

# Motore di ricerca

Da Wikipedia, l'enciclopedia libera.

Un **motore di ricerca** (in inglese *search engine*) è un sistema automatico che analizza un insieme di dati spesso da esso stesso raccolti e restituisce un indice dei contenuti disponibili classificandoli in base a formule statistico-matematiche che ne indichino il grado di rilevanza data una determinata chiave di ricerca.

Uno dei campi in cui i motori di ricerca trovano maggiore utilizzo è quello dell'information retrieval e nel web.

Indice

1

Motori di ricerca per il web

1.1

Le fasi

1.1.1

Catalogazione

1.1.2

Risposta

1.2

Risultati sponsorizzati

2

Raffinazione della ricerca

3

Motori di ricerca più utilizzati

4

Prospettive di sviluppo

4.1

Il web 2.0 e il futuro dei motori di ricerca

5

Voci correlate

6

Altri progetti






7


































Collegamenti esterni












## Motori di ricerca per il web

Esistono numerosi motori di ricerca attivi sul web. Il più utilizzato, su scala mondiale (con un indice che supera gli 8 miliardi di pagine), è Google; molto usati anche Live e Bing (motori di ricerca di Microsoft), Yahoo!, Ask. Da segnalare il tentativo di creare il primo motore di ricerca europeo, Quaero concorrente di Google con una iniziativa franco-germanica. Il progetto, stimato attorno ai 400 milioni di dollari, è stato abbandonato dopo pochi mesi per la rinuncia da parte della compagnia tedesca.

Fra i motori di ricerca nati in Italia quelli maggiormente utilizzati nel nostro paese sono Arianna, attivo nel portale Libero e Virgilio. Tuttavia non sono veri motori di ricerca giacché si limitano a riutilizzare Google. Arianna e altri ne evidenziano chiaramente il logo, mentre Virgilio ne usa i risultati senza evidenziarne la fonte, limitandosi solo ad aggiungere alcuni propri risultati sponsorizzati.

Anno	Motore	Evento
1993	 Aliweb	Lancio
1994	 WebCrawler	Lancio
	 Infoseek	Lancio
	 Lycos	Lancio
	 Spenki	Fondazione

1995	 AltaVista	Lancio
	Magellan	Lancio
	 Excite	Lancio
	 SAPO	Lancio
1996	 Dogpile	Lancio
	 Inktomi	Fondazione
	 HotBot	Fondazione
	 Arianna	Fondazione
	 Multisoft	Fondazione
	 Virgilio	Fondazione
	 Ask.com	Fondazione
1997	 Northern Light	Lancio
	 Yandex	Lancio
1998	 Google	Lancio
1999	 Alltheweb	Lancio
	 Naver	Lancio
	 Teoma	Fondazione
	 Vivísimo	Fondazione
	 superEva	Lancio
2000	 Baidu	Fondazione
	 Exalead	Lancio
2001	 Multisoft	Chiusura
2003	Info.com	Lancio
2004	 Yahoo! Search	Lancio finale
	 A9.com	Lancio
2005	 Windows Live Search	Lancio finale
	 Ask.com	Lancio
	 GoodSearch	Lancio
2006	 Wikiseek	Fondazione
	 Quaero	Fondazione
	 Ask.com	Lancio
	 Windows Live Search	Lancio
	 ChaCha	Lancio beta
	 Guruji.com	Lancio beta
2007	 Wikiseek	Lancio

	 Wikia Search	Lancio
2008	 Cuil	Lancio (chiuso)
	 Ecocho	Lancio
	 DuckDuckGo	Lancio
2009	Wolfram Alpha	Versione sperimentale
	 Bing	Lancio
	 Coozila!	Lancio
2010	 iAlgae	Lancio
2012	 Volunia	Lancio
2012	 Ideao	Lancio beta
2013	 istella	Lancio
	 Aoohe	Lancio

La maggior parte dei motori di ricerca che opera sul web è gestito da compagnie private che utilizzano algoritmi proprietari e database tenuti segreti. Esistono comunque diversi tentativi di dar vita a motori di ricerca fondati sul software libero, alcuni esempi sono:

- Lucene
- Nutch
- Wikia Search
- YaCy
- OpenIndexer

## Le fasi

Il lavoro dei motori di ricerca si divide principalmente in tre fasi:

- **analisi** del campo d'azione (tramite l'uso di crawler appositi);
- **catalogazione** del materiale ottenuto;
- **risposta** alle richieste dell'utente;

### Catalogazione

Dopo l'analisi delle pagine, a seconda di criteri che variano da motore a motore, alcune di esse vengono inserite nel database e nell'indice del motore di ricerca.

