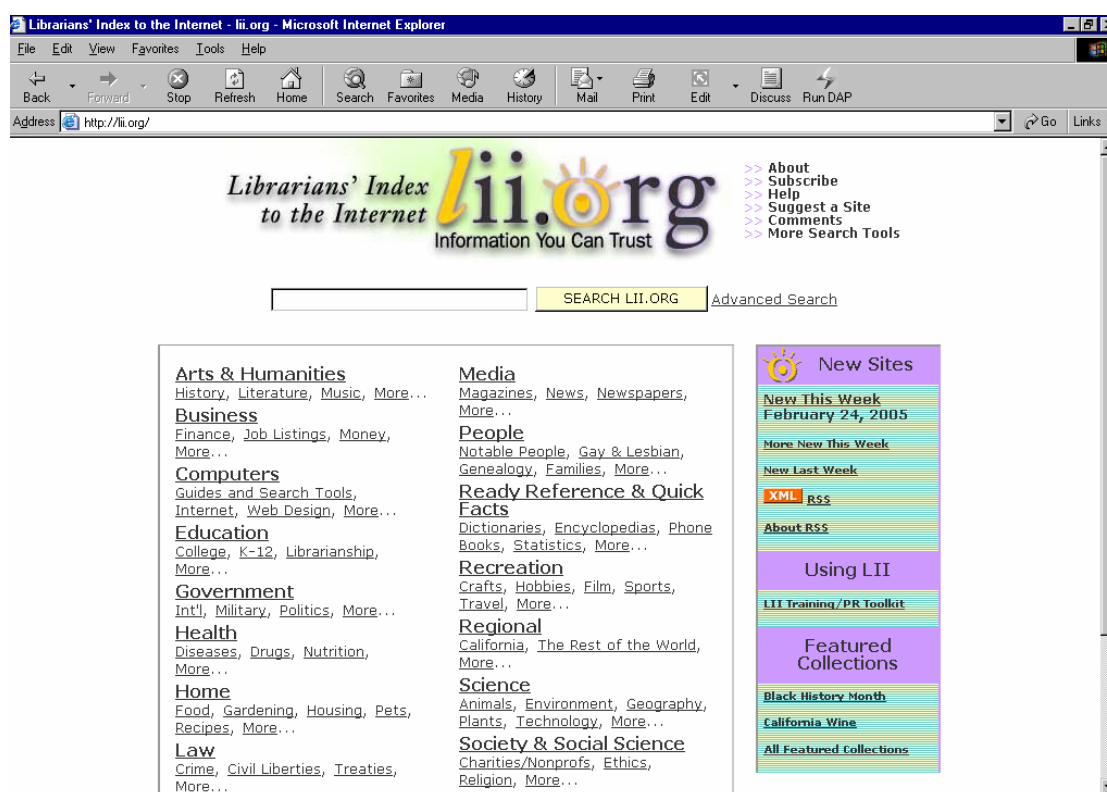


Fig. 5. Librarians' Index to the Internet – bibliotecă virtuală de „uz general”

Biblioteci virtuale de specialitate (orientate spre anumite domenii):

Project Gutenberg (www.gutenberg.org) - beletristică

National Academies Press (www.nap.edu) - știință

Free books for doctors (www.fb4d.com) - medicină

The free management library (www.managementhelp.org) - management

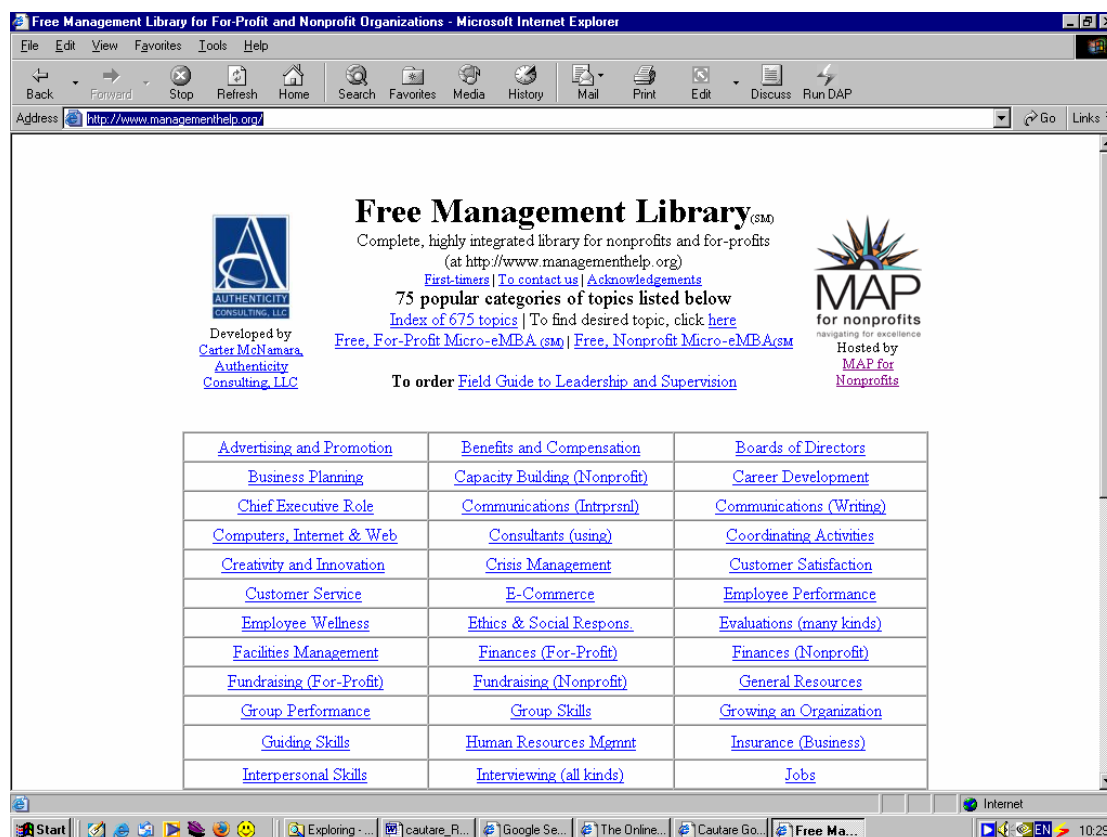


Fig. 6. *Free Management Library – bibliotecă virtuală de specialitate*

iv. Invisible (deep) Web

Reprezintă colecții de informații online stocate în baze de date accesibile pe Web, dar care, din diferite motive, nu sunt indexate de motoarele de căutare tradiționale. Aceste servicii pot fi întâlnite și sub numele de servicii speciale de căutare (specially search engines). Exemple de astfel de colecții sunt:

ProFusion (www.profusion.com)

Invisible-web (www.invisible-web.net)

Complete Planet (www.completeplanet.com)

Resource Discovery Network (www.rdn.ac.uk)

Direct Search (<http://www.freepint.com/gary/direct.htm>)

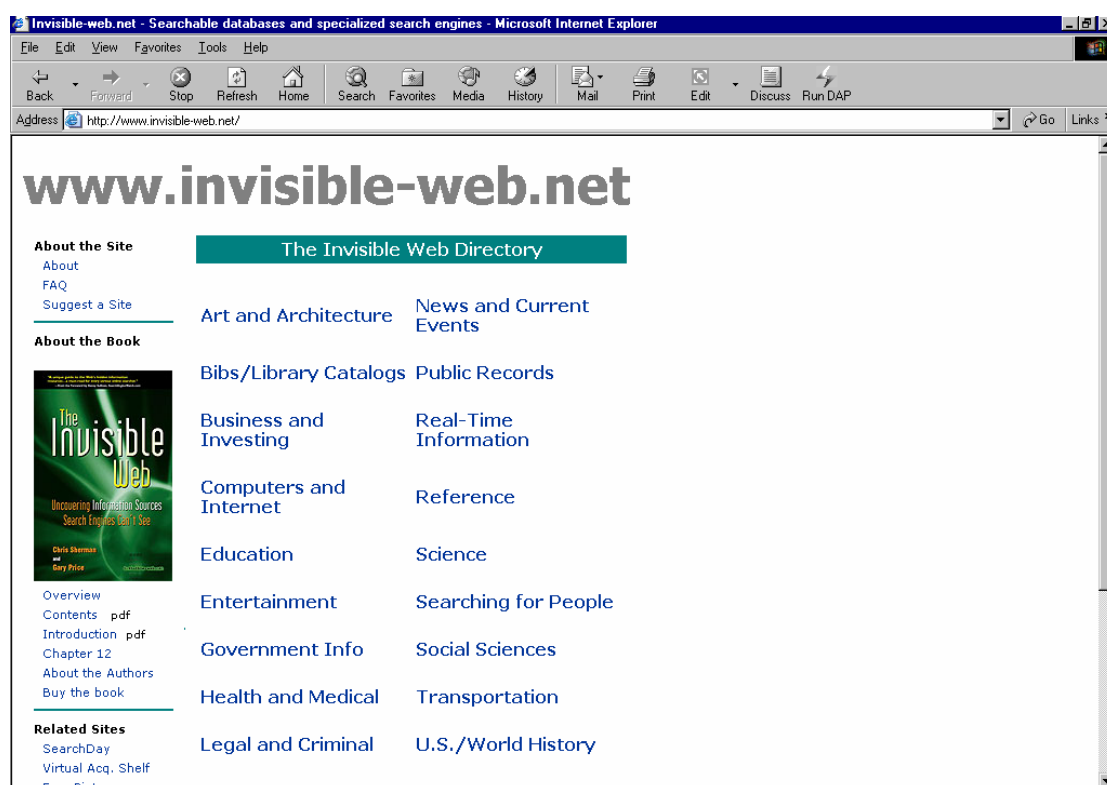


Fig. 7. Invisible Deep Web

v. Motoare de meta-căutare (metasearch engine)

Un motor de meta-căutare este o unealtă de căutare care trimite cerea dumneavoastră simultan către mai multe motoare de căutare clasice, servicii de directoare web și uneori care spre colecții de tip invisible web. După colectarea rezultatelor căutării de la toate serviciile apelate, motoarele de meta-căutare elimină legăturile aflate în duplicat și, conform unui algoritm propriu combină rezultatele într-o singură listă, acordându-le și o rată de relevanță.

Datorită faptului că majoritatea motoarelor de meta-căutare iau în considerare numai primele 10-20 de legături din lista furnizată de fiecare motor de căutare apelat, ne putem aștepta la rezultate excelente ale căutării, din punctul de vedere al relevanței.

Din păcate însă, nu toate motoarele de căutare și serviciile de directoare web suportă o serie tehnici de căutare avansată (cum ar fi căutarea după fraze sau folosirea operatorilor logici). Din acest motiv, dacă căutarea se face folosind aceste tehnici avansate, motorul de metacăutare nu va primi nici un rezultat din partea motoarelor apelate ce nu folosesc astfel de tehnici de lucru.

Din cele menționate mai sus, reiese clar că motoarele de metacăutare NU au propriile baze de date web, ci le folosesc pe cele ale serviciilor apelate. Tot din acest motiv, motoarele de meta-căutare nu permit înscrierea manuală a unei pagini (site) în baza de date, lucru posibil în cazul motoarelor de căutare clasice.

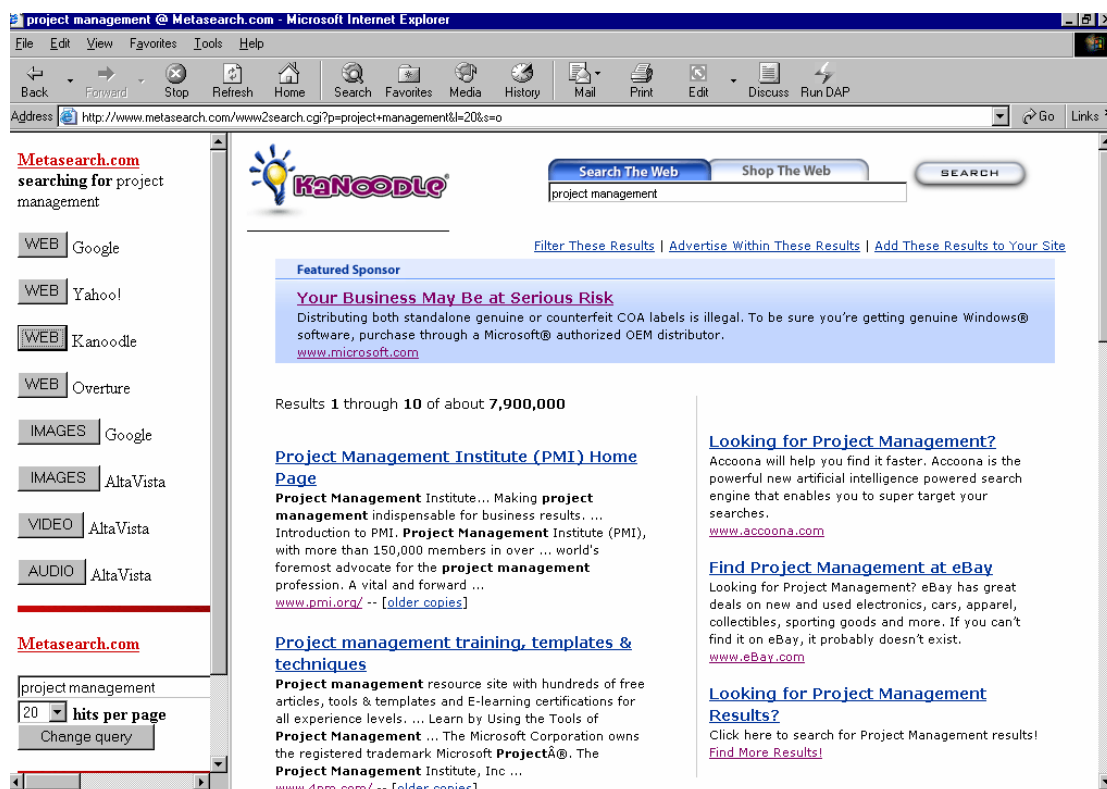


Fig. 8. Motorul de meta-căutare metasearch.com

Unele dintre cele mai cunoscute motoare de meta-căutare sunt:

Metasearch (www.metasearch.com)
ez2Find (www.ez2find.com)
Vivisimo (www.vivisimo.com)
MetaCrawler (www.metacrawler.com)
InfoGrid (www.infogrid.com)
Infonetware (www.infonetware.com)
iBoogie (www.iboogie.tv)

vi. Utilitare de căutare de tip desktop

Un tip special de motoare de meta-căutare sunt utilitarele de căutare de tip desktop (client side search software). Practic, spre deosebire de motoarele de meta-cautare online, aceste utilitare sunt programe ce pot fi descărcate de pe Internet și instalate pe propriul calculator (local). Principiul lor de funcționare este oarecum similar cu al motoarelor de meta-căutare.

