

# **Usare al meglio i motori di ricerca**

Paolo Attivissimo

Settembre 2017

# 1. Come funziona un motore di ricerca: le cose da sapere

Un *motore di ricerca*, come Google o Bing, è un grande sistema informatico che esplora metodicamente e automaticamente tutte le parti accessibili di Internet e ne cataloga e archivia il contenuto. Questo catalogo e questo archivio vengono aggiornati costantemente e sono consultabili dagli utenti comuni visitando il *sito* del motore di ricerca (per esempio Google.com).

Ogni motore di ricerca è composto da due elementi fondamentali: il software “esploratore”, denominato *spider* (“ragno”, dato che esplora il *Web*, che in inglese significa “ragna-tela”), e l'*algoritmo*, che decide i criteri d'importanza con i quali vengono presentati agli utenti i risultati dell'esplorazione.

L'“esploratore” ha vari nomi: per esempio, *Googlebot* per Google, *Bingbot* per Bing e Yahoo, *BaiduSpider* per Baidu o *YandexBot* per Yandex.<sup>1</sup> L'algoritmo, invece, di solito non ha un nome, tranne nel caso di Google, che lo chiama *PageRank* perché determina il rango (*rank*) delle pagine del Web.

Il funzionamento esatto di questi componenti è un segreto commerciale, ma i documenti brevettuali pubblicati da Google, per esempio, gettano un po' di luce sull'argomento: i link presenti nelle pagine esplorate da Googlebot contribuiscono fortemente a determinare l'importanza di una pagina nei risultati del motore di ricerca.

Il concetto di fondo è che **se una pagina Web viene linkata da un'altra, vuol dire che la pagina linkata è considerata importante**: è una sorta di voto di approvazione (o, per analogia con Facebook, è una sorta di “*Mi piace*”). Più “voti” ha un link, più è probabile che sarà in cima alla classifica dei risultati del motore di ricerca.

A sua volta, anche l'importanza della pagina che *ospita* il link è determinante. Per esempio, se il sito di una testata giornalistica autorevole linka una data pagina Web, quella pagina Web sale di rango nella classifica dei risultati del motore di ricerca. Se un blog di un utente qualsiasi linka quella stessa pagina, l'effetto promozionale è molto più debole.

In parole povere: **più una pagina è linkata da altre pagine importanti e autorevoli del Web, più quella pagina diventa a sua volta importante ed emerge in cima ai risultati presentati dal motore di ricerca.**

Il meccanismo è simile a quello usato dalle riviste scientifiche: più un articolo viene citato da riviste autorevoli, più viene considerato meritevole. Infatti il documento fondamentale

<sup>1</sup> Una classifica degli spider più importanti di Internet e della portata della loro capacità esplorativa è disponibile presso [www.incapsula.com/blog/know-your-top-10-bots.html](http://www.incapsula.com/blog/know-your-top-10-bots.html).

che descrive il funzionamento dell'algoritmo di Google parla proprio di citazioni in stile accademico e s'intitola "*The PageRank Citation Ranking: Bringing Order to the Web*" (liberamente reperibile su Internet usando, guarda caso, un motore di ricerca). Un altro documento importante che descrive il funzionamento di Google è "*How does Google collect and rank results?*", pubblicato nel 2006 dal *Librarian Center* di Google.<sup>2</sup> Si può consultare inoltre il brevetto statunitense 6285999, "*Method for node ranking in a linked database*" a nome di Lawrence Page, datato 9 gennaio 1998.<sup>3</sup>

Accanto a questo criterio fondamentale, tuttavia, quasi tutti i motori di ricerca ne usano anche altri, che **personalizzano i risultati in base a chi li richiede o al luogo nel quale vengono richiesti**. Uno degli scopi di questa guida è evitare questa personalizzazione, che può rendere irreperibili o nascondere i risultati che in realtà interessano maggiormente a un utente.

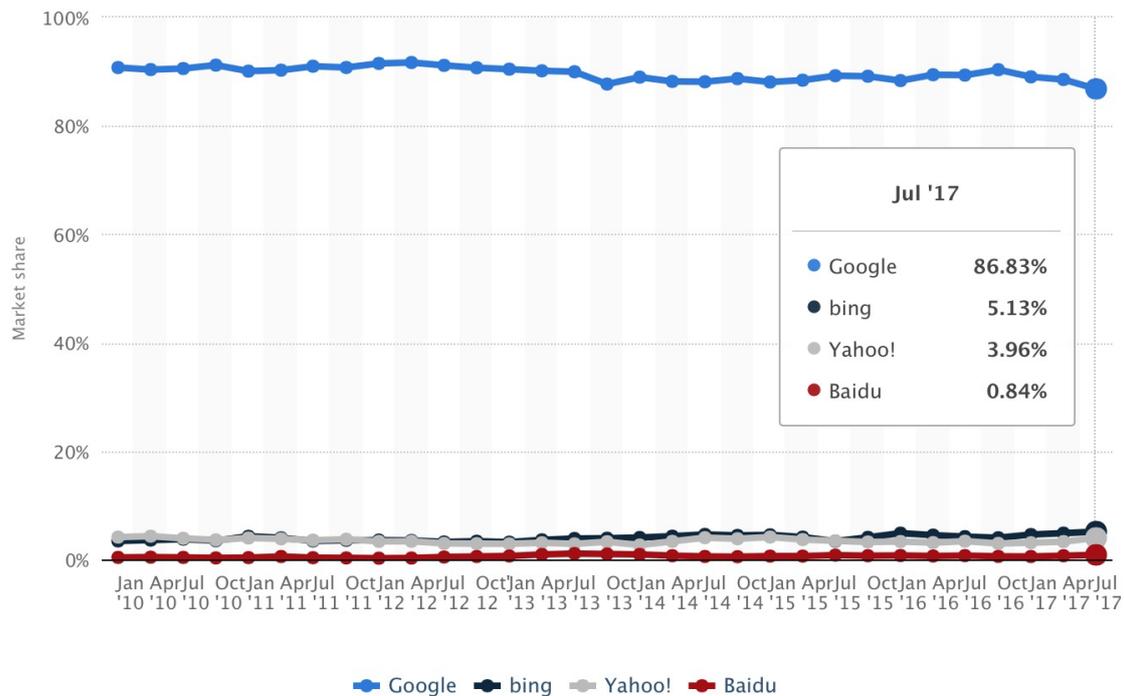
## 1.1. Google, gigante incontrastato ma generalista

In termini di quota di mercato, Google sovrasta di gran lunga tutti gli altri motori di ricerca (Figura 1)<sup>4</sup> e offre un maggior numero di opzioni di ricerca selettiva. Per questo motivo è opportuno dedicare a questo colosso buona parte della presente discussione dei motori di ricerca, perché **in pratica l'universo dei motori di ricerca è costituito da Google più le briciole**. Tuttavia gli altri motori possono, a volte, essere più efficienti su temi specifici, per cui non vanno trascurati completamente.

