

PRETRAŽIVANJE INFORMACIJA NA INTERNETU: kako odabrati pravi pretraživač?



*mr. sc. Jadranka Stojanovski
Knjižnica Instituta Ruđer Bošković*



o čemu će biti riječi

- “pogreška” u naslovu: što je Internet, a što Web
- “vidljivi” i “nevidljivi” Web
- veličina i druge karakteristike Web-a
- kratka povijest kratke povijesti pretraživača
- proces pretraživanja
- značajke dobrog pretraživača
- pregled i usporedba nekoliko pretraživača
- metapretraživači
- budućnost



Koja je razlika između Interneta i Web-a?

- Internet - globalna mreža miliona računala povezanih mrežama s TCP/IP protokolom (1999 dosegnut broj od 200 M korisnika iz više od 100 zemalja)
- World Wide Web - sistem poslužioca koji podržavaju posebno formatirane dokumente (HTML) koji sadrže veze prema drugim dokumentima, kao i grafiku, zvuk i video
- nisu svi Internet poslužiocci dio Web-a



Definiranje Web-a

- statične web stranice “ručno” rađene (Web dizajner) i stavljene na Web poslužioc kako bi bile dostupne svima koji posjećuju poslužilac;
- dinamičke Web stranice stvara računalo koristeći skripte (CGI, Java, Perl) - skripte su posrednik između zahtjeva korisnika (statična web stranica) - “front end” i baze podataka - “back end”



Statičke Web stranice

“vidljivi” Web

npr. <http://lib.irb.hr>

- ručno proizvedene
- generičke informacije
- većina se može indeksirati
- $\sim 10^9$ Web stranica danas

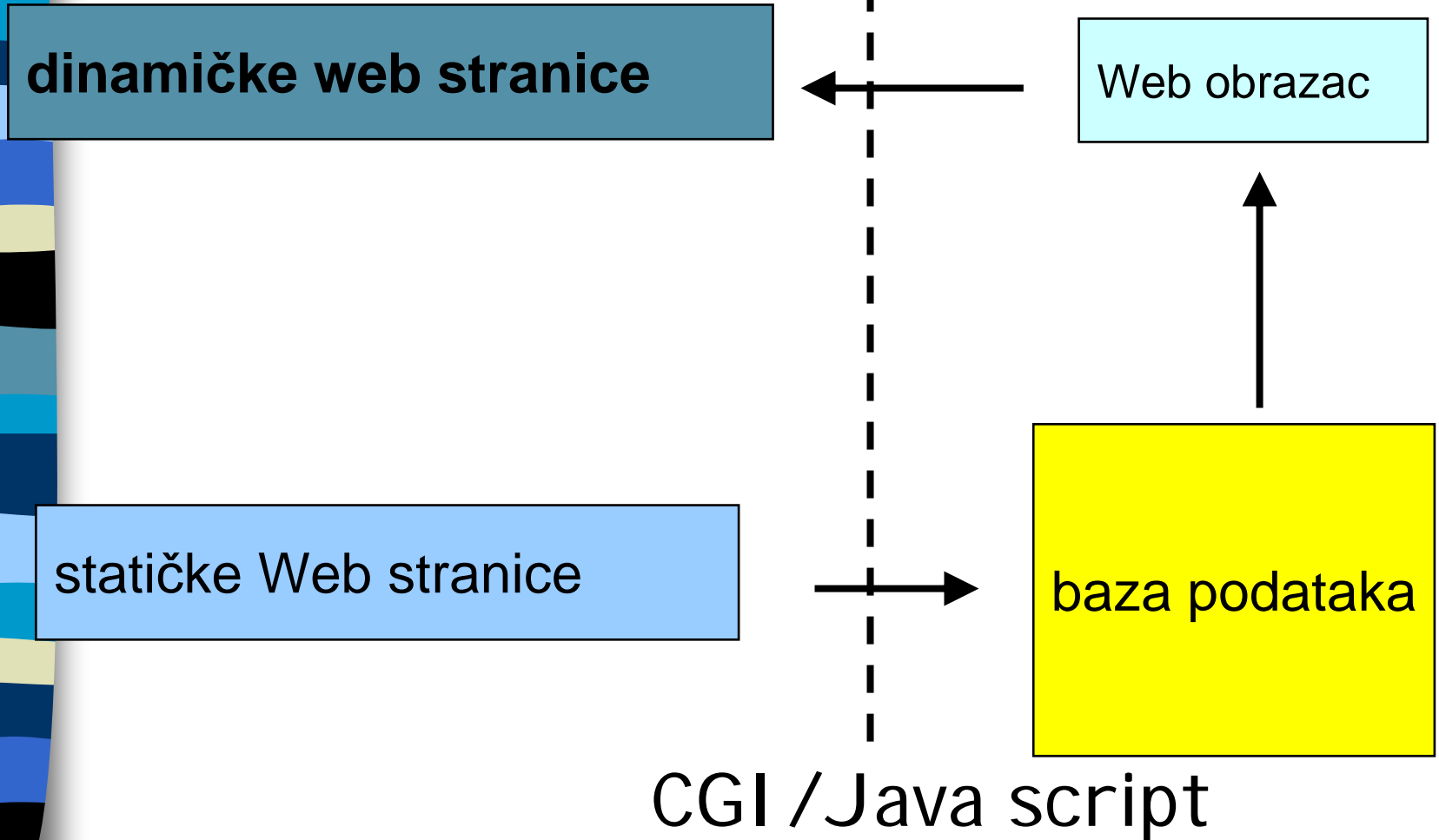
Dinamičke Web stranice

“nevidljivi” Web

npr. <http://bib.irb.hr>

- računalno proizvedene
- prilagodljive informacije
- ne mogu se indeksirati
- još nije provedeno znanstveno istraživanje koje bi im odredilo veličinu

Kako se stvaraju dinamičke stranice?





Veličina “vidljivog” Web-a

prosinač 1997

320 M Web stranica

veljača 1999

800 M Web stranica (preko 6
Tb tekstualnih podataka)

siječanj 2000

1G Web stranica

povijest pretraživača Web-a

- tko se još sjeća Archie-ja i Veronice?
- 1993. g. postoje dva Web pretraživača: ALIWEB i World Wide Web Wanderer (prvi Web robot) - slabo indeksiranje, sporo pretraživanje, linearne baze
- **WebCrawler** - april 1994. - prvi uspješan pretraživač -prvi indeksirao cijeli tekst dokumenta (Univ. Washington)
- slijede **OpenText, Lycos** (Carnegie Mellon Univ.) i dr.
- **KOMERCIJALIZACIJA** - 95. g. većina poslužioaca su akademski, vojni ili vladini, a već 1996. g. dvije trećine Interneta je komercijalno
- **Infoseek** prvi prepoznaje komercijalnu stranu Interneta (1996.)



Alta Vista - kraj 1995 - prva omogućila primjenu Booleovih operatora, nametnula se kao standard drugim pretraživačima

Excite

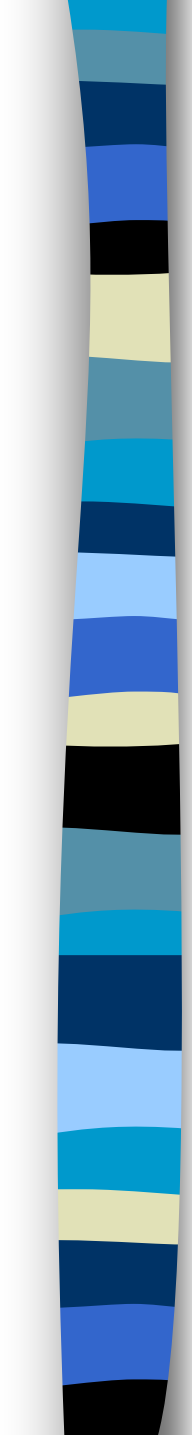
HotBoot (1996.) - naprednije sučelje - pojednostavljivanje Booleovog pretraživanja korištenjem izbornika, moguće pretraživanje prema zemljopisnoj lokaciji

KONSOLIDACIJA - 1997. pojavljuje se **Northern Light** - ne financira se od reklama, kao svi drugi, nego ima potpuno novi pristup

nude se “specijalne zbirke”

pokriva jednako ili čak veće područje nego do tada vodeći pretraživači (Alta Vista)

mogućnost razlikovanja filmova o Jamesu Bondu i kemijskih veza (chemical bond)

- 
- “Meta-pretraživanje” i alati za predmetno pretraživanje - meta pretraživači omogućavaju istovremeno pretraživanje nekoliko pretraživača
 - prvi meta pretraživači pojavili su se 1995. (**MetaCrawler, SavvySearch, Inference Find**)
 - komercijalizacija ih je potisnula u drugi plan
 - “Business model” - polako većina pretraživača prelaze u taj model



“Portalizacija” i personalizacija

- portali po svojoj namjeni žele biti **početna točka** s koje će korisnik pristupati mreži
- prvi portali pojavili su se 1994 , a nudili su ih
 - Internet provider-i (AOL),
 - pretraživači (search sites)
 - preglednici (browsers) - Netscape i Netcenter, te Microsoft i MSN
- nudeći najrazličitije usluge, kao što su elektronička pošta, besplatni prostor za web stranice, forume, pretraživače, online trgovinu i dr., zapravo se manipulira korisnikom



Lawrence i Giles, Nature, July 1999

Količina informacija

- “vidljivi” Web procijenjen je na **2.8 M poslužioca** (ukupno 16 M) - 2500 nasumično odabranih je analizirano
- Web je procijenjen na **8 M stranica**
- **6 Tb tekstualnih podataka**, 3 Tb slika (ukupno 15 Tb)
- 83% Weba je komercijalnog, **6% znanstvenog i edukativnog**, a 1.5% pornografskog sadržaja
- samo 34% poslužioca koristi META tag-ove (0.3% sadrži metapodatke Dublin Core-a) - nedostatak standardizacije!

