Motori di Ricerca

presente e futuro prossimo

Il quadro presente

Paolo Ferragina, Università di Pisa

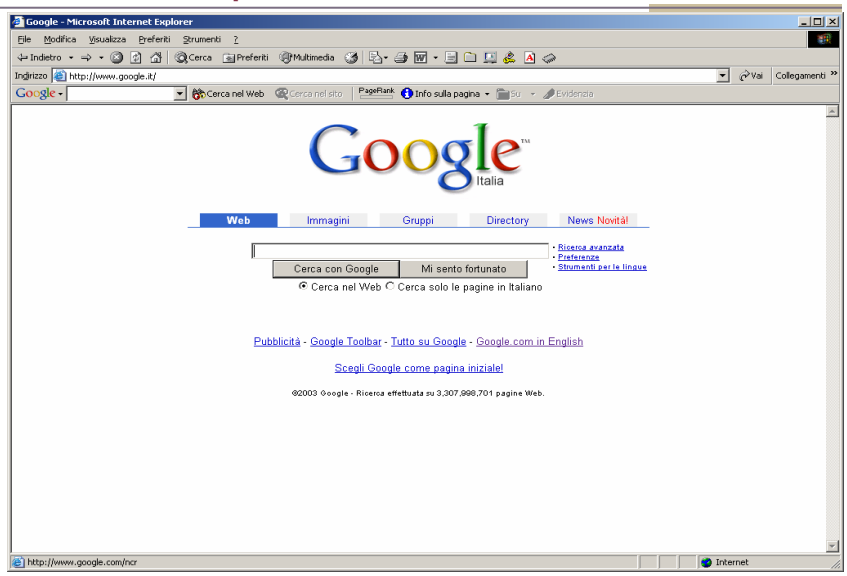
Fino a pochi anni fa...

- Yahoo (migliore del 1995)
- Inktomi (migliore del 1997)
- Altavista (migliore del 1999)
- Lycos, Excite, Northern Light,...

✓ Oggi Google (60%), Yahoo (28%), Msn (12%), Ask (6%). Ogni utente visita più motori di ricerca per le sue query.

Paolo Ferragina, Università di Pisa

Il motore più famoso ...



Cosa non è Google

- Indice su tutti i documenti disponibili sul Web
  - Nessun motore lo è
- Credibile in ogni cosa che ci segnala
  - Non esiste controllo sulla pubblicazione delle pagine
- Perfettamente aggiornato
  - Non riesce a seguire le modifiche giornaliere (milioni di pagine)
- Protetto da contenuto offensivo
  - Dispone di un meccanismo di *filtering*, ma non sicuro al 100%

Cosa è oggi Google

- Alcuni dati interessanti (NY Times, Aprile 2003):
  - Più di 1000 persone
  - 54,000 server - 100,000 processor - 261,000 dischi
  - xMld pagine, 200 milioni query/giorno (60% del totale)
  - 300 milioni di dollari di fatturato 2002 (750 nel 2003 ?)
  - "google" è la parola più utile del 2002 [American Dialect Society]
- Un nuovo scenario di:
  - Business: tra i pochissimi a fare molti profitti !
  - Gestione ed estrazione della conoscenza: non solo Web
  - Problemi matematici interessanti:
    - Qualità risposte, Efficienza, Copertura del web
    - Nuove applicazioni (news,prodotti), Nuovi domini (audio,video)

Google: Il modello di business in 2 iniziative

- *Search services* via la *Google search appliance*
  - Soluzione hardware+software per un motore di ricerca in ambito intranet o singolo website
  - Hardware fissato e quindi limitati problemi di sviluppo e mantenimento del software
  - Per ora disponibile soltanto in USA e Canada (??)
- *Advertising programs* (100.000 sottoscrittori)
  - *AdSense*: Un sito può fornire spazio sulla sua pagina; le pubblicità da visualizzare vengono scelte da *AdSense* in funzione dei contenuti della pagina così da rivolgersi a probabili clienti. Il sito riceve un pagamento in funzione del numero di click sul banner.
  - *AdWords*: Una società può scegliere quanto pagare al giorno/mese e indicare le *parole chiave* che descrivono il suo business. Un *banner* viene visualizzato da Google all'atto di ricerche per quelle parole chiave, e la società paga in funzione del numero di click ricevuti.

Google: altre notizie...

- Il nome deriva dalla parola *GOOGOL*, coniata da un bambino americano di 9 anni per riferirsi al numero 10^{100}
- Un po' di storia:

  [1996-97] Esce il primo prototipo (BackRub).

  [1998-99] Nasce Google, risponde a 10,000 Qpg → 3Ml Qpg

  [2000] 1Mld pagine e 60Ml Qpg

  [2001] 2Mld pagine e 100Ml Qpg, ricerche limitabili a 26 linguaggi.
  Introduce Image e File type search, Usenet dal 1981, Google Catalog.

  [2002] 2,5Mld pagine, ricerche limitabili a 40 linguaggi.
  Intoduce AdWords, Google news, Web API, Froogle, Google Labs.

