

Disclaimer

Questo volume non intende essere una ricerca esaustiva sull'argomento Google. Questo dev'essere considerato come un volume introduttivo: l'argomento Google è molto più vasto e include troppi aspetti troppo diversi tra di loro per essere inclusi sotto un unico titolo.

I capitoli qui trattati non coprono quindi gli argomenti più importanti in assoluto, ma trattano quegli aspetti più interessanti secondo l'autore.

Le informazioni qui contenute sono da ritenersi accurate al momento della pubblicazione. Queste informazioni possono cambiare in ogni momento. Invitiamo il lettore a verificarle prima di utilizzarle.

In particolare il metodo con cui Google assegna il PageRank ai siti e i criteri che usa per valutarli, sono soggetti a variazioni rapide e imprevedibili. Si consiglia di effettuare delle ricerche e verificare che quanto scritto in questo volume sia ancora attuale.

L'autore, o chi collegato a esso, non possono essere ritenuti responsabili dei danni causati dall'utilizzo di informazioni contenute in questo volume.

A chi è rivolto questo libro?

A tutti!

A partire da mia zia, che alla tenera età di 77 anni ha deciso di scoprire cos'è "...questo Internet di cui si parla tanto" , fino al programmatore che mangia codice C/C++, respira Internet e dorme con un manuale di protocolli TCP/IP sotto il cuscino.

Questo libro serve a chi:

Vuole **capire** cos'è Google, un motore di ricerca e Internet.

Ha bisogno di **trovare informazioni** sia a livello **professionale** che personale.

Deve migliorare la visibilità di un sito tramite reali tecniche di **Search Engine Optimization**.

Fa **pubblicità on-line** e necessita tecniche pratiche e comprovate per ottenere il massimo da **AdWords**.

Indice

Disclaimer	1
A chi è rivolto questo libro?	2
Indice	3
Introduzione	5
1. Google per principianti	7
2. Tecniche avanzate	15
3. Gli altri servizi di Google	20
Google Toolbar	20
GMail - Google Mail	21
Google Desktop Search	23
Google Images	25
Le News	25
Google groups e Usenet	26
Froogle - lo shopping comparato	26
Traduzioni on-line	27
Google Answers	27
Google Alerts	28
AdSense	29
Google Local	30
Google Catalogs	30
Account	31
Google Sets	31
Google Labs	32
Google Calculator	33
Google Games	34
4. Search Engine Optimisation	36
5. AdWords	48
Introduzione	48
La struttura di AdWords	48
Non dimentichiamoci del resto	50
Come funziona AdWords	50
Il primo consiglio	50
Tracking	51
Approccio a AdWords	52
I gruppi di annunci	52
Come scrivere un annuncio	53
Matching	53
Nuove keyword	54

Le dimensioni contano	56
Keyword illegal?	56
Powerposting	56
Il click through ratio	57
Sindacare o non sindacare	58
Il budget giornaliero	58
Ottimizzare gli annunci	59
Cost per acquisition (CPA) e budget	59
Essere i primi?	60
Conclusioni	61
6. Storia di Google	62
Perché Google?	62
L'inizio di un lungo cammino	62
In cerca di un acquirente	63
Toccati da un angelo	63
Il Googleplex in un garage	64
Di nuovo on the road	64
Google non è più beta	65
L'innovazione è nel DNA	65
Google assume	67
Google si espande	68
Nuovo capo, nuova direzione	68
Informazioni senza barriere	69
7. Il futuro	70
Quali saranno le sfide di Google?	70
Perché non open-source?	71
Google pubblica?	72
Competitor	73
8. Google Matematicamente	77
Conclusioni	83
Glossario	85
Appendice I. Google Suggest - Google dalla A alla Z	89
Appendice II. Codice sorgente per calcolare il PageRank	94
Appendice III. Risorse on-line	99
pc-facile.com	99
Google Groups	99
Search	100
Search Engine Optimization	100
AdWords	101
Link generici	101
Nota finale	102

Introduzione

Perché questo libro?

Mettiamoci subito d'accordo su una cosa: su Internet, da qualche parte, esiste l'informazione che stai cercando. Sempre. Non importa cosa tu stia cercando: il testo di una canzone, un libro che non è ancora pubblicato o i piani per la bomba atomica (e ci sono davvero i piani della bomba atomica!). C'è. Sta a te trovarla!