Lycos Top Search Terms: Jan. 16-22, 2000

- 
1. Sex *
 2. Pokeman
 3. MP3 *
 4. Weather
 5. Chat *
 6. Pokemon*
 7. Britney Spears
 8. Games *
 9. Maps
 10. Music
 11. Jokes
 12. Jobs
 13. Lyrics
 14. Horoscopes
 15. Pictures
 16. Chat Rooms
 17. Cars
 18. Lunar Eclipse
 19. Wallpaper
 20. Warez *
 21. Movies
 22. American Music Awards
 23. Horoscope
 24. Lingerie
 25. Jennifer Lopez*
 26. Dogs
 27. Travel
 28. Backstreet Boys
 29. Greeting Cards
 30. Search Engines (Yahoo*)
 31. Airlines
 32. Wrestling *
 33. WWF
 34. Pamela Anderson
 35. Screensavers
 36. ICQ
 37. Dictionary
 38. Furniture
 39. Mariah Carey
 40. Real Estate
 41. Data Recovery
 42. Personals
 43. Screen Savers
 44. SNES Roms
 45. Online Movie Review
 46. Winzip
 47. Song Lyrics
 48. Dragon Ball Z
 49. Bikini
 50. Free Stuff



neki Inktomi WebMap podaci

- broj hostova 6.400
- broj mirror-a 1.450
- broj site-ova 4.950

- .com 54.68%
- .net 7.82%
- .edu 6.69%
- .org 4.35%
- .gov 1.15%
- .mil 0.17%

najučestalije ime stranice: **contact.html**

najdulje ime stranice:

<http://www.tax.taxadvice.taxation.irs.taxservices.taxrepresentation.taxpayerhelp.internalrevenueservice.audit.taxes.com>

- postotak dokumenata na engleskom jeziku 86.55%
- postotak dokumenata na francuskom jeziku 2.36%
- postotak dokumenata na nizozemskom j. 0.54%

Lawrence i Giles , Nature, July 1999

Pretraživači

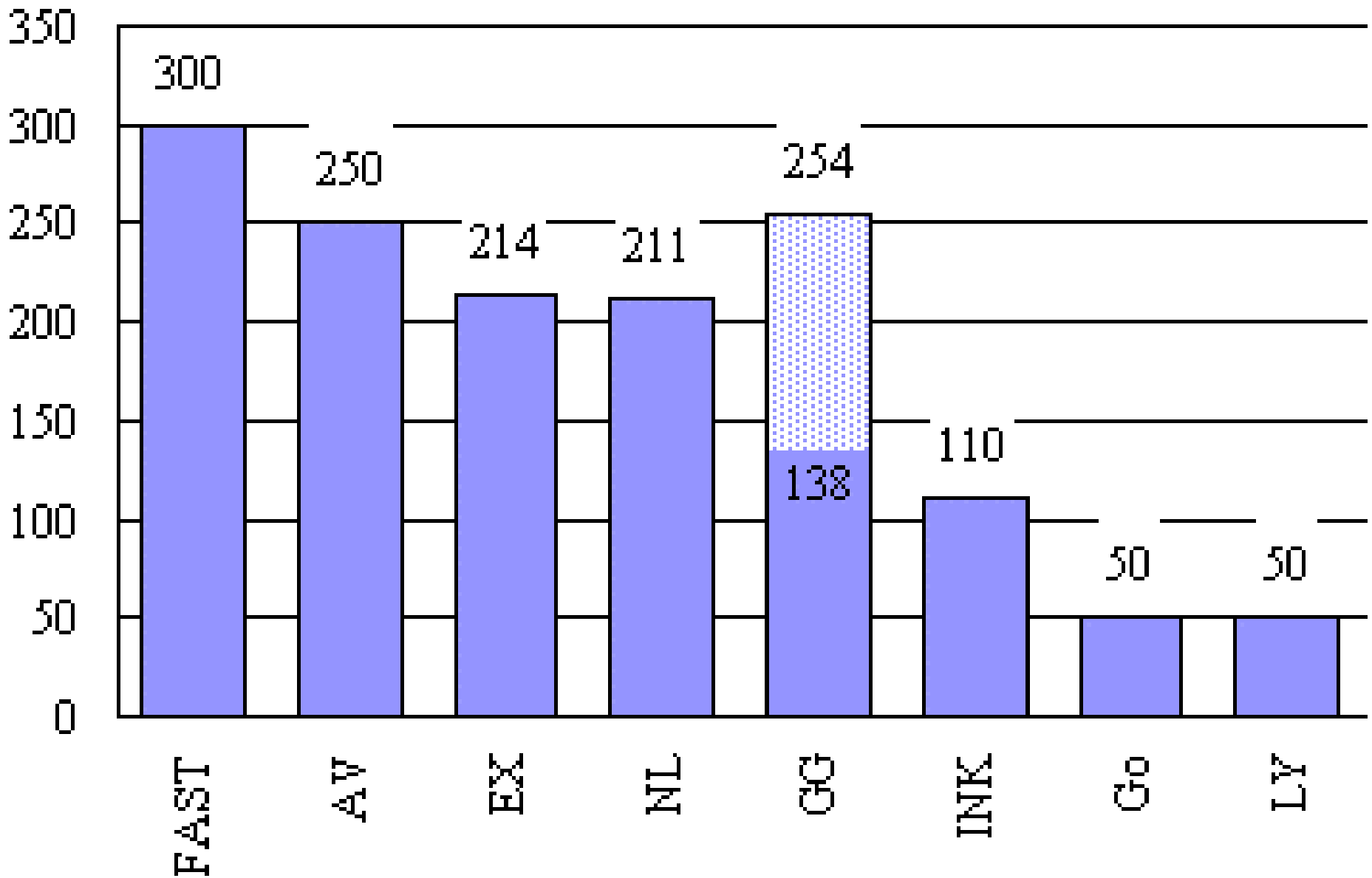
- 11 najvažnijih pretraživača: AltaVista, EuroSeek, Excite, Google, HotBot, Infoseek, Lycos, Microsoft, Northern Light, Snap i Yahoo
- nijedan od pretraživača **ne indeksira više od 16% Web-a:**
NL 16%, AV 15.5%, Snap 15.5%, HotBot 11.3%, MSN 8.5%, Infoseek 8%, Google 7.8%, Yahoo 7.4%, Excite 5.6%, Lycos 2.5%, EuroSeek 2.6%, HotBot 2.2%
- bolje su indeksirani “popularni” poslužiocci, isto tako US poslužiocci, kao i komercijalni
- indeksiranje novih i promijenjenih stranica traje mjesecima (186 dana u prosjeku!)