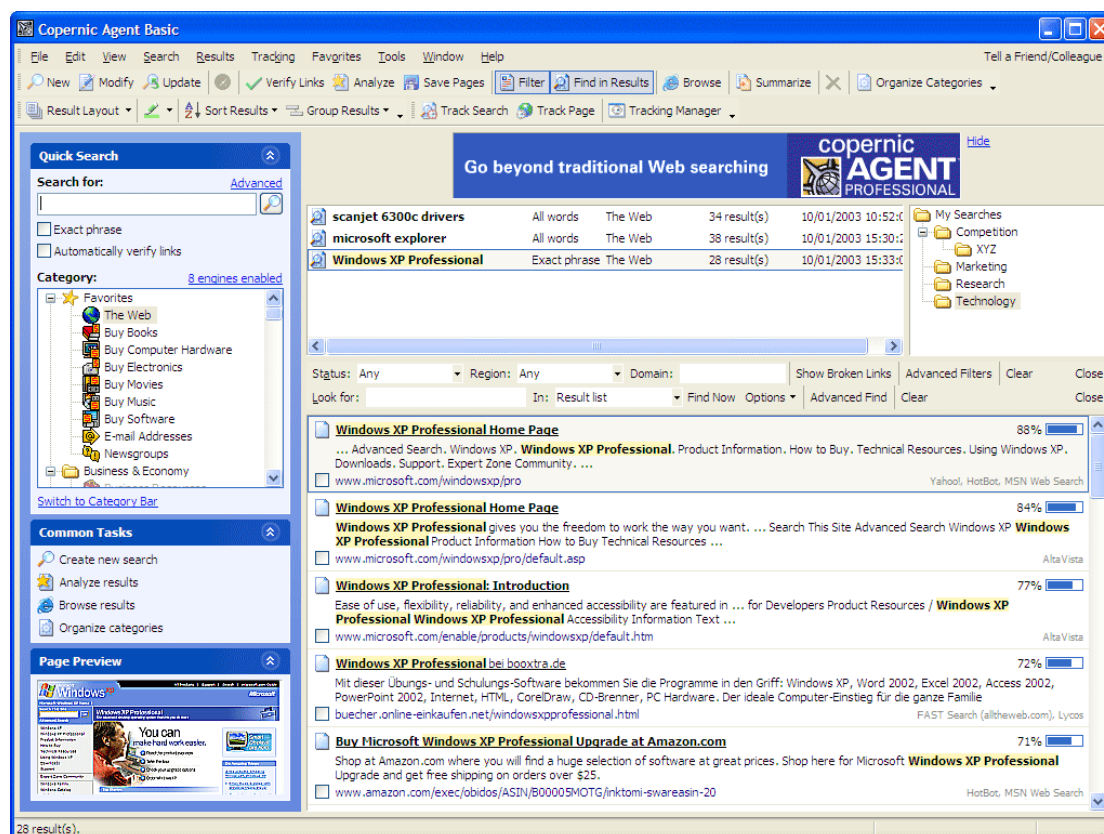


Fig. 9. Utilitare de căutare de tip desktop – Copernic

Câteva dintre cele mai utilizate utilitare de acest fel sunt:

Google Desktop (desktop.google.com)

Copernic (www.copernic.com)

Arrow Search (www.rt-software.co.uk/arrow_search/)

WebFerret (www.ferretsoft.com/download.htm)

ProtoSearch (www.npccenterprises.com/products/protosearch2.shtml)

Ca o concluzie generală, motoarele de meta-căutare sunt unelte complementare excelente, însă datorită unor considerente (număr relativ mic de legături luate în considerare de la fiecare serviciu apelat, imposibilitatea de a folosi eficient tehnici avansate de căutare etc) NU pot elimina necesitatea folosirii unor motoare de căutare clasice.

III. CE UNELTE DE CĂUTARE SĂ FOLOSIM ?

1. **Dacă cunoașteți exact ce tip de informație doriți să aflați, este recomandată folosirea motoarelor de căutare.** De exemplu pentru a găsi informații despre viroze sau despre președinții României puteți folosi motoarele de căutare, deoarece informația nu este suficient de specifică pentru a fi subiectul unui arbore de directoare, dar este domeniul este suficient de bine conturat pentru a nu fi necesară recurgerea la un motor de meta-căutare.

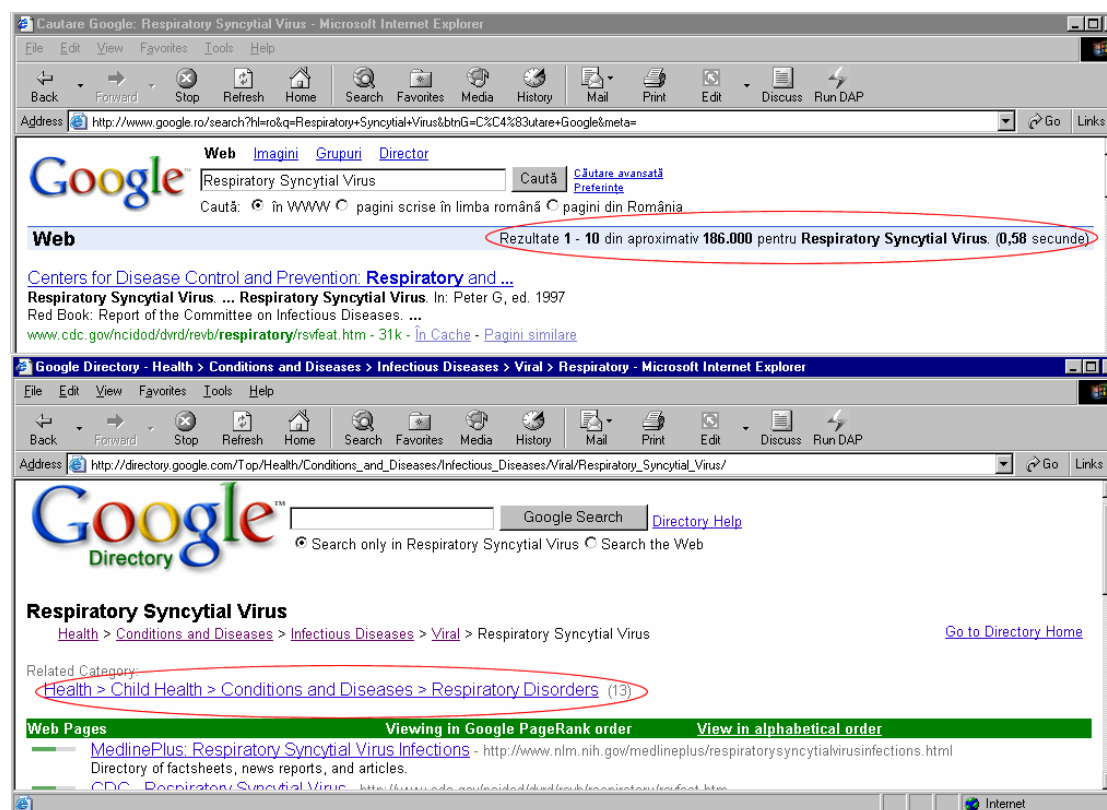


Fig. 10. Utilizarea motoarelor de căutare (186000 de legături) versus servicii de directoare (13 legături) în cazul informațiilor despre virusul sincițial respirator

2. Dacă veți căuta **informații despre un anumit termen sau o combinație obscură de termeni și doriți o analiză exhaustivă a temei analizate,** este de preferat **recurgerea la motoare de meta-căutare,** ce pot apela simultan mai multe servicii web (10 -12 motoare de căutare sau directoare web).

De exemplu, căutarea după termenul vibrion (vezi Vibrionul holerei), Yahoo a returnat 7790 legături dintre care chiar prima legătură se referea la altceva decât vibrionul holerei.

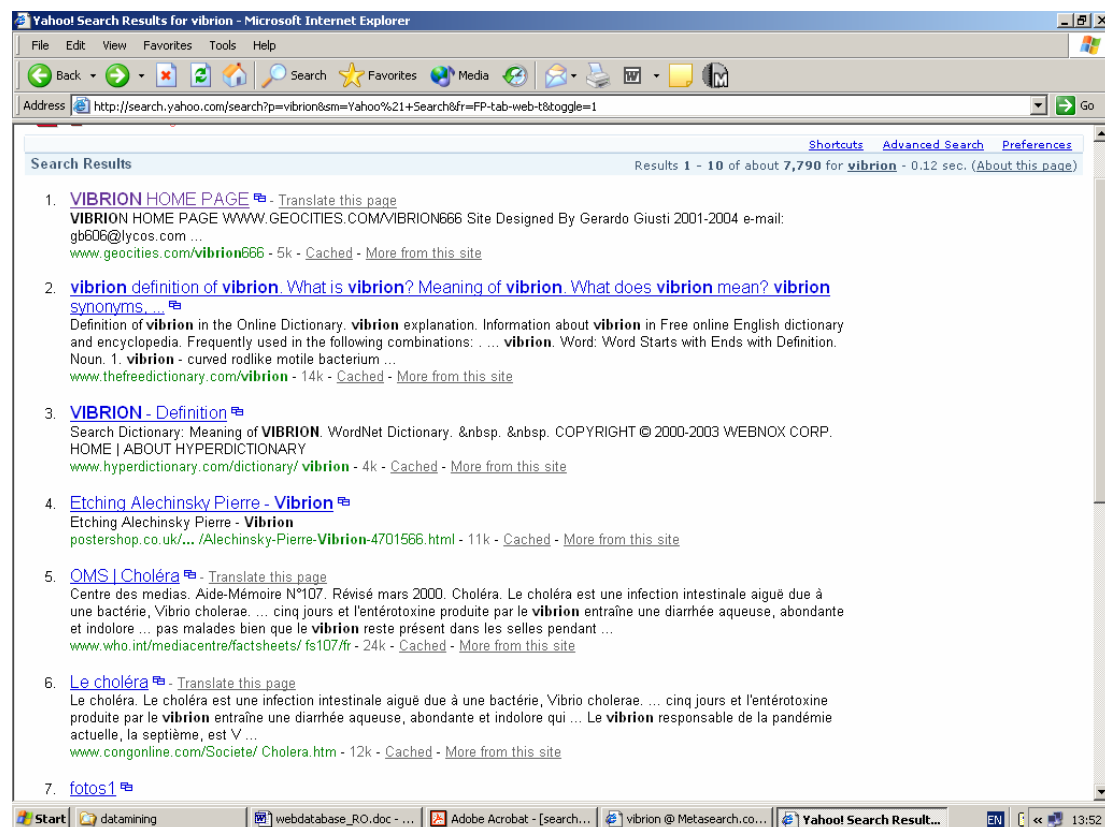


Fig. 11. Rezultatele căutării enunțate mai sus în urma utilizării serviciului Yahoo

În contrapondere, aceeași căutare pe metasearch.com, a returnat 5340 legături, prima referință fiind legată chiar de definirea termenului vibrio.

De remarcat că metasearch.com recomandă și sinonimul vibrio ca posibil termen de căutare.