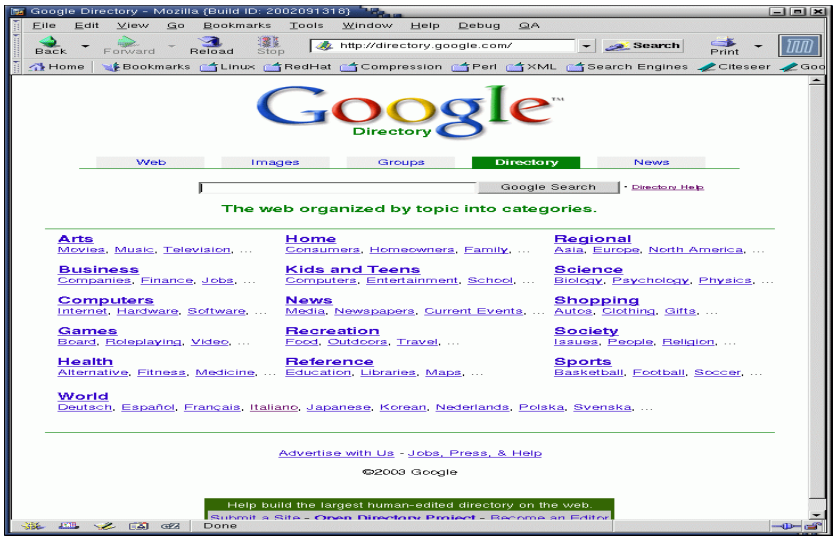
  [2003] 3Mld di pagine, più linguaggi supportati.
  Il programma di business raggiunge i 100,000 sottoscrittori e viene promosso in Italia. Introduce Google AdSense, Local Search.

Motori di Ricerca

presente e futuro prossimo

Altre funzionalità di Google

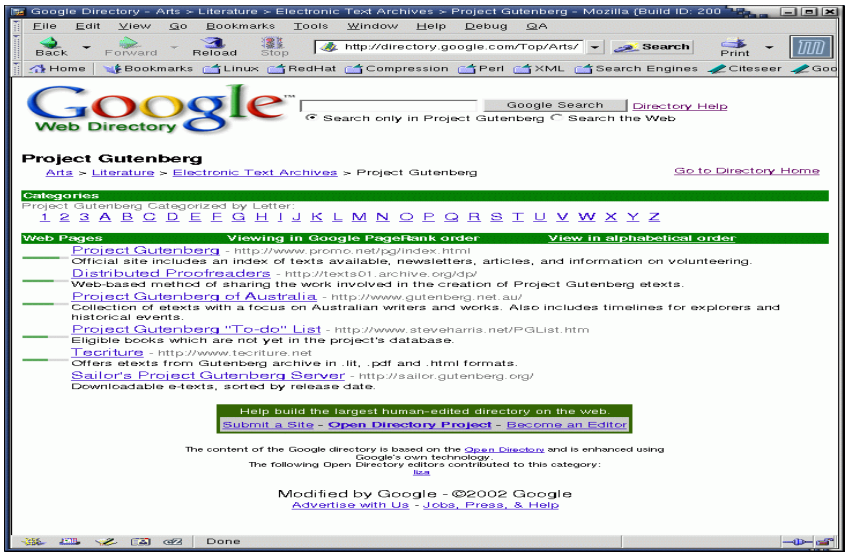
L'archivio *dmoz: open directory project*



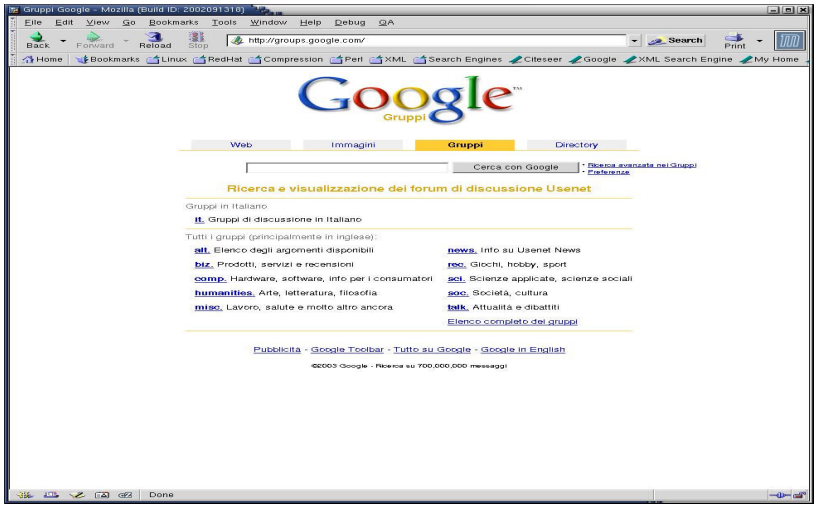
La directory di Google : *dmoz*

- Raccolto e mantenuto da un gruppo di volontari
  - Siti, non pagine, attentamente selezionati e classificati
- Trade-off
  - Ridotta estensione rispetto all'archivio delle pagine di Google
  - Maggiore qualità delle risposte
- Tipo di ricerche:
  - Per parole chiave, ammette anche *inurl* e *intitle*
  - Per navigazione basata sul soggetto
- Ordine delle risposte:
  - Per *Pagerank* oppure per *ordine alfabetico*