Per quante definizioni possano esistere, Internet ha un solo scopo: lo scambio d'informazioni. Non c'è altro. Io ho un'informazione e la rendo disponibile. Sta a te trovarla. Se sono bravo posso cercare di pubblicizzarla, ma non sarà mai l'informazione a venire a cercare te, starà a te trovare lei. Solo che, con il volume d'informazioni disponibili, trovare non è facile: saper cercare diventa quindi una necessità.

Un libro medio ha **300 pagine**, l'enciclopedia Britannica contiene oltre **65.000 articoli**, Amazon.com, il più grosso negozio di libri virtuale ha in catalogo circa **300.000 libri**. Google ha invece indicizzato oltre **8 miliardi di pagine**.

Neanche moltiplicando il numero di pagine medie di un libro al catalogo di Amazon riusciamo ad avvicinarci al volume di pagine indicizzate da Google – e secondo molti queste rappresentano solo una frazione di ciò che è realmente disponibile su Internet. Esiste un vero e proprio El Dorado di informazioni là fuori, cerchiamole!

È vero, la maggior parte di ciò che troviamo è spazzatura, una spazzatura che ci nasconde i gioielli custoditi da Internet. In gergo tecnico questa spazzatura viene definita: rumore di fondo, come centinaia di voci che parlano in una stanza e ci impediscono di ascoltare l'Aristotele che davvero ci interessa. E nell'era dell'informazione il rumore di fondo è talmente assordante che rischiamo di annullare i benefici di questo grande strumento che si chiama Internet; esattamente come lo spam rischia di affogare l'e-mail.

L'abbondanza di dati, parole e numeri, la caratteristica che ha reso grande Internet, potrebbe essere la causa della sua distruzione. Internet ha tre scogli difficili da superare. Se ci riesce nulla potrà arrestare al sua diffusione.

Nell'era dell'informazione queste tre categorie di problematiche sono: la distribuzione delle informazioni - dove si affrontano i problemi legati alla diffusione dei contenuti (in questo campo troviamo i software P2P, tra cui spicca il modello BitTorrent, i feed RSS, l'e-mail, le newsletter...), l'autenticazione dell'origine - cioè non solo come avere la certezza che l'autore delle informazioni sia davvero chi dice di essere, ma anche poter attribuire un valore a questo autore e alle sue informazioni (ricordiamo che fenomeni come lo spam sono causati da questa mancanza di autenticazione) e la ricerca.

Google ha cercato di trovare una soluzione al problema della ricerca e nell'affrontare questo ha dato una risposta a come assegnare un valore alle informazioni in rete. Questi due problemi, come vedremo, sono legati alla radice.

A oggi Google è universalmente riconosciuto come il motore di ricerca per eccellenza, cioè quello che di gran lunga sa meglio trovare le pagine davvero rilevanti per noi. È talmente sicuro di sé che ha un bottone “Mi sento fortunato” che non ci mostrerà tutti i risultati, ma ci porterà direttamente a quello che lui considera contenga ciò che ci interessa.

Già, è così sicuro di sé. E molte volte ha ragione di esserlo.

Il primo lettore di questo libro sono io. Ero bravino a trovare le informazioni che mi servivano. Poi ho dato per scontato che ci sapevo fare e ho smesso di imparare. Ma da qualche tempo mi sono accorto che altri trovano cose che io neanche sapevo sistemare. Ho deciso di documentarmi. Questo è il risultato delle mie ricerche.

1. Google per principianti

Le basi della ricerca sono molto semplici. Si inseriscono alcune parole, dette keyword, in un motore di ricerca e questo cerca, tra i documenti che ha immagazzinato, quali secondo lui sono quelli per noi più interessanti data la query (somma delle keyword) inserita. Nient'altro.

Negli anni ho scoperto una cosa: cercare è un'arte, non una scienza.

Per chi ha una mente predisposta alla matematica o all'informatica, i computer, nella loro rigorosa logica, sono più facili da capire. Per le persone le cui predisposizioni sono invece ben lontano dal freddo razionalismo, questi consigli possono dare un aiuto a capire quanto sia ricca di informazioni Internet.

Convenzione Le query saranno sempre indicate tra parentesi quadre: [query]. Questo perché le virgolette hanno un significato preciso all'interno di una query. Queste parentesi non devono essere incluse quando si digita una query.

Scelta delle keyword

La scelta delle keyword è fondamentale. Tenete a mente questi piccoli consigli quando fate una ricerca:

Come prima cosa provate con le **keyword ovvie**. Se cercate informazioni su Picasso usate [Picasso], non [pittori].

