

# I motori di ricerca

- Un'informazione "disordinata" del web pone di fronte a due tipi di problemi: la **reperibilità e l'affidabilità**
- I motori "corrano" ininterrottamente attraverso il World Wide Web catalogando tutte le pagine che incontrano
  - In un'apposita casella è possibile digitare uno o più termini chiave e chiedere al programma di effettuare una ricerca, terminata la quale compariranno sullo schermo gli indirizzi di tutti i siti che contengono quella parola e che noi potremo comodamente visitare cliccando su di essi

# Ricerca per termini

- **La ricerca attraverso un motore di ricerca per termini** è molto comoda nel caso di **nomi propri** o nel caso che le informazioni che cerchiamo siano caratterizzabili con termini molto specifici
  - **Google ([www.google.it](http://www.google.it)) e Altavista ([www.altavista.it](http://www.altavista.it))**

# Ricerca per argomenti

- La ricerca si svolge scegliendo tra le categorie disponibili quella che si avvicina di più al nostro caso.
- La **ricerca sistematica** avviene su cataloghi ragionati di risorse, suddivisi e organizzati, le cosiddette **web directories**
- La ricerca quindi non è più meccanica bensì è il risultato di una decisione umana
- Un catalogo ragionato di questo genere si basa su una sorta di "albero del sapere" che bisogna percorrere partendo da categorizzazioni più generali per arrivare via via verso quelle più specifiche
  - Yahoo ([www.yahoo.it](http://www.yahoo.it))
  - Virgilio ([www.virgilio.it](http://www.virgilio.it))

# Criteri di ricerca

- Internet mette a disposizione informazione di natura molto eterogenea, raggiungibile attraverso canali diversi
- Troppe informazioni è come non averne nessuna
- ***che cosa vogliamo trovare***
- ***che tipo di informazione vogliamo trovare***

# Criteri di ricerca

- ***informazione strutturata***
  - *Attraverso Internet sono accessibili banche dati specializzate, contenenti informazioni fortemente strutturate come il catalogo di una biblioteca, atlanti stellari, informazioni geografiche ecc.*
  - *non tutte le informazioni che ci servono sono rintracciabili presso banche dati strutturate*
- ***informazione "libera"***
  - Istituzioni, privati, aziende, ecc. **hanno ritenuto utile rendere accessibili informazioni attraverso pagine web**

# Criteria di ricerca

- **informazione "ordinata"**
  - **di una banca dati è più puntuale e affidabile** ma, pur essendo raggiungibile attraverso Internet, **non è integrata nella grande ragnatela del Word Wide Web.**
- **'informazione "disordinata"**
  - **di Word Wide Web è più difficile da valutare** dal punto di vista dell'affidabilità e spesso per accedervi siamo costretti a navigazioni lunghe e qualche volta frustranti. **E' però un'informazione sterminata.**

# Operatori logici

- AND
- OR
- NOT

# Altri parametri

- -> **NEAR** (per es.: **Federico NEAR Fellini**)  
Questo operatore, posto in mezzo a due termini, impone la presenza di entrambi i termini specificati all'interno delle risposte trovate. Il funzionamento è analogo a quello dell'operatore AND anche se si comporta in maniera molto più restrittiva. In questo caso verranno visualizzate solo le pagine che contengono entrambe le parole chiave (escludendo, nel nostro esempio, le pagine contenenti solo la parola 'Federico' o 'Fellini').

- -> **domain** (per es.: **domain:it**) Serve a delimitare la ricerca ad un dominio preciso (nel nostro caso quello italiano): in tal modo si riesce a circoscrivere la ricerca ad una sola nazione o ad una sola tipologia di sito. I domini usati sono quelli normali usati su Internet, quindi le sigle nazionali (.it, .fr, .uk...), il suffisso '.com' (domini commerciali), '.gov' (domini governativi americani), '.edu' (Università), '.org' (organizzazioni senza scopo di lucro)...