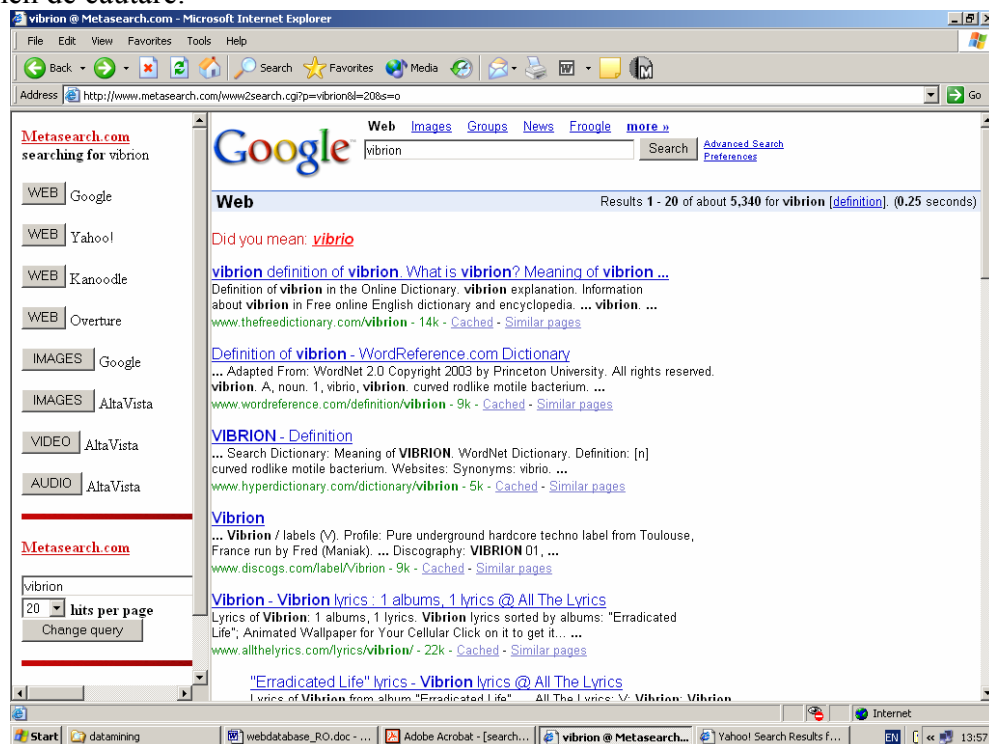


Fig. 12. Rezultatele căutării enunțate mai sus în urma utilizării motorului de căutare Google

3. Dacă doriți să căutați **informații generale despre subiecte cu grad ridicat de popularitate**, este indicată folosirea **serviciilor de directoare web**. Astfel dacă doriți informații despre astmul bronșic în general, este mult mai indicat să folosiți serviciul de directoare Yahoo, în care puteți naviga cu ușurință, informația este bine organizată și există și facilități de căutare prin arborele de directoare.



Fig. 13. Yahoo Directory

4. Dacă doriți, **informații pertinente legate de anumite domenii ale literaturii, științei** etc este preferabilă folosirea **serviciilor tip bibliotecă virtuală**.

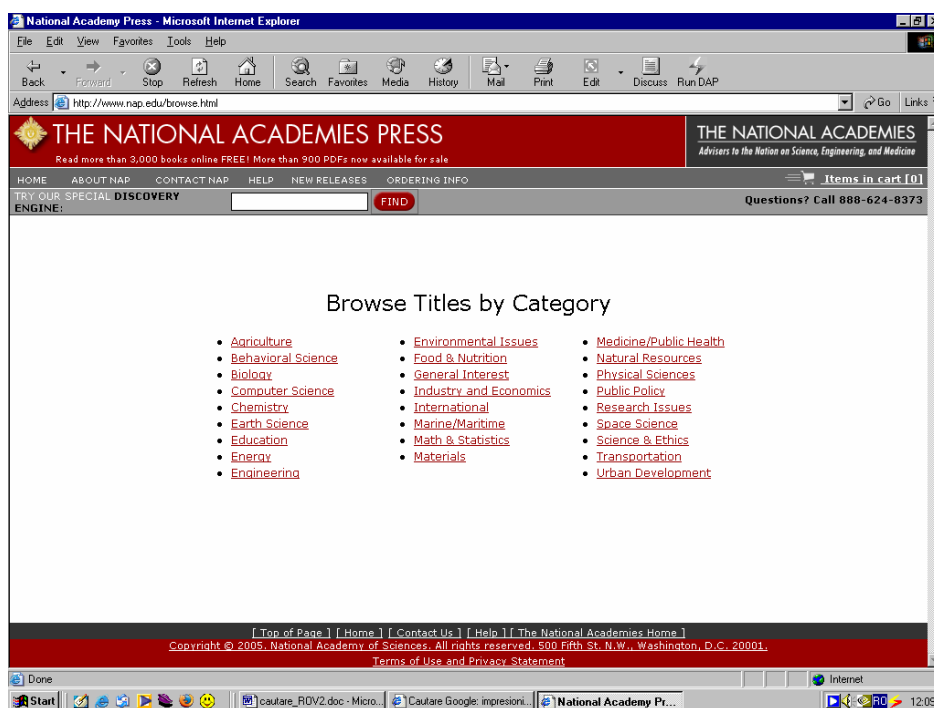


Fig. 14. National Academy Press – bibliotecă științifică virtuală

5. Dimpotrivă, dacă doriți **informații în timp real sau informații al căror conținut se schimbă dinamic** (știri, prognoze meteo, mersul trenurilor) folosiți **baze de date web specializate** (invisible or deep web).

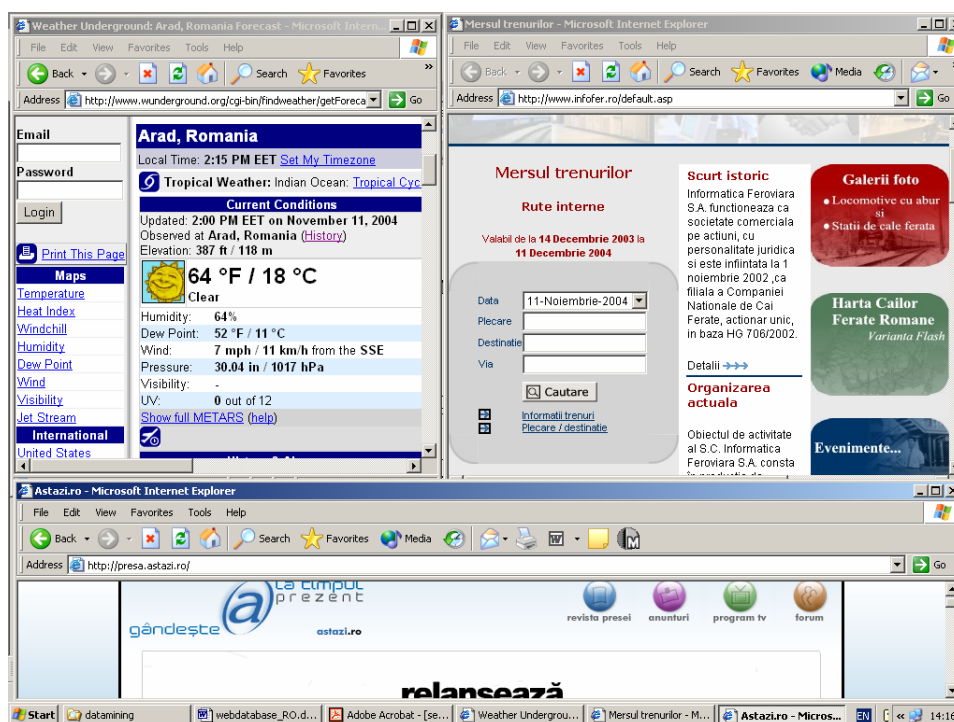


Fig. 15. Modalități de utilizare a serviciilor oferite de diverse baze de date de tip Invisible Web

Există un număr de câteva mii de motoare de căutare, sute de motoare de meta-căutare, servicii de directoare web sau baze de date specializate. Alegerea din start a celei mai potrivite modalități de căutare a informației poate face diferența dintre o cautare cu succes a informației dorite și o experiență frustrantă datorită pierderii de timp și a posibilelor confuzii.

Mai mult, în ultimul timp mulți dintre deținătorii de astfel de servicii au ales să își diversifice activitatea furnizând mai multe servicii sub aceeași interfață utilizator. Astfel s-a cristalizat un nou concept, acela de **PORTAL WEB**.

Portalul Web

Sub acest termen se înțelege în general un site WEB ce oferă o gamă largă de servicii și resurse online, cum ar fi: e-mail, forumuri de discuții, motoare de căutare, directoare web, comerț electronic etc (a se vedea www.yahoo.com, unul dintre cele mai cunoscute servicii de căutare, transformat într-un portal). Dacă primele portaluri (cum ar fi AOL –Amerca Online) furnizau la început acces pe Internet și servicii de e-mail, mai nou majoritatea motoarelor de căutare tradiționale s-au transformat în portaluri Web, ce oferă o mare varietate de servicii, în dorința de a-și lărgi astfel baza de utilizatori.

The screenshot shows the InfoGrid Master Internet Map Index website. The browser window is titled "InfoGrid(tm) Master Internet Map Index - Microsoft Internet Explorer". The address bar shows "http://www.infogrid.com/". The page has a search bar at the top with the text "Search the Internet Using: InfoGrid - All the Web". Below the search bar, there are several sections: "infogrid: free | no-popups | no-advertising | non-commercial | metasearch | web-directory", "sign-up or enter ActivatorMail address then password to login:", "direct e-mail logins: HotMail | Yahoo | Excite | AOL | InfoGrid | Anti-Spam | Anti-Virus | More...", "most popular news: jennifer lopez, lindsay lohan, britney spears, nasa, usher, pamela turner, avil lavigne, lottery, internal revenue, the apprentice, tsunani, bachelorette, academy awards, earthquakes, pamela anderson", and "featured most popular: Google | Yahoo | MSN | Passport | Amazon | eBay | Drudge | More...". There are two line graphs showing "Dow 28-Feb 3:59pm (C)Yahoo!" and "Nasdaq 28-Feb 3:59pm (C)Yahoo!". Below the graphs is a table of Market Indices. The table has columns: Index, Last, Change, and %Change. The data is as follows:

Index	Last	Change	%Change
Dow 30	10766.23	-75.37	-0.70%
S&P 500	1203.60	-7.77	-0.64%
Nasdaq Comp.	2051.72	-13.68	-0.66%
Russell 2000	634.06	-3.47	-0.54%
S&P Midcap 400	666.85	-3.47	-0.52%

Below the table is a link "view: indices | top volume | top price | top declines | nyse | nasdaq | amex | add | news". The right side of the page features a "Yahoo! News: Top Stories" section with headlines like "Bin Laden Enlisting Al-Zarqawi for Attacks (AP)", "Death Toll in Iraqi Bombing Rises to 120 (AP)", "Vatican: Pope's Condition Is Improving (AP)", "Lebanon Government Resigns After Protest (AP)", "White House Must Charge or Free Suspect (AP)", and "Prosecution Portrays Jackson As Molester (AP)".

Fig. 16. Portalul InfoGrid – de remarcat multitudinea serviciilor oferite (servicii de căutare, directoare Web, e-mail, chat, news etc)

IV. UTILIZAREA MOTOARELOR DE CĂUTARE

Așa cum am văzut sunt două modalități mai importante prin care un utilizator poate interoga o bază de date web: folosind căsuța de interogare a unui motor de căutare (așa-numita tehnică type-in-a-box), respectiv selectând anumite opțiuni dintr-o structură arborescentă de directoare.

Acest capitol va analiza facilitățile oferite de prima variantă, de altfel cea mai utilizată cale de regăsire a informațiilor pe Web, cu scopul de a vă ajuta să sesizați diferențele dintre diferitele motoare de căutare, devenind astfel utilizatori mai eficienți ai acestui tip serviciu.

Putem deosebi patru pași în utilizarea unui motor de căutare:

1. “traducerea” nevoilor dumneavoastră de informare în termeni cheie introduși în căsuța de căutare a motorului folosit
2. folosirea unor modalități de combinare a termenilor de căutare, pentru a furniza calculatorului (serverului) cât mai multă informație despre ceea ce doriți să căutați
3. analiza rezultatelor căutării
4. “rafinarea” termenilor de căutare și repetarea interogării, dacă rezultatele obținute nu vă satisfac

Acești pași pot fi repetați de cîte ori este nevoie, până când ați găsit ceea ce vă interesa sau până când veți decide că ați epuizat resursele indexate de motorul de căutare folosit și este necesar să încercați un alt tip de serviciu.

NOTĂ: Tehnicile de lucru prezentate mai jos pot fi utilizate în cazul majorității serviciilor de căutare disponibile pe Web. Există însă posibilitatea ca o anumită modalitate de lucru să NU fie disponibilă în cadrul unui anumit motor de căutare. De aceea vă recomand călduros ca, înainte de a începe lucrul cu un motor de căutare, să vizitați legătura spre secțiunea HELP (Ajutor) a acelui motor, pentru a identifica în detaliu tehnicile de căutare disponibile.