Un esempio di interrogazione



L'archivio dei *newsgroup* *(Usenet)*



- Centinaia di migliaia di topics, secondo una gerarchia
- Oltre 800 milioni di messaggi, raccolti dal 1981

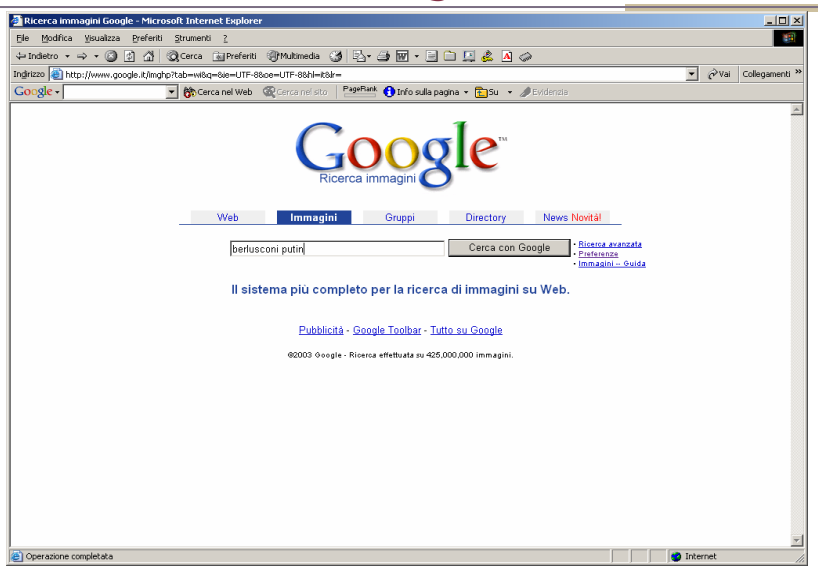
I gruppi di *Usenet*

- Raccolta di messaggi dal 1981, circa 845 milioni
- Tipo di ricerche:
  - Per parole chiave
  - Per navigazione basata sul soggetto
    - comp.sys computer system
    - humanities letteratura
  - Con sintassi speciale:
    - intitle: o author:
    - group:comp*

Esiste una interfaccia per ricerche avanzate

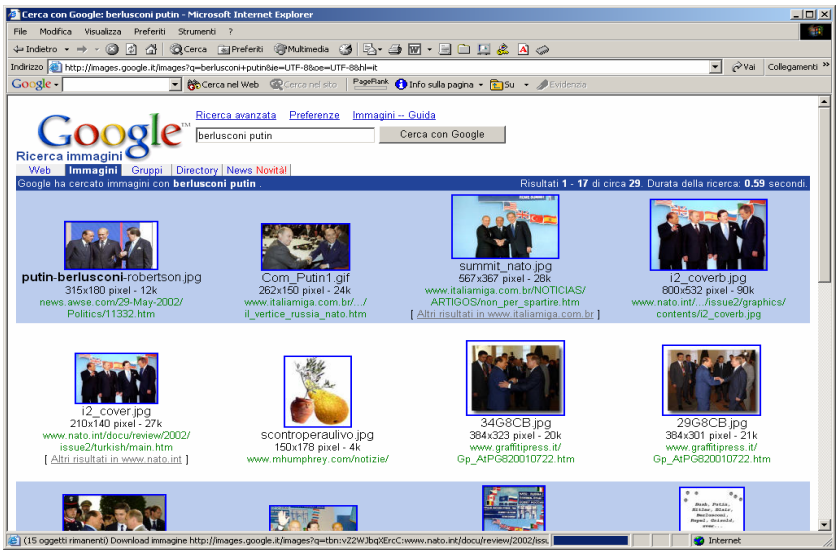
- La ricerca per data qui è molto precisa e utile !!
- E' una sorgente di informazioni validissime su i più disparati soggetti, ottenibili dalle discussioni tra utenti

L'archivio delle immagini



- Circa 800 milioni di immagini dal Web

Un esempio di interrogazione



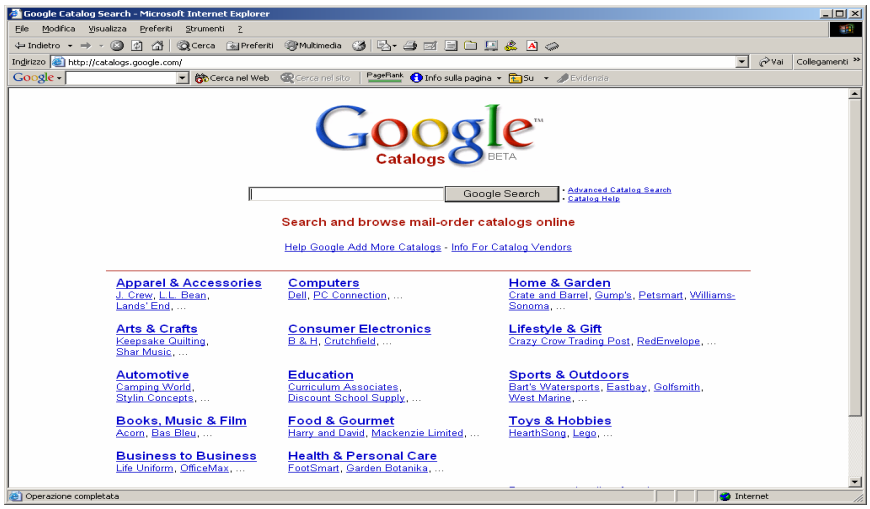
L'archivio delle immagini (contd)