La parte testuale archiviata durante la fase di analisi verrà in seguito analizzata per fornire le risposte alle ricerche degli utenti. Molti motori di ricerca sul web rendono anche disponibile una copia dei dati testuali di ogni pagina archiviata per quando la risorsa originale sia irraggiungibile: questa funzione è detta copia cache.

### Risposta

Rispondere alle richieste degli utenti implica la necessità di elencare i siti in ordine di rilevanza rispetto alla richiesta ricevuta.

Per stabilire la rilevanza di un sito vengono cercati nel database quei documenti che contengono la parola chiave inserita dall'utente, dopodiché ogni motore di ricerca sfrutta propri algoritmi per classificare le pagine, controllando, per esempio, quante volte le parole chiave vengono ripetute, quanti link riceve quel documento, in quali punti della pagina sono poste le parole chiave, quanti siti del database contengono link verso quella pagina, o quante volte un utente ha visitato quel sito dopo una ricerca.

## Risultati sponsorizzati

I motori di ricerca forniscono anche risultati sponsorizzati, ovvero mostrano in maggiore evidenza nelle SERP (*Search Engine Result Pages*, Pagine dei risultati dei motori di ricerca) siti web di aziende che pagano per risultare tra i primi risultati quando si cercano termini (detti keyword o parole chiave) che sono in relazione all'ambito di competenza dell'azienda stessa. I risultati sponsorizzati dei motori possono apparire anche sui siti che partecipano al loro programma di affiliazione. In particolar modo, Google permette di far apparire nelle proprie SERP (chiaramente distinti dai risultati "naturali") risultati a pagamento comprati con il programma AdWords. In aggiunta a questo offre anche un servizio di sponsorizzazione che si rivolge a tutti i siti che hanno determinati requisiti, chiamato AdSense. Google AdSense (spesso abbreviato con Google AS) usa le capacità del motore di ricerca di interpretare il tema della pagina in cui è posizionato l'apposito codice per fornire annunci a tema. Yahoo! Search ha annunciato l'arrivo di un programma analogo chiamato Panama.

## Raffinazione della ricerca

La possibilità di raffinazione della ricerca varia da motore a motore, ma la maggior parte permette di utilizzare operatori booleani: ad esempio è possibile cercare "*Ganimede AND satellite NOT coppiere*" per cercare informazioni su Ganimede inteso come pianeta e non come figura mitologica.

Su Google e sui motori più moderni è possibile raffinare la ricerca a seconda della lingua del documento, delle parole o frasi presenti o assenti, del formato dei file (Microsoft Word, PDF, PostScript, ecc.), a seconda della data di ultimo aggiornamento, e altro ancora. È anche possibile cercare contenuti presenti in un determinato sito, ad esempio "*Ganimede site:nasa.gov*" cercherà le informazioni su Ganimede presenti sul sito della NASA.

Su Exalead si trova una parte speciale per raffinare la ricerca più intuitivamente.

## Motori di ricerca più utilizzati

- Google, oltre ad essere di gran lunga il più usato direttamente, è anche quello con il maggior numero di siti che ne utilizzano il database, anche a causa del programma di sponsorizzazione "AdSense per la ricerca" che appartiene al più vasto programma AdSense.
- Yahoo!
- Windows Live Search (*Live Search* è il motore creato dalla Microsoft ed usato da Microsoft Network, meglio noto come MSN).
- Ixquick (*Ixquick* è uno dei pochi motori di ricerca al mondo a garantire la privacy, poiché non conserva né registra gli ip di chi effettua ricerche).
- Ask.com, usato dalla versione italiana di Excite per un rapporto di partnership.
- FileByType è una raccolta di moduli di ricerca basati sulle categorie.

Le prime quattro tecnologie proprietarie sono utilizzate da una quantità sterminata di provider e di metamotori di ricerca (il cui nome deriva proprio dal pescare i propri risultati da più motori, come il metamotore incorporato nel portale Excite). A più riprese Microsoft ha provato a comprare Yahoo!, cosa che avrebbe portato ad un ulteriore accorpamento e riduzione delle tecnologie proprietarie in campo, a causa della fusione di Live (il motore della Microsoft) e Yahoo! Search. L'ultimo rifiuto di Yahoo è riportato dal Wall Street Journal in data 6/5/2007 (a fronte di un'offerta di circa 50 miliardi di dollari da parte della società fondata da Bill Gates).

