

Ricerca

La ricerca di informazioni in Rete

La ricerca di informazioni in Rete e' svolta adoperando molteplici strumenti che sono comunemente definiti motori di ricerca. In realta' quando cerchiamo in Rete un'informazione, forse senza rendercene conto⁴⁶, utilizziamo una serie di strumenti dotati di caratteristiche e funzioni molto diverse fra loro e tra i quali solo pochi possono essere propriamente classificati nella categoria dei "motori di ricerca".

Secondo la nostra interpretazione, i principali strumenti di ricerca si possono suddividere in quattro classi distinte: le comunita' in Rete; i

bookmarks o directories; gli agenti intelligenti o strumenti push; ed infine gli autentici Motori di Ricerca.

Per comunita' in Rete si intendono i newsgroups⁴⁷, le chat⁴⁸, le mailing-lists⁴⁹, e tutto cio' che si usa in genere per interloquire con altri soggetti umani. Spesso e' in questi ambiti che si trovano le notizie piu' fresche e aggiornate, o le riflessioni piu' originali rispetto a un argomento di pubblico dibattito. Cio' che contraddistingue la comunita' da altri ambiti, ormai sempre piu' chiusi e gerarchici, e' un generoso quanto impressionante spirito collaborativo - sicuramente all'impronta della solidarieta' e della cooperazione orizzontale - e che deriva storicamente dagli atteggiamenti ed atmosfere di aperto sperimentalismo maturate in seno alla scena anni '70 delle prime comunita' hacker che coltivavano e attuavano il (bi)sogno di praticare una reale socializzazione dei saperi⁵⁰. Oggi troviamo un esempio di questo spirito di co-partecipazione nelle situazioni in cui qualcuno in difficolta' lancia un appello di aiuto - perlopiu' richieste di natura tecnica per risolvere problemi di funzionamento hardware o software - e subito riceve risposte esaurienti da persone altamente qualificate. Le comunita' - come pratica di comunicazione orizzontale - , nella maggior parte dei casi, rappresentano anche un tool, o meglio, un contesto in cui reperire informazioni ad hoc o risolvere esigenze specifiche⁵¹.

Un tentativo classico per reperire informazioni rimane comunque quello di improntare una ricerca sulla base di cataloghi gia' esistenti, la

cosiddetta collezione dei siti bookmarks, frutto delle segnalazioni di precedenti navigazioni. Oppure si può procedere facendo riferimento a servizi analoghi di tipo pubblico online: le directories dove sono raccolti, divisi per categorie di interesse, numerosi siti specializzati in settori anche molto specifici e raccolti in cataloghi⁵² (dalla letteratura cyberpunk all'immersione subacquea, dalle sostanze psicoattive all'esoterismo). Rientrano nella categoria delle directories sia la proliferazione odierna dei portali sia il caso di alcuni famosi servizi, sempre di directories ma pubblicizzati come motori di ricerca: l'italiano Virgilio o l'internazionale Yahoo⁵³. Pur integrando nei loro servizi anche possibilità di reperimento dati nello stile e nelle modalità dei motori di ricerca classici, in realtà forniscono principalmente una raccolta di siti divisi per categorie, catalogata ed etichettata da apposite redazioni in tutto e per tutto simili ad eclettiche redazioni di un'agenzia di informazione presente in altri media e la cui composizione raggiunge perfino le centinaia di navigatori professionisti.

Alcuni cybernaviganti particolarmente smaliziati preferiscono invece ricorrere a tecnologie più sofisticate e impegnative: le tecnologie push o agenti intelligenti, cioè tutti quegli strumenti che permettono, una volta opportunamente configurati, di cercare per conto proprio l'informazione dispersa nel Web partendo da una rotta iniziale di navigazione (anch'essa tracciata dall'utente oppure personalizzata sui suoi bisogni da parte del servizio che offre tale tecnologia) e scandagliando gli abissi del cyberspazio con una rete costruita ad hoc e nelle cui trame digitali rimane

impigliato solo il materiale auspicato, già setacciato alla fonte da eventuali indesiderata. Si tratta, a questo proposito, dei gettonati servizi informativi di carattere finanziario, notiziari politici e sportivi che provvedono a far arrivare alla scrivania multimediale e telematica di accreditati giornalisti e businessmen, ad esempio, bollettini quotidiani sull'andamento delle borse o anche solo del tempo meteorologico⁵⁴.