- 800 milioni di immagini prese dal Web e indicizzate per parole chiave, solitamente molto specifiche
- I risultati includono un preview dell'immagine, dimensione, URL
- Problemi: 1 parola → molti risultati, alcune parole → nessuno !!!
- Ricerche: esiste una sintassi speciale
  - intitle: o inurl: o site:
  - filetype:jpg

Esiste una interfaccia per ricerche avanzate

- Può essere utile per le vostre presentazioni o articoli

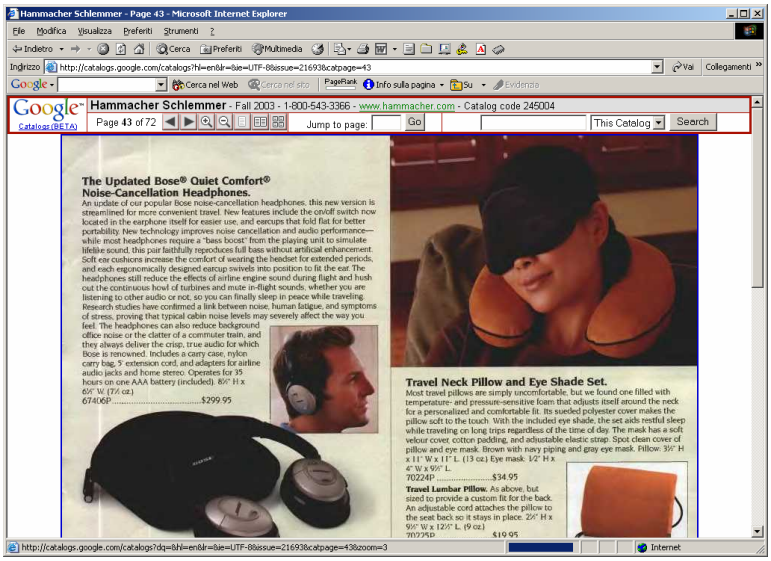
I cataloghi



- Oltre 6000 cataloghi digitalizzati
- Ricerca per parole chiave o navigazione
- Catalogo (anche vecchio) visualizzabile, o sito del venditore

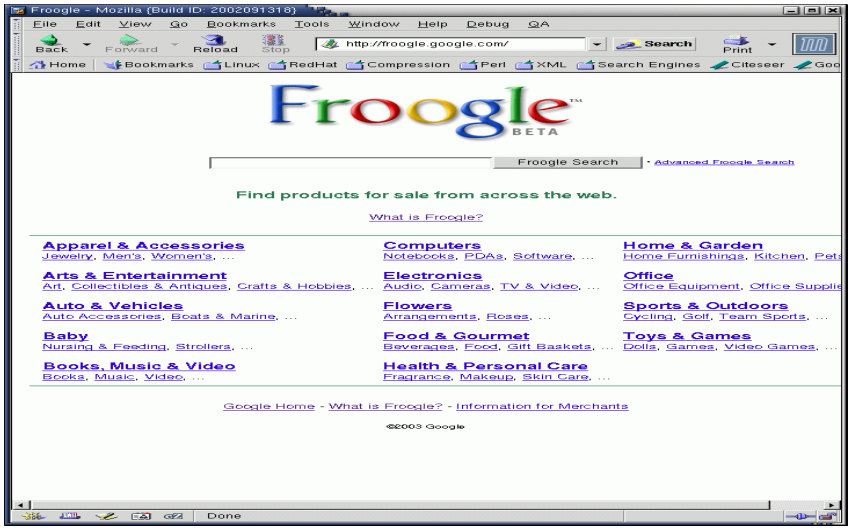
Cercando "baby bed"

Cliccando su una immagine....



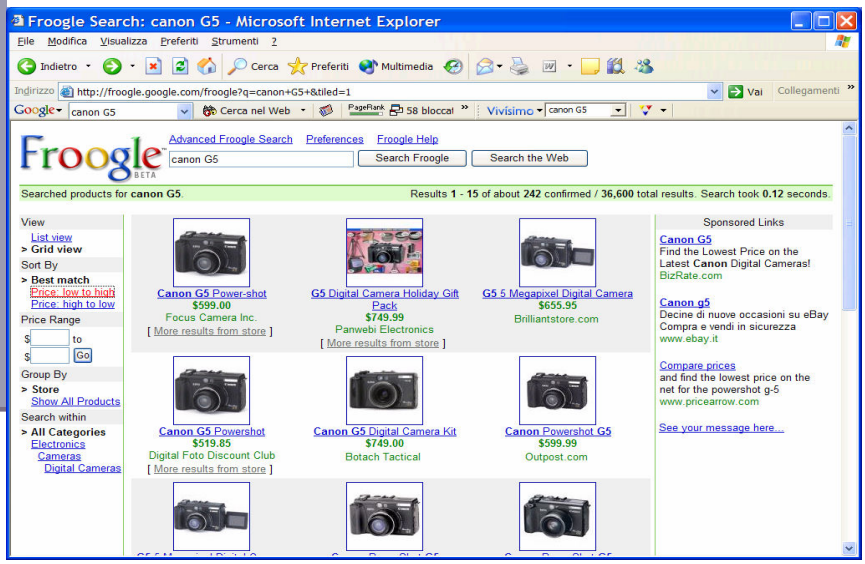
- ... basta scorrere l'immagine!!