Lawrence i Giles , Nature, July 1999

- prema tom istraživanju Web najbolje “pokriva” Northern Light, slijede Snap i Alta Vista
- svih 11 pretraživača pokriva oko **42%** Web-a
- prekrivanje među pretraživačima je slabo - metapretraživači!
- pokrivenost Weba mogla bi se popraviti dodavanjem novih snažnijih računala i poboljšavanjem algoritama
- veličina baze nije najvažnija za kvalitetno pretraživanje, važnije je kvalitetno sortiranje po relevantnosti

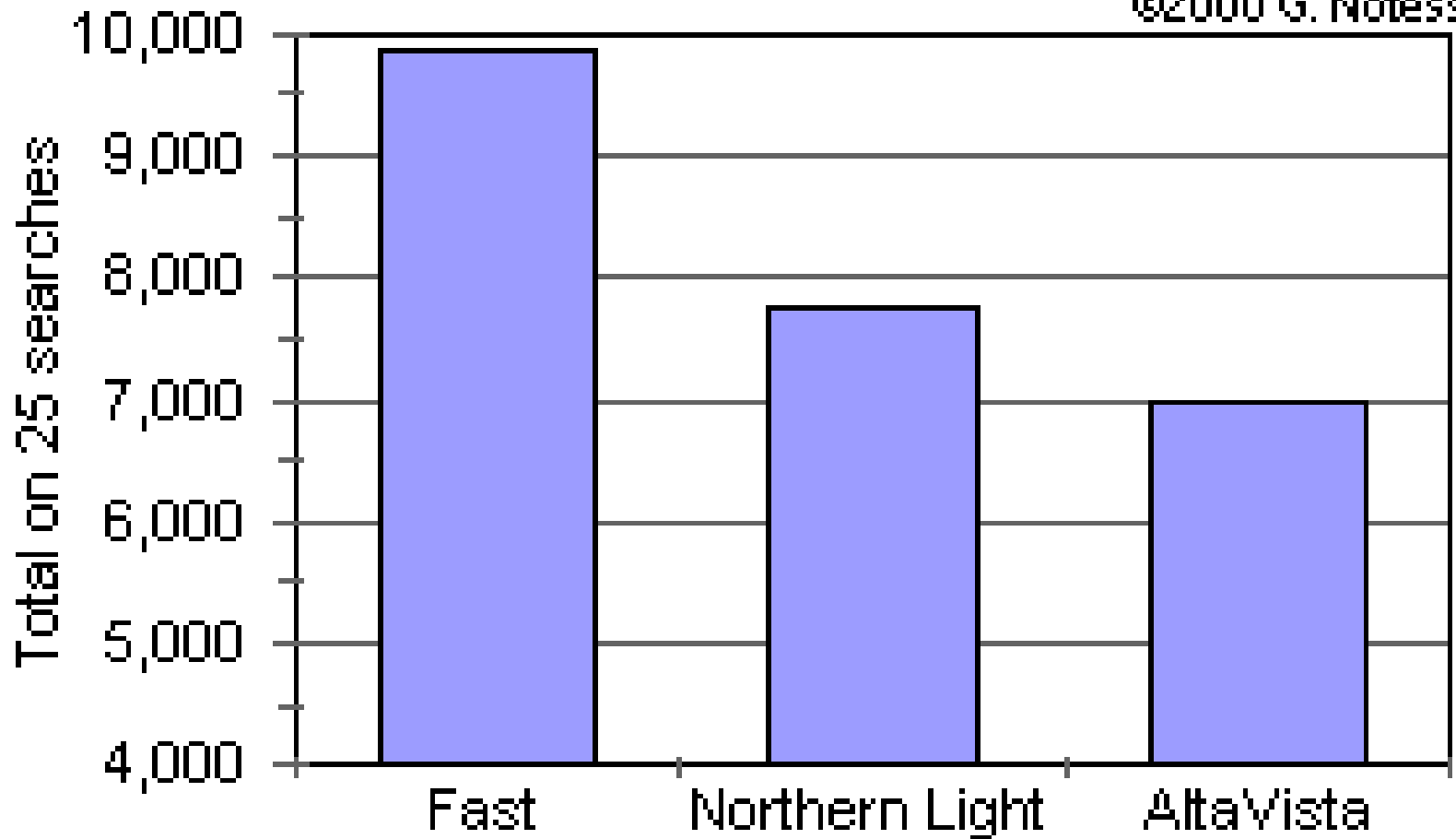
Millions of Web Pages Indexed



Three Largest Search Engines

Compared Feb. 21, 2000

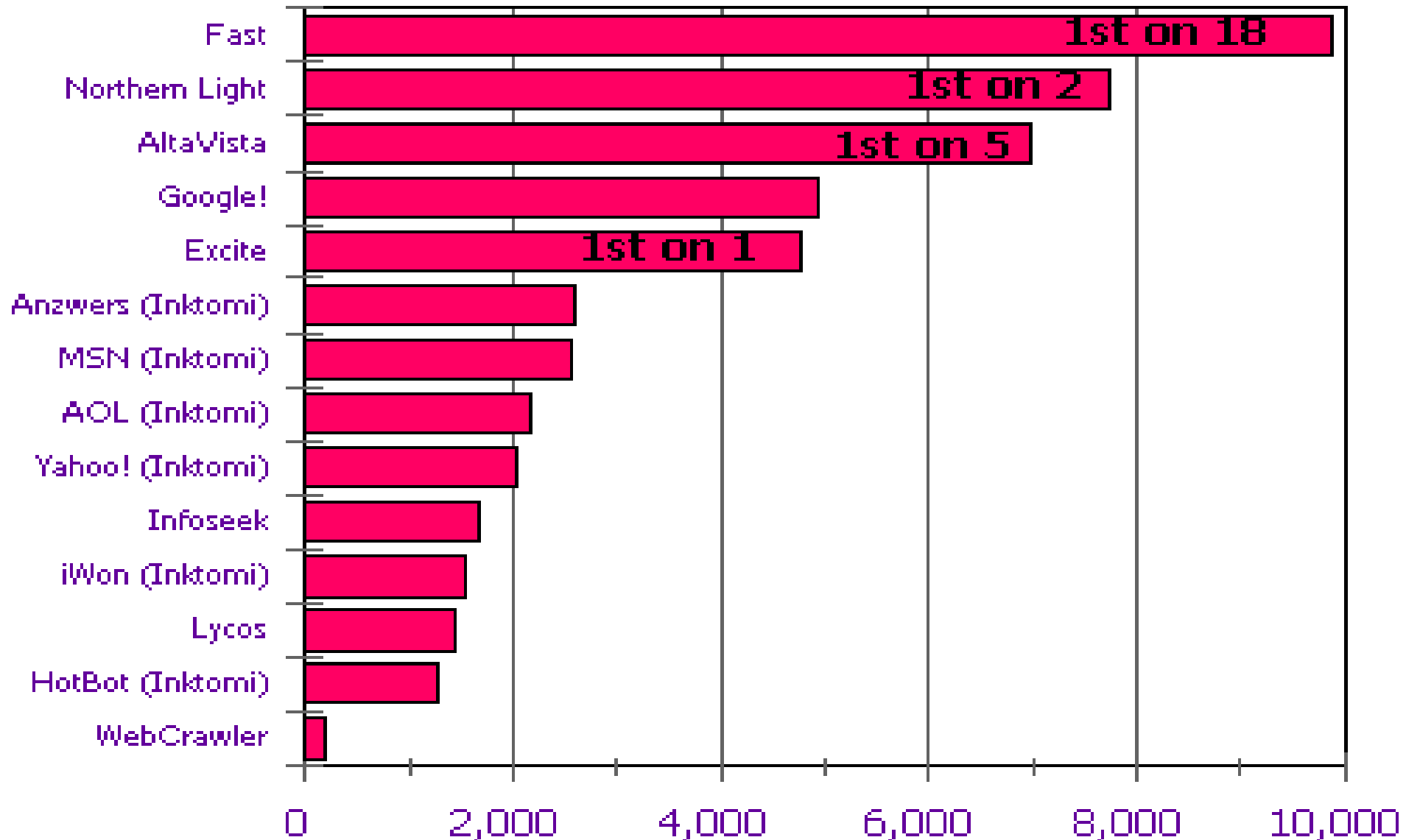
©2000 G. Notess



Total Hits from 25 Searches

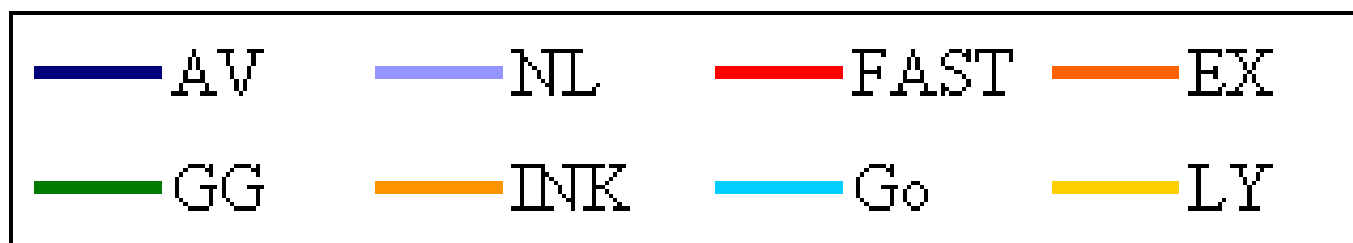
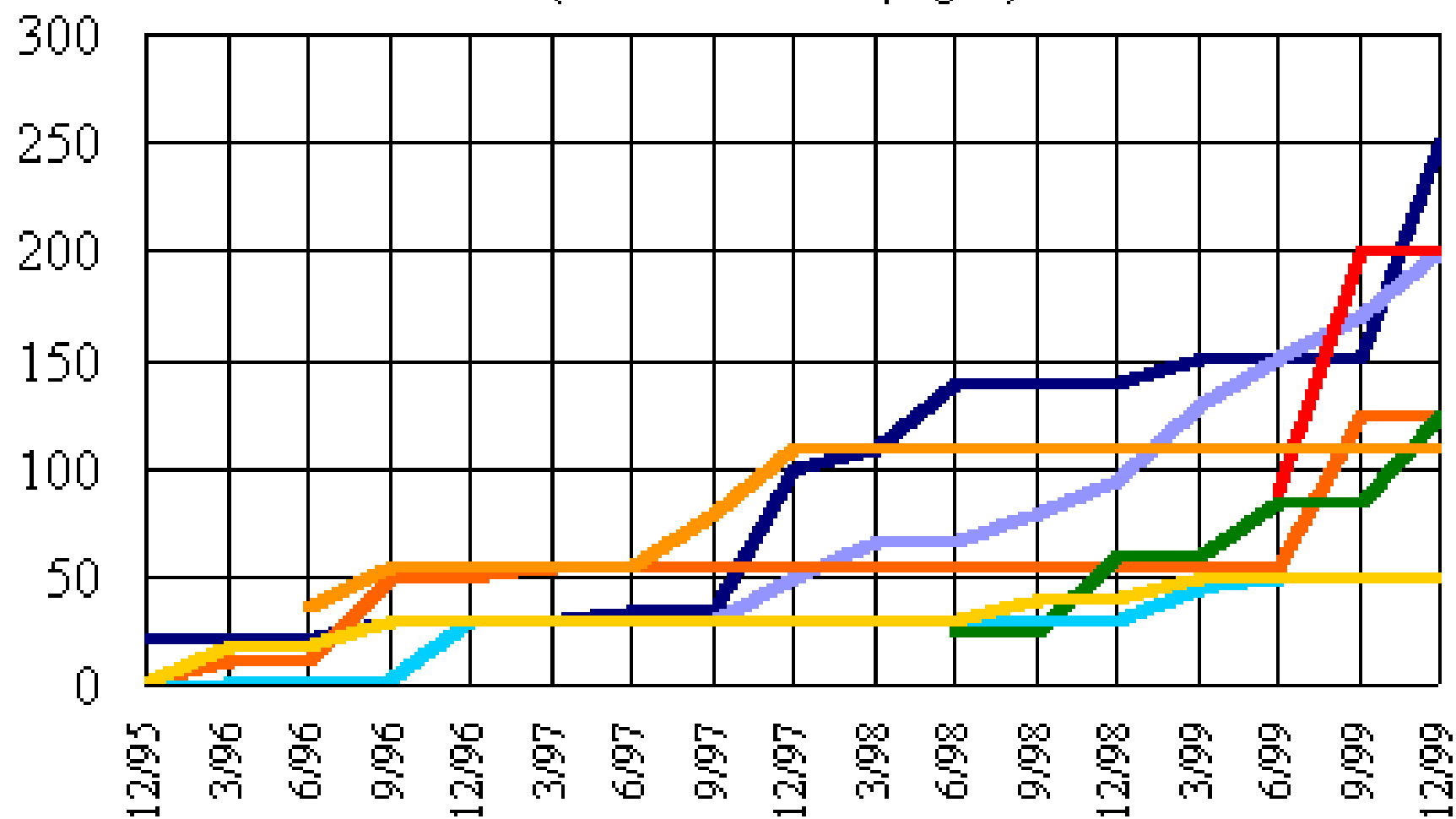
Feb. 21, 2000

©2000 G. Notess



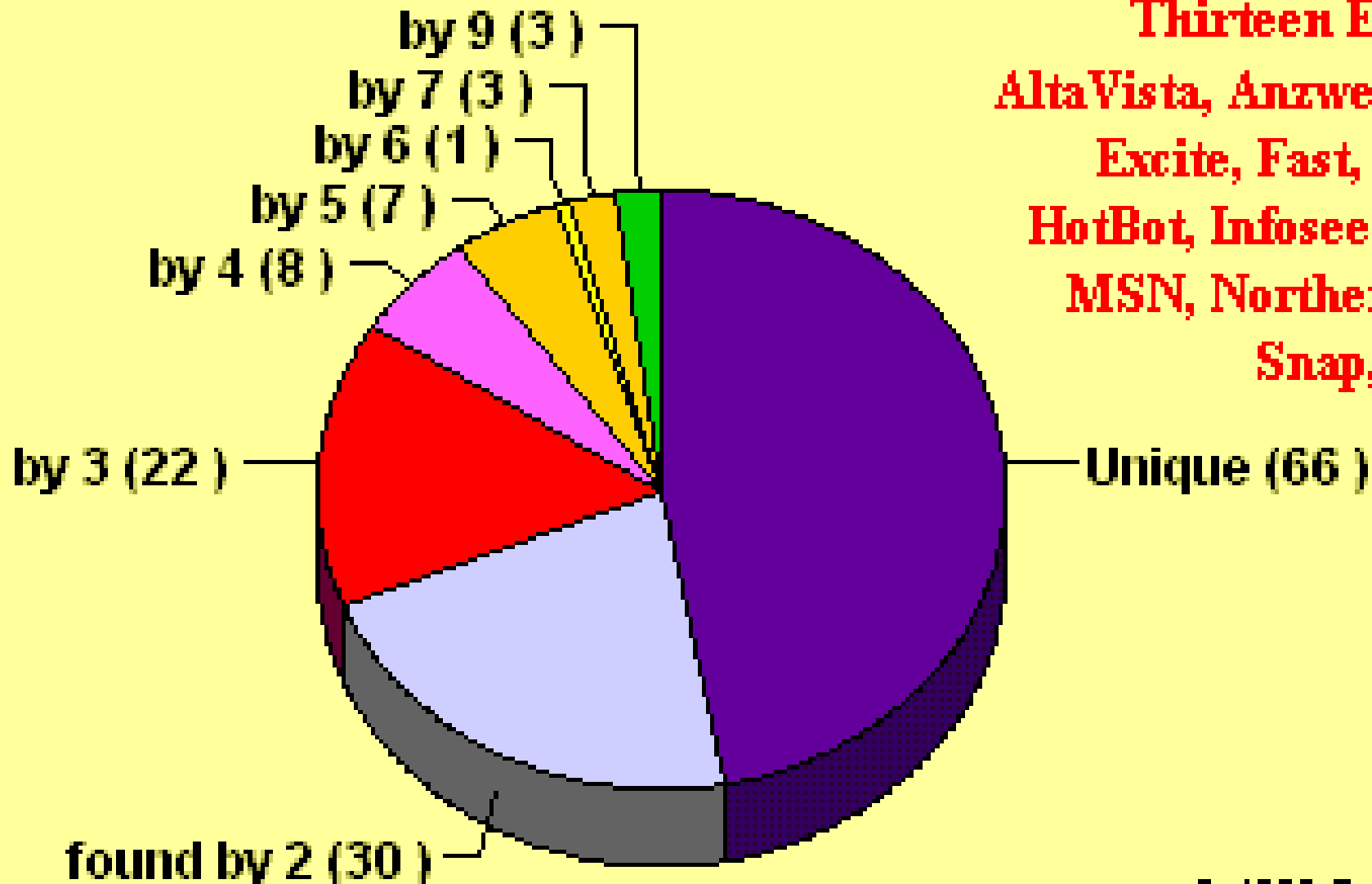
Search Engine Sizes

(millions of web pages)