1. “Traducerea” nevoilor dumneavoastră de informare în termeni cheie

Tehnologia de azi a motoarelor de căutare nu ne permite, din nefericire, să formulăm necesitățile noastre de informare așa cum am dori. Nu putem scrie în căsuța de căutare a motorului google.ro, de exemplu „Vreau să cumpăr un calculator”. Secretul constă în a folosi termeni de căutare cât mai simpli și mai specifici cu putință. Puteți face asta urmând sugestiile de mai jos:

i. - gândiți-vă la ce doriți să căutați și focalizați-vă asupra a ceea ce doriți să obțineți. Probabil că nu doriți să achiziționați orice tip de calculator și de oriunde. De aceea în locul formulării „Vreau să cumpăr un calculator ieftin” veți putea spune mai degrabă „Vreau să cumpăr un calculator Pentium IV la preț bun din România”. Gândiți-vă cât mai detaliat la ceea ce vreți să aflați. Pentru a căuta cu succes informația, însă, trebuie să fiți în același timp cît mai concisi.

ii. – “spargeți” interogarea dumneavoastră inițială în cît mai multe fațete, în încercarea de a o reduce la numai câțiva termeni cheie. Dacă doi termeni sunt similari, este bine să îi combinați într-unul singur (vezi termenii calculator și Pentium IV). Aceasta va simplifica căutarea focalizând-o pe ideile cu adevărat importante. Astfel conform exemplului nostru, ar putea fi suficient să folosim termenii **vînd, calculator, Pentium IV, preț**. (termenul preț este un cuvînt specific românesc, deci ar putea permite

excluderea paginilor care nu sunt în română și subliniază ideea de tranzacție comercială; termenul Pentium IV poate înlocui eventual și termenul calculator, riscând cel mult să aflăm eventual și prețul unor procesoare Pentium IV).

Aceaste trucuri vor simplifica căutarea, focalizând-o asupra ideilor importante.

Odată ce ați urmat acești doi pași aveți deja termenii de bază ai căutării – termeni cheie ce pot fi suficient de ilustrativi pentru a exprima întreaga idee de la care ați pornit.

Din nefericire lucrurile nu sunt întotdeauna atât de simple. Există încă o serie de alte aspecte ce trebuie luate în considerare atunci când alegeți termenii cheie după care veți efectua căutarea:

Sinonimele

Va fi de un real folos dacă veți utiliza sinonime pentru fiecare termen de căutare. Urmînd această idee puteți încerca și „ofer” pentru „vînd”, „PC”, „computer” pentru „calculator”, „PIV” pentru „Pentium IV” etc. Cu cât veți găsi mai multe sinonime pentru termenii folosiți și le veți utiliza în căutare, cu atât va crește șansa de a găsi exact ceea ce vă interesează.

În ceea ce privește utilizarea sinonimelor în căutare trebuie însă subliniat faptul că este bine să aveți incluse în căutare aproximativ același număr de sinonime pentru fiecare termen, cu excepția cazului în care doriți să „dați greutate” mai mare unui anumit termen, tehnică pe care o vom discuta mai târziu.

Omonime

Trebuie acordată o maximă atenție cuvintelor care au mai multe sensuri (cum ar fi veselă/veselă sau Ion/ion). Computerul nu va înțelege diferența și nu poate specifica contextul în care este folosit cuvîntul. Astfel, o căutare simplă după cuvîntul ion ne va returna atât pagini ce prezintă informații ioni (ca noțiune din domeniul chimiei), dar și posibile pagini personale ale unor oameni avînd prenumele de Ion sau chiar instituții care au ca sigla cuvîntul ION (de exemplu Institute of Navigation).

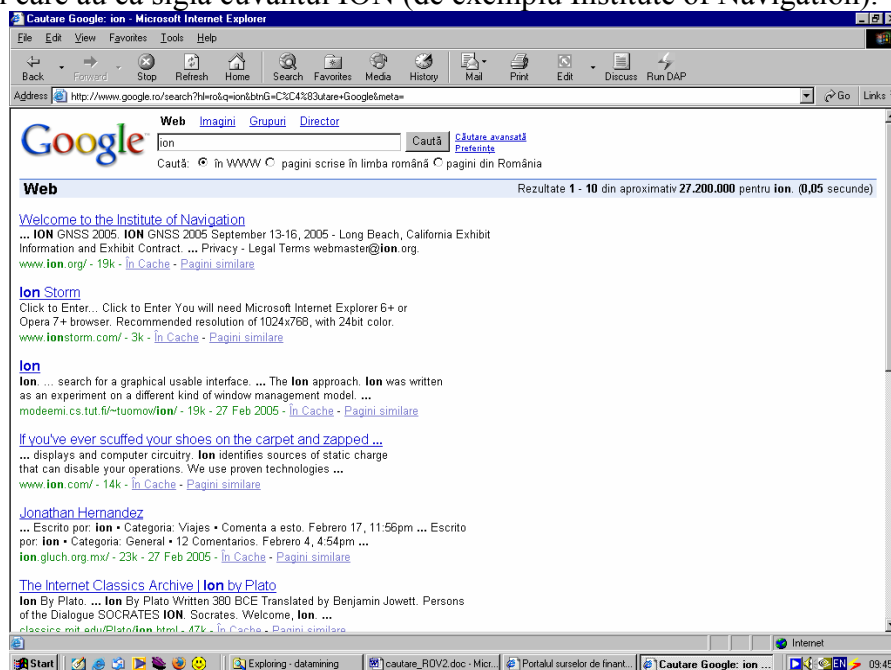


Fig. 17. Influența omonimelor asupra relevanței rezultatelor căutării

Dialectele unei anumite limbi

Dacă termenii folosiți sunt scriși diferit în diferite dialecte ale unei limbi, căutarea trebuie să țină cont de toate aceste forme. Să luăm un exemplu din limba engleză și diferitele sale dialecte (engleza americană, britanică, canadiană): cuvinte ca color/colour sau catalog/catalogue vă vor permite să selectați numai pagini aparținând vorbitorilor unui dialect sau altuia. Dacă doriți includerea tuturor paginilor ce conțin toate formele cuvântului, va trebui să includeți ambele variante ca termeni cheie.

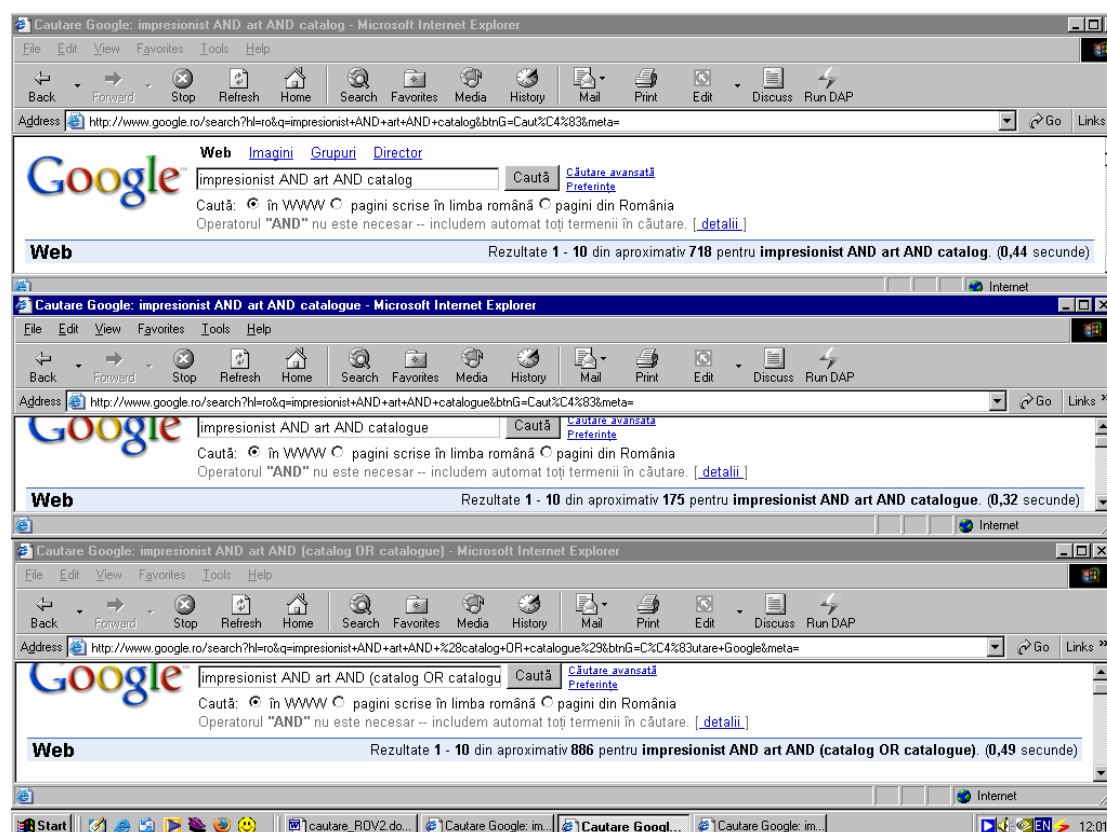


Fig. 18. Influența dialectelor unei limbi asupra rezultatelor căutării

Trunchierea cuvintelor

Această tehnică urmărește cuvintele ce pot avea forme multiple (cum ar fi **woman** și **women** în engleză la singular și plural). Puteți regăsi în urma căutării ambele cuvinte, prin folosirea unor simboluri ca de exemplu * (**wom*n**).

Din nefericire alte motoare de căutare folosesc simboluri diferite pentru a semnifica trunchierea cuvintelor, lucru care trebuie luat în considerare. De asemenea, anumite motoare de căutare, cum este Lycos nu acceptă decât așa numita tehnică "right-truncation", adică trunchierea NUMAI la sfârșitul cuvântului (de exemplu **child\$**, pentru forme ca **children**, **child**, **childlike**, **childhood**).

Această tehnică de lucru permite folosirea celei mai simple forme ale cuvântului împreună cu simbolurile pentru trunchiere, pentru a regăsi toate formele în care apare cuvântul respectiv.

2. Modalități de combinare a termenilor de căutare

Motoarele de căutare vă pun în mod obișnuit la dispoziție posibilitatea de a specifica o serie de relații între termenii de căutare. Specificarea acestor relații poate ajuta serverul să ordoneze lista de rezultate într-o ordine mult mai relevantă pentru dumneavoastră (cu alte cuvinte crește șansa ca primele legături din lista cu rezultate ale căutării să fie exact ceea ce căutați).

Puteți face acest lucru folosind următoarele tehnici: căutarea după fraze, folosirea logicii booleene (operatori logici), folosirea logicii pseudo-booleene, creșterea importanței unui termen (așa numitul term-weighting)

Căutarea după fraze

Este una dintre cele mai performante tehnici de combinare a termenilor de căutare, fiind recomandată folosirea ei de câte ori este posibil. Atunci când stabiliți termenii după care veți efectua căutarea, dacă este importantă a numită ordine a cuvântului în construcția frazei, este suficient să puneți secvența respectivă în ghilimele (" "), simboluri folosite de majoritatea motoarelor de căutare pentru căutarea după se poate fraze. Atunci când poate fi folosită, căutarea după fraze este una dintre cele mai puternice modalități de creștere a relevanței rezultatelor căutării.

Exemplu: Dacă veți dori să citiți online cartea „The comedy of errors” de William Shakespeare, probabil veți obține cele mai bune rezultate introducând în fereastra de căutare următoarea secvență: „**William Shakespeare**” „**The comedy of errors**” **online**, așa cum puteți vedea în exemplul de mai jos:

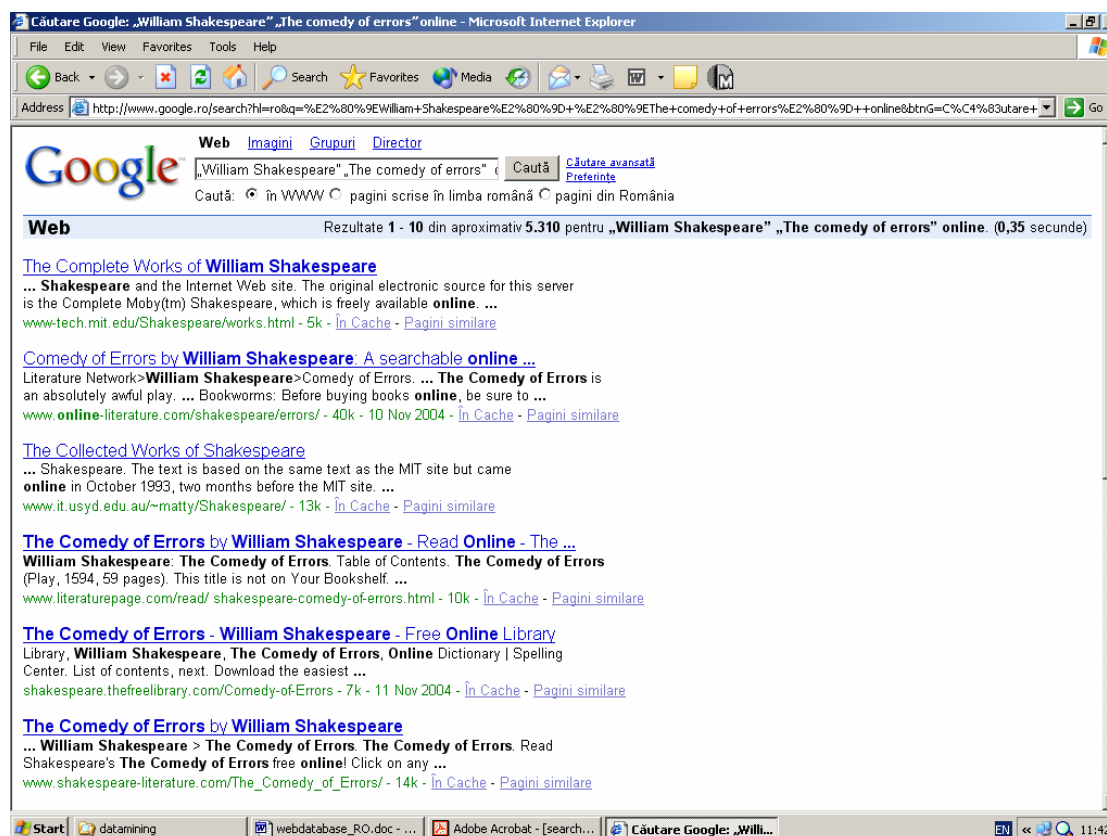


Fig. 19. Căutarea după fraze

Logica booleană (Folosirea operatorilor logici)

Chiar dacă formularea din titlu pare menită să vă sperie, nu vă lăsați înșelați – folosirea acestei tehnici de căutare este extrem de simplă. Practic sunt folosite cuvintele (operatorii logici) AND, OR și NOT (atenție, scrise cu MAJUSCULE), pentru a indica bazei de date web existența unei anumite relații între termenii după care se efectuează căutarea.