Acquisti on-line...



- Froogle = Google + frugal
- Altri approcci noti: shopping robots (Jango)

Un esempio di interrogazione



Notizie USA e internazionali

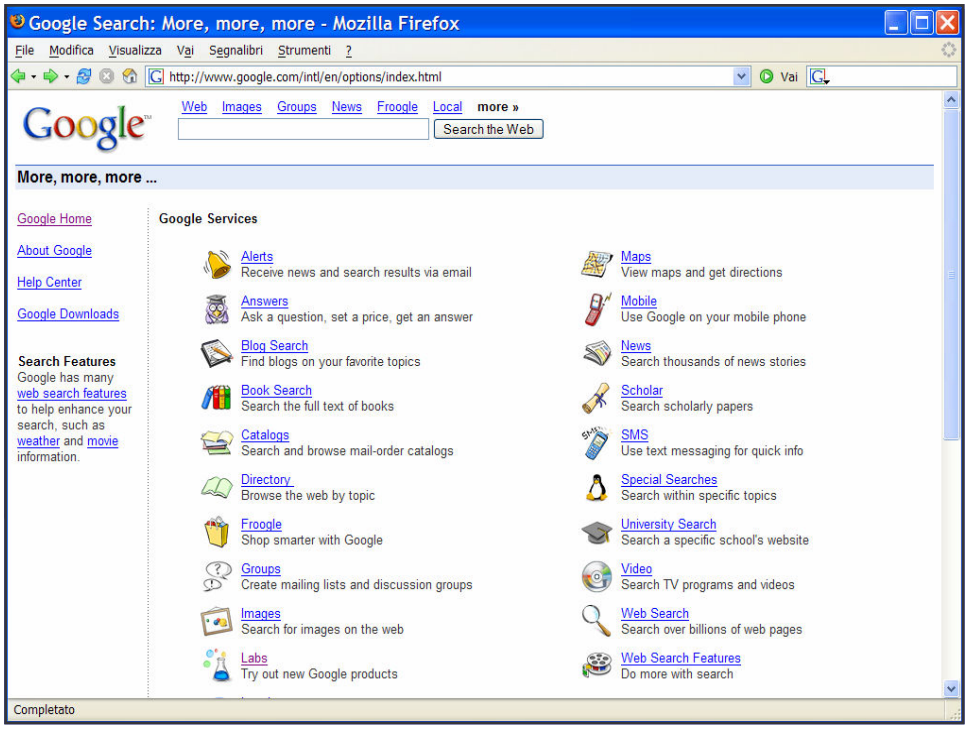
Google news

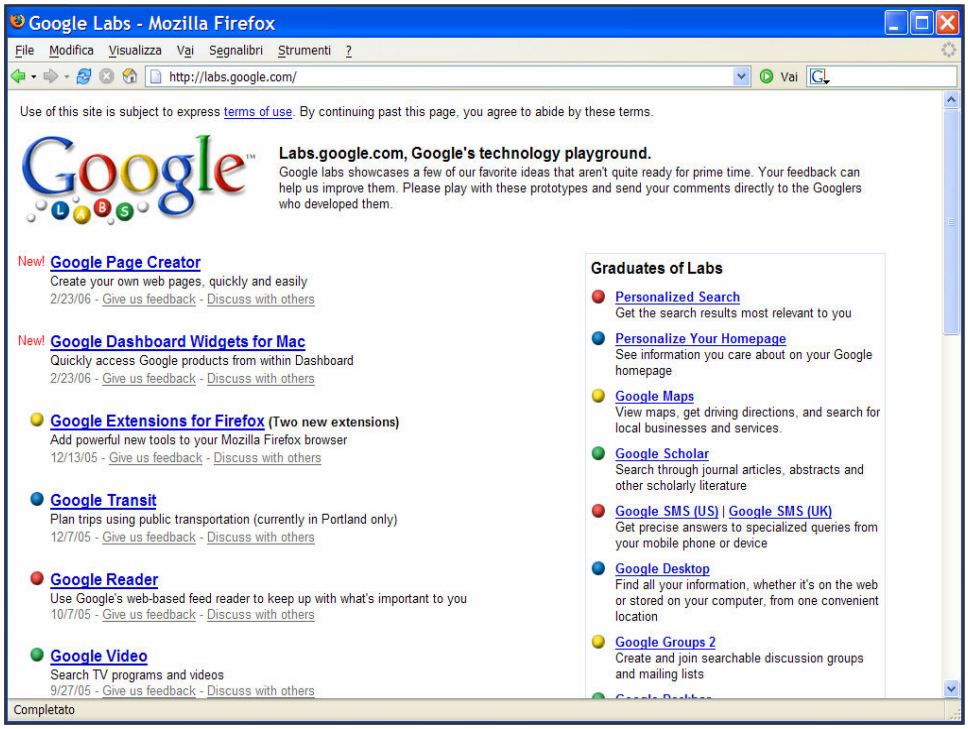
- Oltre 4500 sorgenti di informazione

Esistono versioni per vari paesi

- I risultati includono:
  - Notizie vecchie o recenti (fino a 1 minuto fa...)
  - Indicazione della provenienza
  - Raggruppate per soggetto o storia
  - Presentazione stile Rivista o Quotidiano
- Tutto eseguito in maniera automatica
  - Loro dichiarano "senza influenze politiche o personali"
    .... noi osserviamo che operano "senza un oggetivo filtro" !!!
- E' sorprendente nella sua efficienza, efficacia e ampiezza !!
- Esistono altre proposte, come quella di Yahoo (offre un free alert !!)

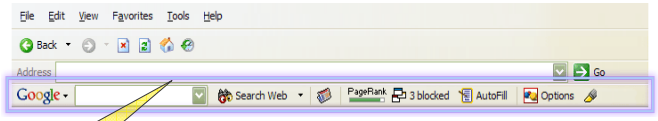
La versione italiana


Google Search: More, more, more - Mozilla Firefox
File Modifica Visualizza Vai Segnalibri Strumenti ?
http://www.google.com/intl/en/options/index.html
Vai
Google
Web Images Groups News Froogle Local more »
Search the Web
More, more, more ...
Google Home
About Google
Help Center
Google Downloads
Search Features
Google has many web search features to help enhance your search, such as weather and movie information.
Google Services
Alerts
Receive news and search results via email
Answers
Ask a question, set a price, get an answer
Blog Search
Find blogs on your favorite topics
Book Search
Search the full text of books
Catalogs
Search and browse mail-order catalogs
Directory
Browse the web by topic
Froogle
Shop smarter with Google
Groups
Create mailing lists and discussion groups
Images
Search for images on the web
Labs
Try out new Google products
Maps
View maps and get directions
Mobile
Use Google on your mobile phone
News
Search thousands of news stories
Scholar
Search scholarly papers
SMS
Use text messaging for quick info
Special Searches
Search within specific topics
University Search
Search a specific school's website
Video
Search TV programs and videos
Web Search
Search over billions of web pages
Web Search Features
Do more with search
Completato


Google Labs - Mozilla Firefox
File Modifica Visualizza Vai Segnalibri Strumenti ?
http://labs.google.com/