2 Il documento è stato rimosso dal sito di pubblicazione originale (Google), ma ne persiste copia presso [web.archive.org/web/20110928082425/http://www.google.com/librariancenter/articles/0512\\_01.html](http://www.google.com/librariancenter/articles/0512_01.html).

3 Disponibile presso <http://www.google.com/patents/US6285999>.

4 Fonte: [www.statista.com/statistics/216573/worldwide-market-share-of-search-engines](http://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines).



© Statista 2017

Figura 1: quote di mercato dei vari motori di ricerca.

Per esempio:

- **Google** eccelle nel fornire i cosiddetti “dieci link blu”, ossia una pagina scarna nella quale sono elencati i dieci risultati che con maggiore probabilità soddisfano la richiesta dell’utente, e offre la possibilità di effettuare ricerche estremamente complesse.
- **Bing** è maggiormente specializzato di Google nella copertura degli argomenti più popolari e offre più grafica e informazioni direttamente sulla pagina iniziale; offre inoltre risposte a formule matematiche complesse tramite Wolfram Alpha (sito specializzato nella risoluzione di quesiti di statistica e matematica).
- **Yahoo! Search** (*search.yahoo.com*) usa Bing e Google, ma ne modula diversamente i risultati.
- **Baidu** (*baidu.com*) è specializzato nelle ricerche in cinese.
- **Yandex** (*yandex.ru*) copre le ricerche in russo; esiste anche in versione ucraina (*yandex.ua*), inglese (*yandex.com*), turca (*yandex.com.tr*), kazaka (*yandex.kz*) e bielorusa (*yandex.by*).
- **DuckDuckGo** si distingue perché, a differenza di Google e degli altri motori di ricerca, non personalizza i risultati in base all’utente ed è molto più rispettoso della privacy;<sup>5</sup> inol-

<sup>5</sup> Alcune forti motivazioni per preoccuparsi della schedatura sistematica delle ricerche effettuate dagli utenti ad opera di Google e altri motori di ricerca, con implicazioni a lungo termine sia personali sia

tre fornisce risultati che sono un ibrido di quelli forniti da Yandex, Yahoo, Bing, Wolfram Alpha e Yummly (motore specializzato in alimentazione, cucina e ricette, che consente ricerche in base a ingredienti, diete, allergie, prezzo, sapori e “impara” i gusti dell’utente). Inoltre è disponibile sul *dark web* tramite Tor.

– **Startpage.com**, come DuckDuckGo, è un motore di ricerca che non raccoglie dati personali; diversamente da DuckDuckGo, è basato soltanto su Google.

## 1.2. Quello che un motore di ricerca non mostra

È importante tenere presente che *Google e i motori di ricerca non sono Internet*, anche se è vero che per la maggior parte degli utenti quello che Google non vede (o non mostra) in pratica non esiste. C’è un’enorme parte di Internet che sfugge completamente a una ricerca tramite Google o quasi tutti altri i motori di ricerca.

Infatti Google vede e *indicizza* (ossia cataloga) fondamentalmente tutte le pagine del Web e tutti i documenti che sarebbero accessibili a un normale utente dotato di *browser* (Internet Explorer, Firefox, Chrome, Safari). La parte di Internet che è accessibile a un comune motore di ricerca come Google è denominata *Web di superficie* (*surface Web*) e corrisponde grosso modo a tutte le pagine Web che sarebbero visibili a un utente che passasse la vita a cliccare su ogni link in ogni pagina. Google non può vedere/indicizzare:

- il contenuto *non pubblico* dei principali social network (Facebook, Twitter, Ask.fm, Instagram.com, LinkedIn, Pinterest);
- il contenuto delle pagine in formato Flash o Silverlight, nelle quali il testo è reso graficamente invece di usare caratteri;
- le pagine Web che includono uno speciale codice di divieto contenuto in un file di nome *robots.txt* che viene pubblicato da ciascun sito;<sup>6</sup>
- le pagine oscurate dal “*diritto all’oblio*” secondo le disposizioni dell’Unione Europea;
- il *deep web*;
- il *dark web*.

politiche, sono descritte in questo articolo: [www.osnews.com/story/24867/DuckDuckGo\\_The\\_Privacy-centric\\_Alternative\\_to\\_Google](http://www.osnews.com/story/24867/DuckDuckGo_The_Privacy-centric_Alternative_to_Google).

6 Esiste anche l’opposto, ossia la *sitemap*, che è un file, pubblicato da ciascun sito, che contiene un elenco di indirizzi (URL) di contenuti del sito che il gestore del sito *vuole* che vengano visti e indicizzati da un motore di ricerca. La *sitemap* consente inoltre di includere informazioni supplementari su ciascun indirizzo (quando è stato aggiornato l’ultima volta, con che periodicità viene aggiornato e quanto è importante rispetto agli altri indirizzi del sito). Si rivela particolarmente utile per consentire ai motori di ricerca di catalogare il contenuto delle pagine in formato Silverlight o Flash, normalmente non indicizzate.

### 1.2.1. Diritto all'oblio

A maggio 2014 l'Unione Europea ha imposto a Google di rimuovere dai propri risultati i link che ledono il cosiddetto “*diritto all'oblio*” se ne riceve richiesta dalla parte lesa.<sup>7</sup> La rimozione avviene se i diritti di privacy del richiedente superano le esigenze dell'interesse pubblico ad accedere alle informazioni in questione. Di conseguenza, chi consulta Google dall'Europa (Svizzera compresa) ottiene talvolta risultati incompleti o parzialmente oscurati in base a questa norma.

Questa norma non si applica necessariamente agli altri motori di ricerca.

### 1.2.2. Deep Web e Dark Web

Qualunque contenuto che sfugga all'esplorazione compiuta da un motore di ricerca è definito *Web sommerso* (*deep Web*). Se un contenuto è accessibile soltanto dopo aver digitato qualcosa in una casella di ricerca di un sito (come nel caso di un dizionario online, le cui definizioni sono accessibili solo se si immette il lemma cercato), è *deep Web*.

Qualunque contenuto di un sito della pubblica amministrazione accessibile solo dopo aver immesso un riferimento in una casella (nome dell'azienda nei registri del commercio, numero di brevetto o nome dell'inventore nell'archivio dell'Ufficio Brevetti Europeo, per esempio) è *deep web*. Il catalogo di una biblioteca è *deep Web*. Questi contenuti non sono stati nascosti *volutamente* all'occhio dei motori di ricerca: semplicemente sono inaccessibili a causa di limitazioni tecniche. Quello che è importante sottolineare è che **il contenuto del *deep Web* non è necessariamente illecito e pericoloso come viene spesso asserito dai *media*.**

All'interno di questo *deep Web* c'è una parte, relativamente piccola, che invece è stata nascosta *intenzionalmente* ai motori di ricerca e non è accessibile tramite i normali browser: questa parte è denominata *Web oscuro* (*dark Web*). Un esempio piuttosto noto di questo *dark Web* è la rete Tor, che è una rete anonima alla quale si accede soltanto usando un particolare browser, denominato *Tor* e scaricabile presso [Torproject.org](http://torproject.org). Questa è la parte di Internet spesso associata ad attività illecite per via dell'anonimato quasi totale che consente.

