

Motori di Ricerca

presente e futuro prossimo

Valutazione dei Risultati

Paolo Ferragina, Università di Pisa

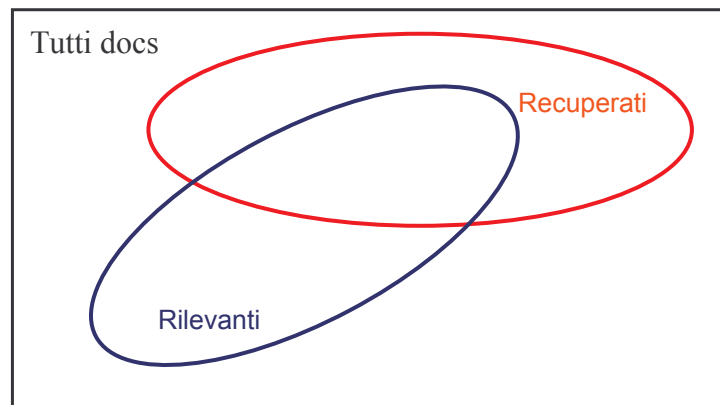
Quanto è "buono" un motore di ricerca?

Alcune misure di valutazione:

- **Costruzione:**
 - Velocità nell'indicizzazione
 - Spazio occupato dall'indice
 - Copertura del Web
- **Modifica:**
 - Frequenza e ampiezza delle modifiche
- **Interrogazione:**
 - Velocità nel produrre le risposte
 - "Rilevanza" dei risultati: *precisione e completezza*

Paolo Ferragina, Università di Pisa

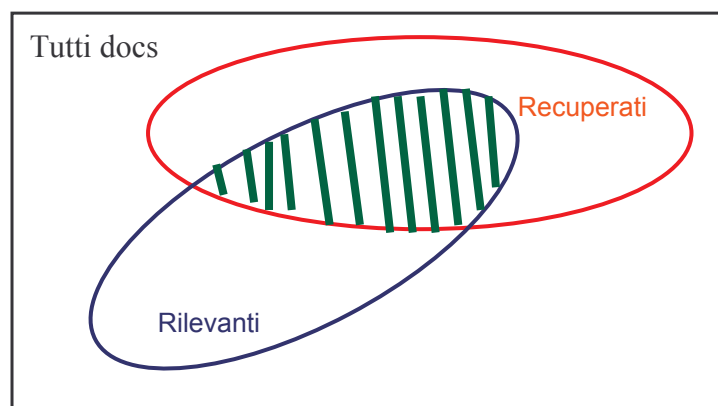
Scenario generale



Paolo Ferragina, Università di Pisa

Approccio classico: Precisione vs. Completezza

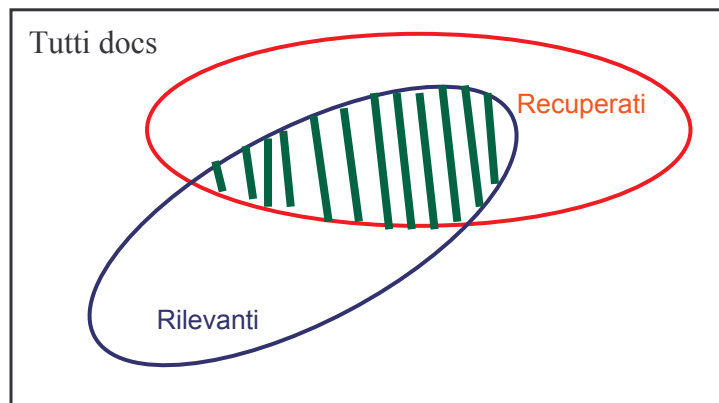
- **Precisione**: % documenti recuperati che sono *rilevanti*
 - Quanta "spazzatura" abbiamo recuperato



Paolo Ferragina, Università di Pisa

Approccio classico: Precisione vs. Completezza

- **Completezza**: % docs rilevanti che sono *recuperati*
 - Quanta "informazione" abbiamo recuperato