Vai
Use of this site is subject to express terms of use. By continuing past this page, you agree to abide by these terms.
Google
Labs.google.com, Google's technology playground.
Google labs showcases a few of our favorite ideas that aren't quite ready for prime time. Your feedback can help us improve them. Please play with these prototypes and send your comments directly to the Googlers who developed them.
New! Google Page Creator
Create your own web pages, quickly and easily
2/23/06 - Give us feedback - Discuss with others
New! Google Dashboard Widgets for Mac
Quickly access Google products from within Dashboard
2/23/06 - Give us feedback - Discuss with others
Google Extensions for Firefox (Two new extensions)
Add powerful new tools to your Mozilla Firefox browser
12/13/05 - Give us feedback - Discuss with others
Google Transit
Plan trips using public transportation (currently in Portland only)
12/7/05 - Give us feedback - Discuss with others
Google Reader
Use Google's web-based feed reader to keep up with what's important to you
10/7/05 - Give us feedback - Discuss with others
Google Video
Search TV programs and videos
9/27/05 - Give us feedback - Discuss with others
Graduates of Labs
Personalized Search
Get the search results most relevant to you
Personalize Your Homepage
See information you care about on your Google homepage
Google Maps
View maps, get driving directions, and search for local businesses and services.
Google Scholar
Search through journal articles, abstracts and other scholarly literature
Google SMS (US) | Google SMS (UK)
Get precise answers to specialized queries from your mobile phone or device
Google Desktop
Find all your information, whether it's on the web or stored on your computer, from one convenient location
Google Groups 2
Create and join searchable discussion groups and mailing lists
Completato

Google è tanto altro ancora

- Dizionari, anche per dialetti: *~parola* define:*parola*
- Elenchi telefonici, pagine gialle,... [anche reverse search]
- Stocks:*identificatore_società* [Yahoo finance]
- Weblogs: "an online site that keeps running commentary and associated links, updated daily."
  - radio.weblogs.com, www.blogspot.com, www.blogger.com



Track user behavior !!

Occhio alla pagina di Google !!