- -> **host** (per es.: **host:www.microsoft.com**)  
Con questo operatore si specifica di cercare le pagine contenute in un ben determinato sito. Se per esempio si volessero trovare notizie ufficiali della Microsoft sui virus delle macro, si potrebbe utilizzare una stringa di questo genere 'macro NEAR virus host:www.microsoft.com' (naturalmente senza gli apici). Questo operatore serve quindi a trovare un documento in un sito non dotato di motore di ricerca interno.

- -> ***title*** (per es.: **title:Biancaneve**) Con questo operatore la parola o la frase indicata sono cercate esclusivamente nel titolo della pagina. Normalmente le ricerche vengono effettuate su tutto il testo delle pagine e questo può generare risultati inutili. Supponiamo di dover cercare la favola di Biancaneve: usando questo operatore, saremo sicuri di trovare solo i siti realmente interessanti (non credo che una pagina col titolo 'Biancaneve' tratti di qualcos'altro!).

- -> ***link*** (per es.: **link:http://www.microsoft.com**) Questo operatore serve più che altro a togliersi una curiosità. Chiunque ha creato un proprio sito vorrebbe sapere se qualcun altro nel Web ha creato un link con la propria Home Page. Inserendo l'esatto Url della propria pagina si riesce a scoprire da quante e quali pagine venga richiamata la propria.

- -> **text** (per es.: **text:Biancaneve**) Questo operatore permette di concentrare la ricerca solo sul testo visibile delle varie pagine, ignorando invece ciò che si trova all'interno dei tag html. Infatti, normalmente, se nel codice html fosse presente la parola che stiamo cercando, tale pagina verrebbe visualizzata, anche se poi il testo cercato non fosse visibile. Usare questo operatore riduce la lunghezza della lista ottenuta ed evita inutili attese per il caricamento di un sito che non ha alcun interesse.
- Le funzioni di Altavista no sono poi limitate alla ricerca delle sole pagine Web ma si estendono anche a quelle di messaggi dei newsgroup. Per fare ricerche nell'archivio dei messaggi postati sui newsgroup basta inserire le parole chiave e selezionare il check box <discussion group>. L'archivio su Altavista è limitato alle due settimane precedenti.
- Per ricerche più approfondite conviene rivolgersi al sito <http://groups.google.com>

- -> **from** (per es.: **from:darty@amicopc.com**)  
Questo operatore individua i messaggi inviati da un utente di cui sia noto l'indirizzo di posta elettronica.
- Bisogna ricordarsi però che per difendersi da spam mailer o per garantirsi un minimo di anonimato, sovente chi scrive ai newsgroup usa un indirizzo fasullo o comunque modificato - per es. con l'introduzione di parole o simboli (per es.:darty@NOSPAMamicopc.com).

- -> ***subject*** (per es.: **subject:masterizzatori**) Con questo operatore si visualizzeranno solo i messaggi che hanno avuto nel subject la parola inserita (nel nostro caso solo i messaggi che avevano come subject la parola 'masterizzatore')

- -> **newsgroup** (per es.: **newsgroup:it.comp.aiuto**) Con questo operatore si limita la ricerca, mediante parole chiave, ad un determinato newsgroup.
- Oltre a tutti questi operatori, la pagina di ricerca avanzata di Altavista mette a disposizione altri strumenti per affinare la ricerca.
- Nei campi **From** e **To** si può indicare al motore di ricerca di concentrare la propria attenzione alle sole pagine che sono state modificate dopo una determinata data e prima di un'altra.
- Invece il campo **Ranking** serve a determinare un criterio per l'ordinamento dell'elenco. Inserendo una parola i documenti che la contengono verranno visualizzati per primi.

# Ricerca avanzata

- Ricerca di una frase esatta
- Un certo formato di file
- Cerca entro una certa data
- Cerca in un certo luogo della pagina (titolo, corpo del testo, ...)
- Cerca solo in certi domini (.com, .org, ...)

# Ricerca specifica

- Cerca pagine simili
  - Cerca pagine simili alla pagina
- Cerca collegamenti
  - Cerca pagine che contengano collegamenti alla seguente pagina