



Web

Images

Groups

Directory

Elementi di Informatica

Google Search

I'm Feeling Lucky

[Advanced Search](#)
• [Preferences](#)
• [Language Tools](#)

Motori di ricerca

New! [Get the Google Search Appliance for your company.](#)

[Advertise with Us](#) - [Search Solutions](#) - [News and Resources](#) - [Jobs, Press, Cool Stuff...](#)

©2002 Google - Searching 2,073,418,204 web pages

Obiettivi della lezione

- Cercare documenti su Internet
- Cos'è un motore di ricerca
- Come si interrogano i motori di ricerca
- Varietà di motori di ricerca

“Un motore di ricerca perfetto è
paragonabile alla mente di Dio”

Sergey Brin, co-inventore di Google

<http://www.technologyreview.com/articles/05/01/issue/ferguson0105.asp?p=0>

Cercare un libro

- I miei libri
- I libri dei miei amici
- I libri nella libreria sotto casa o della mia città
- I libri nelle biblioteche pubbliche d'Italia
- Ricerca sul mio PC
- Ricerca peer-to-peer
- Ricerca su Amazon
- Ricerca su Web

Interrogazioni

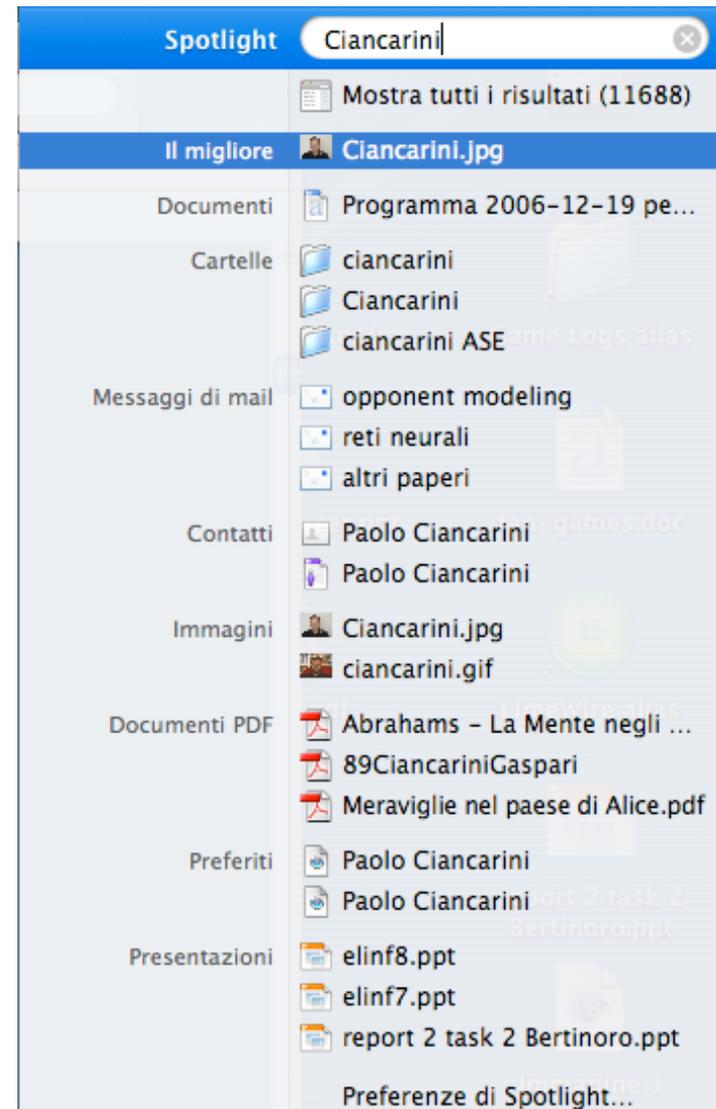
- Un'*interrogazione* (query) è una domanda, non necessariamente in forma di domanda
- Si possono interrogare i database oppure i motori di ricerca
- Esempi di interrogazione:
 - Quante canzoni dei Beatles vennero scritte da John Lennon?
 - Quali canzoni John Lennon scrisse per i quattro di Liverpool?
 - Canzone Beatles autore John Lennon
 - John Lennon autore canzoni 1960-1970

Risposte

- La risposta ad un'interrogazione può essere
 - Nessuna
 - Una risposta univoca
 - Un insieme di risposte

Cercare in un computer

- Molti sistemi operativi offrono la funzione di ricerca di un file per nome
- Spotlight su MacOs cerca anche stringhe **all'interno** di un file



Interrogare un database

- Risultati di una query al db delle riviste a cib.unibo.it

http://serials.cib.unibo.it/cgi-ser/start/it/spogli

LIBRI PERIODICI ARTICOLI BIBLIOTECHE ISSN EDICOLA

Catalogo Articoli (Spogli Riviste)

[HELP](#)

Le informazioni sugli articoli di fonte ISI sono coperte da copyright

La ricerca *find articoli where authors phrase all words 'Ciancarini P' sort by level,fasc_key/DESCEND, pagina_ini_num/ASCEND* ha restituito 13 riferimenti

Selezionare un intervallo oppure rifai la ricerca sulle Annate Precedenti

Hai selezionato spogli inseriti dal **1998** al **2005**

Per ulteriori informazioni selezionare i riferimenti di interesse.

1. Ciancarini, P; Rossi, D; Vitali, F
[Designing a document-centric coordination application over the Internet](#)
INTERACTING WITH COMPUTERS, 13(6), 2001, pp. 677 - 693
2. Ciancarini, P; Franze, F; Mascolo, C
[Using a coordination language to specify and analyze systems containing mobile components](#)
ACM TRANSACTIONS ON SOFTWARE ENGINEERING AND METHODOLOGY, 9(2), 2000, pp. 167 - 198
3. Ciancarini, P; Giovannini, A; Rossi, D
[Mobility and coordination for distributed Java applications](#)
ADVANCES IN DISTRIBUTED SYSTEMS, 1752, 2000, pp. 402 - 425
4. Ciancarini, P; Fogli, D; Gaspari, M
[A declarative coordination language](#)
COMPUTER LANGUAGES, 26(2-4), 2000, pp. 125 - 163
5. Ciancarini, P; Vitali, F; Mascolo, C
[Managing complex documents over the WWW: A case study for XML](#)
IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 11(4), 1999, pp. 629 - 638
6. Ciancarini, P; Mascolo, C
[Using formal methods for teaching software engineering: a tool-based approach](#)
ANNALS OF SOFTWARE ENGINEERING, 6, 1998, pp. 433 - 453
7. CIANCARINI P; RIZZI A; VITALI F
[AN EXTENSIBLE RENDERING ENGINE FOR XML AND HTML](#)
Computer networks and ISDN systems, 30(1-7), 1998, pp. 225 - 237
8. ARBAB F; CIANCARINI P; HANKIN C
[SPECIAL ISSUE ON COORDINATION LANGUAGES FOR PARALLEL PROGRAMMING](#)
Parallel computing, 24(7), 1998, pp. 987 - 988
9. ARBAB F; CIANCARINI P; HANKIN C
[COORDINATION LANGUAGES FOR PARALLEL PROGRAMMING](#)
Parallel computing, 24(7), 1998, pp. 989 - 1004
10. CIANCARINI P; GORRIERI R
[COORDINATION MODELS, LANGUAGES AND ARCHITECTURES](#)
Science of computer programming, 31(2-3), 1998, pp. 175 - 177
11. CIANCARINI P; MAZZA M; PAZZAGLIA L
[A LOGIC FOR A COORDINATION MODEL WITH MULTIPLE SPACES](#)
Science of computer programming, 31(2-3), 1998, pp. 231 - 261
12. CIANCARINI P; TOLKSDORF R; VITALI F; ROSSI D; KNOCHÉ A
[COORDINATING MULTIAGENT APPLICATIONS ON THE WWW - A REFERENCE ARCHITECTURE](#)
IEEE transactions on software engineering, 24(5), 1998, pp. 362 - 375
13. CIANCARINI P; CIMATO S; MASCOLO C
[ENGINEERING FORMAL REQUIREMENTS - AN ANALYSIS AND TESTING METHOD FOR Z-DOCUMENTS](#)
ANNALS OF SOFTWARE ENGINEERING, 3, 1997, pp. 189 - 219

Tempo impiegato dalla query: 0.23 secondi.
Tempo occorso per la creazione del report: 0.06 secondi.
Tempo totale: 0.29 secondi.