L'unico motore con una tecnologia proprietaria in qualche modo affiancabile come utenza ai quattro big è il cinese Baidu [www.baidu.com](http://www.baidu.com) (la cui inferiorità tecnologica è palese, ma che attinge ad un bacino di utenza tanto vasto quanto in crescita, sebbene appartenente ad una sola nazione).

Parte delle descrizioni dei siti presentate nei risultati di ricerca dai principali motori sono importate da DMOZ - ODP (<http://dmoz.org>). ODP, acronimo di Open Directory Project, non è un motore ma una open directory (basata su listing e recensioni fatte da esseri umani, anche se esistono alcuni meccanismi automatici per eliminare i siti estinti). È stata creata da Netscape, a sua volta comprata da AOL nel 1998 per quasi 25 miliardi di dollari, ed appartiene tuttora ad AOL (che è la divisione internet di Time Warner, divisione nella quale Google ha una modesta partecipazione azionaria).

## Prospettive di sviluppo

Le più recenti innovazioni nella produzione di algoritmi e di sistemi di Information Retrieval si basano sull'analisi semantica dei termini e sulla conseguente creazione di reti semantiche. Lo stesso Google ha adottato sistemi per la prevenzione dell'errore e la contestualizzazione dei risultati.

È lecito prevedere che nel giro di alcuni anni i motori di ricerca baseranno le proprie tecnologie sia sull'analisi quantitativa dei contenuti (le parole in sé), sia soprattutto su quella qualitativa (il senso delle parole). I motori di ricerca saranno, ad esempio, in grado di distinguere il senso della parola "*pesc*a" a seconda di quale sia il contesto in cui la parola è contenuta (capire se sia il frutto, la disciplina sportiva, o altro). Per muoversi in questa direzione Google ha acquisito Oingo (un tempo noto come "il motore dei concetti") e la tecnologia dell'azienda che lo creò, la Applied Semantics.

## Il web 2.0 e il futuro dei motori di ricerca



*Per approfondire, vedi **web 2.0**.*

La nuova frontiera dei motori di ricerca è il web 2.0. Adottando questa logica molti motori e directory internazionali puntano a una maggiore partecipazione degli utenti nella creazione dei contenuti dei motori di ricerca, in modo da eliminare qualsiasi ricorso a spider o a link sponsorizzati. Con questa logica gli utenti possono segnalare essi stessi i link e decidono se dare o meno popolarità ai siti segnalati.

Il web 2.0 punta anche a una facilità di utilizzo più rapida, veloce ed intuitiva grazie all'utilizzo di Metamotori (o motori di ricerca multipli) come:

- Ixquick
- MetaCrawler

## Voci correlate

- [.htaccess](#)
- [Crawler](#)
- [Egosurfing](#)
- [Googlebombing](#)
- [Google Hacks](#)
- [HTML](#)
- [Indicizzazione \(motore di ricerca\)](#)
- [Information Retrieval](#)
- [Metamotore](#)
- [Ottimizzazione \(motori di ricerca\)](#)
- [PageRank](#)
- [Protocollo di esclusione robot](#)
- [Landing page](#)
- [Query](#)
- [G2p](#)
- [Motore di ricerca distribuito](#)
- [Lista di motori di ricerca](#)
- [Web 2.0](#)

## Altri progetti

-  **Commons** ([//commons.wikimedia.org/wiki/Pagina\\_principale?uselang=it](https://commons.wikimedia.org/wiki/Pagina_principale?uselang=it)) contiene immagini o altri file su **motore di ricerca** ([//commons.wikimedia.org/wiki/Category:Internet\\_search\\_engines?uselang=it](https://commons.wikimedia.org/wiki/Category:Internet_search_engines?uselang=it))

## Collegamenti esterni

- *Motore di ricerca* (<http://thes.bncf.firenze.sbn.it/termine.php?id=37517>) in *Tesaurus del Nuovo Soggettario* (<http://thes.bncf.firenze.sbn.it/>), BNCF, marzo 2013.



Categorie: Internet | Motori di ricerca | [altre]

- 
- Questa pagina è stata modificata per l'ultima volta il 9 apr 2014 alle 13:17.
  - Il testo è disponibile secondo la licenza Creative Commons Attribuzione-Condividi allo stesso modo; possono applicarsi condizioni ulteriori. Vedi le Condizioni d'uso per i dettagli. Wikipedia® è un marchio registrato della Wikimedia Foundation, Inc.