## 1.3. La bolla di filtraggio (*filter bubble*)

Nel tentativo di offrire ai propri utenti un servizio sempre più efficiente, i motori di ricerca ricorrono a una sofisticata *personalizzazione* dei risultati offerti. Questo vuol dire che **un utente può ottenere risultati differenti da un altro.**

<sup>7</sup> [https://support.google.com/legal/contact/lr\\_eudpa?product=websearch](https://support.google.com/legal/contact/lr_eudpa?product=websearch).

Questa personalizzazione attinge per esempio alle informazioni accumulate sull'utente dal motore: la sua ubicazione, la cronologia delle sue ricerche, i suoi clic passati, i suoi interessi dedotti dalle ricerche precedenti, il tipo di dispositivo e di browser utilizzato, la lingua usata e molto altro ancora.

Dato che questo approccio privilegia i risultati che si conformano ai bisogni e ai punti di vista correnti dell'utente, si forma quindi intorno a ciascun utente una "bolla di filtraggio" (*filter bubble*) che lo separa dalle informazioni che non concordano con il suo punto di vista o sono al di fuori dei suoi interessi (per come li ha dedotti Google). È un ghetto culturale e ideologico autorinforzante che impedisce una visione obiettiva e pertanto va contrastato se si desiderano risultati di ricerca non falsati.

## 1.4. Google: filtraggio volontario

A volte può essere necessario o utile eliminare dai risultati di ricerca quelli che hanno contenuti non adatti al pubblico ai quali vanno mostrati (per esempio immagini violente o oronografiche) e che compaiono perché la parola o frase cercata ha un significato che la lega a questo genere di contenuto (gergo sessuale, doppi sensi, eccetera).

In questo caso Google offre l'opzione *SafeSearch*, che **tenta** di non mostrare questi contenuti. Per attivarla, si va in [www.google.com/preferences](http://www.google.com/preferences), si sceglie *Attiva SafeSearch* e si clicca su *Salva*. Per fissare quest'opzione, si può usarla dopo che si è entrati nel proprio account Google e cliccare su *Blocca*: in questo modo solo chi sa la password dell'account potrà disattivare questo filtro.

## 2. Google: ricerca avanzata

Prima di addentrarci nelle tecniche per eludere questa "bolla di filtraggio" è opportuno riepilogare brevemente le principali opzioni utili per affinare i risultati offerti da Google. Queste opzioni, sommate con le tecniche citate, permetteranno di indurre Google a mostrarci quello che cerchiamo invece di quello che Google cerca di farci trovare.

Le opzioni di ricerca avanzata di Google sono consultabili in due modi fondamentali:

- accedendo alla pagina apposita ([https://www.google.com/advanced\\_search](https://www.google.com/advanced_search)) e compilandone i campi in base alle specifiche esigenze;<sup>8</sup>
- immettendo nella casella di ricerca della pagina standard di Google alcuni *operatori*, che possono essere combinati fra loro per ottenere ricerche ancora più selettive.<sup>9</sup>

<sup>8</sup> Le opzioni sono descritte presso <https://support.google.com/websearch/answer/35890?hl=it>.

<sup>9</sup> Elenco completo degli operatori: <https://support.google.com/websearch/answer/2466433?hl=it>.

## 2.1. Ricerche testuali avanzate

Va ricordato che Google non fa distinzione fra maiuscole e minuscole, con l'unica eccezione dell'operatore *OR*, che deve essere scritto in maiuscolo. A parte questo, digitare *mario bernasconi*, *Mario Bernasconi* o *MARIO bernasconi* non fa alcuna differenza.

### 2.1.1. Virgolette

Se si immette in Google una serie di parole, Google restituirà le pagine più significative che contengono quelle parole *in qualunque ordine e anche in punti separati della pagina*. Se invece si desidera ottenere un elenco delle pagine che contengono l'esatta sequenza di parole, occorre racchiuderle tra virgolette.

Per esempio, per cercare le pagine che citano una persona è opportuno digitare il suo nome e cognome fra virgolette: "*Giovanni Nembrini*", altrimenti Google troverà anche le pagine che parlano di altre persone che hanno lo stesso cognome ma un nome differente.

Google è comunque abbastanza sofisticato da riconoscere automaticamente, anche senza le virgolette, i nomi delle persone celebri e i titoli di film, libri e canzoni. Le virgolette sono particolarmente utili quando si cerca una persona poco nota oppure una citazione esatta di una frase poco conosciuta (per esempio un brano di un articolo di giornale).

### 2.1.2. Asterisco

Se si vuole cercare una frase e le sue varianti si può usare l'asterisco come carattere jolly. Per esempio, "*non \* di lor ma guarda e passa*" chiede a Google di elencare i siti che contengono la citazione dantesca sia nella sua versione corretta (con *ragioniam* al posto dell'asterisco), sia in quella popolare (con *ti curar* al posto dell'asterisco).

### 2.1.3. Alternative (OR)

In una ricerca di testo si può specificare un'alternativa usando *OR* (scritto in maiuscolo): per esempio, scrivere *divorzio "Brad Pitt" OR "Angelina Jolie"* farà emergere nei risultati le pagine che contengono la parola *divorzio* e il nome esatto *Brad Pitt* e anche le pagine che contengono la parola *divorzio* e il nome esatto *Angelina Jolie*.

### 2.1.4. Esclusione (-)

Si può chiedere a Google di non includere nei risultati una parola specifica: basta precederla con un segno "-". Per esempio, *calcio -campionato* escluderà dalla ricerca i siti che

parlano del campionato di calcio e privilegerà quelli che parlano dell'elemento chimico; *salsa -danza* escluderà buona parte dei siti dedicati al ballo e farà emergere quelli di cucina. Viceversa, *salsa -cucina* darà come risultato siti dedicati al ballo e non elencherà quelli dedicati alla salsa come prodotto culinario.

## 2.2. Ricerca limitata a un sito o a un dominio

È possibile ordinare a Google di limitare la propria ricerca a un sito specifico, usando l'operatore *site* seguito dai due punti. Per esempio, "Giovanni Nembrini" *site:rsi.ch* (senza spazio dopo i due punti) limiterà l'ambito della ricerca alle pagine Web accessibili del sito della RSI che contengono quel nome abbinato a quel cognome, oppure "coniglio d'amministrazione" *site:facebook.com* cercherà esclusivamente in Facebook questo frequente errore di battitura.