Ma allora in che cosa consiste un vero e proprio Motore di Ricerca, l'oggetto ancora misterioso dal fascino macchinico e dal nome ingegneristico di questo approfondimento...?

Il Motore di Ricerca

Il motore di ricerca è un insieme di strumenti hardware e software funzionali alla ricerca dell'informazione: procedimento complesso e composto da più fasi.

La prima fase è costituita dall'individuazione dei documenti su cui effettuare la ricerca di informazioni e è svolta dai cosiddetti programmi robot o spider. È un'attività molto impegnativa dal punto di vista delle risorse di connettività e di spazio disco, tanto più che l'informazione in Rete è spesso destinata a mutare contenuto o indirizzo oppure addirittura a scomparire. È questo forse uno dei lati deboli dei motori di

ricerca poiché i limiti di spazio disco e di connettività obbligano i motori a dover prendere in esame solo una parte della Rete, ignorandone la geografia globale. Sommarariamente, esiste una serie di url⁵⁵ di partenza dai quali - di link in link⁵⁶ - sono recuperati molti documenti fino al limite massimo di spazio disco occupabile e, connettività permettendo, in un lasso di tempo accettabile. L'informazione così trovata sarà quindi oggetto di visite periodiche da parte dello spider, in un ciclo che si aggira sulle 3-4 settimane⁵⁷. Il limite, pur fisiologico della porzione di Rete monitorata e che è quantificabile nelle migliori delle ipotesi nel 10-15% presunto⁵⁸, è una delle ragioni per cui i motori si prestano all'ipotetica accusa di controllo o addirittura censura dei contenuti della Rete. In base a quali considerazioni si procede all'esclusione di materiale: considerazioni culturali oppure interessi di marketing? Va pur detto, a onore del vero, che, escludendo alcuni motori di ricerca famosi come Virgilio e Yahoo ma che in realtà non sono altro che directories create da un lavoro redazionale e quindi basato su meccanismi umani di discrezionalità di scelta del materiale esaminato, la maggior parte dei motori classici fornisce con un comando al loro interno (del tipo ADD URL / ADD SITE) la possibilità di aggiungere un sito qualsiasi all'elenco delle liste da reperire e indicizzare. In altre parole tali comandi conferiscono al motore un meccanismo di relativa apertura all'utenza finale.

Una volta reperiti i documenti si passa alla fase successiva:

l'indicizzazione. L'indicizzazione è svolta in maniera funzionale, attraverso complessi algoritmi⁵⁹ che tengono conto della rilevanza e della

frequenza dell'informazione contenuta nel documento da catalogare per essere immesso in un database. L'obiettivo finale di indicizzazione consiste infatti nella creazione del database, una raccolta-catalogo che dovrà essere interrogabile attraverso interfacce user-friendly dal navigatore in cerca di informazioni. Ciò che, al di là di questi processi sotterranei, infine si manifesta all'utente finale è la lista dei documenti stimati più rilevanti rispetto all'interrogazione effettuata, secondo il criterio del ranking⁶⁰.

I motori principali

Senza voler essere provinciali, iniziamo con due esempi "italiani":

Virgilio⁶¹, "il bello di Internet", un motore di ricerca in versione directory fra i più pubblicizzati e conosciuti; e Arianna, storicamente "il motore di ricerca italiano". Il primo, essendo estremamente selettivo nella sua parte di catalogo scelto da una redazione, è anche molto scarso di informazione. Viceversa Arianna, essendo un motore di ricerca nel senso classico ora descritto, copre una grande fetta di Rete: oltre 5 milioni di pagine Web indicizzate e la capacità di soddisfare oltre cinquecentomila interrogazioni al giorno in maniera efficace.