Operatori booleani nelle query

- Autore = Rossi **and** editore = Bianchi
- Autore = Rossi **or** editore = Bianchi
- Autore = Rossi **and not** editore = Bianchi

Cercare in un database vs cercare nel Web

- La ricerca in un database (di solito memorizzato su un singolo computer) consiste nell'analizzarne il contenuto per scoprire quali dati soddisfano l'interrogazione
- Idealmente, questa definizione vale anche per il Web
- Ma il Web è molto grande e distribuito su milioni di computer

Cercare nel Web

- Il Web è grande
 - Google indicizza decine di miliardi di documenti (alla fine del 2006)
 - Alcune stime dicono che conosce meno del 20% del Web
 - Ancora più enorme è l'universo dei dati e documenti che NON sono su Web (ma su computer connessi a Internet)
- Il Web è dinamico
 - Vengono aggiunte milioni di pagine quotidianamente
 - Circa un quarto delle pagine esistenti cambiano quotidianamente
 - Nel dominio .com, i siti in media cambiano il 50% delle pagine ogni 10 giorni
- Il Web è aperto
 - Tutti possono pubblicare in ogni momento
 - Non tutto il Web è raggiunto dai motori di ricerca
- Il Web ha un grande valore, e cresce nel tempo
 - Ogni giorno centinaia di milioni di persone interrogano Google
 - Quel che non c'è oggi, può esserci domani; quel che c'è oggi, rimane comunque da qualche parte anche se viene cancellato dall'autore

Cosa si trova con un motore

- Pagine HTML e altro
 - Ma non tutto quel che c'è sul Web (e tanto meno su Internet)
- Tipi di file
 - Parecchi, anche multimediali
 - Ma non tutti
- Tantissimi duplicati

Come si cerca un documento

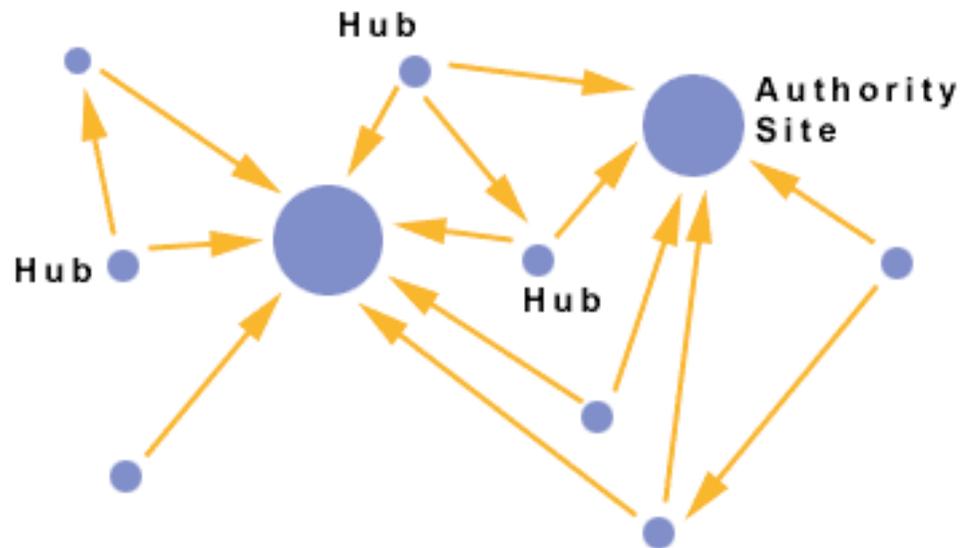
- **Documento** = *sequenza* di parole e link
- **Interrogazione** = *insieme* di parole, per ottenere uno o più link di pagine "rilevanti"
- **Risposta** = *insieme* di pagine che "soddisfano" l'interrogazione

Come si cerca un'immagine

- **Immagine** ~ *documento non testuale*
- **Interrogazione** = *insieme* di parole, per ottenere uno o più link di immagini "rilevanti"
- **Risposta** = *insieme* di immagini che "soddisfano" l'interrogazione

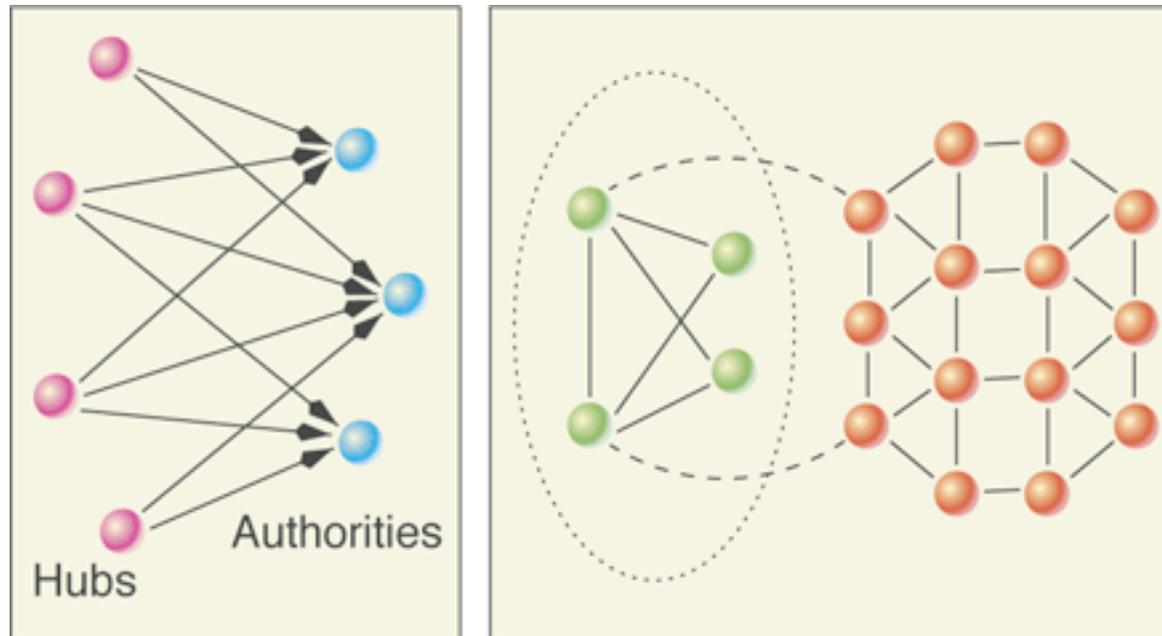
La forma del Web

- Il Web è un "grafo non connesso"
- **Hub**: pagina che punta a molte autorità
- **Autorità**: pagina puntata da molti hub



Cluster (insiemi) di pagine

- Un insieme di pagine sullo stesso argomento ha di solito pagine fortemente connesse: questo fatto si può sfruttare per identificarlo (clusterizzazione)
- Un altro metodo sfrutta schemi tipici di hub e autorità



Tag HTML per aiutare i motori

<HEAD> L'elemento delimitato dai tag <HEAD> </HEAD> contiene il *nome del documento* e altri dati utili di classificazione

<TITLE> L'elemento delimitato dai tag <TITLE> </TITLE> definisce il *titolo della pagina*, visualizzato dal browser nell'intestazione di pagina. In caso di salvataggio lato browser il tag TITLE dà il nome al file. Il contenuto di questo elemento è usato da molti motori di ricerca per indicizzare la pagina e ricavare parole chiave di accesso.