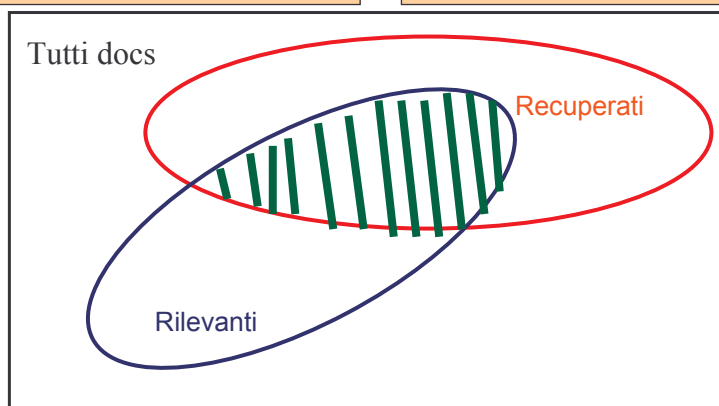


Paolo Ferragina, Università di Pisa

Precisione vs. Completezza

$$\text{Precisione} = \frac{|\text{RilRecuperati}|}{|\text{Recuperati}|}$$

$$\text{Completezza} = \frac{|\text{RilRecuperati}|}{|\text{Ril in Collezione}|}$$

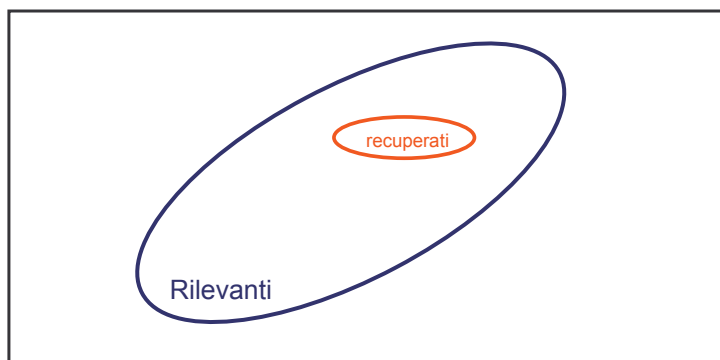


Paolo Ferragina, Università di Pisa

Precisione vs. Completezza

$$\text{Precisione} = \frac{|\text{RilRecuperati}|}{|\text{Recuperati}|}$$

$$\text{Completezza} = \frac{|\text{RilRecuperati}|}{|\text{Ril in Collezione}|}$$



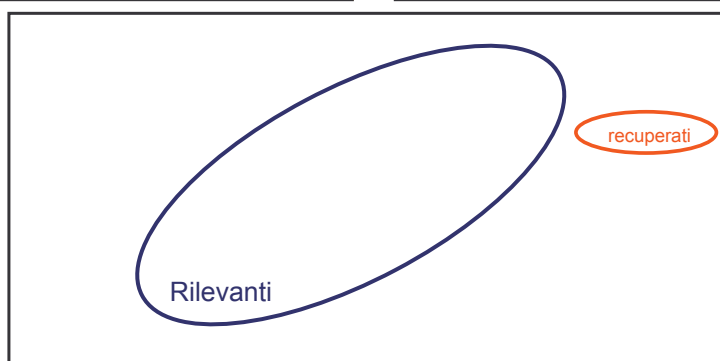
Paolo Ferragina, Università di Pisa

Altissima precisione, bassissima completezza

Precisione vs. Completezza

$$\text{Precisione} = \frac{|\text{RilRecuperati}|}{|\text{Recuperati}|}$$

$$\text{Completezza} = \frac{|\text{RilRecuperati}|}{|\text{Ril in Collezione}|}$$



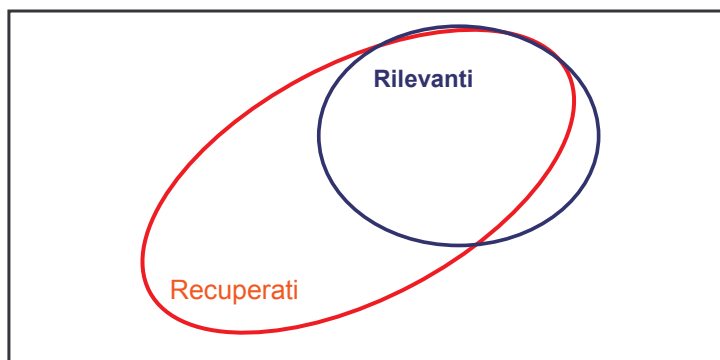
Paolo Ferragina, Università di Pisa

Bassissima precisione, bassissima completezza

Precisione vs. Completezza

$$\text{Precisione} = \frac{|\text{RilRecuperati}|}{|\text{Recuperati}|}$$

$$\text{Completezza} = \frac{|\text{RilRecuperati}|}{|\text{Ril in Collezione}|}$$