Si può anche chiedere a Google, con questo stesso operatore, di limitare l'ambito della ricerca alle pagine Web dei siti di un certo paese o di uno specifico tipo di sito Internet. Per esempio, per cercare soltanto nei siti svizzeri<sup>10</sup> si può digitare in Google l'argomento cercato, seguito da *site:.ch* (sempre senza spazio per i due punti).

## 2.3. Pagine che linkano una data pagina

Se stiamo indagando su una particolare pagina del Web, possiamo chiedere a Google di elencarci quali pagine del Web la citano, per esempio per valutare la popolarità o la risonanza ottenuta dalla pagina sotto indagine o per risalire all'origine di una notizia seguendo la catena di pagine linkate fra loro. Questo si ottiene usando l'operatore *link:* (senza spazio dopo i due punti), come in:

*link:http://www.cancer.gov/cancertopics/factsheet/Risk/cellphones*

## 2.4. Ricerche cronologiche o per intervalli numerici

Digitando una parola o frase seguita da un intervallo di numeri è possibile ottenere un elenco delle pagine che contengono quella parola o frase e che, secondo l'interpretazione di Google, si riferiscono a quell'intervallo di numeri.

Per esempio,

<sup>10</sup> In realtà non proprio tutti i siti svizzeri usano il suffisso *ch*: molti siti usano i suffissi *.com* e *.org*. Inoltre a partire dal 2015 è disponibile anche un suffisso *.swiss*.

## *“moria di pesci” 1980...1990*

restituisce pagine che contengono la frase esatta *“moria di pesci”* e si riferiscono al decennio dal 1980 al 1990.

Un'altra maniera di usare quest'opzione è immettere un intervallo di prezzi.

### **2.5. Ricerche per tipo di file**

È possibile restringere una ricerca a uno specifico tipo di file (per esempio un documento Word o un foglio di calcolo Excel) usando l'operatore *filetype:* (incluso il due punti) seguito, senza spazi, da uno dei seguenti tipi di file: PDF, PS (PostScript), DWF (Autodesk), KML (Google Earth), XLS (Excel), PPT (PowerPoint), DOC (Word), RTF (testo), SWF (Flash).

Per esempio,

*"curriculum vitae" site:aziendademo.com filetype:pdf*

restituisce tutti i documenti in formato PDF ospitati in modo accessibile sul sito *Aziendademo.com* e contenenti l'esatta sequenza di parole *“curriculum vitae”*.

## **3. Ricerca e recupero di documenti rimossi o modificati; ricerca storica**

Se un documento o una pagina Web precedentemente pubblicati su un sito sono stati poi rimossi o modificati, oppure sono inaccessibili a causa di un numero eccessivo di visitatori, spesso è possibile recuperarne lo stesso il contenuto utilizzando le *copie cache* e le copie archiviate.

### **3.1. Copia cache di una pagina o di un sito**

Quando Google esplora il Web per indicizzarlo, registra nei propri server una copia di ogni singola pagina esplorata. Questa è la *copia cache*, che in Google è consultabile digitando in Google la parola *cache*: seguita (senza spazi) dal nome del sito o dall'indirizzo della pagina oppure cliccando sul triangolo incluso in un risultato di ricerca.

La copia cache è utile perché mostra com'era una data pagina quando Google l'ha visitata e quindi costituisce una sorta di “eco” temporanea, che persiste anche se la pagina originale viene cambiata o addirittura rimossa. Questo consente di andare indietro nel

tempo a vedere, per esempio, qual era il contenuto di un sito oscurato, sequestrato, alterato oppure rimosso.

Questa copia è utile anche quando desideriamo esplorare un sito senza lasciare traccia del nostro interessamento. Ogni normale visita a un sito, infatti, lascia all'amministratore del sito una serie di informazioni tecniche che permettono di identificare il visitatore. In alcuni casi può essere sconsigliabile, per esempio, che un giornalista sfogli, da un computer connesso alla rete della redazione, un sito sul quale sta indagando: i titolari del sito potrebbero accorgersi della visita e dedurne che è in corso un'inchiesta.

La copia cache è accessibile anche nei normali risultati di ricerca cliccando sul triangolo visualizzato accanto al singolo risultato e scegliendo la voce *Copia cache*.

Oltre alla copia cache di Google c'è anche quella di Coral (Coralcdn.org) e c'è l'archivio di Archive.org. Per consultarli tutti contemporaneamente si possono usare i siti *Cachedpages.com* e *Cachedview.com*.

Tipicamente la copia più recente è quella nella cache di Google; quella di Coral è leggermente meno recente; e infine le copie multiple custodite presso Archive.org possono andare indietro nel tempo anche fino agli anni Novanta del secolo scorso.

Esiste anche Archive.is, che però conserva solo le pagine che vengono archiviate dagli utenti.

Per esempio, per ottenere un elenco dei siti sequestrati dalle autorità italiane si può usare l'operatore *intitle* di Google per cercare tutti i siti che hanno come titolo l'espressione standard usata in Italia in questi casi, ossia "*sito sottoposto a sequestro*". La sintassi della ricerca diventa quindi:

*intitle:"sito sottoposto a sequestro"*

A questo punto si sceglie un sito sequestrato fra quelli elencati da Google e si immette il suo link in *Cachedpages.com* o *Cachedview.com*: solitamente Archive.org ne contiene una copia risalente a prima del sequestro che, consente di fare ipotesi sulle ragioni del sequestro stesso e anche di determinarne la data approssimativa.

## **4. Google: ricerca e recupero di documenti confidenziali**

Talvolta le autorità commettono l'errore di pubblicare su Internet, presso indirizzi non pubblicizzati, documenti che sono sotto embargo. Scaduto l'embargo, annunciano l'indirizzo al pubblico.

Tuttavia Google vede e indicizza anche quasi tutti i documenti pubblicati sui siti Web, anche se si trovano presso indirizzi non linkati o annunciati, per cui i documenti sono comunque già reperibili prima della fine dell'embargo se si immettono in Google le parole chiave pertinenti.

I documenti usano spesso diciture come *embargo* oppure *embargoed until* seguito da una data, *do not disseminate*, *strictement confidentiel*, *ne pas diffuser*, *ne pas dévoiler*, *streng vertraulich*, *streng vertraulich und rechtlich geschützt*.

La ricerca di queste espressioni, specialmente se accompagnata da una data, dall'operatore *site:* che restringa la ricerca a siti di un ente o di un'azienda e dall'operatore *filetype:PDF* e/o *filetype:DOC* (di norma i documenti vengono pubblicati come file PDF o DOC), produce spesso frutti inattesi.

Un esempio limite è dato da questa ricerca:

*intext:password intext:username filetype:xls inurl:.xls*<sup>11</sup>

che trova i file contenenti elenchi di password incautamente custoditi in una cartella pubblica accessibile a Google.