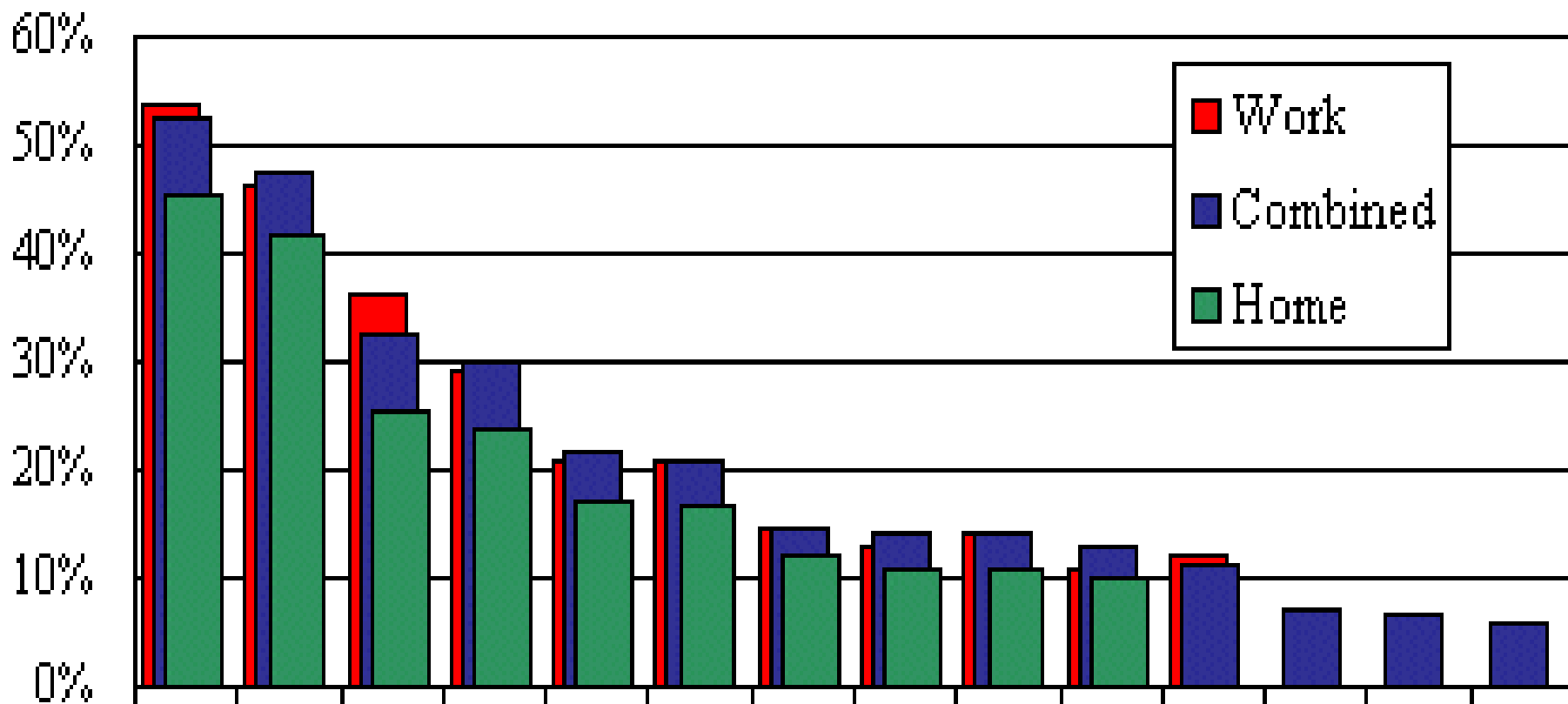
Overlap of Five Searches

142 Total Unique Hits on 9/10/99



Thirteen Engines:
AltaVista, Anzwers, AOL
Excite, Fast, Google!
HotBot, Infoseek, Lycos
MSN, Northern Light
Snap, Yahoo!

Audience Reach - Oct. 99



	YH	MS N	NS	GO	LY	EX	AV	SP	HB	LS	GT	AJ	DH	G2 N
Work	53.9	46.1	36.1	29.0	21.0	20.8	14.6	12.9	14.1	10.9	11.9			
Combined	52.6	47.4	32.5	29.8	21.8	21.0	14.7	14.0	14.1	13.1	11.3	7.2%	6.5%	5.9%
Home	45.6	41.8	25.3	23.7	17.2	16.6	12.1	10.8	10.8	9.9%				



Razlike između pretraživača i direktorija

Tematski pretinci (subject directories)

- zadana lista Web poslužioca
- sastavljeni i uređeni od strane ljudi
- kategorizirani prema predmetu/području
- selektivni

Pretraživači (search engines)

- “spider” koji istražuje mrežu
- indeks/baza podataka Web poslužioca
- softver za pretraživanje



“Spider”-i, “crawler”-i ili roboti

- programi koji traže po Web-u nove stranice koje će biti dodane pretraživaču i koji identificiraju već poznate poslužioce koji su se promijenili
- *spider*-i prikupljaju **informacije o sadržaju stranica** i time “hrane” dio pretraživača koji se bavi indeksiranjem
- programiraju se za *širinu* i *dubinu* do koje će ići
- mnogi spideri posjećuju znatno češće poslužioce koji su **često** posjećeni i one do kojih vodi veliki broj linkova



Program za indeksiranje i indeks

- neki pretraživači deklariraju indeksiranje svake riječi sa svake stranice - upitno
- većina indeksira "title" i URL
- metatag-ovi se najčešće indeksiraju, ali ne uvijek
- neki pretraživači ne indeksiraju *frame*-ove
- različiti načini indeksiranja dovode do različitih rezultata pretraživanja, čak kada su informacije o poslužiocu registrirane



programi za pretraživanje

- funkcioniraju vrlo slično: bilo koja Web stranica koja sadrži riječi ili izraze koji se **podudaraju** s korisnikovim upitom biti će uključena u listu rezultata koju će korisnik vidjeti na zaslonu
- redanje po **relevantnosti** je ono što je vrlo različito - određuje ga algoritam koji analizira npr. učestalost izraza koji je korisnik postavio kao upit, popularnost i sl.



Razlike među pretraživačima

- veličina baze podataka
- učestalost ažuriranja baze podataka
- broj pronađenih dokumenata
- sučelje za pretraživanje
- mogućnosti pretraživanja
- prikaz rezultata pretraživanja
- brzina pretraživanja
- pomoć koju nude
- “dodatna ponuda”
- dizajn i dr.

Strategija pretraživanja

- **formuliranje pitanja**
- utvrđivanje važnih pojmova, **konceptata** unutar tog pitanja
- prepoznavanje **izraza** (search terms) koji opisuju te pojmove, koncepte
- uzimanje u obzir **sinonima** ili varijacija izraza
- čitanje **uputa** na stranici pretraživača
 - pomoć
 - dokumentacija naprednog pretraživanja
 - učestalo postavljana pitanja FAQ
- pripremanje **logike** pretraživanja
 - jednostavna Booleova logika
- **pročišćavanje** dobivenih rezultata (refine)
 - sužavanje - AND, Proximity, preciznije kw
 - proširivanje - OR, sinonimi



dobra strategija pretraživanja



ispravno korišćenje Booleovih operacija



poznavanje osobina i mogućnosti pretraživača



zadovoljavajući rezultat

postoje i dodatne tehnike koje mogu još više poboljšati rezultate



dodatni trikovi

- korištenje CTRL-F kako bi našli traženi tekst u dokumentu
- najbolje rezultate pretraživanja možemo spremiti u bookmark
- moguće je skratiti desni dio URL-a, pa onda napraviti pretraživanje po URL ili Lokaciji
- ako pretražujemo ustanovu možemo pokušati preko URL-a
www.ime_ustanove.domena
- korištenje *wildcard-a*
- korištenje listanja po relevantnosti (Alta Vista i ***most important***, Excite i Infoseek i ***more like this***)



Pretraživači

<http://nippur.irb.hr/hrv/searchglobal.html>

ili odabir Pretraživanje s prve stranice knjiž.

- Alta Vista (<http://www.altavista.com>)
- Excite (<http://www.excite.com>)
- FAST (<http://www.alltheweb.com>)
- Google (<http://www.google.com>)
- HotBot (<http://www.hotbot.com>)
- Infoseek (<http://infoseek.go.com>)
- Northerl Light (<http://www.northernlight.com>)
- (Lycos (<http://www.lycos.com>))



Tipiča stranica pretraživača sadrži:

1. prozor za unos upita (*query box*)
2. mogućnosti definiranja (ograničavanja) pretraživanja
3. veza na napredno pretraživanje
4. pomoć
5. dodaci
6. reklame
7. direktoriji (tematski pretinci) / kanali (chanel)



6

Reference. Community. News.
Health has a Homepage WebMD CLICK HERE

Search Smarter

1

SEARCH

Look for:

all the words

Date:

anytime

Language:

any language

Pages Must Include:

image MP3

video JavaScript

Return Results:

10

full descriptions

2

3


[News Headlines](#) [White Pages](#) [Stock Quotes](#) [Free Email](#) 5
[Email Addresses](#) [Yellow Pages](#) [Discussion Groups](#) [Free Homepages](#)

HOTBOT DIRECTORY

- [Arts & Entertainment](#)
Movies, Music, Television ...
- [Business & Money](#)
Investing, Jobs, Industries ...
- [Computers & Internet](#)
Hardware, Internet, Software ...
- [Games](#)
Videogames, Role-Playing ...
- [Health](#)
Medicine, Fitness, Alternative ...
- [Home & Family](#)
Family, Kids, Houses ...
- [News & Media](#)
Online, Newspapers, Weather ...

- [Reference](#) 7
Libraries, Education, Maps ...
- [Regional](#)
US, Europe, Asia ...
- [Science & Technology](#)
Technology, Social Sciences ...
- [Society](#)
Government, People, Religion ...
- [Sports](#)
Football, Soccer, Basketball ...
- [Travel & Recreation](#)
Autos, Food, Travel ...

guru.com™
Empowering Freelancers!

 Download the HotBot
Neoplanet Browser

[About Us](#) | [Advertise](#) | [Jobs](#) | [Privacy Statement](#) | [Feedback](#)
[Text-only version](#)

[Test the HotBot Beta site.](#)

Search Resources
[Free Downloads](#)
[Research Service](#)

4



dodatne usluge (pored osnovnog pretraživanja)

- pregledavanje po **tematskim kategorijama**
- zasebna pretraživanja slika, zvuka i videa, FTP
- **prevođenje** dokumenata
- **elektronička pošta**
- prostor za **Web stranice**
- prilagodljive **portale** (“My ime_pretraživača”)
- obiteljske **filtere** i sl. (Family filter ili Adult content filter)
- online **trgovina**
- podaci o novostima, dionicama, kompanijama, ljudima, kartama, vremenu, putovanjima...