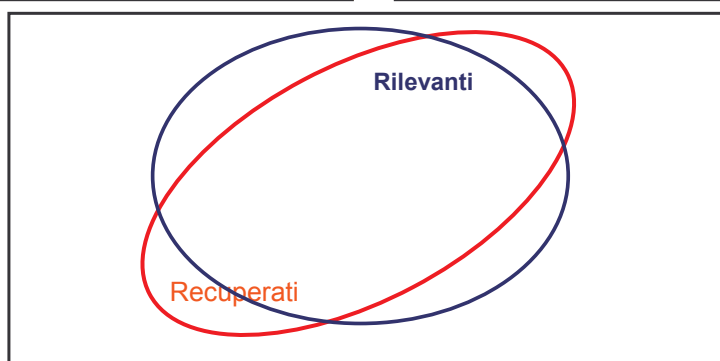
Paolo Ferragina, Università di Pisa

Alta completezza, bassissima precisione

Precisione vs. Completezza

$$\text{Precisione} = \frac{|\text{RilRecuperati}|}{|\text{Recuperati}|}$$

$$\text{Completezza} = \frac{|\text{RilRecuperati}|}{|\text{Ril in Collezione}|}$$

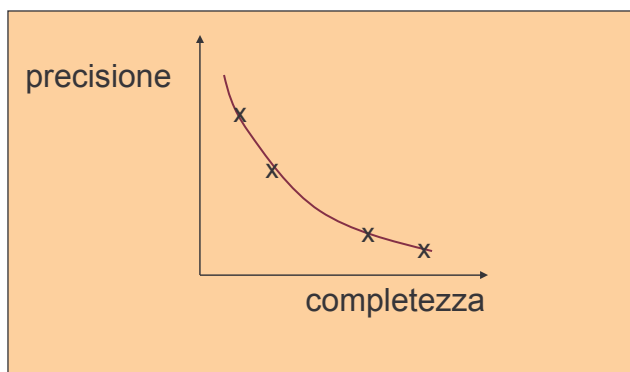


Paolo Ferragina, Università di Pisa

Alta completezza e precisione

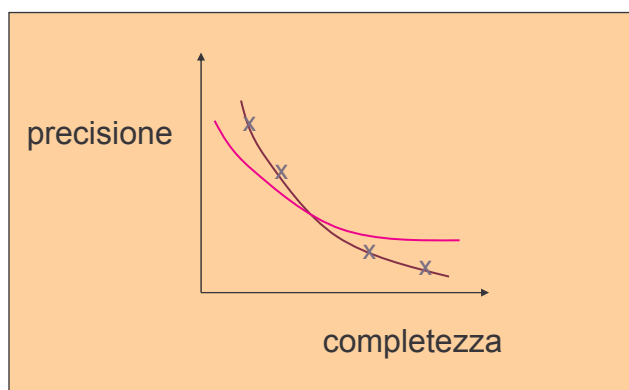
Trade-off

- Si misura la Precisione:
 - Per 10,20,30,... risultati [asse-x diverso a seconda della query]
 - Per 10%,20%,30% risultati [asse-x normalizzato]
 - a diversi livelli di Completezza [asse-x e y normalizzati]
- *Nota:* è una MEDIA su numerose interrogazioni