Pensate a quali parole potrebbero essere sulla pagina che cercate. [Hotel Firenze] darà risultati più rilevanti di [un bel posto dove trascorrere il fine settimana]. Non strutturate le vostre query sotto forma di domanda: è più probabile che troviate una pagina che contenga la domanda piuttosto che la risposta che cercate. Ad esempio, la seguente query [dove posso fare domanda per ottenere un visto per l'Egitto], non vi sarà utile quanto questa query [domanda richiesta visto Egitto].

Cercate di essere il più **specifici** possibile. [automobili modellino Ferrari] vi sarà di più aiuto di [vecchi giocattoli].

Se la ricerca non dà i risultati sperati provate a cercare dei **sinonimi** delle keyword che state usando.

Cercate di pensare a una frase che potrebbe comparire sulla pagina e inseritela proprio come la potreste trovare. Ad esempio per cercare il titolo di una canzone che ho sentito in radio dico a Google di cercare tutte le parole che ho sentito in radio e aggiungo la keyword "lyrics" (che in inglese significa "testo" riferito a una canzone).

Se siete bloccati con le keyword, potete utilizzare questo Google Sets, un servizio che vi aiuta a costruire gruppi di oggetti con proprietà comuni. Ad esempio per dare aria ad una ricerca sulle arti marziali potete inserire keyword come karate, judo, jiu-jitsu, kung fu e kendo e Google vi restituirà una lista di arti marziali con cui potrete sviluppare il vostro lavoro.

<http://labs.google.com/sets/>

AND

La prima cosa da imparare è che l'operatore AND è automatico in Google. In altre parole lui cercherà tutte le parole che voi inserite. In effetti Google farà di più: cercherà tutte le parole da voi inserite dando maggiore importanza a quelle pagine che contengono tutte le parole nell'ordine in cui voi le avete inserite.

La query [iraq] da dei risultati ben inferiori (in termini qualitativi, non in termini numero di risultati) alla query [iraq guerra golfo].

Invece la query [iraq guerra golfo] darà gli stessi risultati di [guerra golfo iraq], ma questi ultimi saranno ordinati in modo diverso.

Quando effettuate una ricerca è importante cercare di pensare a una frase che potrebbe essere contenuta nella pagina che cercate e inserire le parole che compongono quella frase.

Per cercare, ad esempio, il significato di una parola molto specifica a un determinato settore che non conosco e che non trovo sui dizionari specializzati, uso una query come segue: [definizione xxx] oppure [glossario xxx]. Dove xxx è il termine che mi interessa.

Il mio ragionamento è che una pagina che mi spiega il significato di una parola includerà anche la parola "definizione" o "glossario". Se avessi inserito solo il termine è molto probabile che avrei ottenuto solo pagine in cui viene utilizzato il termine nel corpo del discorso.

Parole comuni

Google ignora le parole ed i caratteri comuni, denominati anche "stop words", scartando automaticamente termini come "http" e ".com" nonché alcune singole cifre e lettere, dal momento che questi termini non aiutano a restringere il campo di ricerca ma anzi contribuiscono a rallentare i tempi di ricerca. È quindi inutile inserire queste parole all'interno di una query.

Per Google query come [iraq guerra golfo] e [iraq la guerra golfo] sono equivalenti.

Vedremo dopo come fare in modo che queste parole vengano cercate lo stesso quando è importante che lo siano.

Massimo 10 parole

Una query può avere una lunghezza massima di 10 parole. Parole inserite oltre il limite non verranno considerate da Google.

Lettere maiuscole/minuscole o accenti

Google non fa distinzione tra lettere minuscole e maiuscole poiché considera tutte le lettere come minuscole. Ad esempio, digitando [google], [GOOGLE] e [GoOgLE] si ottengono sempre gli stessi risultati.

Stando alla documentazione ufficiale di Google, “per impostazione predefinita, Google non riconosce accenti o altri segni diacritici. In altre parole, se si digita [Munchen] e [München] si otterranno gli stessi risultati.”

In realtà provando a fare ricerche per queste due parole si otterranno risultati molto diversi tra di loro. Ancora più strano, è però il fatto che la stessa ricerca, limitando i risultati a specifiche lingue, da risultati ancora diversi. In caso di ricerche importanti è consigliabile provare più modi di scrivere la stessa parola.