Motori di Ricerca

presente e futuro prossimo

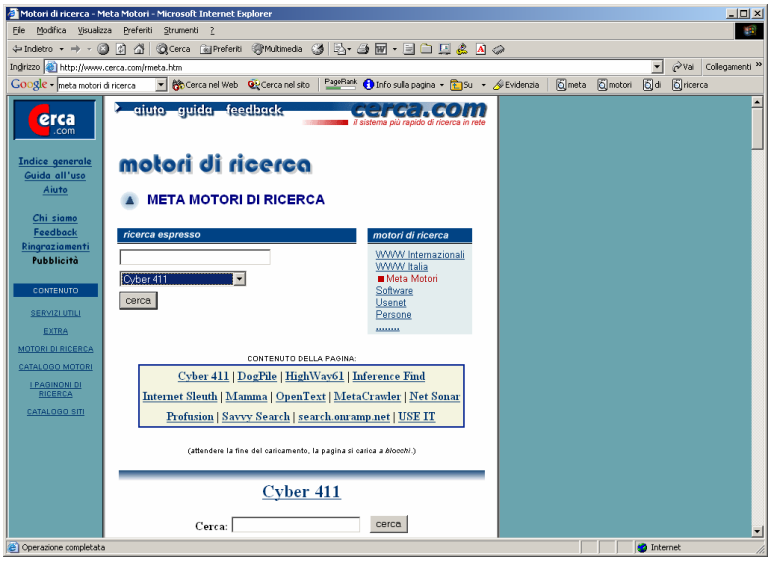
Altre interessanti proposte

Funzionalità: Il quadro corrente

- Migliorare il "ranking" delle pagine: Teoma e molti altri....
- Migliorare la copertura del Web: "Meta" motori di ricerca
- "Supporto all'utente" nella ricerca dei risultati che meglio soddisfano le sue interrogazioni !
  - Suggerimenti sulle parole da cercare (AskJeeves)
  - Suggerimenti su *oggetti* visti da "utenti simili" (Amazon,Epinions)
  - Categorizzazione delle risposte (il fu NorthernLight)
    - Testuale (Vivisimo, Copernic) o grafico (Kartoo)