Operatorul AND – prin folosirea acestui operator între doi termeni de căutare, transmiteți bazei de date o cerere de genul: „Vreau să găsesc toate paginile care conțin OBLIGATORIU ambii termeni, INDIFERENT de ordinea în care apar în pagină. Dacă o pagină conține NUMAI unul dintre termeni, nu mă interesează”. În cazul nostru, căutarea propusă anterior, poate fi formulată astfel: **vând AND calculator**. Dacă am folosi numai unul dintre termeni, rezultatele căutării ar fi cu siguranță nerelevante.

Operatorul OR – folosirea acestui operator indică motorului necesitatea unei căutări mai extinse, cu returnarea tuturor legăturilor ce conțin unul SAU altul dintre termenii căutării. Operatorul OR este foarte util în specificarea sinonimelor pentru termenii doriți.

Exemplu: **vând AND PC OR computer OR calculator**

Un aspect important de remarcat este felul în care am folosit operatorii AND și OR pentru a combina termenii căutării în scopul creșterii șansei de a obține rezultate cât mai relevante. Astfel este indicată folosirea operatorului **OR între sinonimele termenilor căutării**, respectiv a operatorului **AND între termeni diferiți ca semnificație**.

Relevanța rezultatelor va crește dacă vom fi și mai specifici, ca în exemplul

Exemplu: **vând AND PC OR computer OR calculator AND „magazin online”**

următor:

Observați combinarea folosirii operatorilor logici cu tehnica de căutare după fraze (termenul specificat în ghilimele).

Uneori ne este de folos și gruparea termenilor în paranteze, pe baza semnificației lor (sinonime). Exemplu de mai sus devine astfel:

Exemplu: **vând AND (PC OR computer OR calculator) AND „magazin online”**

Operatorul NOT - operatorul poate fi folosit în anumite circumstanțe, cum ar fi îndepărtarea din rezultate a unei semnificații generate de un termen omonim. În cazul omonimelor veselă= ustensile de bucătărie, veselă=voioasă, putem exclude unele rezultate nerelevante printr-o căutare de genul:

Exemplu: (**femeie OR persoană**) AND (veselă NOT bucătărie).

Vom exclude astfel rezultatele cu legături spre pagini cu informație nerelevantă ca în exemplul „...este genul de persoană care cumpără multă veselă de bucătărie”. Nu uitați de gruparea termenilor negației în paranteze !

Un alt exemplu este excluderea anumitor contexte legate de unul din termenii căutării, ca în exemplu

Exemplu: „**Leonardo Da Vinci**” NOT (picturi AND tablouri).

O astfel de căutare indică faptul că dorim alte informații legate de Leonardo Da Vinci decât cele legate de pictura sa.

Deoarece tehnicile prezentate mai sus reprezintă concepte esențiale pentru căutarea pe Web, este foarte important ca, înainte de a merge mai departe, să vă asigurați că ați reținut corect când și cum trebuie utilizați operatorii logici (booleani): operatorul AND - pentru **combinarea** termenilor căutării cu scopul de a **focaliza căutarea** pe subiectul dorit de noi; operatorul OR - în special **în cazul sinonimelor**, pentru a **mări aria de căutarea** prin posibilitatea de a regăsi pagini care să conțină unul sau mai multe sinonime ale termenului dorit; operatorul NOT - pentru **excluderea anumitor contexte** legate de unul din termenii căutării.

NOTĂ: Atenție ! Nu toate motoarele de căutare vă permit folosirea acestei tehnici, în timp ce alte motoare de căutare o folosesc ca tehnică implicită, prin adăugarea unui AND „invizibil” între termenii folosiți pentru căutare (așa-numitul AND/OR invizibil). Acest lucru va avea ca efect afișarea în primul rând a legăturilor care conțin TOȚI termenii și abia apoi a legăturilor ce conțin doar o parte din termenii căutați. (pentru aceste motoare *vând calculator ≈ vând AND calculator*)

Logica pseudo-booleană

Deoarece folosirea operatorilor logici AND, OR și NOT poate părea destul de complicată, multe motoare de căutare permit utilizarea așa-numitului „limbaj natural” pentru formularea scopului căutării, limbaj ce implică folosirea a două simboluri (+ și -) pentru operații “pseudo-booleene”.

Astfel, dacă veți adăuga semnul + înaintea unui termen de căutare sau înaintea unei fraze incluse în ghilimele, aceasta va însemna că respectivul termen TREBUIE să se regăsească neapărat în toate legăturile returnate. Se poate observa că are o utilizare destul de apropiată cu a operatorului logic AND. Poate fi folosit cu succes dacă doriți de pildă informații despre o anumită persoană.

Exemplu: + “**Leonardo da Vinci**” painter, sculptor, architect, engineer

O astfel formulare a căutării ar putea crește șansele de a găsi în cadrul aceleiași legături mai multe informații despre opera de o mare diversitate, a lui Leonardo da Vinci.

În urma unei astfel de căutări, motorul va returna TOATE legăturile din baza de date ce conțin numele **Leonardo da Vinci**, și le va ordona în lista cu rezultate în funcție de ceilalți termeni pe care aceste pagini îi conțin sau nu.

Folosirea semnului ” – ” înaintea unui termen de căutare (sau a unei fraze) va însemna că acesta NU trebuie să se regăsească în conținutul niciuneia din paginile (legăturile) returnate. Are deci o utilizare similară operatorului logic NOT.

Creșterea importanței anumitor termeni (term weighting)

Un ultim pas legat de combinarea termenilor căutării este stabilirea importanței lor în cadrul operațiunii de căutare. Tehnica se bazează pe sublinierea importanței anumitor termeni în comparație cu restul termenilor folosiți. Prin această tehnică puteți comunica motorului de căutare că anumiți termeni de căutare sunt mai importanți decât alții. Pentru exemplul considerat de noi, legat de cumpărarea unui calculator, s-ar putea să considerăm termenul „Pentium IV” ca fiind foarte important, în sensul că dorim să cumpărăm nu orice fel de calculator ci unul de ultimă generație. Diferitele motoare de căutare utilizează, evident, tehnici diferite pentru a face aceasta. Ca regulă generală însă, cu cât avem un număr mai mare de sinonime pentru un anumit termen de căutare, cu atât acesta va fi considerat mai important de către motorul de căutare, evident din punct de vedere al ratei de relevanță.

Astfel dacă dorim să cumpărăm un calculator neapărat Pentium IV, este bine să formulăm căutarea în felul următor:

Exemplu: **vând AND calculator AND (“Pentium IV” OR PIV OR „Intel PIV”)**

Tehnici de limitare a ariei de căutare

Una dintre cele mai puternice facilități oferite de majoritatea motoarelor de căutare este capacitatea de a controla ce pagini Web (situri) vor fi sau nu luate în considerare într-o căutare.

i. Căutarea în anumite locații (situri).

Să presupunem că dorim să aflăm o serie de informații despre astm, folosind **numai** resursele aflate la dispoziție pe situl National Institute of Health (nih.gov). Vom folosi în cadrul motorului de căutare Google comanda **site:** , combinată cu tehnicile și operatorii menționați anteriori:

Exemplu: **site:nih.gov asthma**

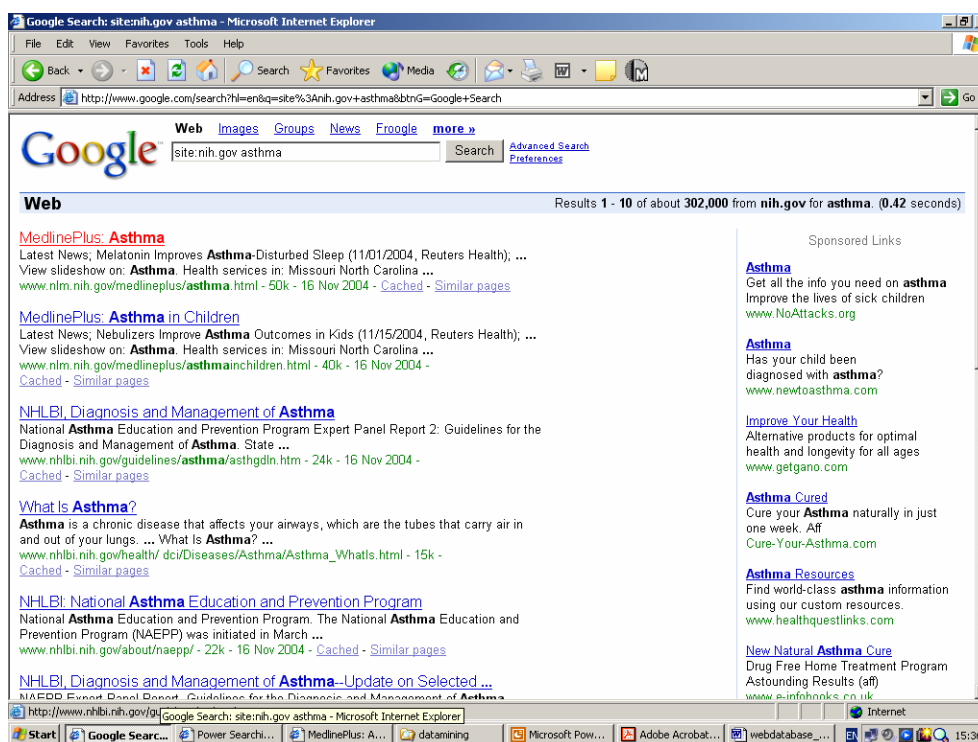


Fig. 20. Tehnici de limitare a ariei de căutare - Căutarea în anumite locații

Ca rezultat, Google va lista toate paginile ce conțin cuvântul „asthma” indexate din domeniul nih.gov. Vom afla astfel că există în cadrul NIH instituții specializate numite National Institute of Allergies and Infectious Diseases (NIAID), respectiv National Heart, Lung and Blood Institute (NHBLI) ce au câte un subdomeniu (site) propriu (niaid.nih.gov, nhlbi.nih.gov), putând astfel să restrângem și mai mult căutarea.