Ortografia

Il correttore ortografico di Google verificherà automaticamente la vostra query per vedere se state utilizzando la versione più comunemente utilizzata per scrivere una parola. Se calcola che la probabilità di ottenere più risultati con un altro modo di scriverla sia alta vi chiederà se modificare la query.

Siccome il correttore ortografico di Google si basa sul numero di ricorrenze di tutte le parole su Internet, Google è in grado di suggerire modi di scrivere la parola che non sono presenti sui dizionari comuni.

Alcune persone usano questo sistema per verificare l'ortografia di parole di cui hanno dei dubbi. Se Google suggerirà un altro modo di scrivere la parola inserita, allora ci sono buone probabilità che questa parola sia sbagliata.

OR

L'operatore OR dice a Google di cercare una keyword oppure l'altra è indifferente. La query [Hotel Firenze OR Roma] troverà quindi tutti gli hotel di Firenze e Roma anche se le due città non appaiono nella stessa pagina.

Apostrofo

Cercare [dell'orto] e [dell orto] comporta risultati diversi.

Ricerche booleane

Google supporta solo un limitato numero di operatori booleani (in realtà si tratta dei soli operatori AND e OR) e anche questi sono supportati piuttosto male. Per cui se siete degli amanti di questo tipo di logica riuscirete a lavorarci poco.

Ecco alcuni esempi di ricerche che funzionano su Google:

[x AND y]

[x OR y]

[x AND (y OR z)]

[(x OR y) AND (z OR w)]

[x AND (y OR z OR w)]

[x AND (y OR z) AND w]

Pregiamo notare che:

l'operatore "OR" dev'essere scritto in maiuscolo

l'operatore "AND" può essere omesso

le parentesi non servono; sono state incluse qui solo per chiarezza

A dimostrazione dell'ultima affermazione portiamo ad esempio questa query che, su Google, non funziona:

[(x AND y) OR (z AND q)]

site:

L'operatore site è uno dei più comodi e utili tra gli operatori di Google.

È purtroppo abbastanza facile trovare siti il cui search interno non è molto buono per cui cercare qualcosa è un incubo. Una soluzione è quella di usare Google come motore di ricerca.

Infatti Google fa talmente bene questo lavoro che adesso offre gratuitamente questo servizio ai siti che lo richiedono.

L'operatore "site:" specifica a Google di cercare solo tra le pagine di un determinato sito.

[guerra golfo iraq site:www.corriere.it] cercherà solo tra le pagine del Corriere della Sera.

Attenzione, questo operatore funziona bene su siti che hanno dei grossi archivi che vengono aggiornati raramente e che non dispongono di una funzione di ricerca interna adeguata. Siti il cui contenuto cambia velocemente, come ad esempio nel caso dei forum, non daranno dei buoni risultati in quanto

Google non sarà mai aggiornato quanto il motore di ricerca interno.

Questo operatore ha inoltre il vantaggio di trovare documenti male indicizzati da un sito o che non compaiono nel search interno. Addirittura, un documento non più linkato può essere ancora presente negli indici di Google, come pure un documento che è stato cancellato dal proprietario del sito (gli esempi migliori potete trovarli sul sito della Microsoft).

Cancellare qualcosa, che è stato reso pubblico su Internet, è molto molto difficile. Fate sempre attenzione a cosa divulgate: è difficile tornare indietro.

L'operatore “+”

Un modo per far sì che anche le parole comuni siano incluse nella ricerca è di anteporre il segno “+” prima della parola stessa.

Ad esempio la query [Guerre Stellari Episodio 1] non cercherebbe il numero “1”. Inserendo invece la query nel seguente modo [Guerre Stellari Episodio +1] si ottiene il risultato desiderato.

Preghiamo notare che c'è uno spazio prima del segno “+”, ma NON c'è dopo.

Le virgolette

Questo è un altro modo per indicare che tutte le parole vanno incluse nella query. Quindi [Guerre Stellari Episodio +1] e ["Guerre Stellari Episodio 1"] sono equivalenti.

Ma le virgolette sono molto più potenti e dicono a Google di cercare la query esattamente come è stata inserita. In altre parole, mentre nel primo caso Google cerca una pagina contenente le quattro keyword dando una rilevanza maggiore alle pagine che hanno le keyword in quest'ordine, nel secondo caso Google cercherà esattamente quella frase.

Attenzione quindi a non abusare delle virgolette o non otterrete alcun risultato.

L'operatore “-”

Il segno “-” prima di una parola ha l'effetto contrario al “+” indicato prima.