Paolo Ferragina, Università di Pisa

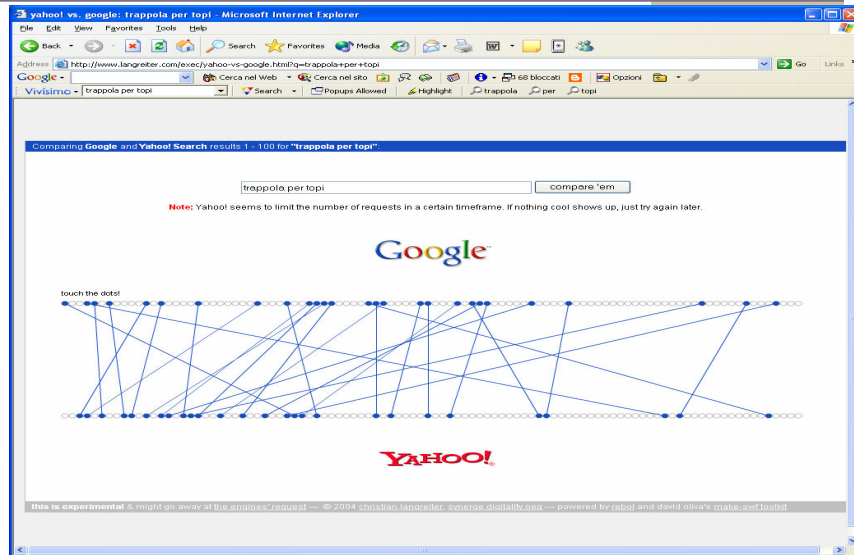
Difficoltà per il web



- ⊗ Sul Web non conosciamo la “completezza”, quindi guardiamo soltanto ai primi 10÷100 risultati. Su questi si gioca la “partita” !!

Paolo Ferragina, Università di Pisa

Ognuno sceglie il suo Ranking !