Un altro esempio altrettanto estremo è il seguente:

*"Your password is \* Remember this for later use"*

## 5. Google: ricerca selettiva per tipo di pubblicazione

Google ha dei "sotto-motori" di ricerca specialistici:

- *Google Libri* (*books.google.com*) limita la propria ricerca ai libri e alle riviste che Google ha scansionato;
- *Scholar.google.com* è dedicato alle ricerche nella letteratura accademica;
- *News.google.com* è dedicato alle notizie e consente di scegliere la zona, il paese e la testata da seguire; permette ricerche per data e per fonte e altro ancora all'interno dei siti che si occupano di notizie (siti di giornali o altre testate che Google ritiene giornalmisticamente accettabili).

11 Al posto di *xls* si può usare anche *csv*.

## 6. Google: ricerca meno personalizzata

Per ridurre la personalizzazione e il filtraggio che Google applica ai risultati di ricerca ci sono due metodi di base: si può uscire dall'account Google e si può chiedere a Google la *depersonalizzazione* dei risultati.

### 6.1. Navigazione privata: quasi inutile

Spesso si crede che si possa eliminare il filtraggio personalizzato effettuato da Google se si usa la *navigazione privata*, ossia una particolare modalità del browser (accessibile per esempio in Firefox e Safari scegliendo *File > nuova finestra privata*, in Internet Explorer scegliendo *Strumenti o Sicurezza > InPrivate Browsing* e in Chrome scegliendo l'icona delle impostazioni e poi *Nuova finestra di navigazione in incognito*).

Tuttavia questa modalità di navigazione serve soltanto a non lasciare memoria dei siti visitati sul computer o in generale *sul dispositivo che usiamo* (durante la navigazione privata i siti visitati non vengono aggiunti alla cronologia, per esempio), non accumula *cookie* nuovi, che vengono eliminati alla chiusura della finestra di navigazione privata, ma fa poco o nulla per anonimizzare la navigazione, che è infatti perfettamente rilevabile e registrabile da parte del sito visitato e dal fornitore d'accesso alla Rete.

### 6.2. Uscita dall'account Google

Consultare Google quando non si è nel proprio account Google *riduce* la personalizzazione ma non la elimina del tutto: restano per esempio le personalizzazioni su base linguistica e geografica e (in parte) quelle basate sulla cronologia delle ricerche precedenti. Inoltre l'uscita dal proprio account implica la scomodità di non poter accedere ai servizi di Google (per esempio Gmail) fino a quando non si rientra nell'account.

Per uscire dal proprio account Google si va a Google.com, si clicca sull'icona circolare della foto dell'utente (in alto a destra) e si sceglie *Esci*.

### 6.3. Depersonalizzazione

Anche se si esce dal proprio account Google, i risultati di ricerca di Google continuano a essere parzialmente personalizzati sulla base delle ricerche precedenti. Per disabilitare questa personalizzazione si può visitare la pagina <http://www.google.com/history/optout>

dopo essere usciti dall'account e fare clic sul selettore accanto alla frase *Attività di ricerca svolte senza avere eseguito l'accesso attivo*. Quest'impostazione vale per tutti gli utenti che utilizzano quel browser e quel computer fino a quando qualcuno esegue l'accesso al proprio account.

Anche questo metodo, tuttavia, non è completamente risolutivo.

## 7. Google: aggiramento del diritto all'oblio UE

Alcuni esempi di ricerche che portano a risultati coperti dal diritto all'oblio previsto dalle norme UE sono stati pubblicati fino a fine 2016 presso [Hiddenfromgoogle.com](http://Hiddenfromgoogle.com) (che ora non esiste più ma che è archiviato presso [Archive.org](http://Archive.org))<sup>12</sup>. Anche la BBC ha deciso di pubblicare quelli che riguardano pagine del proprio sito.<sup>13</sup>

Un esempio per tutti: grazie a [Hiddenfromgoogle.com](http://Hiddenfromgoogle.com) sappiamo che c'è un articolo di *Der Spiegel* del 1995<sup>14</sup> che è stato rimosso dai risultati di Google in base al diritto all'oblio, per cui se si cerca in Google "Fred Anton" (tra virgolette) insieme a *scientology* o *spiegel* da un computer che si trova in UE (o in Svizzera), l'articolo di *Der Spiegel* che lo cita<sup>15</sup> non compare tra i risultati. In compenso viene presentato un avviso: "Alcuni risultati possono essere stati rimossi nell'ambito della normativa europea sulla protezione dei dati".

Tuttavia è sufficiente **cambiare la localizzazione del computer** da Svizzera a Hong Kong o a un'altra località extraeuropea, oppure **usare un altro motore di ricerca**, come Bing, Yahoo o DuckDuckGo, per far comparire il link all'articolo tra i risultati.

## 8. Ricerche eludendo le personalizzazioni geografiche

I motori di ricerca personalizzano i risultati anche in base all'ubicazione geografica dell'utente: un esempio particolarmente vistoso è, in Google, quello dell'oscuramento dei risultati ai quali è stata applicata la normativa sul "diritto all'oblio", che si applica a tutti gli utenti che si trovano in Europa ma non a quelli extraeuropei.<sup>16</sup>

12 <https://web.archive.org/web/20161106094334/http://hiddenfromgoogle.afaqtariq.com:80>.

13 *Google removes 12 BBC News links in 'right to be forgotten'*, di Edwin Lane, 19/8/2014 (<http://www.bbc.com/news/technology-28851366>); <http://www.bbc.com/news/technology-28311217>.

14 *Wie tausend Metastasen*, 15/5/1995 (<http://www.spiegel.de/spiegel/print/d-9183695.html>).

15 *Sekten - „Wie tausend Metastasen“*, 15.05.1995 (<http://www.spiegel.de/spiegel/print/d-9183695.html>)

16 Inizialmente Google applicava il diritto all'oblio solo agli utenti europei che visitavano le versioni europee di Google (per esempio [Google.fr](http://Google.fr) o [Google.it](http://Google.it)), mentre gli europei che consultavano [Google.com](http://Google.com) o altre versioni extraeuropee di Google non venivano interessati. Ora Google ha esteso il diritto oblio a tutti i suffissi (TLD).

Per contrastare questa forma di “*geopersonalizzazione*” si può ingannare Google (o un altro motore di ricerca) facendogli credere di trovarsi in una località diversa da quella reale, per esempio in un altro paese o in un altro continente.

Esistono varie tecnologie, gratuite o a pagamento, che consentono di ottenere questo effetto (*proxy server, VPN, tunnelling*), ma tutte si basano su un unico principio: affacciare il proprio computer (o tablet o smartphone) a Internet tramite un altro computer che si trova fisicamente altrove.