<META> Un elemento denotato dal tag META contiene *suggerimenti* ai motori di ricerca su come catalogare la pagina:

```
<META name="keywords" Content="varie, parole, chiavi">
```

```
<META name="author" content="Paolo Ciancarini">
```

```
<META name="generator" content="Word2004">
```

Ricerche su Web

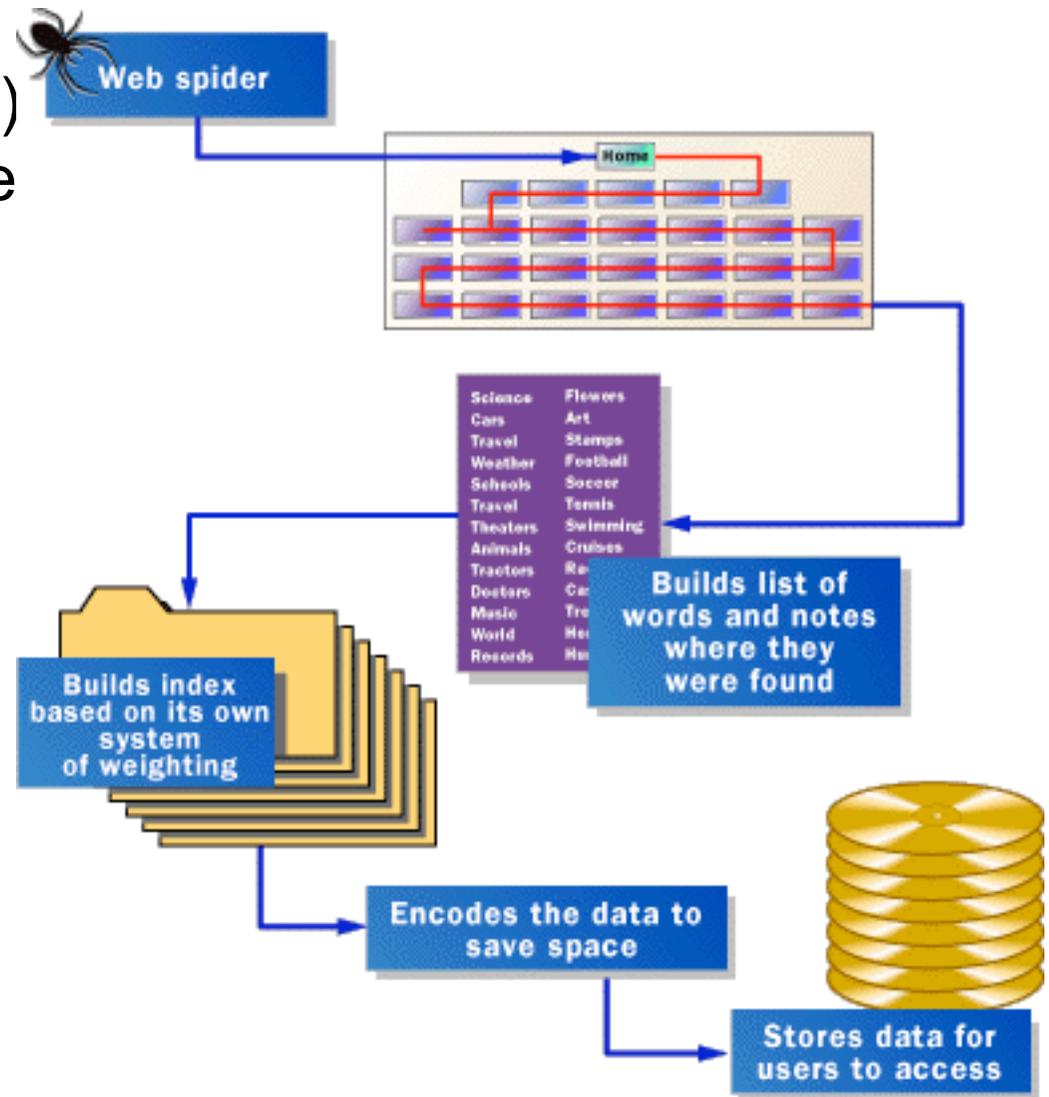
- Conoscere un motore di ricerca (specifico)
 - Interfaccia: Comandi base e avanzati
 - Funzioni: Cosa si può fare (o forse no)
 - Copertura: Cosa c'è e cosa non c'è
- Conoscere i (vari) motori di ricerca
 - Diversi domini di documenti
 - Diversi modelli di catalogo e di risposte

www.philb.com/engines.htm

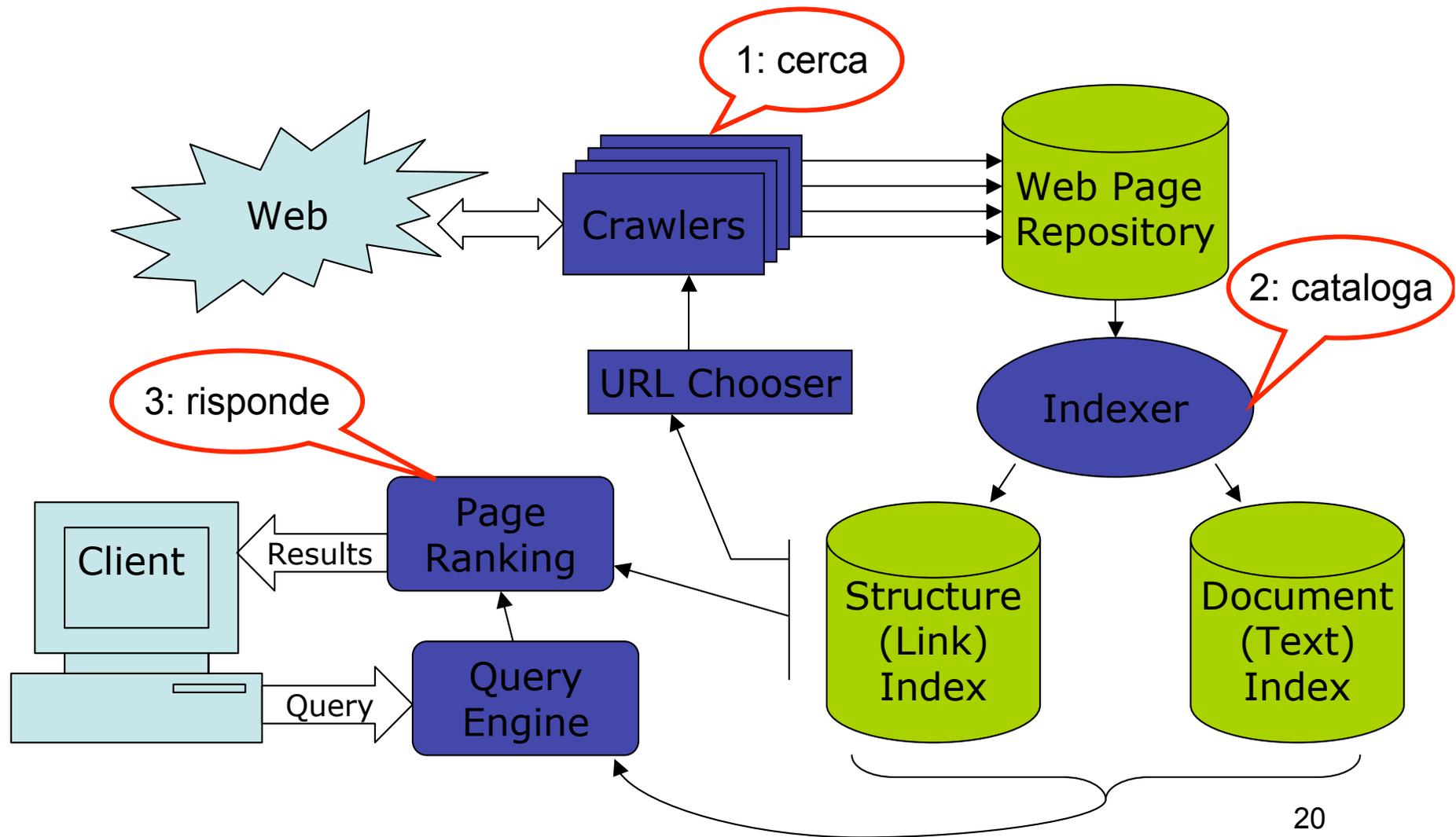
Motore di ricerca

Un motore di ricerca su Web

1. Naviga il Web (**crawling**) e legge e copia le pagine che trova
2. Classifica le pagine trovate (**indexing**)
3. Risponde alle interrogazioni (**answering**)