Un utilizzo classico dell'operatore “-” è per vedere quali siti linkano un determinato sito.

["www.pc-facile.com" -site:www.pc-facile.com]

Le virgolette indicano che anche “.com” (cioè una parola comune va inclusa nella query). Il segno meno invece dice a Google che “www.pc-facile.com” non deve essere nel sito www.pc-facile.com, altrimenti i risultati indicherebbero anche tutte le pagine interne del sito www.pc-facile.com, ma che ci siano link

interni ad un sito lo sappiamo già.
Comodo, no?

Altre ricerche interessanti

Le seguenti ricerche possono aiutarvi:

[fedex xxx]: Per tenere traccia di un vostro pacco che viaggia con la Federal Express senza dover visitare il loro sito.

[ups xxx]: Per tenere traccia di un pacco che viaggia con l'UPS senza dover visitare il loro sito.

[usps xxx]: Per tenere traccia di un pacco che viaggia con la United States Parcel Service senza dover visitare il loro sito.

[patent xxx]: Per consultare l'ufficio brevetti Americano - xxx rappresenta il numero di un brevetto che conoscete

keyword1 keyword2 ["parent directory"]: (in questo caso le parentesi quadre fanno parte della query).

Per cercare un software o un mp3 senza dover passare da un programma P2P o dai nefasti siti warez.

Provare per credere!

Ricerca di radici di parole (stemming)

Per garantire la massima precisione nei risultati, Google non supporta la ricerca di radici di parole ("stemming") né le ricerche con caratteri "jolly" (*wild card*). In altre parole, Google cerca solo le parole esattamente come digitate nel campo di ricerca. Se si inserisce [googl] o [googl*], il risultato non sarà "googler" o "googlin". In caso di dubbio, provare ad usare sia il singolare che il plurale; ad esempio "linea aerea" e "linee aeree".

Da notare che questo è valido per tutte le lingue tranne l'Inglese; lingua che invece supporta lo stemming.

L'operatore "*"

I caratteri jolly non funzionano come siamo abituati. Normalmente l'asterisco, e il punto di domanda, sostituiscono uno o più caratteri all'interno di una parola. Ad esempio [port*] troverà "porta", "porte", ma anche "portone" e "portacenere".

Google lo interpreta invece in un altro modo, utilizzando l'asterisco come fosse una parola intera. Per trovare quindi "a caval donato non si guarda in bocca", possiamo usare la query [a caval * * * * in bocca]. Come potete vedere si possono utilizzare più asterischi all'interno di una query.

L'operatore “~”

Mettere “~” prima di una parola dice a Google di cercare anche sinonimi e le parole correlate con la parola che precede. Una ricerca per [cervino ~mappa] troverà anche pagine che includono le parole “passeggiata”, “sentiero” e “bussola”.

La ricerca dei sinonimi troverà alle volte anche i plurali di una parola e altri varianti grammaticali. La ricerca di prima troverà quindi anche “mappe” e “cartine”.

Attenzione: in realtà questo operatore funziona solo su parole contenute nel dizionario Inglese. Gli esempi qui riportati sono stati presentati solo per chiarezza.

Ricerca per categorie

La directory Google è uno strumento molto utile per effettuare ricerche limitate su un argomento specifico. Ad esempio, per visualizzare solo risultati relativi al pianeta Saturno, è sufficiente selezionare “Saturno” nella categoria Scienze > Astronomia della directory Google. In questo modo si evita che vengano visualizzati risultati relativi alla macchina Saturn, al gioco omonimo o altri risultati che possano essere correlati al termine Saturno.

Le ricerche all'interno di categorie specifiche permettono di restringere il campo di ricerca e di visualizzare rapidamente solo le pagine desiderate.

<http://directory.google.com/>

I punti deboli di Google

Nonostante la grande rilevanza dei risultati di Google, anche questo motore ha i suoi limiti. Oltre a quelli già indicati in precedenza aggiungiamo che:

Google indicizza solo i primi 101KB di HTML di una pagina e circa i primi 120KB di un documento PDF.

Google ha problemi a cercare caratteri speciali come: &, %, #, @, \$, £, €, +, -, /, *, =, <, {, [(),], }, >, :, ;, ! e ?

Solitamente ignora questi caratteri, ma sta cominciando a introdurre il supporto per alcuni come:

I/O.

Le note in formato anglo-sassone: A_, A#, B_, B#, C_, C#, D_, D#, E_, E#, F_, F#, G_, G#.