Paolo Ferragina, Università di Pisa

<http://www.langreiter.com/exec/yahoo-vs-google.html?q=...>

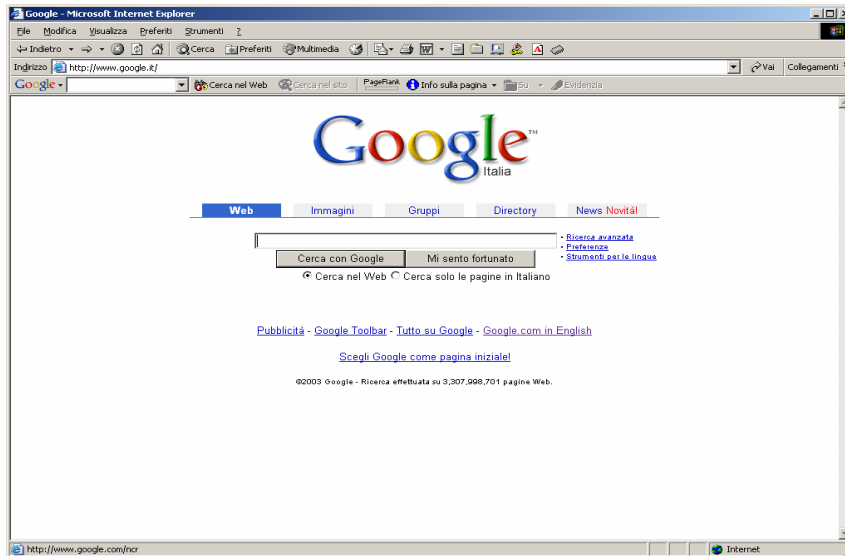
Motori di Ricerca

presente e futuro prossimo

Formulare query su Google

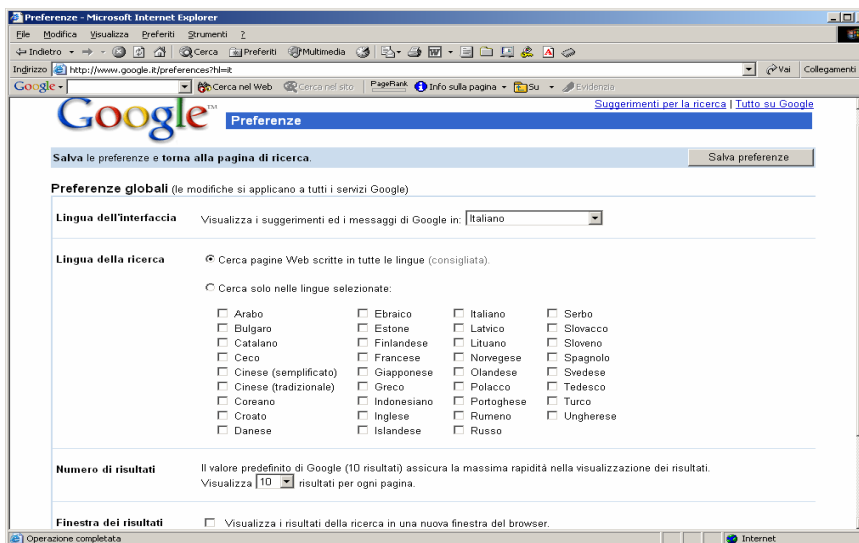
Paolo Ferragina, Università di Pisa

Una interfaccia essenziale



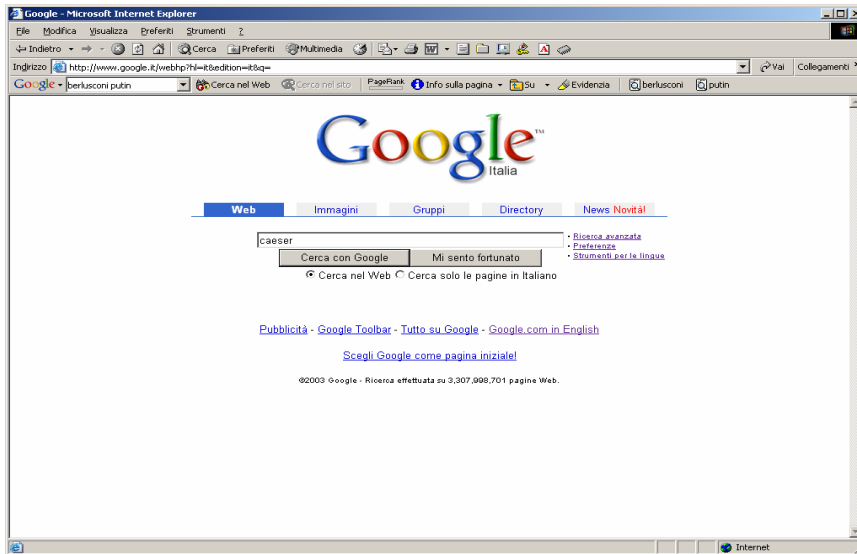
Paolo Ferragina, Università di Pisa

Le preferenze dell'utente



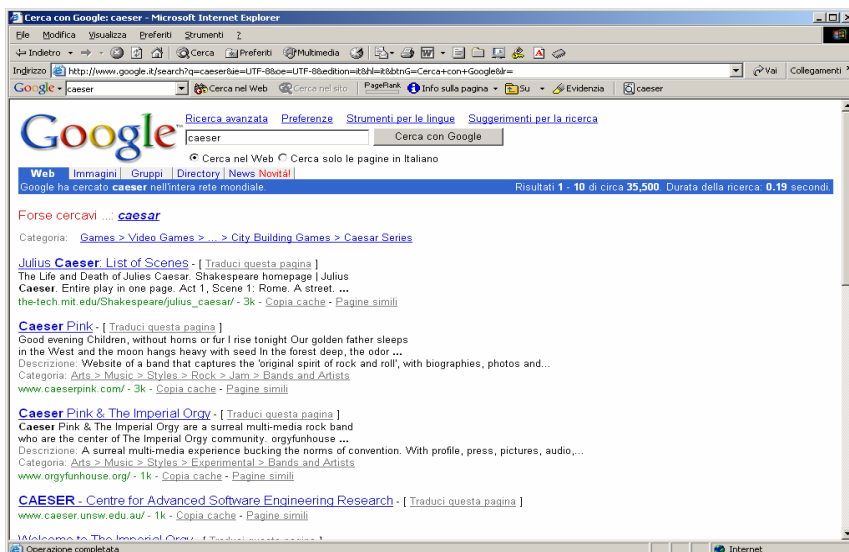
Paolo Ferragina, Università di Pisa

Una ricerca: "Presentazioni inglesi su Cesare" !



Paolo Ferragina, Università di Pisa

Correzione ortografica "caeser" → "caesar"



Paolo Ferragina, Università di Pisa

I nuovi risultati per caesar

Cerca con Google: caesar - Microsoft Internet Explorer

Indirizzo: <http://www.google.it/search?hl=it&ie=UTF-8&oe=UTF-8&edition=it&q=caesar&spell=1>

Google - caesar

Ricerca avanzata Preferenze Strumenti per le lingue Suggerimenti per la ricerca

caesar

Cerca nel Web Cerca solo le pagine in Italiano

Google ha cercato **caesar** nell'intera rete mondiale. Risultati 1 - 10 di circa 1,990,000. Durata della ricerca: 0.27 secondi

Ricerca solo i risultati in Italiano?