Architettura di un motore

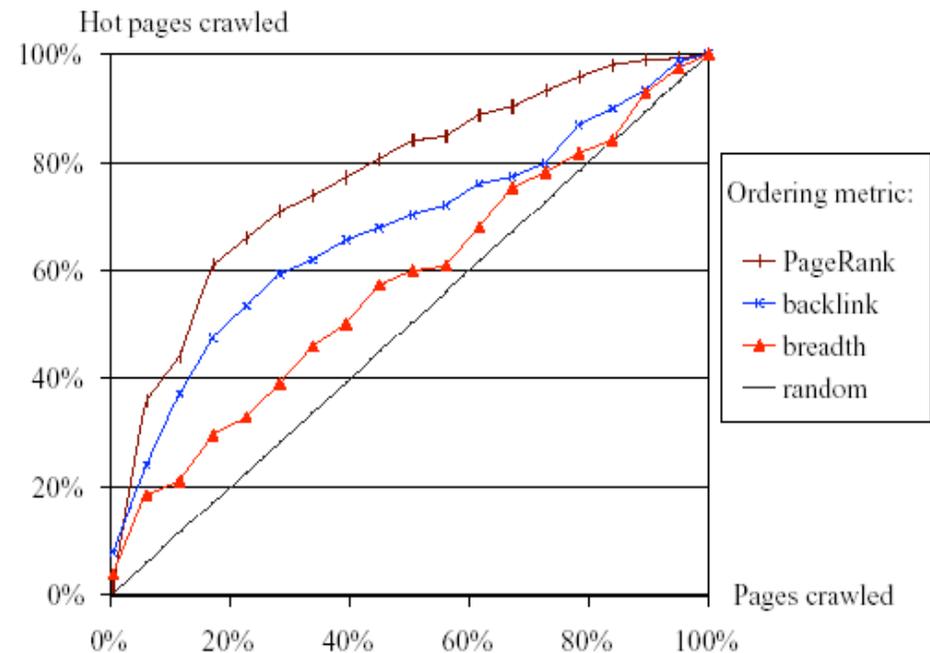


Crawling

- Il crawling consiste nel visitare il maggior numero di pagine Web
- Inizia da un insieme di pagine "radice"
- Dai link in queste pagine si raggiungono nuove pagine
- Le pagine vengono copiate in un database
- Quando si ferma il crawling?
 - Quando non ci sono più pagine
 - Dopo che un certo numero di pagine viene raggiunto
 - Dopo che viene raggiunta una certa "profondità"
- Quando si ricaricano le pagine?
 - Le pagine che cambiano più di frequente vengono visitate più spesso
 - Tutte le pagine vengono visitate con la stessa frequenza

La selezione delle pagine

- L'ordine di crawling è importante
- L'insieme delle URL da visitare viene pre-ordinato
- Metriche di ordinamento:
 - Popolarità (PageRank)
 - Contesto del sito: es. .com più importante di .biz
 - Query driven: quanto una pagina soddisfa certe interrogazioni predefinite



Indexing (catalogazione)

- Ipotesi: la rilevanza di un documento per chi lo cerca è correlata ai termini che contiene
- Un documento d si può rappresentare con una sequenza di termini: $d = ((t_1, k_1), (t_2, k_2) \dots (t_n, k_n))$ dove t_i indica un termine della pagina e k_i il numero di volte che t_i si trova in d
- I termini vanno pesati, perché quelli che ricorrono in molti documenti (es. “the”) sono meno rilevanti dei termini che appaiono in pochi documenti

Catalogazione

- Nelle pagine Web oltre ai termini del testo ci sono altri elementi: i tag di formattazione HTML si usano per pesare l'importanza dei termini che contengono
 - Termini nei titoletti e in neretto più importanti del testo di paragrafo
 - Inclusione del testo dei link che puntano alla pagina, che spesso ne danno una buona descrizione
 - Catalogo: indice invertito, cioè insieme di liste invertite, una per ciascuna parola
 - Associa ogni parola ad una lista ordinata di locazioni (pagina e posizione entro la pagina)

Problemi

- Problema dell'abbondanza
 - Il Web è troppo grande, e ad ogni interrogazione vengono restituite troppe pagine (provate a giocare a Googlewack!)
- Problema dell'irrilevanza
 - Il Web contiene un sacco di immondizia
 - Alcune pagine dicono bugie o contengono termini per il solo scopo di ingannare i motori di ricerca
- Molte pagine non sono autodescrittive
 - Gran parte delle pagine NON descrivono se stesse
 - Alcune pagine contengono pochi termini
 - Alcune pagine contengono termini non correlati con il loro contenuto
- Manipolazioni
 - Alcuni autori alterano le pagine per qualche scopo di solito commerciale in modo da far sì che i motori di ricerca le cataloghino in modo speciale

Answering

- Anche un'interrogazione q si può rappresentare con una sequenza di termini: $q = (t_1', t_2', \dots t_m')$
- Quindi una risposta è un insieme di documenti D "*abbastanza somiglianti*" a q
- Ipotesi: un documento d abbastanza somigliante a q soddisfa l'autore dell'interrogazione q
- Misure di similarità usate per restituire documenti ad un'interrogazione
 - Distanza euclidea: prodotto scalare dei vettori $d \circ q$
 - Coseno: angolo tra i due vettori

Ordinamento (ranking) delle risposte

- Il motore come ordina le pagine in risposta ad una interrogazione?
- Assunzioni sui link:
 - Un link dalla pagina A alla pagina B è una raccomandazione della pagina B da parte dell'autore di A
 - Se due pagine sono connesse da un link, probabilmente riguardano lo stesso argomento
- Ordinamento dei link
 - Indipendente dall'interrogazione: ogni pagina ottiene un voto in base ai collegamenti che la puntano
 - PageRank
 - Dipendente dall'interrogazione
 - HITS

PageRank

- Idea: le pagine cui puntano molti link sono migliori
 - Ma non tutti i link sono uguali
 - I link di una pagina importante debbono contare di più di quelli di una pagina poco importante
- PageRank base $R(A)$ della pagina A :

$$R(A) = \sum_{(B,A) \in G} \frac{R(B)}{\text{outDegree}(B)}$$

- $\text{outDegree}(B)$ = numero di link su B
 - La pagina B contribuisce al voto di tutte le pagine cui punta
- Modello del navigatore casuale
 - Una persona naviga casualmente seguendo i link di pagina in pagina
 - Il PageRank di A è proporzionale alla frequenza con cui il navigatore casuale incontra A

Google Bombing

- **Google Bombing**: Creazione di numerose pagine che puntano ad una pagina specifica, in modo che certe interrogazioni ottengano quella pagina come risposta
- Esempio: "Miserabile fallimento"
- Questo "attacco" funziona sfruttando il Page Ranking