Si possono usare i caratteri speciali &, + e _ una o più volte nel mezzo o alla fine di una parola. Ad esempio: A+, a_, C++, net__, page_count, i++++, a&b&c, i&&, “Johnson +& Johnson”. Non si può però usare il “+” di fronte ad una parola o numero. Ad esempio +i, “++numero”.

Per le ultime sul modus operandi di Google, non c'è di meglio che la fonte stessa:
<http://www.google.it/intl/it/help/basics.html>

2. Tecniche avanzate

Le tecniche viste fino ad ora sono sufficienti per la maggior parte degli utenti. La potenza di Google, però, non si limita a queste, ma permette di fare delle ricerche molto più mirate e molto più specifiche.

Far comprendere la potenza degli strumenti che indicherò qui non è facile.

Utilizzerò, come traccia per gli esempi, la sicurezza. Questo è un tema solitamente difficile da spiegare, ma cercherò di far capire quanto sia facile ottenere informazioni su Internet con un “servizio pubblico”, semplice e alla portata di tutti, come Google. Spero gli amministratori di sistema e gli addetti alla sicurezza utilizzino queste informazioni per proteggere meglio i nostri dati. Dati che consapevolmente o inconsapevolmente lasciamo a ogni nostro passaggio.

Nascondere queste tecniche con la speranza che questo ci proteggerà è particolarmente ingenuo: le informazioni qui citate sono facilmente reperibili su Internet con - toh - proprio Google.

intitle:

Questo operatore dice a Google di restringere la ricerca a pagine che contengono i termini indicati nel tag <title> di una pagina - cioè quelle parole che si leggono nella barra blu in cima al browser. La query [intitle: login password] dice a Google di cercare pagine che contengano la parola login nel titolo e la parola password nel resto della pagina.

Se vogliamo cercare più parole all'interno del titolo della pagina possiamo usare l'operatore [allintitle:]. Ad esempio usando [intitle: login intitle: password] equivale a [allintitle: login password].

Per quanto sembri banale, molti webmaster non si premurano di camuffare neanche il titolo delle pagine che possono essere attaccate da un cracker o possono contenere informazioni sensibili. Cercare la parola “password” in questo modo dà quindi sorprendenti risultati.

inurl:

Operatore che specifica che i termini devono essere contenuti nell'URL del sito. Ad esempio [inurl: passwd] troverà solo pagine che contengono la parola “passwd” nell'URL.

Come nell'esempio precedente, la ricerca di più parole all'interno dell'URL avviene tramite l'operatore [allinurl:]. Ad esempio [allinurl: etc/passwd] cercherà URL che contengono “etc” e “passwd”. Lo slash (“/”) tra le parole è ignorato da Google.

filetype:

Google è in grado di fare ricerche non solo all'interno delle pagine web, ma anche di file. La query [filetype:xls site:www.xyz.it bilancio] cerca file con l'estensione .xls (estensione di Excel) all'interno del sito www.xyz.it e che contengono "bilancio" nella pagina o nel nome del file Excel.

Sareste sorpresi di sapere il quantitativo di bilanci e stipendi di grandi società che si possono trovare!

I formati leggibili da Google sono:

- Adobe Portable Document Format (pdf)
- Adobe PostScript (ps)
- Lotus 1-2-3 (wk1, wk2, wk3, wk4, wk5, wki, wks, wku)
- Lotus WordPro (lwp)
- MacWrite (mw)
- Microsoft Excel (xls)
- Microsoft PowerPoint (ppt)
- Microsoft Word (doc)
- Microsoft Works (wks, wps, wdb)
- Microsoft Write (wri)
- Rich Text Format (rtf)
- Shockwave Flash (swf)
- Text (ans, txt)

Provate con la seguente query: [filetype:mdb mdb]. Potreste essere sorpresi nel vedere quante persone linkano apertamente il proprio database Access. Google non sarà in grado di leggere all'interno del database, ma non ci vorrà molto a un malintenzionato per aprirlo.

Conclusione. Internet non è un luogo sicuro dove depositare i segreti aziendali: non lasciatevi trarre in inganno dalle password. Ricordatevi che per ogni protezione è disponibile su Internet un software per crackarne la password. Se non volete che sia trovato non è sufficiente nascondere un file: NON mettetelo su di una rete pubblica!

link:

Questo operatore indica a Google di cercare pagine che hanno un link verso un determinato sito. Ad esempio [link:www.pc-facile.com] cercherà tutte le pagine che hanno un link verso <http://www.pc-facile.com>.