Alcuni esempi di servizi che offrono questa forma di “teletrasporto” virtuale sono Hide-my-ass.com, Tunnelbear.com e Freedome.<sup>17</sup>

In pratica, quando li usiamo, navighiamo in Rete come consueto, con l’unica differenza che il nostro luogo geografico di ingresso in Internet, agli occhi dei servizi di Internet e quindi anche a quelli dei motori di ricerca, è il luogo in cui si trova il *servizio*, non quello in cui si trova il nostro computer o tablet o smartphone. In molti casi è possibile scegliere esplicitamente il paese dal quale si vuol simulare di provenire: questo consente, fra l’altro, di accedere ai servizi che sono disponibili soltanto per gli utenti di quel paese (video “geobloccati”, siti oscurati dalle autorità, dati nascosti dal “diritto all’oblio”, e altro ancora).

Il motore di ricerca restituisce quindi i risultati di ricerca che otterremmo se ci trovassimo nel paese in cui risiede il servizio di proxy/VPN/tunnelling che stiamo usando. In molti casi l’utente può selezionare a piacimento questo paese.

Nel caso di Google, se sommiamo le depersonalizzazioni precedenti a questo ricollocamento geografico fittizio riusciamo a ridurre al minimo il controllo sui risultati di ricerca esercitato da questo motore.

## **8.1. Verifica dell’elusione**

Per verificare che un servizio di proxy/VPN/tunnelling stia simulando correttamente una localizzazione differente da quella effettiva si possono usare servizi come *Speedtest.net* oppure *Whatismyip.com*, che restituiscono a video l’informazione di localizzazione presunta.

## **8.2. Sicurezza, tracciamento e anonimato**

Non tutti i servizi di proxy/VPN/tunnelling effettuano una *cifratura* dei dati trasmessi, per cui non è detto che la trasmissione sia protetta contro le intercettazioni da parte di provi-

<sup>17</sup> Presso [https://www.f-secure.com/en/web/home\\_global/freedome](https://www.f-secure.com/en/web/home_global/freedome).

der ostili (problema frequente in alcuni paesi con regimi poco democratici) o di intrusi (problema frequente nelle reti Wi-Fi pubbliche). Se si desidera questo genere di protezione, occorre consultare le specifiche del servizio e verificare che sia fornito.

Inoltre, se si naviga usando un dispositivo mobile dotato di GPS, è possibile che alcuni siti (come per esempio Google) possano attingere alla posizione geografica indicata dal GPS e quindi rilevare la vera ubicazione dell'utente nonostante l'uso di sistemi di simulazione della localizzazione. Per questo è consigliabile disattivare il GPS del dispositivo durante la navigazione con localizzazione simulata.

## 9. Tor come strumento di anonimato

Tor Browser, disponibile gratuitamente presso [Torproject.org](http://Torproject.org), è un browser concepito per la navigazione sicura e anonima che ha molte funzioni utili al nostro scopo, come la cifratura forte e l'anonimizzazione della provenienza del traffico.

In estrema sintesi, quando si usa Tor tutto il traffico di dati scambiato con Internet viene cifrato molto pesantemente e viene scomposto in parti, ciascuna delle quali viene affidata a un *nodo* Tor (un computer abilitato a ricevere e ritrasmettere i dati degli utenti di Tor Browser). Il nostro flusso di dati viene quindi passato ripetutamente da un nodo all'altro, rendendo estremamente difficile risalire alla sua origine reale, prima di uscire su Internet e raggiungere il sito desiderato. Per maggiore sicurezza è consigliabile combinare l'uso di Tor con l'impiego di una VPN.

Usando Tor si beneficia della localizzazione fittizia prodotta dai nodi Tor; tuttavia normalmente non è possibile scegliere quale paese di provenienza si desidera simulare.

## 10. Ricerche private nel Web e nel *dark Web*: Tor

Tor Browser consente di esplorare in modo riservato, sostanzialmente senza lasciare tracce, sia il Web tradizionale sia il *dark Web*. Si rivela quindi molto utile per “sopralluoghi” virtuali a siti da effettuare senza allertare i titolari oppure per lo scambio di comunicazioni riservate in ambienti a rischio. È il sistema usato da Edward Snowden per comunicare dati riservati dell'NSA al *Washington Post* e al *Guardian*.

L'intera navigazione è cifrata e anonimizzata a vari livelli; tuttavia vanno evitati i comportamenti che possono rendere identificabile l'utente di Tor. Per esempio, è inutile usare Tor per accedere al proprio account su Amazon o su Google, perché il fatto stesso di accedervi con nome utente e password identifica il titolare dell'account. Inoltre il traffico cifrato di Tor può risaltare rispetto al traffico normale e quindi identificare l'utente: per esempio,

se l'utente è l'unico che adopera Tor in una rete aziendale, l'amministratore della rete non farà fatica a dedurre chi è (non ne conoscerà le attività, ma capirà di chi si tratta).

Una volta avviato il browser Tor, si può visitare *The Hidden Wiki*, che è l'indice dei servizi nascosti all'interno della rete Tor (l'indirizzo è variabile ma è reperibile su Google). Fra questi servizi ci sono anche motori di ricerca, come *Torch*, che tentano di catalogare questa parte di Internet. Gli indirizzi dei siti della rete Tor sono caratterizzati dal suffisso *.onion* e da un nome costituito da lettere che non formano parole di senso compiuto.

## 11. Simulazione di altri dispositivi

Alcuni motori di ricerca e alcuni siti di Internet restituiscono risultati differenti anche in base al tipo di browser o sistema operativo o dispositivo utilizzato. Un esempio piuttosto celebre è quello dei siti per le prenotazioni dei voli, che offrono tariffe differenti in base al fatto che l'utente stia usando un Mac oppure un PC; un altro è costituito dai siti-truffa, che mostrano contenuti innocui se visitati con un normale computer ma presentano delle pagine-esca se visitati con uno smartphone.

Si può eludere questa personalizzazione impostando un parametro del proprio browser, lo *user agent*. Quando un browser visita un sito, gli passa infatti una serie d'informazioni sulla propria natura (nome e versione del browser, tipo di sistema operativo, dispositivo, lingua, eccetera). Il sito si fida delle informazioni ricevute, per cui è possibile impostare il browser per far credere al sito di essere visitato da un dispositivo diverso da quello reale (per esempio un iPhone quando in realtà si sta usando un PC Windows).

La procedura d'impostazione varia da browser a browser: le istruzioni specifiche sono disponibili nella guida d'uso dei rispettivi browser. In Safari, per esempio, si va in *Preferenze - Avanzate* e si attiva *Mostra menu Sviluppo nella barra dei menu*. Fatto questo, la barra dei menu acquisisce una voce supplementare, *Sviluppo*, dalla quale si può scegliere lo *user agent* da simulare.

## 12. Google: monitoraggio automatico (Google Alerts)

Presso <https://www.google.com/alerts> è possibile definire uno o più allarmi automatici contenenti una parola o una frase: quando Google incontra una pagina o un documento contenenti quella parola o frase, ne segnala l'indirizzo mandando una mail all'utente.

Per esempio, per tenere d'occhio le novità su un modello di automobile della marca Tesla si può immettere "*Tesla model 3*" in Google Alerts.