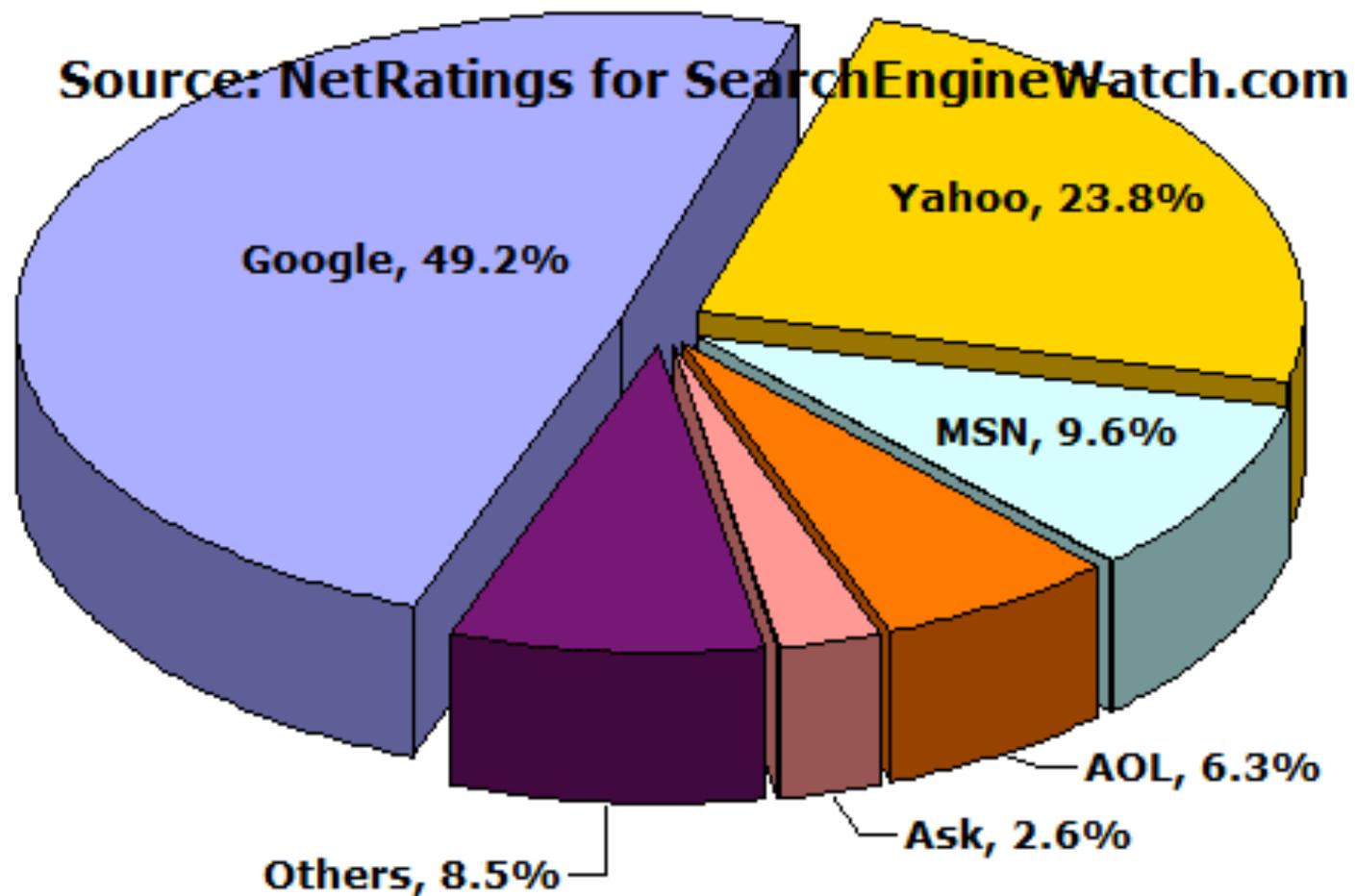
Come mai motori diversi danno risposte diverse?

- Diverso crawling (con tempi e punti di partenza diversi)
- Diverso indexing
- Diverso answering (es. PageRank è un brevetto degli inventori di Google)
- Tempi diversi di interrogazione (il Web cresce continuamente!)

Partizionare il database

- Un'altra tecnica per migliorare la precisione di answering consiste nel partizionare il database delle pagine, e offrire agli utenti motori “specializzati”
- Esempio: Google offre
 - Google Scholar
 - Google.it, google.fr, ecc.
 - Google News
 - Google Groups
 - Google Print

Popolarità dei motori (utenti USA, 7-06)





- Google.com indicizza (fine 2006)
 - Oltre 5×10^{10} documenti
 - Html, txt, pdf, doc, ps, ppt
 - Oltre 1×10^9 immagini
 - PNG, GIF, JPEG
 - Oltre 1×10^9 articoli di newsgroup
 - Tutto USENET dal 1980
 - Oltre 4500 fonti in linea di notizie
 - Lasso temporale gratuito: un mese

www.googleguide.com/feature_history.html

L'interfaccia di Google

The image shows a screenshot of the Google search interface with various elements annotated with letters A through P. The search bar contains the text "google address +mountain view". The search results are displayed below the search bar, including a tip, news results, and a search result for "Google Corporate Information: Google Offices".

A - Navigation menu (Web, Images, Groups, News, Froogle, Local, more »)

B - Search button

C - Advanced Search Preferences link

D - Google logo

E - Search bar

F - Search results header (Web, Results 1 - 10 of about 87,500 for google address +mountain view. (0.31 seconds))

G - Tip: Find maps by searching for a street address with city or zip code

H - News results for google

I - Google logo

J - Footer (Advertise with Us - Business Solutions - Services & Tools - Jobs, Press, & Help ©2004 Google - Searching 4,285,199,774 web pages.)

K - Google Corporate Information: Google Offices

L - 1600 Amphitheatre Parkway Mountain View, CA 94043. ... Turn right onto CHARLESTON Drive past Landings Drive; Turn left into driveway at the "Google" sign at ...

M - http://www.google.com/corporate/address.html - 15k - Cached - Similar pages

N - [More results from www.google.com]

O - [More results from www.google.com]

P - [More results from www.google.com]

Articolazione dei DB Google

- Web
 - Tutto il mondo
 - Italia
 - Pagine in Italiano
- Images
- Groups
 - Tutti
 - Solo quelli in Italiano
- News
 - 22 versioni "regionali"
- Froogle - per cercare merci
- Local - per localizzare geograficamente le ricerche
- altro

Google News

- Le news di Google hanno uno span temporale di 30 giorni
- Versioni regionali: Argentina, Australia, Austria, Canada (English), Canada (French), China, Chile, France, Germany, Hong Kong, India, Italy, Japan, Korea, Mexico, New Zealand, Spain, Switzerland (French), Switzerland (German), Taiwan, United Kingdom, United States

Google Books books.google.com

- Google indicizza interi libri e risponde alle query anche usando il loro contenuto
- Indicizza libri sotto diritto d'autore: accesso solo a parti limitate
- Indicizza libri fuori diritto di autore: accesso completo (e a volte scaricamento gratuito!)