Ceramiche Caesar, produzione di grès fine porcellanato ...
Ceramiche Caesar nasce nel 1988 per la produzione di grès fine porcellanato, un materiale ceramico ad altissima resistenza, non smaltato, colorato direttamente ...
www.caesar.it/ - 3k - Copia cache - Pagine simili

Julius Caesar: The Last Dictator - [Traduci questa pagina]
A Biography of Gaius Julius Caesar and all major contemporaries in the Last Days of the Roman Republic. JULIUS CAESAR: THE LAST DICTATOR ...
Descrizione: Comprehensive full-length biographies of Julius Caesar, all major contemporaries (Pompey, Sulla, Marius, ...
Categoria: Society > History > ... > Ancient > Rome > People > Caesar, Julius
heraklia.kwst.com/ - 16k - Copia cache - Pagine simili

caesar - center of advanced european studies and research - ...
Descrizione: Technologie- und Forschungszentrum in Bonn. Die Seite ist bis auf Pressemitteilungen nur in englischer...
Categoria: World > Deutsch > ... > Gründer- und Technologiezentren
www.caesar.de/ - 35k - Copia cache - Pagine simili

The Julius Caesar Site - [Traduci questa pagina]
The Julius Caesar Site. Bust of Julius Caesar, from the British Museum from The Art of the Romans by HP Walters (1911) ...
Descrizione: Texts, sources and analogs, and student projects.
Categoria: Arts > Literature > ... > Works > Plays > Tragedies > Julius Caesar
www.perseus.tufts.edu/JC/ - 2k - Copia cache - Pagine simili

The Internet Classics Archive | Caesar by Plutarch - [Traduci questa pagina]

Paolo Ferragina, Università di Pisa

Raffinare le risposte: domini .edu

Cerca con Google: caesar site:edu - Microsoft Internet Explorer

Indirizzo: <http://www.google.it/search?hl=it&ie=UTF-8&oe=UTF-8&edition=it&q=caesar+site%3Aedu&hl=it>

Google - caesar

Ricerca avanzata Preferenze Strumenti per le lingue Suggerimenti per la ricerca

caesar site:edu

Cerca nel Web Cerca solo le pagine in Italiano

Google ha cercato **caesar site:edu** nell'intera rete mondiale. Risultati 1 - 10 di circa 155,000. Durata della ricerca: 0.25 secondi

The Julius Caesar Site - [Traduci questa pagina]
The Julius Caesar Site. Bust of Julius Caesar, from the British Museum from The Art of the Romans by HP Walters (1911) ...
Descrizione: Texts, sources and analogs, and student projects.
Categoria: Arts > Literature > ... > Works > Plays > Tragedies > Julius Caesar
www.perseus.tufts.edu/JC/ - 2k - Copia cache - Pagine simili

The Internet Classics Archive | Caesar by Plutarch - [Traduci questa pagina]
Caesar By Plutarch. Commentary. Many comments have been posted about Caesar. ... He was in love with Pompeia, Caesar's wife, and she had no aversion to him ...
Descrizione: Source biography for Caesar's life by Plutarch, from the Internet Classics Archive
Categoria: Society > History > ... > Ancient > Rome > People > Caesar, Julius
classics.mit.edu/Plutarch/Caesar.html - 101k - Copia cache - Pagine simili

The Internet Classics Archive | Works by Julius Caesar - [Traduci questa pagina]
Works by Julius Caesar. The African Wars Translated by WA McDevitte and WS Bohn Read discussion: 3 comments The Alexandrian Wars Translated by WA McDevitte and ...
classics.mit.edu/Browse/Browse-Caesar.html - 7k - Copia cache - Pagine simili
[Altri risultati in classics.mit.edu]

Julius Caesar: List of Scenes - [Traduci questa pagina]
The Life and Death of Julius Caesar. Shakespeare homepage | Julius Caesar. Entire play in one page ... BRUTUS's orchard Act 2, Scene 2 CAESAR's house ...
the-tech.mit.edu/Shakespeare/julius_caesar/ - 3k - Copia cache - Pagine simili

Caesar - [Traduci questa pagina]
Caesar: the lives I can find in ... Caesar was forced to forego his triumph in order

Paolo Ferragina, Università di Pisa

Comporre una query: Gli operatori booleani

- AND logico
 - paolo AND ferragina \equiv paolo ferragina
- OR logico
 - Paolo OR ferragina \equiv paolo | ferragina
- Negazione
 - ferragina -paolo
- Inclusione esplicita
 - +to +be +or +not +to +be