## 13. Google: motori di ricerca personalizzati (CSE)

Presso <https://cse.google.com/cse/all> è possibile definire un *motore di ricerca personalizzato*, ossia un insieme di siti ai quali si desidera limitare le ricerche. Per esempio, si può creare un motore di ricerca che prenda in considerazione soltanto i risultati riguardanti una rosa di siti di medicina ritenuti attendibili.

Il motore è poi consultabile con le consuete modalità di Google: si immette una parola o una frase nella sua casella di ricerca e si visualizzano i risultati.

## 14. Ricerche nei social network

### 14.1. Facebook Graph Search

La funzione *Graph Search* di Facebook è attivabile scegliendo l'inglese come lingua dell'interfaccia di Facebook e consente di cercare foto, luoghi visitati, preferenze personali di chiunque abbia lasciato pubbliche queste informazioni, spesso con risultati imbarazzanti.

Le richieste di ricerca vanno formulate in inglese e seguendo una struttura analoga a quella dei seguenti esempi:

- *people who like [argomento/persona]*
- *photos of [argomento/persona]*
- *photos from [anno] that I liked*
- *people who live in [località]*
- *[argomento] liked by my friends*

Le richieste possono essere combinate, consentendo ricerche precise come queste:

- *Single women who live in [città] and are interested in dating*
- *Photos liked by my friends that are from 2012*

*Facebook Search* (info presso [search.fb.com](http://search.fb.com)) è disponibile per ora soltanto agli utenti che usano Facebook in inglese americano su iPhone, Android e computer e consente ricerche di testo libere.

## 14.2. Instagram

Instagram ha una propria casella di ricerca, ma per effettuare ricerche più mirate si può utilizzare **Picodash.com**, sito a pagamento che mette a disposizione una sezione che permette di cercare e vedere tutte le foto pubbliche che sono geolocalizzate in una data località e di cercare le persone, le parole chiave e gli *hashtag*.

## 14.3. Twitter

L'indice completo dei messaggi pubblici di Twitter è accessibile a tutti presso <https://www.twitter.com/search-advanced>, consente una ricerca che si estende fino alle origini di Twitter, nel 2006, e include moltissime opzioni.

**Twicsy.com**, invece, è un indice delle immagini pubblicate su Twitter e funziona come una sorta di *cache*, archiviando una copia delle immagini e permettendo ricerche per categorie o cronologie.

**Followerwonk.com**, infine, permette di esplorare in lungo e in largo gli account Twitter, per esempio immettendo parole chiave che definiscano una professione o una categoria e poi cercando, all'interno dei profili risultanti, chi ha scritto il maggior numero di tweet, chi ha il maggior numero di *follower*, chi ha l'account più vecchio e chi ha la "*social authority*" più alta. Si possono anche creare grafici che confrontano vari account, per scoprire quanti sono i *follower* condivisi e i *followed* (utenti seguiti dagli account analizzati) in comune, insieme a tante statistiche, come la percentuale di *retweet* o di tweet contenenti un link. Nella sezione *Analyze follower* è possibile generare una mappa della distribuzione geografica e delle fasce d'età dei *follower*, mentre in *Track followers* si può produrre un grafico dell'evoluzione dei *follower*: quanti ne arrivano, quanti se ne vanno, gli orari nei quali sono maggiormente attivi, se sono uomini o donne, le lingue utilizzate e altro ancora. Alcuni di questi servizi possono richiedere un account a pagamento, specialmente se l'utente Twitter da analizzare ha un numero elevato di follower.

## 15. Ricerche di e per immagini

**Tineye.com** consente di cercare immagini sulla base di un'immagine campione ed è in grado di rilevarne anche versioni più grandi, troncate o manipolate.

La ricerca per immagini di Google ([images.google.com](https://images.google.com)) e quella di Bing ([bing.com/?scope=image](https://bing.com/?scope=image)) permettono di cercare immagini in base a una descrizione testuale oppure di sottoporre al motore di ricerca un'immagine campione per scoprire se esistono immagini simili.

**ImageRaider.com** è simile a TinEye, ma offre in aggiunta un servizio di monitoraggio, alertando l'utente quando trova un sito che ha pubblicato un'immagine sufficientemente simile a quella che gli è stata data come campione.

**Karmadecay.com** identifica soltanto le immagini pubblicate sul social network Reddit ed è utile per identificare l'origine di memi e immagini controverse.

**CO Everywhere** ([www.coeverywhere.com](http://www.coeverywhere.com); app disponibile per iPhone e iPad su iTunes presso <https://itunes.apple.com/us/app/co-everywhere-local-news-social/id686737697?mt=8>) è un'app per dispositivi mobili che consente di monitorare l'attività pubblica dei principali social network (Instagram, Facebook, Twitter, FourSquare, Yelp, prevalentemente le immagini) in una data area geografica delineandola sulla mappa con un dito.

## 16. Analisi forense di immagini

Le immagini possono essere analizzate dal punto di vista tecnico tramite siti come **Fotoforensics.com** e **Hackerfactor.com**, ai quali è possibile inviare una foto per scoprire se è stata manipolata aggiungendo, deformando, ricolorando o sottraendo elementi.

Esistono inoltre servizi professionali, come per esempio **Tungstène** (presso [www.exomakina.fr/eXo\\_maKina/Tungstene.html](http://www.exomakina.fr/eXo_maKina/Tungstene.html)), usati dalle testate giornalistiche, che svolgono analisi matematiche più approfondite sulle fotografie e consentono di rivelare ritocchi e correzioni spesso impercettibili anche per l'occhio esperto, come le "clonazioni" di porzioni di immagini.

## 17. Ricerche di video (Youtube)

Youtube offre una serie di filtri nella pagina di ricerca: si inizia a digitare una parola o frase nella casella di ricerca standard comandi selettivi e compare un pulsante *Filtra* che consente di scegliere vari parametri (data di caricamento, tipo, durata, caratteristiche).<sup>18</sup>

Tuttavia esiste anche una serie di comandi testuali molto potenti, che si separano e combinano usando la virgola:

Ricerca solo nel titolo	<i>intitle:argomento</i>
Esigere una certa parola	<i>+parola</i>
Escludere una certa parola	<i>-parola</i>
Ricerca di una frase esatta	<i>"argomento"</i>

<sup>18</sup> <https://support.google.com/youtube/answer/111997?hl=en>.

Carattere jolly	*
Cercare un canale	<i>argomento</i> , channels
Cercare un film	<i>argomento</i> , film
Cercare un programma	<i>argomento</i> , show
Cercare video di meno di 4 min	<i>argomento</i> , short
Cercare video lunghi più di 20 min	<i>argomento</i> , long
Solo video 4K	<i>argomento</i> , 4k
Solo video HD	<i>argomento</i> , HD
Solo video con sottotitoli	<i>argomento</i> , cc
Solo video con licenza CC	<i>argomento</i> , creativecommons
Solo video 3D	<i>argomento</i> , 3D
Solo video in diretta	<i>argomento</i> , live
Solo video panoramici a 360°	<i>argomento</i> , spherical
Solo video caricati nell'ultima ora	<i>argomento</i> , hour
Solo video caricati oggi	<i>argomento</i> , today
Solo video di questa settimana	<i>argomento</i> , week
Solo video di questo mese	<i>argomento</i> , month
Solo video di quest'anno	<i>argomento</i> , year

Questi operatori possono essere combinati separandoli con virgole (per esempio *john oliver, hd, month*).