Prodotti aggiuntivi di Google

- Google Toolbar
- Google Desktop
- Google API
- Google Earth

Interrogare Google

Query	Google.com	Google.it/web	Google.it/italiano	Google.it/Italia
Paolo Ciancarini	14.900	15.100	1.140	1.290
"Paolo Ciancarini"	12.500	12.700	622	816
+Paolo +Ciancarini	15.100	15.100	1.150	1.290
Paolo OR Ciancarini	8.350.000	8.350.000	2.960.000	2.630.000
Filetype:pdf Paolo Ciancarini	461	461	77	138

1 Maggio 2005

Stemming

- Google usa lo stemming
- Stemming: cercare non solo le parole della query, ma anche per parole "simili"
- Esempio: se la query è ***pet lemur dietary needs***, Google cerca anche ***pet lemur diet needs***, e altre varianti
- Ogni variante cercata viene segnalata nei risultati

Usi "strani" di Google

- `labs.google.com`
- `www.google.com/press/zeitgeist.html`
- `douweosinga.com/projects/googletalk`
- `douweosinga.com/projects/visualpoetry/`

Come fa soldi Google

- AdWords Select
- Google Search Appliance
- Google Answers
- Pubblicità Content-Specific

Polemiche su Google

- Soggetto a filtraggio in Cina
- Censura di alcuni siti su google.fr/de
 - `cyber.law.harvard.edu/filtering/google`
- Problemi di privacy col governo americano
- Monopolio e predominio eccessivo

Altri motori

- Google è grande, ma non completo
- Altri motori offrono contenuti diversi e funzioni che Google non ha
- FAST: Lycos & AlltheWeb
- AltaVista
- Inktomi: MSN Search & Hotbot
- Teoma

Confronto su stessa query

Query "ciancarini" (2/5/2005 e 19/1/2006)

- Google: 28.300 - 96.800
- Yahoo: ??? - 50.300
- Altavista: 13.500 - 49.300
- Alltheweb: 12.300 - 45.000
- Teoma: 7.420 - 8.110
- MSN Search: ??? - 4.887

Contare le pagine

- I motori non sanno contare
- I numeri cambiano continuamente
 - Da una pagina all'altra
 - Da un minuto al successivo
- Se occorre fare stime, controllare più motori e provare più volte

Funzioni dei motori

Search Engines	Boolean	Default	Proximity	Truncation	Case	Fields	Limits	Stop	Sorting
Google Review	-, OR	and	Phrase, GAPS	No, but stemming, word in phrase	No	intitle, inurl, link, site, more	Language, filetype, date, domain	Varies, + searches	Relevance, site
Yahoo! Review	AND, OR, NOT, (), -	and	Phrase	No	No	intitle, url, site, inurl, link, more	Language, file type, date, domain	Very few	Relevance, site
Ask Jeeves/Teoma Review	-, OR	and	Phrase	No	No	intitle, inurl	Language, site, date	Yes, + searches	Relevance, metasites
MSN Search Review	AND, OR, NOT, (), -	and	Phrase	No	No	link, site, loc, url	Language, site	Varies, + searches	Relevance, site, sliders
WiseNut Review	- only	and	Phrase	No	No	No	Language	Yes, + searches	Relevance, site
Gigablast Review	AND, OR, AND NOT, (), +, -	and	Phrase	No	No	title, site, ip, more	Domain, type	Varies, + searches	Relevance
Exalead Review	AND, OR, NOT, (), -	and	Phrase, NEAR	Yes	No	title	Language, file type, date, domain	Varies, + searches	Relevance, date

Strategia di interrogazione

- Usare prima interrogazioni semplici
- Spesso bastano due parole
- Poi passare alle frasi
- Se le interrogazioni semplici falliscono
 - Usare le funzioni avanzate
 - Escogitare interrogazioni complesse
- Non c'è penalità per interrogazioni reiterate!

Ricerca per frase

- Usare più parole
 - In passato, il default era OR e le risposte erano date in ordine di ranking su entrambi gli operandi
- Oggi il default è sempre AND
- Il segno "-" equivale sempre a NOT
- La sintassi della funzione OR varia da motore a motore

Google: operatori su pagine

www.googleguide.com/advanced_operators.html

- Inclusione di stringa: "+ *stringa*"
- Sinonimi: "~ *stringa*"
- "*a* OR *b*": unico operatore in MAIUSCOLO
- Intervallo numerico: "*N..M*"
- Pagine che puntano una certa url "link:*url*"

Google: operatori di restrizione

- Restrizione di sito: "site:*domain*"
- Restrizione di tipo di documento: "filetype:*tipo*" (pdf,ps,ppt,doc e altri)
- "related:*url*" pagine simili a quella denotata da *url*
- Allintitle,allintext,inurl,intitle
- "cache:*url*": accede la pagina in cache
- "info:*url*": informazioni disponibili su *url*

OR

- Google: OR maiuscolo
 - `yellowstone bison OR buffalo`
- AlltheWeb: usare ()
 - `yellowstone (bison buffalo)`
- AltaVista: normale
 - `yellowstone AND (bison OR buffalo)`
- Gigablast: normale
 - `yellowstone AND (bison OR buffalo)`
- Teoma
 - `yellowstone bison OR buffalo`
 - `Becomes (yellowstone AND bison) OR buffalo`

Ricerche con campi

- titolo (`title: o intitle:`)
- collegamento (`link:`)
- Sito (`site:`)
- Tipo file (`filetype:`)
- data
 - Non accurato, di solito non rappresenta la data di pubblicazione
 - Serve a tagliar via materiale più vecchio
 - O per trovare documenti più antichi

Tipi di file

PDF	AlltheWeb, AltaVista, Gigablast, Google, MSN Search
MS Word	AlltheWeb, Gigablast, Google, MSN Search
Flash	AlltheWeb
PowerPoint	AlltheWeb, Gigablast, Google, MSN Search
Excel	AlltheWeb, Gigablast, Google, MSN Search
PostScript	AlltheWeb, Gigablast, Google
WordPerfect	AlltheWeb, Google
StarOffice	AlltheWeb

Aspetti temporali

- www.archive.org
 - Tutto il Web dal 1996 a oggi
 - Wayback Machine:motore di ricerca temporale
- Google
 - Da ieri a tre mesi fa, senza indicazione di data
- Gigablast
 - Da ieri a un anno fa, con indicazione di data di cache
- ScrubTheWeb, IncyWincy, & More at
 - searchengineshowdown.com/others/archive.shtml

Note some duplicates are not shown. [See all.](#)

* denotes when site was updated.