Paolo Ferragina, Università di Pisa

Comporre una query: Frasi e 10-word limit

- Phrase search
 - "to be or not to be"
- 10-word limit
 - # totale di parole e *token speciali* inferiore a 10
- Simbolo * di *wildcard* sostituisce parole intere
 - "to be * * to be"
 - Aiuta a saltare parole non filtranti
 - Aiuta a cercare frasi famose, discorsi, canzoni, ...
 - Aiuta a rientrare nel limite di 10

Paolo Ferragina, Università di Pisa

Comporre una query: i *token speciali*

- **intitle:parola** [anche allintitle:]
 - Cerca la *parola* nel titolo di una pagina
- **inanchor:parola** [anche allinanchor:]
 - Cerca la *parola* nel testo di un hyperlink
- **site:hostname**
 - site:rutgers.edu cerca nei siti dell'Università di Rutgers
- **inurl:stringa** [anche allinurl:]
 - inurl:help trova per esempio http://...../help/...../...

Paolo Ferragina, Università di Pisa

Comporre una query: i *token speciali* (contd)

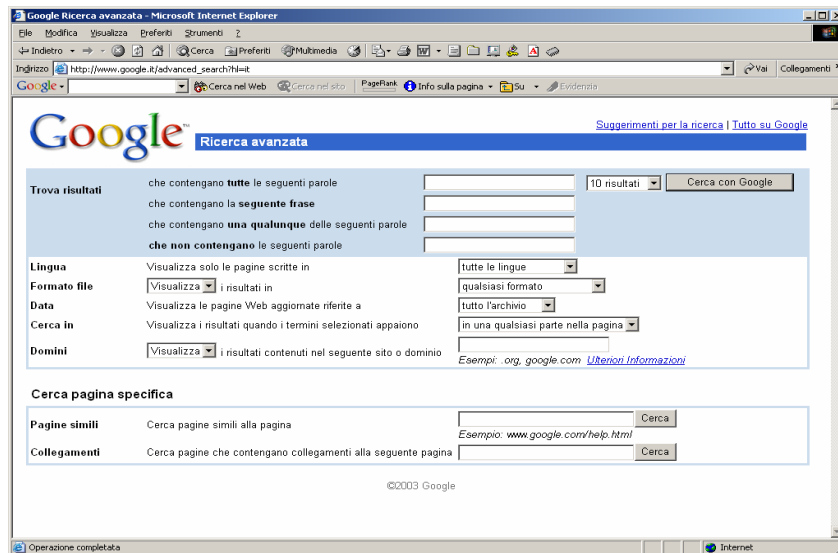
- **link:url**
 - Lista di pagine che puntano a quella indicata
- **filetype:tipo**
 - Ad esempio: ppt, pdf, ps, doc, rtf,...
- **daterange:numero1-numero2** [Calendario di Cesare, 4714 AC]
 - Limita la ricerca alle pagine indicizzate tra la data indicata da *numero1* e la data indicata da *numero2*

“Geri Halliwell” “Spice Girls” daterange:2450958-2450968

Cerca le pagine che trattano del divorzio tra la Halliwell e il suo gruppo in una finestra temporale che va dal 25 Maggio 1998 al 4 Giugno 1998.

Paolo Ferragina, Università di Pisa

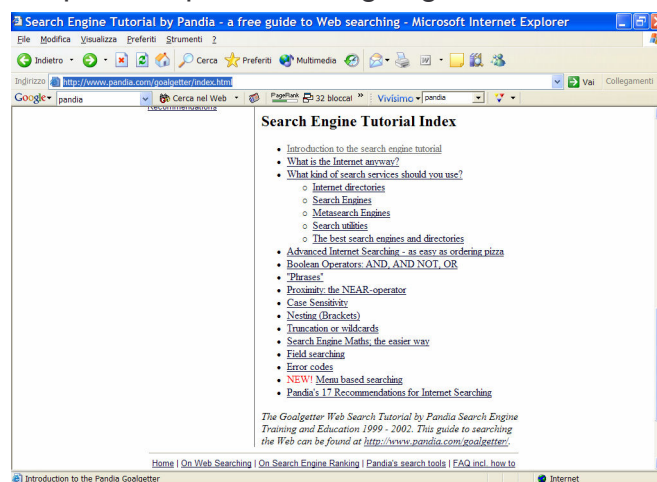
L'interfaccia per le ricerche avanzate



Paolo Ferragina, Università di Pisa

Una reference interessante

- <http://www.pandia.com/goalgetter/index.html>



Paolo Ferragina, Università di Pisa