Să presupunem că dorim să găsim modalități sau ghiduri de tratament al astmului, folosind situl NHBLI. Vom formula căutarea, conform celor învățate anterior:

Exemplu: site:nhlbi.nih.gov asthma AND (treatment OR guidelines OR management)

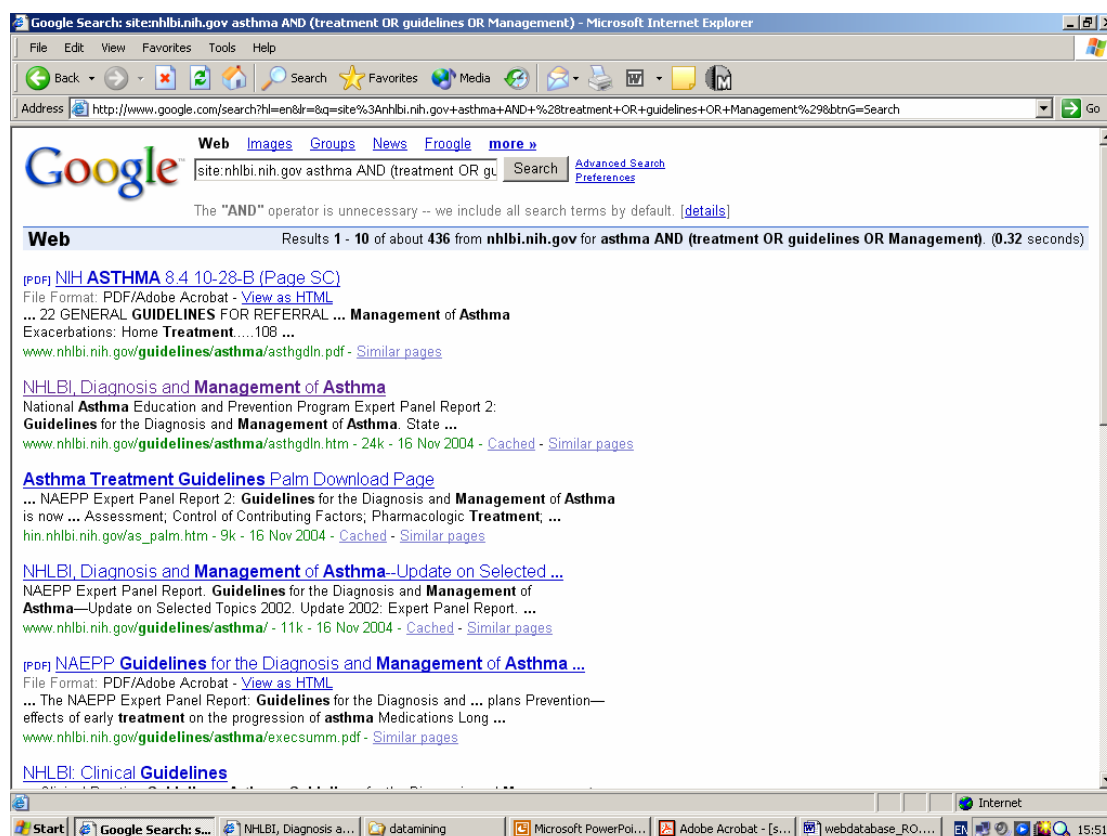


Fig. 21. Tehnici de limitare a ariei de căutare - căutarea în anumite locații + combinarea termenilor de căutare folosind operatori logici

Trebuie remarcat faptul că în acest fel am restrâns lista cu rezultate, *de la peste 300000 de legături la doar 436*.

Dacă vom analiza prima cerere, vom constata că există o serie de legături ce duc spre articole de prezentare generală a bolii (de tipul „What is asthma?” sau “Asthma Basic”). Dacă dorim eliminarea unor astfel de articole putem formula căutarea astfel:

Exemplu: **site:nhlbi.nih.gov asthma -"What is asthma" -basic**

În același fel putem construi expresii care să excludă un anumit sit din rezultatele căutării:

Exemplu: **asthma AND (treatment OR guidelines OR management) -site:nih.gov**

Astfel am căutat modalități de tratament al astmului în alte locații de pe Internet decât nih.gov.

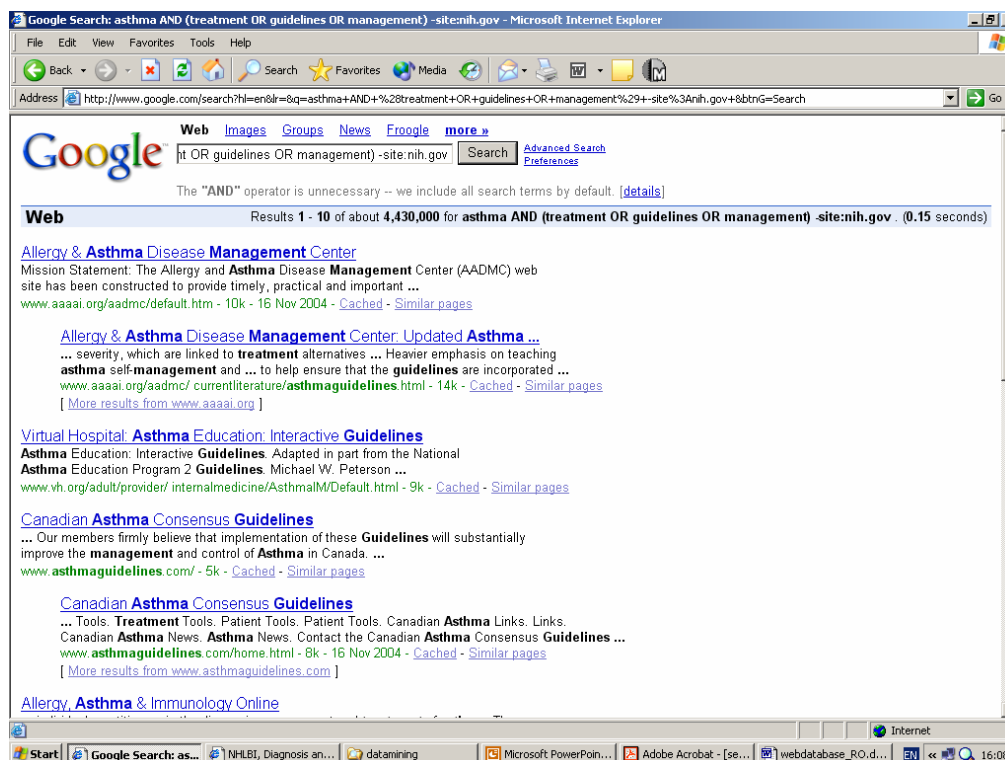


Fig. 22. Tehnici combinate de limitare a ariei de căutare - folosirea operatorului “-”

Dacă, în final am dori să aflăm și opiniile altor surse de informație medicală online asupra tratamentului astmului, ar trebuie să formulăm căutarea astfel:

Exemplu: *asthma AND (treatment OR guidelines OR management) -site:.nih.gov*

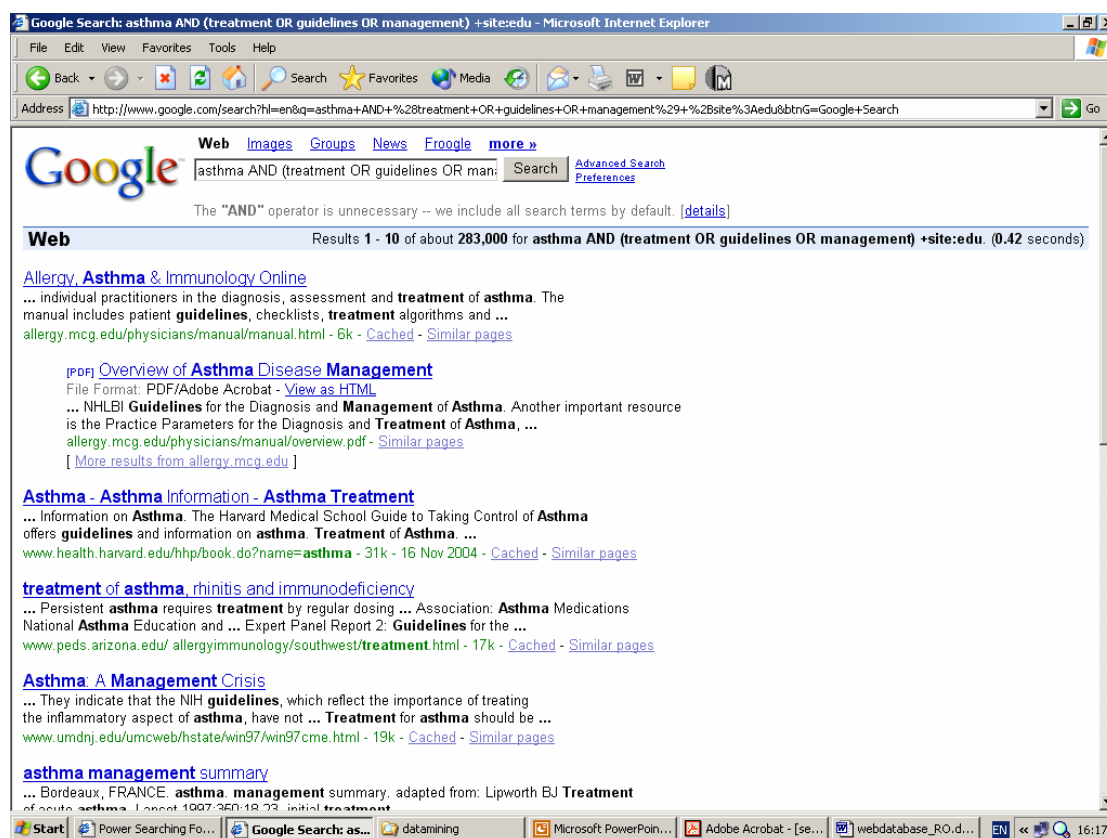


Fig. 23. Tehnici combinate de limitare a ariei de căutare - folosirea operatorului “+”

ii. Căutarea pe baza adresei URL (Uniform Resource Locator)

Uneori dorim să găsim pagini cu un anumit subiect, la o anumită locație. O comandă care ne poate ajuta este comanda **inurl:**
Astfel,

inurl:stanford.edu asthma AND (treatment OR guidelines OR management)

ne permite să regăsim pagini despre tratamentul astmului ce au în adresa URL obligatoriu cuvintele „stanford.edu”.

iii. Căutarea după cuvintele din titlul paginii

Dacă până acum termenii erau căutați oriunde în pagină, putem folosi comanda **title:** (sau **intitle:** în alte motoare de căutare), pentru a forța căutarea termenilor DOAR în titlul paginii.

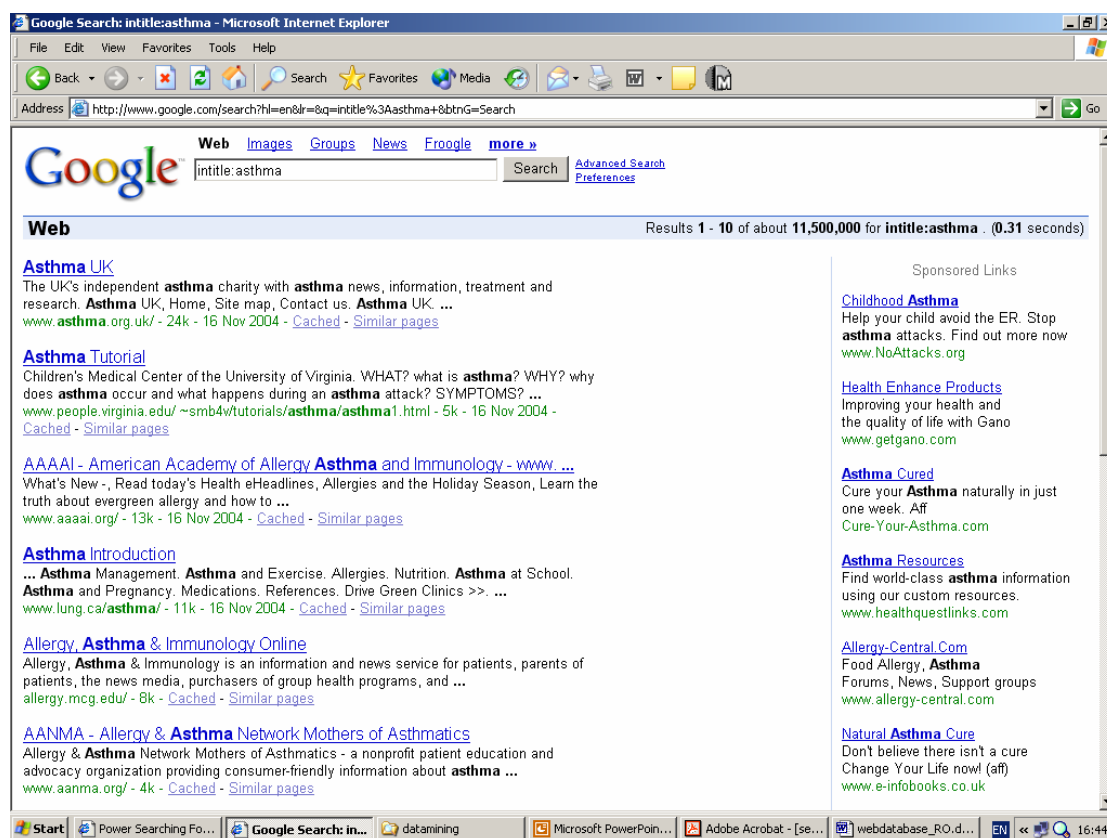


Fig. 24. Tehnici de limitare a ariei de căutare - folosirea operatorului „intitle:”

iv. Căutarea după tipul de fișiere

Să presupunem că dorim să regăsim toate documentele legate de tratamentul astmului ce sunt în formatul pdf (Adobe Portable document format). Putem face aceasta folosind comanda filetype: .