Search Results for Jan 01, 1996 - May 03, 2005

1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
0 pages	3 pages	4 pages	11 pages	22 pages	41 pages	14 pages	33 pages	39 pages	0 pages
	Jan 28, 1997 * Oct 15, 1997 * Dec 11, 1997 *	Jan 23, 1998 Feb 08, 1998 May 01, 1998 Dec 12, 1998 *	Jan 25, 1999 * Jan 28, 1999 Feb 08, 1999 * Feb 09, 1999 Mar 02, 1999 Apr 22, 1999 * Apr 24, 1999 Apr 29, 1999 Oct 12, 1999 * Oct 13, 1999 Oct 23, 1999	Apr 08, 2000 * Apr 24, 2000 May 05, 2000 May 10, 2000 May 12, 2000 * May 19, 2000 * Jun 17, 2000 * Jun 21, 2000 Aug 15, 2000 Aug 19, 2000 Aug 29, 2000 Aug 31, 2000 Oct 17, 2000 Oct 18, 2000 * Oct 19, 2000 * Oct 25, 2000 Oct 27, 2000 * Oct 27, 2000 * Nov 09, 2000 * Nov 19, 2000 Dec 05, 2000 * Dec 11, 2000	Apr 08, 2001 * Feb 01, 2001 * Feb 08, 2001 Feb 26, 2001 * Mar 01, 2001 Mar 02, 2001 Apr 03, 2001 Apr 04, 2001 Apr 05, 2001 Apr 07, 2001 Apr 10, 2001 Apr 11, 2001 Apr 12, 2001 Apr 13, 2001 Apr 14, 2001 Apr 17, 2001 Apr 18, 2001 Apr 19, 2001 Apr 20, 2001 Apr 21, 2001 Apr 22, 2001 Apr 24, 2001 Apr 28, 2001 Apr 29, 2001 Apr 30, 2001 May 01, 2001 May 02, 2001 May 03, 2001 May 04, 2001 May 05, 2001 May 06, 2001 May 07, 2001 May 08, 2001 May 09, 2001 May 11, 2001 May 15, 2001 May 16, 2001 Jun 04, 2001 Jul 21, 2001 * Sep 26, 2001 * Nov 27, 2001 *	Feb 09, 2002 * Jun 02, 2002 Jul 19, 2002 * Aug 01, 2002 * Sep 15, 2002 * Sep 21, 2002 Sep 26, 2002 Oct 16, 2002 Oct 16, 2002 * Nov 22, 2002 * Nov 24, 2002 Nov 25, 2002 Nov 28, 2002 Dec 07, 2002 *	Feb 02, 2003 * Feb 16, 2003 Feb 18, 2003 Mar 09, 2003 * Mar 21, 2003 * Apr 01, 2003 Apr 04, 2003 Apr 08, 2003 Apr 19, 2003 Apr 19, 2003 * Apr 22, 2003 * May 24, 2003 May 30, 2003 Jun 14, 2003 * Jun 24, 2003 Jul 23, 2003 * Aug 04, 2003 Aug 07, 2003 Aug 10, 2003 * Sep 21, 2003 * Oct 07, 2003 Oct 08, 2003 * Oct 19, 2003 * Oct 23, 2003 Oct 26, 2003 Nov 21, 2003 Nov 23, 2003 Nov 29, 2003 Nov 29, 2003 * Dec 01, 2003 * Dec 05, 2003 * Dec 07, 2003 Dec 11, 2003 *	Jan 30, 2004 * Feb 02, 2004 * Mar 21, 2004 * Mar 24, 2004 May 05, 2004 * Jun 04, 2004 * Jun 04, 2004 * Jun 05, 2004 * Jun 06, 2004 Jun 08, 2004 Jun 11, 2004 Jun 12, 2004 Jun 14, 2004 Jun 22, 2004 Jun 23, 2004 Jul 10, 2004 Jul 16, 2004 Jul 21, 2004 Jul 22, 2004 Sep 21, 2004 Sep 24, 2004 Oct 12, 2004 Oct 13, 2004 * Oct 14, 2004 * Oct 15, 2004 Oct 16, 2004 Oct 19, 2004 Oct 23, 2004 Oct 26, 2004 Oct 29, 2004 Oct 31, 2004 Nov 02, 2004 * Nov 08, 2004 * Nov 11, 2004 Nov 16, 2004 Nov 21, 2004 * Nov 25, 2004 * Nov 26, 2004 Nov 28, 2004	

Motori di ricerca specializzati

Motori per ricerca scientifica

- `scholar.google.com` Google Scholar
- `citeseer.ist.psu.edu/cs`
- `scirus.landingzone.nl/other/`

- `www.filedonkey.com` Motore per reti Peer2Peer
- `www.musicplasma.com` Motore musicale
- NetDetective Motore per informazioni personali

Autonomous Citation Indexing

- Un sistema autonomo di indici di citazioni (ACI) crea un indice di citazioni (riferimenti bibliografici) a partire da un insieme di documenti digitali
- I documenti vengono cercati in rete, vengono classificati come "interessanti" mediante analisi automatica del testo, vengono identificate ed estratte le citazioni che contengono, eliminati i duplicati (anche in diversi formati), e infine vengono memorizzati i contesti in cui si trova la citazione
- Esempio: CiteSeer

CiteSeer

- CiteSeer è un motore di ricerca il cui crawler cerca articoli scientifici in rete; quando ne trova uno lo copia e lo converte in testo
- Analizza gli articoli per cercare le citazioni ed il loro contesto, che vengono indicizzati
- L'interfaccia di interrogazione dà accesso al testo completo dell'articolo in vari formati
- Permette di localizzare gli articoli o per parola chiave o per navigazione sui link di citazione

Motori specializzati su una lingua o nazione

- I motori specializzati su lingua o nazione sono di solito directory (redazioni umane che classificano le pagine)
- `www.searchenginecolossus.com`

Motori italiani

- altavista.it
- google.it
- arianna.libero.it
- lycos.it
- tiscali.it
- msn.it
- 11.200
- 5.160
- 950
- 172
- 160
- 548.867 (?)

Numero di risposte alla query “ciancarini”
su siti in Italia, 3/5/2005

Metamotore

- Un **metamotore** è un'interfaccia su più motori, che interroga simultaneamente e di cui riassume e aggrega le risposte
- Sembra una buona idea, ma non funzionano molto bene
- E' una buona idea nel caso di cataloghi di biblioteche www.ubka.uni-karlsruhe.de/hylib/en/kvk.html

Metamotori

- A9.com
- Mamma.com
- Metacrawler
- eu.ixquick.com
- Vivisimo
- Surfswax
- Dogpile
- Ask Jeeves

A9

A9 Search: [Home](#) [Prefs](#) [Toolbar](#) [Sign Out](#)
[Advanced Search](#) Hello Paolo Ciancarini. [Click here](#) if this is not you

Web Results [close](#) Showing 1 - 10 of about 3,540

Paolo Ciancarini
Find me, prof. **Paolo Ciancarini**
Dipartimento di Scienze dell'Informazione - Univ. di Bologna Mura Anteo Zamboni 7, 47127 Bologna - Italy. Tel. ...
[New] <http://www.cs.unibo.it/~cianca/> - 2k Cached [Site Info](#)

Bookmarks for Paolo Ciancarini
Bookmarks for **Paolo Ciancarini**. Info Sources. News. User Friendly the Comic Strip; l'ANSA; Metanews - L'Informazione sull'Informazione; Il Sole 24 Ore ...
[New] <http://www.cs.unibo.it/~cianca/wwwpages/bookmarks.html> - 50k Cached [Site Info](#)

Paolo Ciancarini's personal bibliography
Paolo Ciancarini's personal bibliography. This bibliography is a part of the Computer Science Bibliography Collection.
[New] <http://ilnwww.lra.uka.de/bibliography/Misc/Ciancarini.html> [Site Info](#)

DBLP: Paolo Ciancarini
... **Paolo Ciancarini**: A Perspective on Multiagent Coordination Models. ... 83, EE, **Paolo Ciancarini**, Alexander L. Wolf: Issues In Coordination Languages and ...
[New] <http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/c/Ciancarini:Paolo.html> - 52k Cached [Site Info](#)

SS > NF reviews > Paolo Ciancarini
Paolo Ciancarini. **Paolo Ciancarini**. Search Web for **Paolo Ciancarini** • Google search • Alta Vista search. Papers/Articles. A Formal approach to software ...
[New] <http://www-users.cs.york.ac.uk/~susan/bib/nf/c/palcnm.htm> - 2k Cached [Site Info](#)

Coordination technologies for internet agents
11 Paolo Ciancarini, Coordination

Book Results [close](#) [full](#) Showing 1 - 10 of about 20

Formal Methods for Open Object-Based Distributed Systems (IFIP International Federation for Information Processing)
by Paolo Ciancarini, Alessandro Fantechi, and Roberto Gorrieri (28 February, 1999)

Agent-Oriented Software Engineering
by Ireland) Aose 200 (2000 Limerick, Paolo Ciancarini, Michael J. Wooldridge, and Ireland Aose 200 2000 Limerick (01 April, 2001)

Object-based Models And Languages For Concurrent Systems: Ecoop '94 Workshop On Models And Languages For Coordination Of Parallelism And Distribution. ... ed Papers (Lecture Notes in Computer Science)
by Paolo Ciancarini, Oscar Nierstrasz, and Akinori Yonezawa (30 June, 1995)

Coordination Languages and Models: Third International Conference, Coordination '99, Amsterdam, the Netherlands, April 1999 Proceedings (Lecture Notes in Computer Science)
by Netherlands) Coordination '9 (1999 Amsterdam, Paolo Ciancarini, and Alexander Wolf (01 June, 1999)