Što čini kvalitetan pretraživač?

- mogućnosti pretraživanja
- intuitivnost jednostavnog ili naprednog pretraživanja
- odgovarajuća dokumentacija
- količina dobivenih rezultata
- postotak relevantnih rezultata
- da li su rezultati poredani po relevantnosti
- ažurnost podataka
- nizak postotak *dead link*-ova
- odgovarajući prikaz rezultata (sažetak, URL...)
- očitost i/ili logika redoslijeda prikaza
- korisnost dodatnih značajki pretraživača



[Search](#) [Live!](#) [Shopping](#) [Raging Bull](#) [Free Internet Access](#) [Email](#)

[Own your favorite music at CDNow!](#)

[Click Here!](#)

[Video Game Central!](#)

[Search](#) [Advanced Search](#) [Images](#) [MP3/Audio](#) [Video](#)

[Try our NEW Search Centers](#)

Find this:

any language

Example: **+hiking +trails**

Find Results on: The Web News Discussion Groups Products

[My Live!](#) [Money](#) [Sports](#) ^{Now!} [Women](#) [Health](#) [Travel](#) [Real Estate](#) ^{Now!} [News](#) [Jobs](#) [Translate](#) [Entertainment](#)
[Chats](#) [Message Boards](#) [Free Internet Access](#) [Email](#) [Yellow Pages](#) [People Finder](#) [Directions](#) [Local](#) [Hotels](#)

Breaking News

- ['Peanuts' Artist Shulz Honored](#)
- [Are ISPs Defending Their Users](#)
- [Griffey Slips Into Training](#)
- [Net Patent Squabbles Increase](#)
- [Tedeschi Feels Grammy Effect](#)

Stock Quotes

[Symbol Lookup](#)

Web Directory

[Arts & Entertainment](#)

[Movies](#), [TV](#), [Music](#)...

[Autos](#)

[Classic](#), [Dealers](#), [Manufacturers](#)...

[Business & Finance](#)

[Recreation & Travel](#)

[Food](#), [Outdoors](#), [Humor](#)...

[Reference](#)

[Maps](#), [Education](#), [Libraries](#)...

[Regional](#)

Comparison Shopping

Top Product Searches

- [Palm V Organizer](#)
- [Pokemon](#)
- [Gameboy](#)

Shopping Departments


[Computers](#) [Music](#)
[DVD](#) [Office](#)

alta vista: SEARCH



- vrlo moćno pretraživanje
- jedan od najvećih pretraživača
- usluge prevođenja
- pronalaženje slika
- trude se osigurati kvalitetan index
- trude se biti međunarodni, nisu tako proamerički kao drugi
- mnogo intuitivnije sučelje nego prije
- nekonzistentnost rezultata
- prikazuje samo 10 pogodaka istovremeno
- nema mogućnosti sortiranja (osim po relevantnosti)

altavista: SEARCH

- ZNAČAJKE PRETRAŽIVANJA 
- Booleovi operatori: AND, OR, AND NOT
- operatori približnosti NEAR
- +, -, “”
- ograničavanje na vremenski period, jezik
- **polja:** anchor: applet: domain: host: image: link: text: title: url: like:
- kraćenje: desno i unutarne
- osjetljivost na mala i velika slova (upit s velikim slovima forsirat će takvo pretraživanje)
- upit na prirodnom jeziku (Ask Jeeves)

alta vista: SEARCH



- PRIKAZ REZULTATA
- ŠTO: naslov dokumenta, URL, prve dvije linije teksta dokumenta, jezik, datum i veličina (bytes)
opcije: Translate, Company Fact Sheet, More pages from this site, Related Pages
- KOJIM REDOSLIJEDOM: prema relevantnosti (u Advanced Search može se odrediti “ranking keywords”)
- SUŽAVANJE: ne postoji (trenutno) “refine” opcija, samo kod nekih dokumenata postoji link na “related pages”

alta vista: SEARCH



- NEKONZISTENTNOSTI
- Times out prije nego što je pretraživanje završilo
- Ne može izbrojati - broj na prvom zaslonu nije uvijek pouzdan
- Nekonzistentnost u dijakritičkim znakovima
+éléphant -elephant



[Create your Start Page!](#)

[Choose your favorite photo!](#) »



New Members [Sign Up](#) • Excite Members [Sign In](#) • [Help](#)

My Excite

- [FREE Voicemail/Email](#)
- [Calendar/Address Book](#)

[Make Excite My Start Page!](#)

Today On Excite

Tuesday, February 22, 7:17AM EST

News [Pet Cloning Is Possible](#)

Poll [Cramming Victim?](#)

- [Hot! Auto Slide Shows](#)
- [Have a Nice Day Cards](#)
- [@Home - 24x7 Connection](#)

30% off Grammy® CDs @CDNOW

My Stocks

edit - x

Sponsored by **Charles Schwab**

Get Quotes:

[Full Portfolio](#) | [Find Symbol](#)

Symbol	Price	Change
Nasdaq	4411.74	-137.18
D	10010.50	205.05

My News

edit - x

[Top Stories](#) | [Photos](#) (Feb 22 6:29AM)

- [Report on Medical Mistakes Sought](#)
- [Gore, Bradley Debate in Harlem](#)
- [Bush, McCain Continue To Bicker](#)

[ZDNet Tech News](#) (Feb 22 5:22AM)

- [Back to court for Microsoft, DOJ](#)
- [Crowds flock to PlayStation 2 fest](#)

[More](#)

- [Autos](#)
- [Business](#)
- [Careers](#)
- [Computers](#)
- [Education](#)
- [Entertainment](#)
- [Family](#)
- [Games](#)
- [Health](#)
- [Investing](#)
- [Lifestyle](#)
- [Real Estate](#)
- [Relationships](#)
- [Sports](#)
- [Travel](#)

Free ISP 3 Minute Download

[Shopping](#) [Classifieds](#) [Auctions](#) [Digital Cameras](#)
[Horoscope Shop](#) [Wedding Shop](#) [More...](#)

[People & Chat] [Chat!](#) [3824 People](#) [Meet People](#)
[Share Photos](#) [Find Friends](#) [Voice Chat](#) [More...](#)

Shortcuts

Today's News

- [Free E-Cards](#)
- [Stock Quotes](#)
- [Yellow Pages](#)
- [Free Web Access](#)
- [Desktop Radio](#)
- [Photo Screen Savers](#) **new!**
- [Classifieds](#)
- [Maps & Directions](#)
- [Horoscopes](#)
- [More...](#)

Send Fresh
Dan's Chocolates
dans.com click here

My Weather

edit - x

Enter your Zip Code:

[Airport Delays](#)



- nudi sofisticiranu personalizaciju
 - “more like this”
 - mogućnost izbora “site-a”
 - nudi vrlo relevantne rezultate za vrlo popularne upite
 - News Search osigurava pristup Web verzijama novina
 - brojne referentne baze (rječnici, alamanasi, enciklopedije)
- srednja baza
 - nema kraćenja koje kontrolira korisnik ili pretraživanja po poljima
 - potpuno proamerički
 - Power Search je sve prije nego “powerfull”



- ZNAČAJKE PRETRAŽIVANJA
- Booleovi operatori: AND, OR, NOT (default je OR)
- Napredno pretraživanje: CAN contain, MUST contain, MUST NOT contain
- +, -, ""
- Napredno pretraživanje: ograničavanje na jezik, zemlju i domenu
- polja: Title, URL, Image, Java applets, Links
- kraćenje: nije podržano!
- NEosjetljivost na mala i velika slova



- PRIKAZ REZULTATA
- ŠTO: naslov dokumenta, URL, sažetak i nivo relevantnosti kao postotak (nije objašnjeno kako je izračunat)
(postoji mogućnost prikazivanja isključivo naslova)
- KOJIM REDOSLIJEDOM: prema relevantnosti - prvih deset Web stranica, prvih pet novosti s relevantnim diskusijskim grupama
(može se izabrati samo prvih 40 Web stranica)
- SUŽAVANJE:
 - Excite nudi riječi koje se mogu dodati pretraživanju kako bi se ono suzilo
 - postoji link "more like this"

FAST SEARCH	FTP SEARCH	MP3 SEARCH	PICTURE SEARCH	LISTENING ROOM	FA PREMIER	CONTACT US
<i>300M Pages</i>	<i>100M Files</i>	<i>1,000,000 MP3s</i>	<i>17M Pics & Sounds</i>	<i>Hot New Band Site</i>	<i>Largest Soccer Site</i>	<i>FAST Info</i>

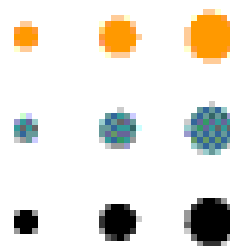
Search for [Search](#)
[help](#)Try our new [Advanced Search](#)