Fra i più particolari in campo "internazionale" è Excite che a tutt'oggi possiede versioni opportunamente disegnate sulle caratteristiche

informative e linguistiche di vari paesi - Australia, Canada, Cina, Francia, Germania, Italia, Giappone, Paesi Bassi, Spagna, Svezia e Regno Unito.

La directory per eccellenza in campo mondiale e' senza dubbio Yahoo, sebbene NorthernLight proponga fra le sue versioni di presentazione dei risultati anche un approccio che, impiegando cartelle (SEARCH FOLDERS) suddivise in 200mila soggetti tematici, mostra i risultati nello stile organizzato ad albero della directory.

Google, uno degli ultimi arrivati ma che ha subito conquistato una grande visibilita' in Rete grazie alla sua grafica semplice e scarna senza dar troppo rilievo ai banners pubblicitari, con l'opzione "I'm feeling lucky" fornisce la possibilita' di trovare immediatamente l'argomento di interesse. Si tratta di una funzione il cui successo e' tanto illusorio quanto fortuito, come illustra il nome, ma che nasconde una delle proprieta' che ha fatto non solo la fortuna di questo giovane motore di ricerca ma anche scuola, rispetto agli altri, in fatto di ranking. La popolarita' di un sito, quindi, il premio consistente nell'essere posizionato in cima ai risultati di ricerca (il tanto ambito top ranking) vengono infatti misurati sul numero di link che questo sito puo' vantare in Internet ovvero sul fatto che tanti altri siti lo abbiano linkato. E' una funzione sicuramente efficace ma che nasconde anche un risvolto perverso: il fatto di premiare sempre e comunque i piu' noti, finendo per escludere - e non a torto si puo' dire censurare - le piccole realta' emergenti e poco

pubblicizzate⁶² che, secondo questa logica a spirale, appaiono destinate a rimanere nell'ombra nella maggior parte dei casi.

Per facilitare la ricerca tra una ormai esponenziale e proliferante mole di informazioni, i motori di ricerca si sforzano di escogitare nuove tecniche.

Infoseek, ad esempio, si e' inventato la possibilita' di ricercare documenti che siano il piu' possibile simili ad uno in particolare⁶³, mentre DirectHits pone in evidenza i siti piu' visitati su un determinato argomento.

HotBot mette a disposizione dell'utente una miriade di variabili di ricerca tra le quali le piu' celebri riguardano le possibilita' di ricercare documenti reperiti e indicizzati, rispettivamente nell'ultimo mese, nelle ultime due settimane o nell'ultima settimana, sviscerando - come dice la traduzione stessa del nome - la possibilita' d trovare materiale "caldo", le notizie piu' fresche e aggiornate.

Altavista, sia nella versione italiana che inglese, e' uno tra i pochi motori che presenta addirittura la possibilita' di utilizzare comandi specifici di ricerca avanzata come si puo' illustrare in alcuni esempi:

"Corvo Rosso": cerca esattamente la stringa "Corvo Rosso" e nulla al di fuori di questa espressione

domain:it: pagine che fanno riferimento solo al dominio .it

link:http://www.shake.it: verifica quanti links secondo il database di

Altavista puntano a tali siti

host:www.ecn.org: pagine che fanno riferimento all'host (nodo o server di Rete) www.ecn.org

image:che: pagine che contengono presumibilmente immagini del che

Altavista propone altri tools aggiuntivi, come la possibilita' di operare una ricerca mediante l'impiego degli operatori booleiani⁶⁴. Si tratta in parole povere di operatori utilizzati in logica matematica per allargare o restringere un campo di definizione. Ad esempio:

corvo AND rosso: cerca "corvo" e "rosso"

corvo AND NOT rosso: cerca "corvo" ma non "corvo rosso"

corvo NEAR rosso: cerca "corvo" e "rosso" dove si presentino non solo nella stessa pagina come per AND ma anche vicini