Agent-Oriented Software Engineering II : Second International Workshop, AOSE 2001, Montreal, Canada, May 29, 2001. Revised Papers and Invited Contributions (Lecture Notes in Computer Science)
by Michael J. Wooldridge, Gerhard Wei, and Paolo Ciancarini (22 March, 2002)

Coordination Languages and Models: First International Conference, Coordination '96, Cesena, Italy, April 15-17, 1996 : Proceedings (Lecture Notes in Computer Science)
by Paolo Ciancarini and Chris Hankin (01 April, 1996)

System programming with logic languages (Technical report / Dipartimento di informatica, Università di Pisa)
by Paolo Ciancarini (1989)

Coordination languages for open system design (Technical report / Dipartimento di informatica,

Image Results [close](#) [full](#) Showing 1 - 20 of about 22



NEW [See More Columns](#)

Click to Select Search Result Columns

- Web
- Books
- Images
- Movies
- Reference
- Yellow Pages
- Your History
- Your Bookmarks
- Your Diary

Did you Know? You can search for "query" by entering the URL a9.com/query

Interfaccia visuale

- Google e altri motori restituiscono sequenze di pagine
- E' possibile aggregare le pagine in strutture astratte che vengono presentate in modo grafico
- La grafica rappresenta di solito una **rete semantica** (concetti e relazioni)

Motori a interfaccia visuale

- Kartoo.com
- Map.net
- Kwmap.net
- Mooter.com

Kartoo

The screenshot displays the Kartoo v4 web browser interface. At the top, the address bar shows "ciancarini" and navigation buttons for "ok", "it", and a globe icon. The left sidebar contains two sections: "Siti Principali" with a list of five URLs (1) www.cs.unibo.it, (2) ig.cs.unibo.it, (3) cs.unibo.it, (4) www.musicshow.it, and (5) www.unict.it; and "Temi" with a list of search terms including "università di bologna", "university of bologna", "scienze della comunicazi...", "giocatori artificiali", "modena and reggio emilia...", "italy vassilis christoph...", "institute of computer sc...", "greece paolo ciancarini", "italy costas courcoubeti...", "maggio 1998", "agenti mobili e coordina...", "paolo", "bologna", "inform", and "fabio". The main area features a network map with nodes and connections. Nodes include "www.musicshow.it", "leoncini.freesevers.com", "fabio", "www.di.unipi.it", "editors", "www.sssub.unibo.it", "paolo", "cs.unibo.it", "www.scacchiceriano.it", "giocatori artificiali", "www.chess.it", "www.cs.unibo.it", "scienze", "bologna inform", "università", "www.unict.it", "ig.cs.unibo.it", and "www.unibo.it". A large white arrow points from the "università" node towards the right. The bottom left corner shows the "KartOO v4" logo with a timer at "00:00:00" and the text "OK!". The bottom right corner features a vertical stack of flags and a green download icon.

Kartoo

The screenshot displays the Kartoo web application interface. At the top, there is a navigation bar with a search box containing the text "ciancarini", and several utility icons including "ok", "en", a globe, and smiley faces. On the left side, there are two lists: "Top Sites" and "Topics".

Top Sites:

- 1) www.amazon.fr
- 2) www.cs.unibo.it
- 3) gd.tuwien.ac.at
- 4) citeseer.ist.psu.edu
- 5) eventseer.idi.ntnu.no

Topics:

- "agent oriented software ..."
- "paolo ciancarini"
- "coordination language"
- "object based"
- "computer science"
- "models and languages"
- "coordination models"
- agent
- software
- oriented
- coordination
- amazon
- engineering
- integrators
- science

The main area of the interface is a network map with a blue background. Nodes are represented by yellow spheres of varying sizes, connected by blue lines. The nodes are labeled with website URLs and topic names. A large white arrow points from the "www.amazon.fr" node towards the right side of the map.

Nodes and Labels:

- www.informatik.uni-trier.de
- issues perspective
- models languages
- citeseer.ist.psu.edu
- coordination
- paper
- catalogue-bibliotheque.imag.fr
- portal.acm.org
- www-users.cs.york.ac.uk
- integrators
- software agent
- oriented
- italy
- www.amazon.fr
- amazon
- engineering
- www.cs.unibo.it
- www.springeronline.com
- eventseer.idi.ntnu.no
- science
- gd.tuwien.ac.at

At the bottom left, there is a logo for "KartOO v4" with a timer showing "00:00:00" and the text "OK!". On the right side, there is a vertical bar with several flags representing different countries and a green arrow icon at the bottom.

Mooter.com



Search the web: World Australia

Moot!

I like to be Mooting in Red Blue

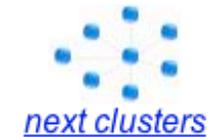
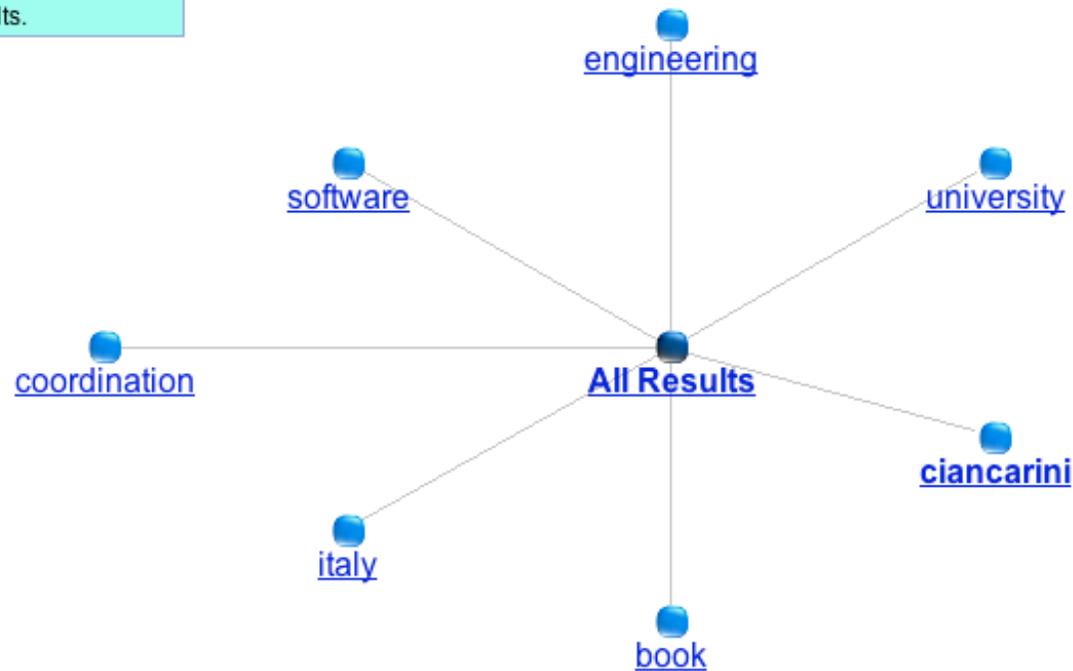
Moot, think, and be happy.

[FAQ](#)

Clusters for the search of [ciancarini](#)

Cluster page [\[1\]](#) [\[2\]](#) [\[3\]](#)

Select the cluster that best describes what you are looking for. This will lead to relevant results.



[next clusters](#)



[I want it ALL!](#)

Esempio di domanda del Web test

- Quali delle seguenti funzioni sono tipiche di un motore di ricerca?
 - Crawling 
 - Indexing 
 - Answering 
 - Ranking 
 - Downloading

Esempio di domanda del Web test

- Se pubblico una pagina Web, Google la cataloga di sicuro?
 - Sì, entro un mese
 - No, se la pagina è crittografata
 - Non è detto: occorre che la pagina sia collegata ad una pagina che Google conosce
 - Non è detto: occorre che la pagina sia collegata da una pagina che Google conosce



Domande?