[Fast Search & Transfer announces world's biggest fastest search engine; 300 million web pages searched in under one half second with highly relevant results](#)

[Submit Your Site](#) | [Add FAST Search to Your Page](#) | [Join the FAST Team!](#)
[FAST, Dell Press Release](#) | [All the Web \(Pictures\)](#)
[Frequently Asked Questions \(FAQs\)](#)

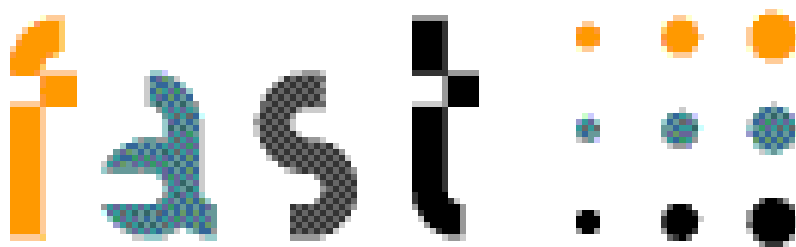


fast

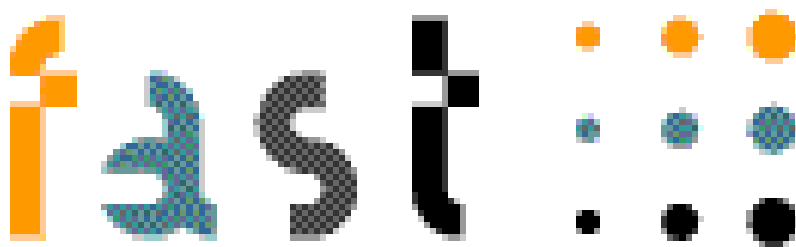


- trenutno najveći pretraživač
- brzina
- nema stop riječi
- novo napredno pretraživanje dozvoljava prikaz do 100 rezultata istovremeno

- baza se ne ažurira redovito
- nema Booleovih operatora, kraćenja i pretraživanja po mnogim poljima



- ZNAČAJKE PRETRAŽIVANJA
- default je “all of the words”, postoji i “any of the words” i “exact phrase”
- +, -, “” (Napredno pretraživanje nudi pretraživanje po nekim poljima: title, link name, url, link to url, s opcijama Should Include, Must Include i Must Not Include
- nema ograničavanja pretraživanja
- kraćenje: nije podržano!
- NEosjetljivost na mala i velika slova



- PRIKAZ REZULTATA
- ŠTO: naslov dokumenta, tri linije teksta i URL
- KOJIM REDOSLIJEDOM: prema relevantnosti
- SUŽAVANJE: ne postoji “refine” opcija



[About Google](#)

Search the web using Google

Google Search

I'm Feeling Lucky

[Add Free WebSearch To Your Site](#)



- jednostavnost sučelja
- brzina!!!
- pretraživanje stranica koje imaju link na zadanu stranicu
- može pretraživati stop riječi
- jedini pretraživač koji nudi pristup stranici (cache) iz vremena kada je indeksirana
- slabe mogućnosti ciljanog pretraživanja
- pretraživanje po linku mora biti egzaktno



- ZNAČAJKE PRETRAŽIVANJA
- default je AND
- “”, +, -
- POLJA: link: (nalazi stranice koje sadrže određeni hiperlink) related: (poziva Google Scout koji nalazi sličan uzorak povezivanja na dani URL)
- ograničavanje: nije podržano
- kraćenje: nije podržano
- NEosjetljivost na mala i velika slova



- PRIKAZ REZULTATA
- ŠTO: naslov dokumenta, URL, prvih nekoliko riječi iz teksta dokumenta, link na prethodno pohranjenu verziju dokumenta
- KOJIM REDOSLIJEDOM: prema broju stranica koje imaju link na stranicu iz rezultata
- SUŽAVANJE: Google scout pronaći će stranice koje su nekako vezane uz nađene stranice (slično "more like this" opciji kod Excite-a)



- jedan od najvećih pretraživača
- dvostruki sistem: Inktomi i Direct Hit baza i imenički sistem Open Directory
- napredne i raznolike mogućnosti pretraživanja
- lagano pretraživanje po poljima
- jasno prezentirane sve mogućnosti
- link: pretraživanje mora biti egzaktno
- veličina baze se smanjuje
- ima stop riječi



ZNAČAJKE PRETRAŽIVANJA

“all of the words”, “any of the words”, “exact phrase”, “the person” i “links to this URL”

Booleovi operatori: AND, OR, NOT

+, -, “”

POLJA: title: domain: depth: feature:(frame, image, applet...), likdomain: outgoingurlext: scriptlanguage: **after:** **before:** within:(3/months)

ograničavanje na vremenski period, jezik i vrstu medija (Javascript, slika, video); Adv. S. još nudi i limitiranje na domenu (npr. .edu, .hr), zemljopisno područje i dubinu stranice

kraćenje: desno i unutarne; “**stemming**” - Adv. S. - gramatičke varijante

osjetljivost na mala i velika slova (upit s velikim slovima forsirat će takvo pretraživanje)



- PRIKAZ REZULTATA
- ŠTO: tri opcije: puni opis, kratki opis i samo URL
puni opis uključuje naslov dokumenta, URL, prvih nekoliko linija teksta dokumenta
- KOJIM REDOSLIJEDOM: prema relevantosti
- SUŽAVANJE: dva načina:
 - unutar već postojećih rezultata
 - "This site only"



SEARCH the Web Images [More...](#)

[Search options](#) | [How to search](#) | [GOguardian™ is off](#)

[My Page](#)

[Home](#)

[Search](#)

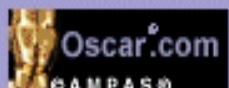
[Communicate](#)

[Shopping](#)



MEMBER SERVICES

- [Free e-mail](#)
- [Personalized start page](#)
- [Register today!](#)
- [Member sign-in](#)



TOP ABCNEWS

- [Michigan Battleground](#)
- [Aircraft Problems](#)
- [Iran Elections](#)

FEATURED SPONSORS

- [Donate Britannica.com's \\$](#)
- [myprimetime.com - Now life](#)

TODAY ON GO NETWORK

[Get Oscar](#) | [Your wildest fantasy](#) | [Lose weight now](#)

[ABCNEWS](#)

[US](#), [world](#), [business...](#)

[ESPN Sports](#)

[NBA](#), [NFL](#), [fantasy...](#)

[Movies and TV](#)

[News](#), [reviews](#), [bios...](#)

[Auction](#)

[Collectibles](#), [art...](#)

[Family](#)

[Pets](#), [babies](#), [45+...](#)

[Music](#)

[CDs](#), [news](#), [bios...](#)

[Autos](#)

[Pricing](#), [specs...](#)

[Food and Drink](#)

[Recipes](#), [dining...](#)

[Real Estate](#)

[Homes](#), [rent](#), [move...](#)

[Broadcast](#)

[Radio](#), [mp3](#), [media...](#)

[Games](#)

[Free games](#), [prizes...](#)

[Romance](#)

[Weddings](#), [Valentine's...](#)

[Careers](#)

[Jobs](#), [advice...](#)

[Health](#)

[Diseases](#), [fitness...](#)

[Shopping](#)

[Products](#), [merchants...](#)

[Computing](#)

[Reviews](#), [screensaver...](#)

[Kids](#)

[Play games](#), [prizes...](#)

[Travel](#)

[Flights](#), [bargains...](#)

[Education](#)

[Teaching](#), [learning...](#)

[Money](#)

[Stocks](#), [markets...](#)

[Women](#)

[Family](#), [beauty...](#)

TOOLS

- [Dictionary](#)
- [Driving directions](#)
- [Maps](#)
- [Translator](#)
- [White pages](#)
- [Yellow pages](#)

QUICKLINKS

- [Downloads](#)
- [Horoscopes](#)
- [Movie times](#)
- [Stock quotes](#)
- [TV times](#)
- [Weather](#)

COMMUNICATE

- [Chat](#)
- [Free home pages](#)
- [GO Invite](#)
- [Message boards](#)

NEW ON GO.COM



- sortira rezultate prema site-u i datumu
- bogati dodaci i sadržaj portala
- brojne referentne baze
- usluge prevođenja (od travnja 1999) - eng, fra, ger, ita, spa, por
- rezultati se mogu suziti
- može pretraživati dijakritičke znakove



infoseek

- nema previše mogućnosti pretraživanja
- mala baza



infoseek

■ ZNAČAJKE PRETRAŽIVANJA

- (Booleovi operatori): nema (default je OR)
- +, - ne u punom značenju, samo za rankiranje, “”
- ograničavanje na domenu, zemljopisnu lokaciju, Infoseek tematsku kategoriju
- **polja**: title:, url:, link:, site:
- kraćenje: nije podržano! (ali ima atomatski “inteligentni plural”)
- osjetljivost na mala i velika slova (upit s velikim slovima forsirat će takvo pretraživanje)



infoseek



■ PRIKAZ REZULTATA

- ŠTO: naslov dokumenta, URL, prvih nekoliko linija teksta dokumenta, datum i veličina (bytes)
- KOJIM REDOSLIJEDOM: prva stranica prikazuje Recommended Sites, iza kojih slijedi lista Infoseek-ovih Directory categories, a zatim prvih deset rezultata
 - mogućnost prikaza samo stranica s određenog site-a
 - mogućnost sortiranja po datumu
- SUŽAVANJE: dva načina:
 - “Find similar pages” - za razliku od sličnih “similar” značajki kod drugih pretraživača, ovo je vrlo upotrebljivo
 - “Search within results” - unošenje dodatnog izraza

NorthernLight.com

Just what you've been searching for.