Di grande utilita' e risparmio di tempo sono infine i meta-motori di ricerca⁶⁵, sistemi piu' semplici dei motori di ricerca poiche' constano di interfacce che vanno ad interagire, rigirando l'interrogazione fatta al metamatore, con decine oppure centinaia di motori di ricerca. Nel

guadagnare in numero di risposte, si perde tuttavia in qualità; c'è il rischio di effettuare un tipo di interrogazione meno raffinata, dal momento che con questi strumenti risulta troppo complesso richiedere un'interrogazione sofisticata come quella illustrata per Altavista. Infatti interrogando più motori aventi regole di interrogazione anche molto diverse tra loro, la ricerca avanzata è drasticamente ridimensionata.

In una panoramica sui motori di ricerca, appare evidente che molti motori condividono lo stesso database dal quale attingono le informazioni da presentare nella lista dei risultati di ricerca finale, pur ordinandoli in maniera diversa. A questo proposito una risorsa condivisa da molti motori di ricerca è infatti il famosissimo Inkitomi66.

Tecniche di ricerca

I motori di ricerca, per loro natura fisiologica e struttura anatomica, rappresentano uno strumento con cui l'utente, cybernavigante esperto o di primo pelo, si interfaccia con la ricerca dell'informazione in Rete.

Spogliata dalla responsabilità della ricerca, la Rete si presenterebbe infatti come un immenso oceano informazionale, un campo a maglie larghe le cui frontiere sono in continuo e fluttuante movimento. Come reperirvi l'informazione? Quali pratiche sperimentare? Bisogna compiere una serie di passaggi pratici per raggiungere il reperimento delle informazioni in una

situazione così complessa, innanzitutto ovviando alle problematiche e ai limiti affrontati finora, nella speranza di usufruire di tutte le opzioni offerte dalla Rete e dall'inventiva dei naviganti stessi.

Una buona ricerca in Rete deve essere performativa, nel senso di fondarsi su un criterio efficientistico, perlomeno nel tentativo di ottimizzare le capacità tecniche del mezzo; ma deve anche essere in grado di integrare l'impiego della tecnologia con la creatività, dal momento che sappiamo che non esiste nessuna verità ma solo combinazioni pragmatiche⁶⁷.

Suggeriamo in questo senso un possibile percorso di ricerca:

- 1) verificare se fra i propri bookmarks esistono già indirizzi utili;
 - 2) ipotizzare un dominio `www.info_da_cercare.it` ma anche i suoi possibili equivalenti `.com`, `.org`, `.net` ecc.⁶⁸;
 - 3) non vergognarsi di chiedere alla comunità di cui si fa parte in Rete, sia essa una lista, un newsgroup o un canale irc, lumi in materia; e' altrettanto efficace provare a cercare altre comunità in Rete che si presume che siano a conoscenza delle informazioni cercate;
 - 4) trovare tutti gli strumenti push in materia noti per poter ricercare il tipo di informazione auspicata;
-

5) sfogliare le directories, i cataloghi e i portali in Rete pertinenti in materia⁶⁹;

6) interrogare tutti i motori di ricerca e i metamotori conosciuti tramite interrogazioni semplici ma anche in modalita' avanzate (e dove possibile ricorrere alla logica booleiana);

7) avvalersi di help online come l'italiano www.vorreisapere.com sul quale si puo' inviare una richiesta ad un operatore in carne ed ossa che, per nostro conto, effettua la ricerca inviandone il risultato per email⁷⁰.

Solo l'esercizio della critica e l'interazione comunicativa, nell'ambito di una pratica collettiva dei saperi, possono forse esorcizzare i rischi di tecno-controllo o censura digitale insiti in un perverso e cattivo impiego dei motori di ricerca in Rete. Nella prospettiva postmoderna di una proliferazione che tende all'infinito dell'informazione su web e della sempre maggiore volatilita' dell'informazione stessa ,rimane praticabile la strategia del giocatore - di chi conosce le regole del gioco - ma ne sa anche stravolgere il senso e capirne la logica profonda.