Questo operatore non è però esaustivo, cioè non vi indicherà proprio tutte le pagine, ma solo un campione. I motivi per questo possono essere tanti e spaziano da una tecnologia non perfetta quanto sostiene Google, a un modo per impedire il reverse engineering dei loro algoritmi.

related:

L'operatore [related:www.pc-facile.com] cerca pagine con contenuto simile, in questo esempio, a pc-facile.com.

Vi può servire se volete trovare altri siti che Google ritiene appartengano alla stessa categoria.

cache:

Vi permette di vedere la copia di una determinata pagina che Google tiene nei suoi archivi. Questo è utile quando il sito o la pagina in questione non sono più disponibili. Molti siti sono stati tolti sul fatto quando, dopo aver negato di aver pubblicato una cosa, questa è stata trovata nella cache di Google.

Ma la cache è ancora più potente. Immaginiamo che le password di un sito siano all'interno di un file .mdb (ho detto immaginiamo). Con questa query [site:www.pc-facile.com filetype:mdb mdb] troveremmo tutti i file .mdb nel sito www.pc-facile.com (è inutile che proviate: non ce ne sono). Se adesso andassimo a prendere questo file lasceremmo un'impronta sul server (cioè una traccia nei loro log). Ma perché lasciare un'impronta se possiamo prendere il file .mdb direttamente dalla cache di Google?

intext:

Usando questa sintassi si indica a Google di ignorare tutto (link, URL, <title>...) e di cercare solo nel testo della pagina.

Ricerche numeriche

Google vi permette non solo di cercare un numero singolo, ma anche una serie di numeri. Cioè, specificando due valori, Google troverà tutti le pagine che contengono numeri tra due valori - [5..11] trova tutte le pagine che contengono numeri tra 5 e 11.

Google, però, non sapendo contare molto bene alla query [pippo 123456] troverà sia pagine contenenti "123456" che "123,456".

Google inoltre permette ricerche numeriche aperte:

[..100] troverà pagine contenenti numeri fino a "100"

[100..] troverà pagine contenenti numeri oltre il "100"

Un utilizzo di questo operatore è dato nelle ricerche dei prezzi degli articoli. La seguente query, ad

esempio, troverà tutte le pagine che parlano di “Lettori DVD” con un prezzo tra i 100 e i 200 Euro.
[Lettori DVD 100..200 Euro]

Quando costruite queste query cercate sempre di indicare l'unità di misura o un indicatore di cosa rappresenta il numero. Se il prezzo è espresso in dollari allora si può utilizzare la seguente query (preghiamo notare che questa sintassi non funziona con altre valute):
[Lettori DVD \$100..200]

Questo operatore vi mostra come sia stato facile per Google creare un servizio come Froogle (<http://www.froogle.com>) partendo dalla tecnologia già sviluppata.

Ma provate anche con questa query... sì, è esattamente ciò che pensate, ma non correte a usare questi numeri per comprarvi un lettore DVD!
Davvero, non fatelo: è un consiglio.
[Visa 4366000000000000..4366999999999999]

Lingua

Google vi permette di selezionare in quale lingua devono essere le pagine in cui deve cercare. È possibile impostare questo dalle proprie preferenze o nelle opzioni avanzate.
In alternativa potete appendere all'URL di ricerca “&lr=lang_it”.

Il potenziale di queste informazioni è altissimo. Una persona capace di sfruttarle ha in mano una vera miniera d'oro e l'unica cura è la prevenzione, cioè anticipare gli attacchi e non presentare un bersaglio da colpire.

Indicherò di seguito alcune query, che si basano sugli operatori citati qui sopra, nella speranza che gli amministratori di sistema che hanno in mano dei miei dati si premurino di verificare che i loro server siano configurati in modo adeguato.

Non è difficile difendersi, ma è necessario sapere contro cosa ci si deve difendere.

Sintassi [Index of]

Alcuni siti sono configurati in modo che il contenuto delle cartelle sia visibile agli utenti che ci navigano. Questo è un modo facile e veloce per recuperare informazioni su di un server.

[intitle: “Index of” admin]

[intitle: “Index of” passwd]

[intitle: "Index of" ".htaccess"]

[intitle: "Index of logs"]

[intitle: "Index of config"]

Sintassi [allinurl:winnt/system32/]

Mostra una lista di siti che danno accesso a cartelle riservate, come la "system32", via web. In alcuni casi è possibile avere accesso al cmd.exe e da qui aumentare i propri privilegi.