Be specific: Instead of **turkey**, use **turkey recipes**.

Search for

Search

Select a source

Search all Northern Light Sources

[Simple Search](#) | [Power Search](#) | [Business Search](#) | [Investext Search](#) | [Stock Quotes](#) | [Search News](#)



Custom Search Folders™

The World Wide Web prioritized, categorized and organized into neat little [Custom Search Folders](#).

Dow Jones Industrials 10219.52 -295.05

Updated February 18, 2000 04:53 PM EST



Today's Headlines

[More headlines ...](#)

[Arizona, Mich. Vote in GOP Race](#)

[NATO Troops Push Back Kosovo Protest](#)

[50,000 Ethnic Albanians Protest](#)

[Iran Reformers Win Parliament Seats](#)

[Unilever To Lay Off 25,000 Workers](#)

News stories updated continuously

What's New !!!



Search Alert Service™

Free personalized news and research email updates covering the topics of your choice.

- [Learn more](#)
- [Create a new alert](#)



Special Editions™

In-depth coverage of major news stories.

Northern Light



StoreRunner
Let's Shop!

Help Center

Accounts

About

Alerts

Portfolio



- usporedo s bazom web dokumenata, NL nudi i pristup posebnoj zbirci
- jedan od najvećih pretraživača
- bogate mogućnosti pretraživanja (Boole, kraćenje, Power Search, Business Search)
- Current News s mogućnosti pretraživanja



- rezultati nisu tako pouzdani kao kod nekih drugih
- nema linka *field searching*, tj. pretraživanja po poljima (makar je to unutar Power Search-a)
- nema operatora približnosti
- samo po 10 pogodaka





- ZNAČAJKE PRETRAŽIVANJA
- Booleovi operatori: default je AND, no, rade svi: AND, OR, AND NOT
- +, -, ""
- POLJA: title: url: text: (pub: company: ticker: recid:> Spec.Coll.)
- ograničavanje prema položaju unutar dokumenta (bilo gdje, u naslovu, URL), datumu, vrsti izvora (komercijalni site, osobna stranica, edukacijski site, neprofitni site), jeziku, zemlji, tematici (Arts, Business, Education, Travel) i vrsti dokumenta (informacije o kompaniji, za prodaju, materijal za izobrazbu, press release, pregled)
- kraćenje: * za više i % za jedno slovo, oboje moguće na kraju i u sredini
- NEosjetljivost na mala i velika slova



- PRIKAZ REZULTATA

- ŠTO: naslov dokumenta, URL, kratki sažetak, jezik, datum i vrsta poslužioca
- KOJIM REDOSLIJEDOM: prema relevantnosti
Power Search omogućuje sortiranje po datumu
- SUŽAVANJE: koristi se Custom Search Folders kako bi se grupirali rezultati sličnih karakteristika (tematika, izvor, tip dokumenta i jezik), klikanje na Folder pokazuje SAMO rezultate te određene kategorije

TODAY

Feb. 21 2000

Get Weather Forecast

Enter ZIP

[Canada](#)


- [Used Car Center](#) Get tips on buying a used car
- [Extreme Multimedia](#) Make the most out of your high bandwidth
- [Time for a vacation?](#) Book the perfect getaway today
- [Rentals Classifieds](#) Moving? Find a new place fast
- [Top News](#) Today's stories

[Lowestfare.com](#)
Vacations

Join Now!

- [Free Online Games](#)
- [Software Downloads](#)

ADVERTISEMENT

Higher paying jobs!

[headhunter.net](#)
[Click Here!](#)

[Yellow Pages](#) | [White Pages](#) | [Maps](#) |
[E-mail Look-up](#) | [Pictures](#) | [Power Search](#) |
[Music Downloads](#) | [Audio](#) | [Video](#)

 Powered by [Search Tips](#)

SHOPPING

- [Shop on Snap](#)
- [Auctions & Classifieds](#)
- [Electronics Shop](#)
- [Flower Shop](#)
- [Gift Shop](#)
- [Buyers Guide](#)

COMMUNITIES

- [Chat, Flirt & Gossip](#)
- [Astrology Chat](#)
- [Romance Chat](#)
- [Start a Club](#)
- [Freaks & Geeks' Club](#)
- [MP3 Club](#)

FAVORITES

- [More on Snap](#)
- [Free Stuff](#)
- [Horoscopes](#)
- [Personals](#)
- [Stock Quotes](#)
- [Weather Center](#)

- [Garden Tips & Tours](#)

- [Weight Loss](#)
[Tips you can use](#)

WEB DIRECTORY

- [Autos](#)
- [Finance](#)
- [Multimedia](#)

NBCi SERVICES

- [500M of Web Storage](#)



- dvojaki: pretraživač i direktoriji
- jedinstvena baza unutar direktorija (16 top level i 48 second-level)



- nije najveća ili najmoćnija verzija Inktomija
- ponekad je spor

- ZNAČAJKE PRETRAŽIVANJA
- Booleovi operatori: AND, OR, AND NOT
- +, -, “”
- POLJA: naslov: link: domain: depth: feature: linkdomain: linkext: scriptlanguage: after: before: within:
- ograničavanje samo u naprednom pretraživanju prema datumu, jeziku, vrsti medija, lokaciji/domeni
- kraćenje: nije podržano!
- NEosjetljivost na mala i velika slova (u naprednom pretraživanju više nije svejedno)

- PRIKAZ REZULTATA
- ŠTO: naslov dokumenta, URL, kratki sažetak
- KOJIM REDOSLIJEDOM: nepoznato!



META-PRETRAŽIVAČI

- DOGPILE (<http://www.dogpile.com>) - brojni pretraživači, Usenet, ima jednostavno i složeno pretraživanje, Boole
- INFERENCE FIND (<http://www.inference.com/ifind>)
- INTERNET SLEUTH (<http://www.isleuth.com>) - izvrstan za specijalizirana pretraživanja, omogućava određivanje kategorije unutar koje se pretražuje (poduzetništvo, računala, obrazovanje, sport...)
- METACRAWLER (<http://www.go2net.com/search.html>) - izvrstan za dobivanje brzog odgovora
- PROFUSION (<http://www.profusion.com>)
- SAVVYSEARCH (<http://www.savvysearch.com>)



Intelligentni agenti

- napredne Booleove mogućnosti, tezaurus, spell checker, *sounds-like*
- rezultati pretraživanja analiziraju se, filtriraju
- COPERNIC (<http://www.copernic.com>) - uklanja duplikate
- BULLSEYE (<http://www.intelliseek.com>) - eliminira duplikate i mrtve linkove, mogućnost ponavljanja pretraživanja svaki sat, dnevno, tjedno i slanja rezultata elektroničkom poštom
- MATA HARI (<http://www.thewebtools.com>) može naučiti jednu grupu naredbi za pretraživanje, a onda to automatski prevoditi za svako pretraživanje



NEW KIDS IN THE BLOCK

- Oinga.com
- Surfwax.com



Savjeti za pretraživanje

- planirajte strategiju prije nego izaberete pretraživač = odlučite što bi bilo korisno koristiti: Booleovu logiku, zagrade, približnost, kraćenje, izraze
- započnite vrlo određeno, a onda proširujte prema potrebi
- ne ustručavajte se isprobati razne pretraživače
(<http://nippur.irb.hr/hrv/searchglobal.html>)
- dobro sročeni upiti spremite s CTRL-C, da ga možete ponovno koristiti u drugim pretraživačima



Zlatna pravila:

- naučite puno o dva do tri pretraživača, ali često koristite i ostale
- ako ne nalazite lako ono što tražite, pokušajte s drugim pretraživačem
- koristite uvijek AND, OR i NOT - ako pretraživač i ne podržava operatore, ignorirat će ih
- za različite vrste upita koristite različite pretraživače



Budući razvoj

- spider - samo kvalitetne stranice (kvaliteta na račun kvantiteta)
- pretraživanje korištenjem prirodnog jezika (sada samo Ask Jeeves i The Electric Monk)
- standardizacija (HTML u XML)
- klasificiranje i “katalogiziranje” web stranica



LOST IN SPACE

Log On

AT THE LIBRARY