

La ricerca in Internet



Motori di ricerca per il Web

- Componente software (utilizzabile online) progettato *per ricercare informazioni (pagine web o HTML)*
- Consente di interrogare (utilizzando *browser*) grandi archivi di pagine web
- Restituisce pagine con collegamenti (*link*) alle pagine dei siti che contengono le informazioni ricercate (dette **SERP: Search Engine Results Page**)
- La qualità di un motore si valuta sulla base della **pertinenza** e **rilevanza** dei risultati

Fasi operative dei motori di ricerca

- Il lavoro dei motori di ricerca si divide principalmente in quattro fasi:
 - **scansione** di pagine web tramite appositi programmi (detti crawler)
 - **analisi** del contenuto delle pagine recuperate
 - **catalogazione** del materiale ottenuto
 - **risposta** all'utente

Motori di ricerca: struttura

Ogni motore di ricerca è composto da cinque componenti fondamentali:

- **crawler**
- **indice**
- **interfaccia**
- **query processor**
- **sistema di ranking**

Motori di ricerca: **crawler**

- I **crawler** (detti anche **robot**, **spider**, **bot**) sono programmi che navigano il Web, esaminando le diverse pagine e seguendo i link in esse presenti
- analizzano le pagine cercando le parole importanti, dette parole chiave (**keywords**)
- memorizzano le keywords in un **indice**
- memorizzano le pagine in un archivio temporaneo (memoria **cache**)

Motori di ricerca: **indice**

- è un enorme database (**archivio**) dove le pagine vengono memorizzate con tutte le parole chiave (**keywords**)
- rappresenta l'archivio che verrà consultato quando verrà richiesta una ricerca tramite il browser

Motori di ricerca: **interfaccia**

- l'**interfaccia** interpreta la richiesta dell'utente generalmente inviata tramite un browser
- organizza e visualizza i risultati della ricerca