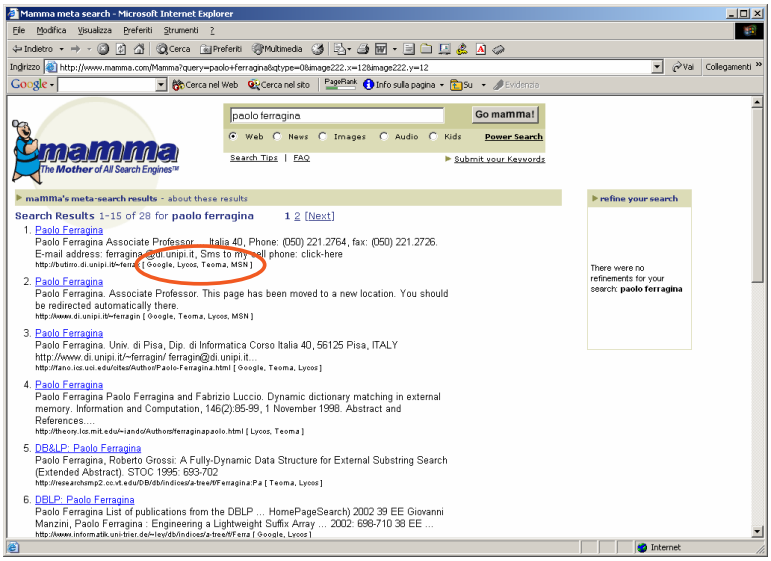
Knowledge Management Systems

Un tipo di Meta-motore

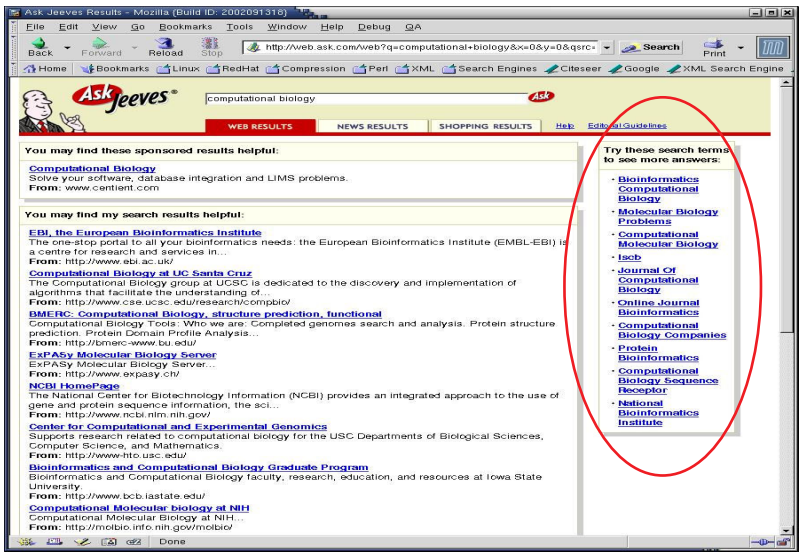


Qualcosa di più sofisticato

La struttura delle risposte



Suggerimenti sui termini: AskJeeves

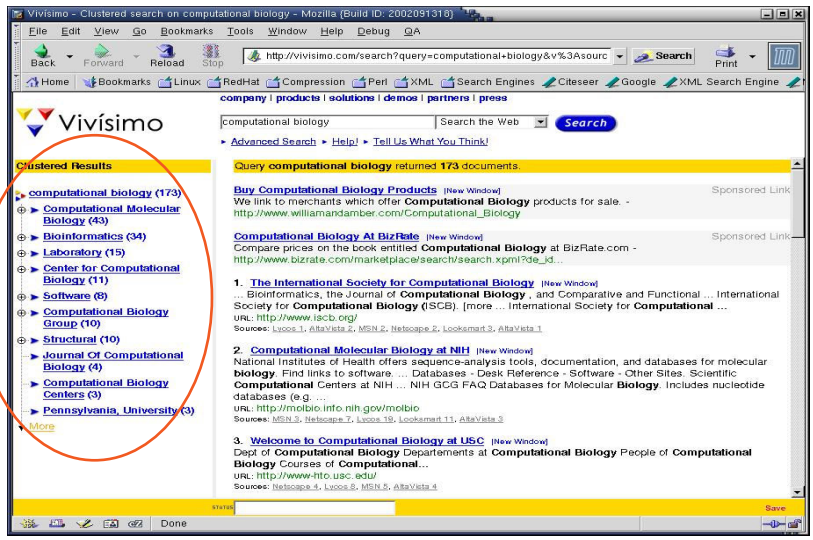


◎ Si appoggia a Teoma, e risponde a "domande" !!

Suggerimenti su oggetti visti: Amazon

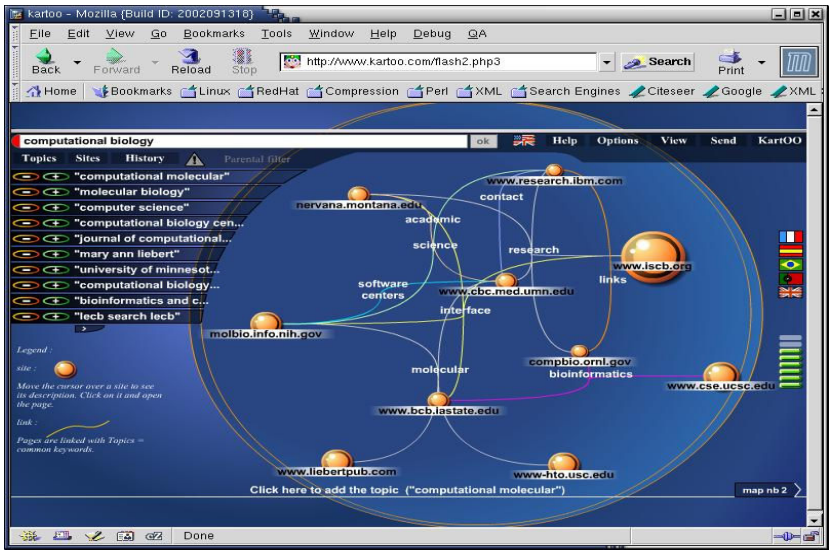
◎ Ora anche ricerche sul testo dei libri ! (anche Google)

Vivísimo: Raggruppare le risposte

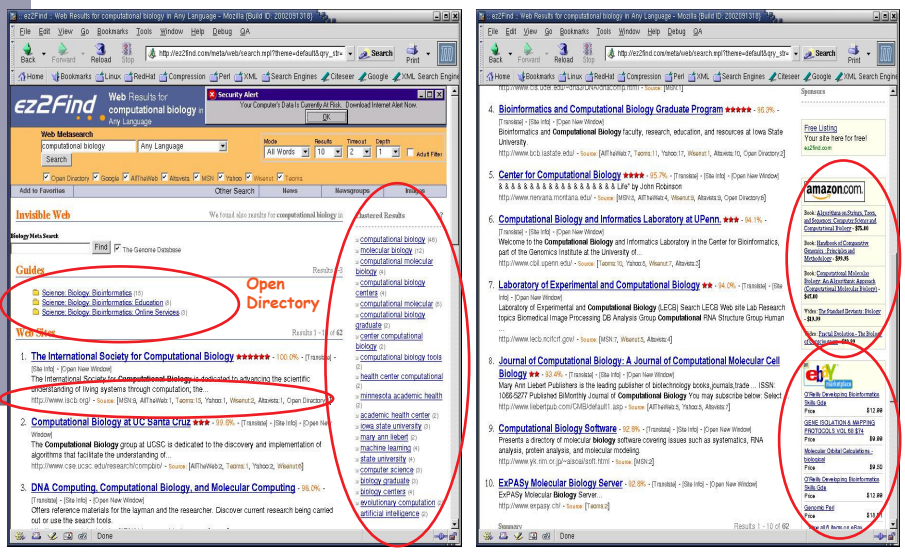


- Offre categorizzazione risultati di FirstGov.com
- Copernic funziona sul nostro desktop

Kartoo: Non solo testo



Un mix interessante: Ez2find



Paolo Ferragina, Università di Pisa

◎ Metasearch + Directory + Cluster dei risultati !!

Un tool interessante