46 Con questo non vogliamo accusare di dabbenaggine la media delle persone naviganti in Rete ma e' pur vero - purtroppo - che si sta affermando un modello per cosi' dire televisivo dell'utilizzo della Rete per cui si svolge automaticamente una serie di funzioni immediatamente utili ma senza riflettere sulla natura e sulle caratteristiche di funzionamento degli strumenti di Rete, senza tenere conto che tutto cio' influisce negativamente sulla qualita' dei risultati ottenuti.

47 I newsgroups sono forum pubblici di discussione usufruibili tramite un newsreader come il famoso FreeAgent che si collega a un newsserver pubblico o piu' comunemente al newsserver del proprio provider di accesso per visionare i forum di eventuale interesse ai quali abbonarsi e partecipare

alla discussione con decine, a volte migliaia di altri utenti. I newsgroups si trovano in tutte le lingue e su i piu' disparati argomenti, fra quelli storici e a volte piu' interessanti e' doveroso segnalare quelli contrassegnati con la gerarchia alt(ernative).

48 Le chat sono tutti quei momenti in cui si realizza una comunicazione in tempo reale con uno o piu' interlocutori umani perlopiu' attraverso servizi Web appositamente creati dallo scopo, ma anche attraverso software molto veloci ed efficaci come l'IRC - Internet Relay Chat - oppure attraenti e flessibili, per le loro caratteristiche multimediali come ICQ - I Seek You. Da notare come in altri casi, acronimi che identificano software o servizi in Rete si debbano sciogliere con la pronuncia inglese ottenendo la descrizione del loro funzionamento.

49 Per iscriversi a una mailing-list, ovvero a un dibattito che si svolge attraverso lo scambio di e-mail fra indirizzi di posta personali, e' necessario inviare una stringa di comando che varia da mailing-list a mailing-list a un programma apposito come Majordomo o Listserver, oppure interfacciarsi ad appositi servizi Web che facilitino questo compito di iscrizione.

50 Sono gli anni, decine sessanta-settanta, prima della privatizzazione delle scoperte dell'industria informatica di base, in cui si elabora una pratica collettiva della ricerca di avanguardia nel settore del software e dell'hardware. Ne e' un emblema la figura di Felseinstein, eretico pioniere

dell'informatica (Raf Valvola Scelsi, No Copyright, Shake Edizioni, Milano1994).

51 In questa sezione verra' approfondito il funzionamento dei motori di ricerca, mentre le possibilita' di ricerca offerte dalle comunita' in Rete, dalle directories e dalla tecnologia push sara' materia di approfondimento di sezioni successivi.

52 Il catalogo, secondo l'interpretazione di Bruno Snell ("La civiltà greca e le origini del pensiero europeo", Einaudi Tascabili), e' uno dei tratti di matrice comune del pensiero occidentale. In particolare fa la sua prima apparizione nella letteratura arcaica dove prima Omero e poi Esiodo lo adoperano nei loro scritti (celebre il "Catalogo delle navi" nel secondo capitolo dell'Iliade e l'elenco delle Muse ne "I lavori e i giorni") utilizzandolo come elemento teorico di conoscenza ma anche come elemento esornativo ed espediente narrativo: l'esigenza di creare un ordine (kosmos) sia mentale che estetico dal chaos dell'universo esistente si manifesta nella scrittura del catalogo, che appunto e' una delle modalita' di ordinare gli oggetti nella forma di lista.

53 E' curioso utilizzare termini come nazionale e internazionale in un approfondimento come questo che trattando di Rete non dovrebbe tenere conto delle frontiere politico-burocratiche: in realta' sin intende per "italiano" un servizio realizzato in lingua italiana o in Italia; e "internazionale" un servizio che per la sua capacita' di trattare materiale

informativo originario di piu' paesi e in piu' lingue puo' essere appunto cosi' definito.