## 18. Ricerche di dati scientifici: Wolfram Alpha

**Wolfram Alpha** ([www.wolframalpha.com](http://www.wolframalpha.com)) non è un motore di ricerca nel senso stretto del termine, perché non scandaglia e non cataloga le pagine di Internet ma usa invece delle fonti accademiche e commerciali selezionate. Tuttavia viene usato per scopi analoghi a quelli dei motori di ricerca tradizionali: in particolare viene adoperato per ottenere risposte a domande complesse.

A patto di sapere l'inglese, lingua nella quale vanno immesse le domande, Wolfram Alpha ha capacità davvero notevoli. Ecco alcuni esempi di domande alle quali sa rispondere: non per nulla viene usato da Siri, l'assistente vocale dei dispositivi mobili Apple.

- Confronti fra oggetti usando l'operatore vs (*Earth vs Mars, France vs Germany*)
- Grafici matematici (*graph sin t + cos (sqrt(3)t)*)
- Localizzazione su Internet (*where am i?* oppure un indirizzo IP)
- Giorni mancanti a un evento (*days until easter*) oppure giorni fra due eventi (*days from today to 5/2/12*)
- Calcoli di età relativa (*How old was Paul McCartney on June 12, 1976?* oppure *Is Mick Jagger older than Bill Wyman?*)
- Generazione e valutazione di password (*8 character password*)
- Parole che iniziano con una certa lettera e finiscono con un'altra (*English words starting with B and ending with F*)
- Classifiche (*What is the third largest country in the world?*)
- Distanze correnti dei pianeti (*distance between Pluto and Jupiter in km* o *distance between Earth and Venus in light seconds*)
- Analisi del proprio profilo Facebook (*http://www.wolframalpha.com/input/?i=facebook%20report#*)

## 19. Ricerca e identificazione di brani musicali

È possibile cercare e identificare brani musicali utilizzando gli appositi motori di ricerca, ai quali si fornisce un campione sonoro (originale o ricantato) del brano desiderato oppure il nome dell'artista e il titolo del brano, digitati o pronunciati ad alta voce nel microfono del computer o del telefonino.

I principali motori di questo genere sono:

- *SoundHound* ([www.soundhound.com](http://www.soundhound.com)), che ha 260 milioni di utenti al mondo ed è disponibile su App Store, Google Play, Windows Phone Store e Blackberry 10 e anche per Mac;
- *Shazam* ([www.shazam.com](http://www.shazam.com)), che ha circa 100 milioni di utenti ed esiste in versione Android, iOS, BlackBerry OS, e Windows.

## 20. Altri archivi e motori di ricerca specialistici

**Informazioni sulle persone.** Intelius.com è una società che offre verifiche d'identità e di reputazione ai consumatori e alle aziende, soprattutto nel campo degli incontri sentimentali e nel lavoro e principalmente per gli Stati Uniti.

**Informazioni sulle aziende.** Zefix.ch e Moneyhouse.ch offrono ricerche d'informazioni in base a nomi e cognomi di persone o nomi di aziende, con dettagli sulle loro attività, in Svizzera.

**Proprietari di siti Internet.** Whois.net e Domaintools.com consentono di sapere chi sono gli intestatari e gli amministratori di un sito Internet a partire dal nome del sito; offrono anche una ricerca cronologica degli intestatari passati.

**Sicurezza informatica.** Shodan.io è un motore di ricerca per computer e altri dispositivi collegati a Internet; consente di trovare tipi specifici di dispositivi e di ottenere informazioni sul loro funzionamento per segnalare o prevenire violazioni di sicurezza.

**Aviazione.** FlightRadar24.com e FlightAware.com tracciano in tempo reale i voli commerciali di tutto il mondo; Airdisaster.com cataloga e archivia gli incidenti aerei; NTSB.gov è l'ente statunitense per la sicurezza dei trasporti che si occupa delle indagini sugli incidenti aerei.

**Leggi.** Findlaw.com offre informazioni legali e archivi di leggi, principalmente statunitensi, australiane, britanniche e canadesi. LexisNexis.com offre uno dei più grandi archivi al mondo di documenti legali statunitensi, francesi, australiani, canadesi, sudafricani e britannici

**Documenti riservati.** Wikileaks.org pubblica dossier e archivi riservati provenienti da governi, eserciti, banche e imprese che rivelano violazioni delle leggi.

**Brevetti.** Presso Espacenet.com è disponibile un motore di ricerca specializzato per la documentazione brevettuale di tutto il mondo.

**Riviste scientifiche.** Moltissime pubblicazioni mediche sono catalogate e cercabili presso PubMed.gov. ScienceDirect.com ospita circa 2500 riviste accademiche e oltre 26.000 libri digitali. IngentaConnect.com e OnlineLibrary.wiley.com sono altri due vasti archivi di pubblicazioni tecniche. CiteseerX (*citeseer.ist.psu.edu*) è un motore di ricerca dedicato al campo delle pubblicazioni scientifiche.

## 21. Ricerche iper-personalizzate: cosa diamo a Google

Google e quasi tutti gli altri motori di ricerca sono gratuiti in base a un patto: l'utente li usa liberamente, ma in cambio cede i propri argomenti di ricerca al motore usato. Questo consente una raccolta impressionante di dati personali.

Per scoprire quali vostri dati personali di ricerca Google acquisisce e conserva, il punto di partenza è <https://myactivity.google.com/myactivity>, da visitare entrando nel proprio account Google.

Qui troverete, radunati **per tutti i dispositivi sui quali usate l'account Google e accessibili solo a Google e a chi ha accesso al vostro account:**

- l'elenco delle parole che avete cercato in Google
- l'elenco dei video visti su Youtube
- le applicazioni usate (come Instagram, Orologio, Mappe, Telefono...)

Nella sezione *Altra attività di Google* (<https://myactivity.google.com/more-activity>), troverete in particolare la cronologia delle vostre geolocalizzazioni, giorno per giorno.

Presso <https://myactivity.google.com/myactivity?restrict=vaa&hl=it>, infine, trovate tutte le registrazioni della vostra voce, archiviate da Google quando avete effettuato una ricerca vocale (per esempio toccando l'icona del microfono oppure dicendo "Ok Google...") o quando avete dettato un messaggio in un'app di Google.

Queste attività sono eliminabili selettivamente o a blocchi/intervalli andando alla voce di menu *Elimina attività per*.