asthma AND treatment OR guidelines OR management filetype:pdf

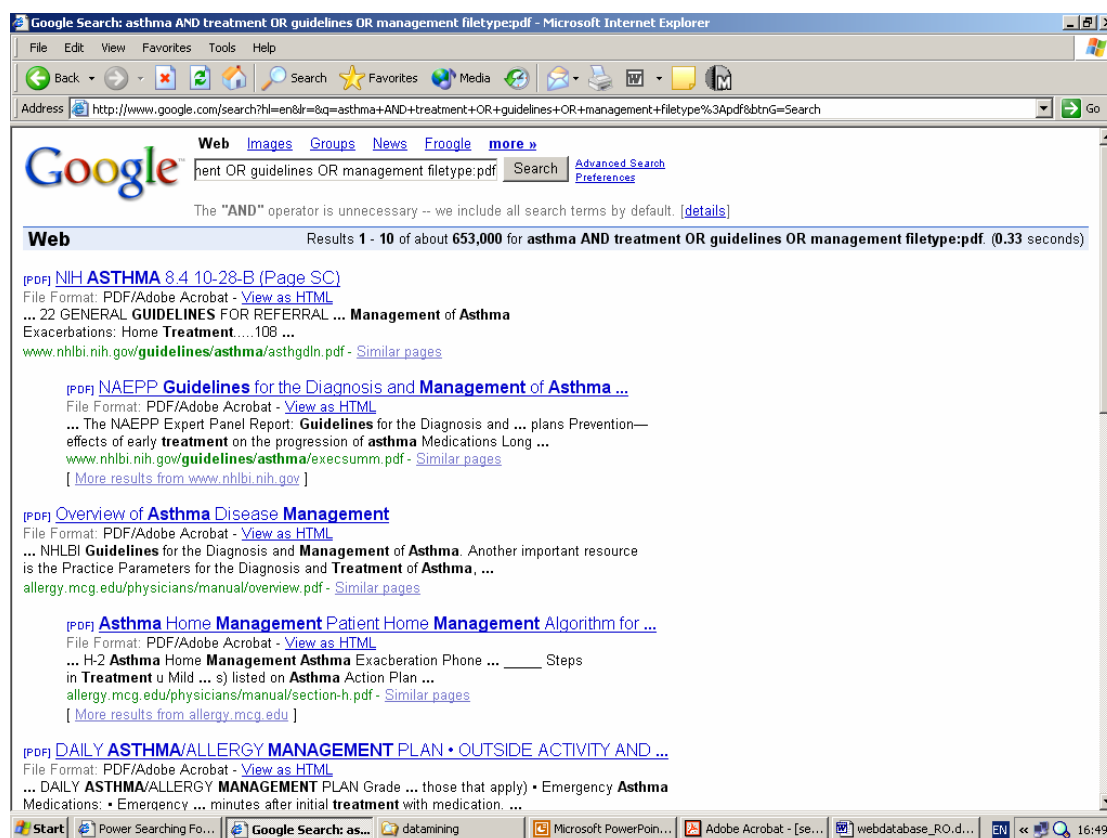


Fig. 25. Tehnici de limitare a ariei de căutare - folosirea operatorului „filetype:”

3. Analiza rezultatelor căutării

Odată ce aveți în față lista cu rezultatele căutării, chiar dacă nu sunt (foarte) relevante, nu dezarmați. Este bine să urmăriți rapid primele 10 – 30 de legături, analizând în primul rând URL-ul, apoi scurtul rezumat afișat pentru fiecare legătură. Dacă informația pare interesantă puteți încerca să vizitați primele 2-5 legături, pentru a vă lămurii dacă informația conțuită este relevantă pentru dumneavoastră.

Dacă nici una dintre acestea nu este relevantă pentru dumneavoastră, **încercați să combinați altfel termenii căutării, să încercați eventual (și) alți termeni de căutare sau, pur și simplu, să încercați folosirea unui alt motor de căutare.**

În general, regula de aur a căutării pe Web poate fi formulată astfel: *cu cât petreceți mai mult timp selectând și combinând termenii de căutare, cu atât veți petrece mai puțin timp analizând lista cu rezultate* (pentru simplul motiv că ele vor avea un grad mai ridicat de relevanță).

Atunci când analizați lista cu rezultate focalizați-vă atenția asupra adreselor (URL) – paginile cu adrese foarte similare provin probabil din cadrul aceluiași site și pot fi, deci neglijate, odată ce ați vizitat acel site.

Paginile ce conțin formulări de genul "Re: XXXXXXXX", provin cel mai frecvent din arhivele unor liste de discuții (e-mail) și pot fi neglijate de cele mai multe ori.

Cele mai folosite pagini sunt cele care conțin ele însele legături spre același subiect sau subiecte similare celui căutat.

O altă modalitate este căutarea unor pagini de tip FAQ (frequently asked questions), legate de subiectul dorit de dumneavoastră. De obicei astfel de pagini

conțin informații importante sau cel puțin o introducere în domeniul de care sunteți interesați.

De ce trebuie să evaluați paginile rezultate în urma căutării pe Web?

- în general aproape orice căutare se soldează o listă ce conține sute, mii sau zeci mii de legături.
- multe pagini nu sunt aduse la zi, riscând astfel să pierdeți mult timp “filtrând” informații deja depășite
- informația publicată pe cele mai multe situri NU este evaluată de către cineva autorizat (“evaluare peer-reviewed”)
- trebuie să aveți în vedere faptul că este extrem de simplu ca oricine să publice pagini de Internet, în acest moment neexistând nici o modalitate de a “controla” în vreun fel “calitatea” respectivei informații.

Modalități de evaluare

Înainte de a deschide o anumită pagină...

Analizați atent adresa paginii (URL):

- este vorba de un site sau de pagini personale? (dacă conține semne ca ~ sau % poate fi vorba de utilizatori sau membri ai unei organizații)
- este numele de domeniu potrivit pentru conținutul paginii respective (de exemplu edu, org, gov – dacă doriți anumite informații științifice din surse autorizate)
- este pagina publicată de o organizație credibilă de a furniza acel tip de informație (de exemplu este credibil să citiți știri de pe o pagina a sitului cnn.com)
- provine acea informație de la o sursă (organizație, agenție, instituție) credibilă în domeniul respectiv (de exemplu, este foarte probabil să găsiți informație medicală credibilă pe situl National Institute of Health – www.nih.gov, informații despre Uniunea Europeană pe paginile sitului www.europa.int etc).

După ce ați deschis o anumită pagină...

- puteți afla cine a scris-o ? (numele autorului, instituția...ați auzit vreodată de acea instituție?); o simplă adresă de e-mail a autorului NU este de regulă suficientă.
- există pe sit dovezi asupra pertinentei informației publicate, respectiv a competenței în domeniu a celui ce a publicat-o ? (căutați pe sit legături de tipul “About us” “Philosophy” “Background” “Biography”)
- este informația de pe pagina respectivă destul de la zi ? (căutați să găsiți în cadrul paginii o informație de genul “last updated: 12.11.2004”

“Indicatori” ai calității unei pagini Web:

- surse bibliografice bine documentate (legături, note de subsol, referințe bibliografice)
- sunt cele de mai sus similare cu cele găsite în publicații tipărite ?
- aigurați-vă că informația găsită nu a fost modificată sau chiar falsificată (există mențiunea unei versiuni tipărite a documentului ? poate fi ea obținută ?
- existența unor legături spre pagini similare

Ce spun alții despre acea pagină (acea legătură) ?

Din fericire, unele motoare de căutare, cum ar fi Google, ne pun la dispoziție o unealtă extrem de eficace pentru a afla câte legături există spre pagina pe care tocmai o analizăm (cu alte cuvinte câte referiri sunt că alții consideră pagina respectivă ca având un conținut suficient de credibil pentru a fi citat). Unealta menționată este comanda **link:**

De exemplu pentru căutarea

link:www.nih.gov

Google va returna peste 59000 de referiri, față de numai 82 în cazul Ministerului Sănătății din România.

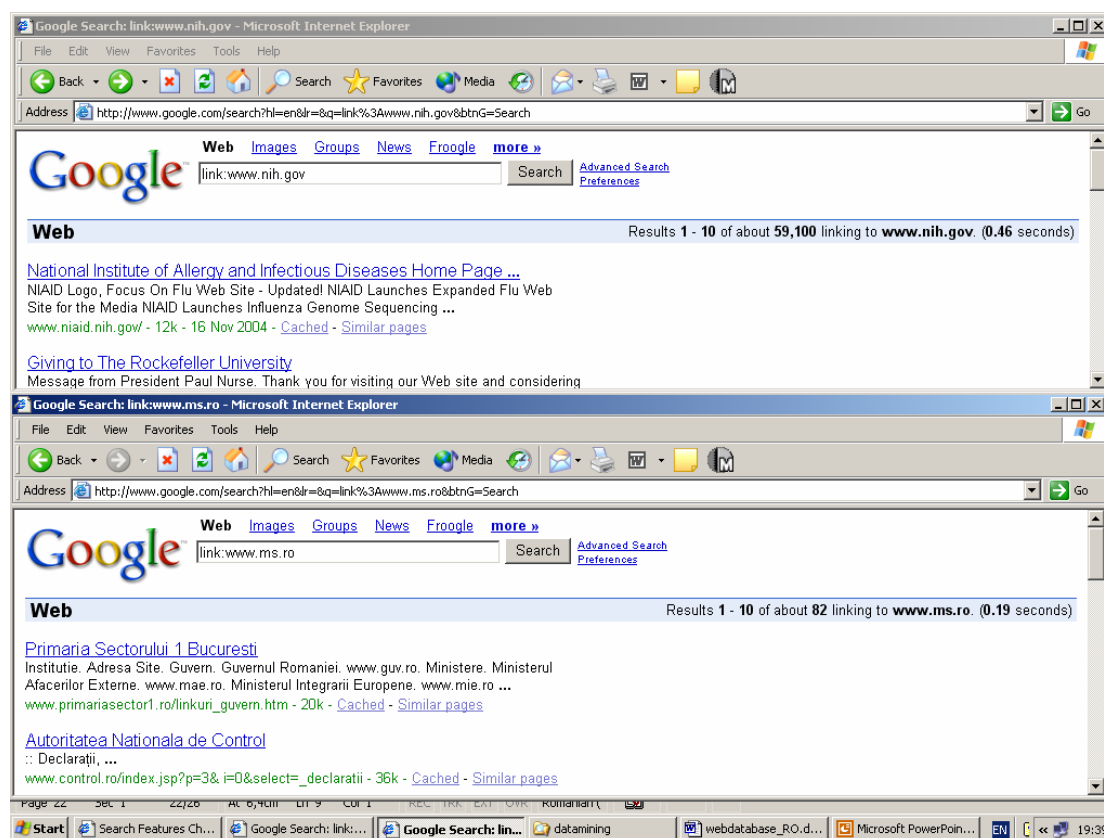


Fig. 26. Evaluarea rezultatelor căutării prin folosirea operatorului „link:”

Întrebați-vă care ar putea fi scopul pentru care a fost publicată această pagină ...

- pentru informarea cu date exacte și fapte a publicului asupra unui anumit subiect ?
- pentru a lămuri unele probleme ?
- pentru a vinde ceva ?
- pentru a discredita o altă sursă de informare ?
- pentru a parodia pe cineva sau ceva... ?

Întrebarea finală pe care trebuie să o puneți este următoarea : îmi este utilă informația pe care am găsit-o în acea pagină ? Servește ea scopului meu ?

4. Continuarea căutării

Așa cum am spus anterior, dacă nici una dintre legături nu este relevantă pentru dumneavoastră, încercați să combinați altfel termenii căutării, să încercați eventual (și) alți termeni de căutare sau, pur și simplu, să încercați folosirea unui alt motor de căutare.

Dacă doriți să schimbați termenii de căutare și sunteți în pană de inspirație în legătură cu identificarea unor noi termeni ce ar fi potriviți, o sugestie bună este să examinați conținutul unei pagini Web la care aveți acces și care conține informația dorită pentru a vedea termenii folosiți (eventual sinonime, fraze etc). Această manevră ar putea fi utilă în identificarea unor noi termeni pentru căutare.

Dacă **lista cu rezultate** este **prea extinsă**, este indicat să **repețați căutarea** folosind **mai mulți termeni** combinați prin operatorul **AND**.

Dacă lista cu rezultate este prea săracă, renunțați la termenii mai puțin importanți și/sau schimbați modul de combinare a lor (modul în care folosiți AND/OR).

Dacă în listă regăsiți foarte multe legături ce conțin informații numai despre anumiți termeni de căutare, în timp ce alți termeni lipsesc din rezultate, aceasta înseamnă că respectivii termeni au prea multă greutate (term-weight) și este indicat să renunțați la o serie de sinonime ale lor. Atenție însă la faptul că prea puțini termeni de căutare folosiți pot scădea specificitatea înseamnă de cele mai multe ori lipsă de specificitate, respectiv un număr mare de legături în lista cu rezultate, de cele mai multe ori lipsite de relevanță.

Trebuie remarcat faptul că anumite motoare (Google de exemplu) permit, odată cu afisarea listei cu rezultate, alegerea unei opțiuni de genul „pagini similare” („similar pages”).

Prin folosirea acestei facilități, puteți restrânge rapid căutarea la un grup relevant de documente. Această opțiune este extrem de utilă dacă aveți probleme cu identificarea unor sinonime ale termenilor căutării. Odată ce ați identificat o pagină cu informații utile, puteți găsi cu ușurință altele similare.