54 Spesso e volentieri si pensa a questi agenti intelligenti nei termini di strumenti molto raffinati e di difficile gestione. In effetti in alcuni casi lo sono, ma esistono applicativi relativamente semplici da utilizzare come Teleport o Wget che configurati in maniera specializzata e puntati su siti di partenza efficaci sono in grado di realizzare report quotidiani di grande efficacia su ogni argomento.

55 URL sta per Unique Resource Location ovvero significa che ad un indirizzo - in questo caso indirizzo web - puo' corrispondere unicamente un solo documento.

56 Una delle caratteristiche principali del Web e' appunto quella di poter effettuare una navigazione ipertestuale che appoggiandosi allo strumento del link permette di passare da un documento a un altro indipendentemente dalla loro localizzazione geografica o di Rete.

57 La periodicitá in questione dipende da vari fattori ed e' direttamente proporzionale alla frequenza con cui viene aggiornato il sito su cui si trova la documentazione reperita dal motore stesso.

58 Presunto nel senso che e' letteralmente e scientificamente impossibile determinare la grandezza della Rete e la quantita' dei documenti ivi

presenti.

59 L'argomento verra' piu' diffusamente trattato nella prossima sezione.

60 Un articolo molto sintetico e chiaro e' quello di Attardi Giuseppe, Ricerca di informazione su Web, in AI*IA Notizie, Vol. 12, 1999, No. 2, p. 24-27.

61 Per visualizzare uno qualsiasi dei motori di ricerca citati in questo approfondimento e' sufficiente anteporre www. e posporre .com oppure .it (nel caso degli esempi italiani) al nome indicato, escamotage linguistico che peraltro mette in risalto il processo di massificazione e commercializzazione di questi strumenti.

62 Si potrebbe dire poco linkate in questo caso: ovvero il link come sinonimo di popolarita'.

63 Dato un sito appena reperito e di comprovato interesse si puo' utilizzare questa funzione del motore per reperire altri siti il cui contenuto informativo e' stimato simile.

64 Da George Boole (1815-1864) scienziato che invento' quella che poi venne chiamata algebra booleana: la base matematica sulla quale si fonda l'algebra di commutazione dell'elettronica digitale che consente di elaborare i valori falso e vero che si riferiscono a proprieta',

proposizioni, enunciati eventi ecc.

65 Un esempio per tutti www.metamonster.com che nella sua versione completa e' in grado di interrogare contemporaneamente piu' di 500 motori di ricerca.

66 www.inkitomi.com

67 Jean-François Lyotard ne *La condizione postmoderna*.

68 Nel momento in cui andra' in stampa questo approfondimento probabilmente sara' stato approvato il nuovo regolamento dell'ICANN - una delle tante istituzioni che regolano Internet - riguardanti i suffissi di dominio di primo livello e che prevede molte nuove possibilita' come .shop per attivita' commerciali e .news per i servizi informativi

69 Il percorso di ricerca puo' essere differente da quello proposto e puo' anch'esso essere di natura ipertestuale, cioe' si potrebbe chiedere nell'ambito della mailing-list a cui si partecipa se esistono directories utili in materia oppure si puo' cercare con un motore di ricerca se esiste un canale push utile allo scopo.

70 Anche se non e' attualmente stimato molto affidabile questo genere di servizi e' emergente e si conquistera' probabilmente in un prossimo futuro una certa notorieta' e quindi una fetta di mercato ragguardevole: negli

States, per esempio, sono già numerosi e molto frequentati i servizi educational, di training e di formazione universitaria online che possono, in un certo senso, anche essere considerati complessi sistemi di ricerca-reperimento e acquisizione dell'informazione. Non c'è da stupirsi della (apparente) gratuità di alcuni di questi servizi poiché viene bilanciata dalla fidelizzazione del nostro indirizzo di posta elettronica che viene presumibilmente associato a una identità - si' di Rete ma corrispondente a un consumatore vero e proprio che ha delle preferenze anche di consumo (e che sono dichiarate tramite la richiesta di ricerca) e che quindi sarà forse oggetto a spam di tipo commerciale.



[Aggiungi un Sito](#)
Gjlg Metamotore, [Gjlg Metasearch](#)