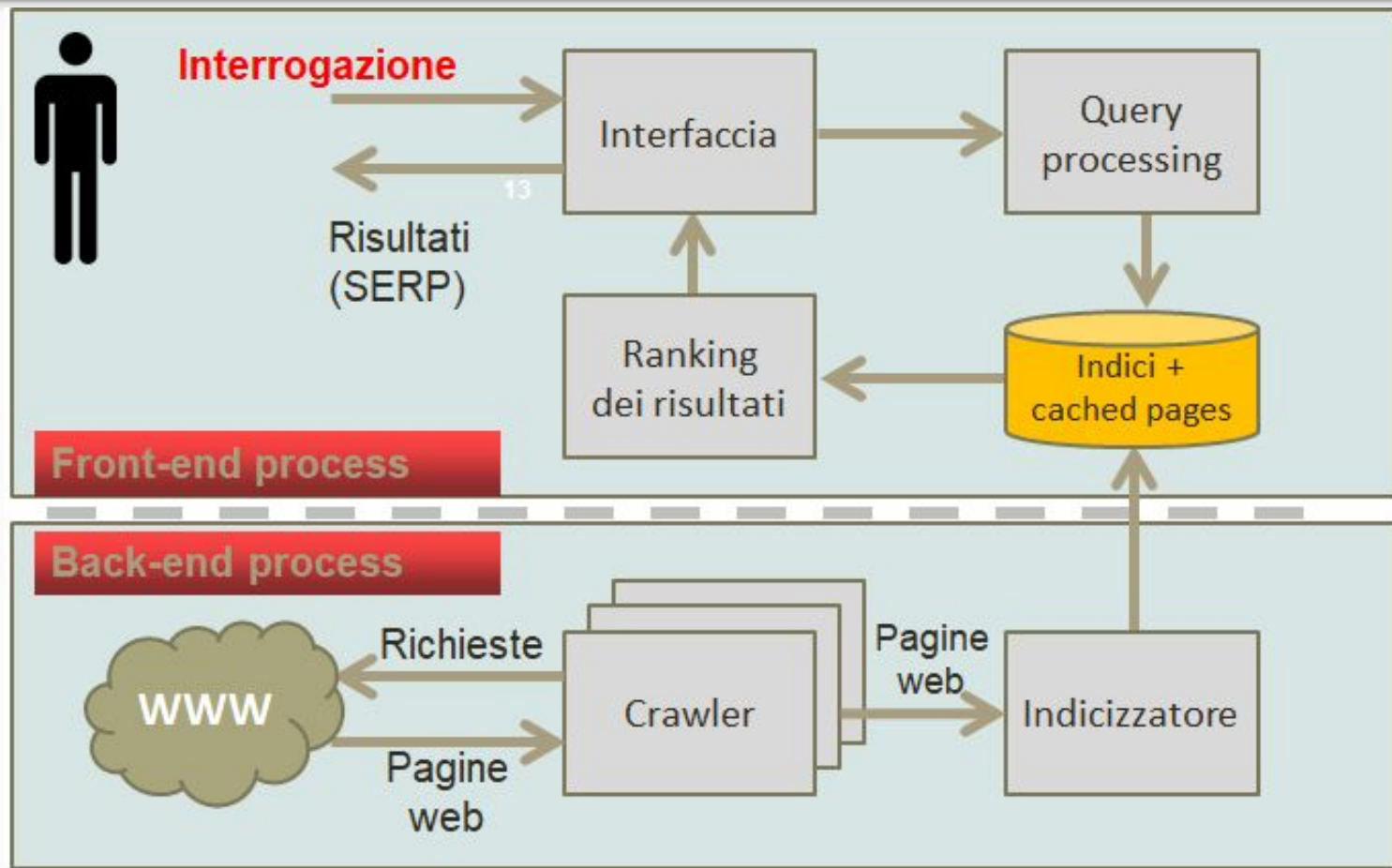
Motori di ricerca: **query processor**

- il **query processor** è un programma che agisce sull'indice per estrarre le informazioni utili in funzione di come è stata formulata la ricerca

Motori di ricerca: **sistema di ranking**

- il sistema di classificazione (**ranking**) ordina i risultati della ricerca in funzione della loro "rilevanza"
- eventualmente filtra risultati non pertinenti
- organizza la pagina di risposta

Motori di ricerca: struttura



Funzionamento dei motori di ricerca

- In risposta alla richiesta dell'utente, i motori di ricerca elencano i siti recuperati in ordine di rilevanza rispetto alla domanda posta
- Per stabilire tale rilevanza, ogni motore di ricerca classifica all'interno del proprio indice le pagine trovate usando i propri algoritmi (*criteri di classificazione*)

Pesi e misure: PageRank

- PageRank assegna un peso a ogni elemento della pagina e la classifica in base
 - al contenuto della pagina (occorrenza dei termini cercati - cioè quante volte compaiono)
 - al numero di link che puntano alla pagina
 - al contenuto delle pagine che linkano alla pagina
- L'idea di fondo è:
 - l' "importanza" di una pagina Web è tanto maggiore quante più sono le pagine Web (a loro volta "importanti") che la linkano

L'algoritmo Page Rank: altri criteri

- Google utilizza centinaia di criteri per il ranking delle pagine (classificazione)
- I criteri sono segreti e in continua evoluzione
- Questi alcuni esempi:
 - Parole chiave nella pagina (titolo, header, testo ...)
 - Età del sito
 - Qualità dei contenuti
 - Ranking fra i siti di analogo argomento
 - Numero di accessi via URL della pagina
 - Numero di accessi da parte di chi cerca
 - Tempo speso sulla pagina dagli utenti



Ricerca con Google

[Come funziona la Ricerca - Come funziona la Ricerca Google](#)

[Google Trends - Hot Searches](#)

<https://trends.google.it/trends/?geo=IT>

[Ricerca avanzata Google](#)

[Ricerca immagini avanzata di Google](#)

[Ricerca video avanzata di Google](#)

[Ricerca Libri avanzata](#)

[Perfezionare le ricerche sul Web - Guida di Ricerca Google](#)

[L'evoluzione della Ricerca](#)

[Google Lens - Cerca quello che vedi](#)

Come trovare l'informazione in rete

- **Browser diversi → spesso risultati diversi**
- **Frase diverse → spesso risultati diversi**
- **Motori di ricerca diversi → spesso risultati diversi**
- **Utilizzo dei Preferiti**
- **Ricerca semplice**
- **Ricerca avanzata**

Browser diversi

- Spesso utilizzando gli stessi termini di ricerca in browser diversi si ottengono risultati diversi
- E' possibile che i risultati siano in parte uguali (primi risultati) e poi differenzino
- Da ricordare che i risultati sono “dinamici”, cioè possono variare (anche notevolmente) a seconda se nel tempo i crawler hanno indicizzato nuovi siti o cancellato pagine obsolete o hanno modificato i criteri di indicizzazione

Frase diverse

- I motori spesso ritornano risultati diversi (o in ordine diverso) se scambiamo la posizione delle keyword (parole) di ricerca
- Conviene a volte provare a sostituire alcune parole con altre più pertinenti o di uso più comune
 - es. frase da ricercare: “Come e dove posizionare delle piante grasse in un appartamento”
 - alternative: “Dove **mettere** piante grasse in **casa**”

Motori di ricerca diversi

- Sappiamo che ogni motore di ricerca utilizza algoritmi diversi per l'indicizzazione ed il ranking (la classificazione) delle pagine e, chiaramente, i risultati possono (e spesso lo sono) diversi
- O almeno è possibile che sia diverso l'ordine (e quindi la priorità) con cui espongono i risultati

Utilizzo dei Preferiti

- Quando facciamo ricerche, ci viene comodo memorizzare i siti interessanti che troviamo, in modo da ritornarci in un secondo momento senza compiere nuovamente la ricerca
- Possiamo quindi utilizzare i Preferiti (o Bookmark - segnalibri) che è possibile gestire in ogni browser
- Da ricordare che ogni browser utilizza propri segnalibri, anche se poi, volendo, possiamo esportarli ed importarli da un browser ad un altro

FINE

Grazie per l'attenzione