Concluzii

Motoarele de căutare sunt probabil cea mai utilizată cale de accesare a informației stocate în bazele de date Web. Principalii pași necesari pentru a utiliza eficient un motor de căutare sunt următorii:

1. Formulați-vă atent nevoia de informare. Încercați să faceți aceasta cât mai simplu și specific cu putință
2. “Spargeți” cererea de mai sus într-o serie de termeni de căutare
3. Găsiți sinonime pentru fiecare termen identificat
4. Atunci când este posibil, recurgeți la fraze (termeni specificați în ghilimele)
5. Combinați sinonimele termenilor de căutare prin intermediul operatorului OR și legați aceste combinații prin intermediul operatorului AND
6. Analizați circa 30 de rezultate înainte de a căuta din nou
7. Dacă este posibil, înainte de a reîncepe căutarea, folosiți pagini cu informație similară celei căutate pentru a identifica noi termeni de căutare și sinonime ale acestora

V. UTILIZAREA SERVICIILOR DE DIRECTOARE WEB

În literatura românescă de specialitate sunt cunoscute și sub numele de **anuale online** sau **repertoare tematice valide**.

Așa cum am spus, serviciile de directoare Web constituie o veritabilă bază de date cu documente clasate pe categorii și subcategorii. Indexate manual sau automat (mai rar), ele sunt foarte mult utilizate pentru căutări simple sau complexe. Un anuar de căutare tematic ghidează utilizatorul printre temele disponibile, înainte de a alege contextul cercetării finale. Contrar motoarelor de căutare, serviciile de directoare Web indexează de obicei site-uri complete, începând de la pagina de primire (Home page).

Această modalitate de lucru este pusă la dispoziție de servicii de căutare extrem de cunoscute, cum ar fi Yahoo sau Magellan. Au avantajul că pun la dispoziția utilizatorului un set de resurse existente despre un anumit subiect colectate, indexate și în același timp, extrem de bine organizate. Astfel dacă un subiect de interes pentru utilizator există deja în arborescența de directoare a serviciului, este foarte probabil ca această cale să fie și cea mai rapidă pentru a găsi informații cu grad ridicat de relevanță.

Utilizarea acestui serviciu este extrem de simplă: utilizatorul nu are decât să examineze lista de opțiuni arborescente pusă la dispoziție și să selecteze domeniul cel mai apropiat de nevoia lui de informare. Prin alegerea acestei opțiuni (legături) va trece apoi la următorul nivel de detaliu al serviciului de directoare unde va avea posibilitatea să selecteze un domeniu și mai apropiat de ceea ce caută. Utilizatorul va putea să continue în acest fel, pe nivele din ce în ce mai ridicate de detaliu, până când va găsi o serie de legături cu informație utilă, fie va ajunge la o „fundătură”, adică o ramură a arborelui de directoare ce nu conține informație relevantă pentru el.

Cu ajutorul butonului "BACK" al browserului poate reveni oricând la un nivel superior al arborelui de directoare pentru a relua căutarea spre o altă zonă a repertoriului tematic.

Căutarea informației în respectivul director poate fi considerată încheiată atunci când utilizatorul s-a întors la meniul de pornire și nu mai are la dispoziție nici o legătură neexplorată deja.

Dacă această modalitate de căutare nu vă satisface, majoritatea directoarelor Web pun la dispoziție și o casuță de căutare rapidă (type-in-box).

Căutarea rapidă în directoarele Web

Trebuie spus că această căutare NU este similară celei folosite pentru motoarele de căutare descrise anterior. *Căutarea se produce într-o colecție mult mai mică de date* (de obicei structura directoarelor și descrierea acestora) și, de aceea, căutarea se face oarecum diferit.

- puteți folosi căutarea rapidă în primul rând pentru a afla dacă există vreun director cu informații despre subiectul de interes pentru dumneavoastră.
- căutarea trebuie să fie cât mai simplă cu putință. De obicei un singur termen ce reflectă cel mai bine subiectul dorit este suficient. Dacă nu găsiți nici un rezultat, încercați un sinonim al termenului anterior.

- în general, nu este recomandată folosirea termenilor multipli sau a combinațiilor de termeni; un director Web este ceva similar unui repertoriu de tip Pagini Aurii – vă gândiți la un domeniu și deschideți la pagina ce conține informații din domeniul respectiv. Dacă nu găsiți nimic, încercați alt termen.
- Folosirea mai multor termeni sau a unor combinații nu duce decât la o listă confuză de rezultate, cu grad scăzut de relevanță. În plus, introduceți termenii de căutare în forma cea mai simplă cu putință și evitați folosirea operatorilor AND, OR sau a frazelor (chiar și atunci când serviciul folosit o permite).

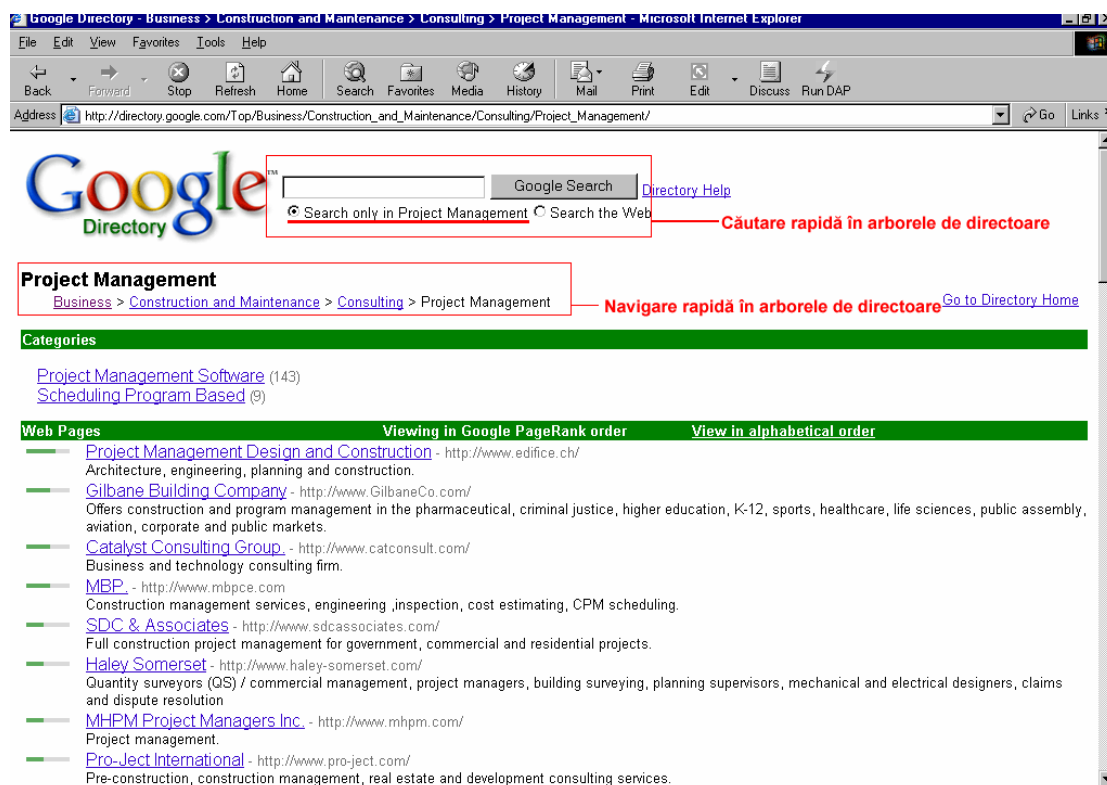


Fig. 27. Modalități de utilizare a serviciilor de directorate Web

Dezavantaje ale directoratelor Web

Deși sunt simplu de utilizat și includ și facilități de căutare, directoratele Web au (cel puțin în acest moment al evoluției lor) câteva dezavantaje ce trebuie subliniate:

- indexarea datelor în directoratele web se face de obicei manual (sub coordonarea unui editor) și nu automat, ca în cazul motoarelor de căutare. Intrările noi se actualizează greu, putând dura chiar luni până să apară în structura de directorate.
- dacă nu ați “ghicit” termenul folosit de editor pentru pagina (tema, subiectul) respectiv, aceasta va fi foarte greu de găsit.
- paginile ce tratează subiecte multiple pot fi regăsite frecvent doar într-un singur arbore de directorate, conform subiectului considerat ca fiind principal, celelalte subiecte fiind în consecință cvasi-ignorate.
- bazele de date sunt semnificativ mai mici decât în cazul motoarelor de căutare (zeci sau sute de mii de pagini pentru directoratele web, yeci de milioane de pagini pentru motoarele de căutare).

Din aceste motive, pentru orice căutare exhaustivă, în special asupra unor subiecte ce nu sunt foarte cunoscute (mediatizate), este bine să se facă cu ajutorul unor motoare de căutare importante, în detrimentul directoarelor Web.

CONCLUZII:

Directoarele Web sunt locuri excelente de pornire în căutarea informației pe Internet, cu condiția să găsiți un arbore de directoare care să trateze tema căutată (subiectul dorit).

Dacă însă tema căutată nu a fost selecționată de editorul serviciului ca „rădăcină” pentru un arbore de directoare, este de preferat să renunțați la folosirea acestui tip de serviciu de căutare și să recurgeți la un motor de căutare.

Ca regulă generală este recomandat să vă începeți căutarea informației pe Internet folosind un arbore de directoare și să recurgeți abia apoi, dacă este necesar, la un motor de căutare, ce dispune de mult mai multă informație indexată în bazele de date, dar necesită o experiență mult mai mare din partea utilizatorului, pentru a putea fi utilizat eficient.

VI. CRITERII DE SELECȚIE A BAZELOR DE DATE WEB

Este evident că bazele de date Web folosite de serviciile de căutare a informației pe Internet NU sunt la fel. Ele diferă una de cealaltă prin următoarele aspecte:

Caracteristici	Observații
Mărimea bazei de date folosite	<i>Directoarele Web</i> folosesc în general baze de date mai mici
	<i>Bazele de date web de mărime medie</i> pot conține pagini pe care altele mai mari le-au scăpat din atenție
	Este greu de căutat eficient în <i>bazele de date Web de mari dimensiuni</i>
Tipul paginilor indexate	<i>Toate tipurile de pagini</i> – baze de date Web de de “uz general”
	<i>Pagini “populare”</i> – servicii utile pentru a găsi informații „fierbinți” (știri, meteo etc)
	<i>Pagini cu informație “țintită”, de înaltă calitate</i> – Servicii utile pentru mediul academic și științific
Metoda de indexare	<i>Full-text</i> – Căutare după termeni specifici (combinație de termeni) - motoare de căutare
	<i>Indexare “manuală”</i> – în servicii de directoare Web
Modul de acces al utilizatorilor la serviciu	<i>Servicii de directoare Web</i> – indicate ca punct de pornire pentru căutare
	<i>Motoare de căutare</i> – mai performante, conțin mai multă informație indexată, mai dificil de folosit
Capabilitățile motorului de căutare	<i>Operatori logici</i> – unele motoare de căutare Nu permit utilizarea acestora
	<i>Operatori pseudo-booleeni</i> – cele mai multe motoare permit folosirea acestei metode de căutare folosind „limbajul natural”
	<i>Căutarea după fraze</i> – una dintre cele mai performante metode de căutare
	<i>Creșterea importanței termenilor de căutare (Term weighting)</i> – dacă poate fi folosită crește semnificativ relevanța rezultatelor
	<i>Trunchierea termenilor</i> – cele mai multe motoare o permit
Prezentarea rezultatelor	<i>Relevanța informației</i> – unele servicii returnează numărul de termeni regăsit în fiecare legătură și calculează un “scor” privitor la relevanța legăturii
	<i>Rezumate</i> – pot diferi ca întindere, în funcție de serviciul de căutare
	<i>Continuarea căutării</i> – unele motoare oferă o opțiune foarte utilă: „pagini similare”

Așa cum se poate observa, există o multitudine de criterii ce caracterizează o bază de date Web, astfel că anumite nevoi de informare necesită folosirea unui anumit tip de bază de date Web, respectiv a unui anumit tip de serviciu de căutare.

În principiu, nu este recomandată folosirea aceluiași serviciu de căutare pentru toate tipurile de căutare pe care le faceți, tot așa cum nu folosiți un singur dicționar pentru orice tip de traducere.

Sopul acestui capitol introductiv este de a vă ajuta să decideți care este „unealta” cea mai potrivită pentru a căuta ceea ce doriți, respectiv de vă însuși modalitățile de bază pentru a folosi eficient serviciul de căutare ales.

BIBLIOGRAFIE

1. **Danny Sullivan** - *Power Searching For Anyone*– Online book, <http://searchenginewatch.com/facts/article.php/2156031>, 2001
2. **Debbie Flanagan** – *Web Search Strategies* - Online book, <http://www.learnwebskills.com/search/main.html>, 2004
3. ******* - *Google Guide* – Online book, <http://www.googleguide.com>
4. **E. Coiera** - *Guide to Medical Informatics, the Internet and Telemedicine*, Chapman & Hall, London, (1997).
5. **E. Coiera** - *Information Economics and the Internet*, Journal American Medical Informatics Association, 7, 215-221, 2000
6. **Jalobeanu M.** - *"WWW în învățământ: Instruirea prin Internet, Cum căutăm și Cum publicăm pe Web"*, Ed. CCD, Cluj-Napoca, 2001