Altre sintassi pericolose

[inurl: admin filetype:txt]

[inurl: admin filetype:cfg]

[inurl: passwd filetype:txt]

[inurl: iisadmin]

[allintitle: index of root]

[allintitle: index of admin]

[allintitle: sensitive filetype:doc]

[allintitle: restricted filetype:mail]

[allinurl: scripts cart32.exe]

[allinurl: phpinfo.php]

Conclusioni

Installare tutte le patch più recenti dei software utilizzati sui server..

Non mettere informazioni, che non possono essere divulgate, su server pubblici a meno che questi server non siano stati protetti adeguatamente da persone esperte in materia di sicurezza.

Disabilitare l'accesso all'indice delle cartelle in modo che questo non sia visibile e che venga presentato al suo posto un documento di default come, ad esempio, "index.htm".

Se trovate link a documenti riservati su Google è possibile chiedere il link sia rimosso. Qui di seguito le istruzioni.

<http://www.google.it/remove.html>

3. Gli altri servizi di Google

Il successo di Google è stato costruito su pochi punti fermi:

L'abilità nell'indicizzare documenti e nel farvi accedere a questi velocemente (MOLTO velocemente). Per i più tecnici questi si traduce nella capacità di suddividere un documento nei suoi elementi di base, in modo che la ricerca sia effettuata tra questi indici invece che nella totalità del documento, e quindi si velocizzi il processo di ricerca.

L'abilità nell'ordinare per rilevanza i risultati delle ricerche. Cioè Google riesce, attraverso una serie di criteri, a determinare se una pagina è più o meno rilevante rispetto a una serie di keyword o di altri parametri impostati dall'utente.

Questi due punti non solo sono il "segreto" di Google, ma sono il cuore delle sue attività: è tutto ciò che fa. E lo fa bene.

Al di fuori della mura di Google sono in pochi a sapere quali siano gli algoritmi che usa, ma ho il sospetto che Google si sia evoluto parecchio dall'esperimento del 1998 all'Università di Stanford. Sono certo che il primo Google fosse molto più semplice e meno raffinato di quanto non lo sia adesso. Eppure, nonostante questo, già i primi tentativi di Brin e Page erano oltremodo migliori di tutta la concorrenza. In realtà penso che parte del suo successo sia dovuto dal fatto che Google sia stato il primo a introdurre una certa sofisticazione nel campo dei motori di ricerca.

Non facciamoci però trarre in inganno: Google è diventato un motore molto sofisticato e di questa sua abilità nel trovare relazioni tra documenti e keyword sta facendo la sua forza.

I servizi che Google offre, di cui una parte sono descritti qui di seguito, sono TUTTI dovuti alla sua abilità nel dare una rilevanza ai risultati della ricerca.

Pensateci: nonostante i nomi particolari non sono altro che ricerche.

Google Toolbar

Il primo servizio di cui scriverò non può essere considerato un servizio alla stregua degli altri offerti da Google. Si tratta della toolbar, ovvero di un piccolo software che si integra con il browser (al momento solo per Internet Explorer 5.5 e versioni più recenti) permettendo di effettuare una ricerca su Google senza dover visitare la home page di Google. Questa toolbar ha anche altre funzioni tra le quali può bloccare i pop-up non richiesti e evidenziare le parole ricercate nella pagina.

Oggi tutti i principali motori e i principali siti offrono ai propri utenti una toolbar che permette di interagire con il sito senza doverlo visitare, ma quando questa fu presentata la prima volta nessuno ne aveva mai sentito parlare: Google fu il primo a introdurre questo strumento.

Adesso, addirittura, alcuni browser (come Firefox) includono di default una toolbar simile a quella di Google tanto che è possibile fare ricerche senza dover mai visitare Google.

Fino a poco fa, però, la toolbar originale aveva un vantaggio che nessun'altra aveva: rivelava il PageRank di un sito. Cosa che nessuno era in grado di fare. Ma adesso anche le toolbar open source sono in grado di comunicare con i server di Google e di mostrare il PageRank del sito che si sta visitando.

Resta il fatto che questo strumento è comodissimo e che è stato Google a introdurlo.

<http://toolbar.google.com/intl/it/>

GMail - Google Mail

Qualche mese fa Google ha sorpreso tutti annunciando di voler offrire un servizio di e-mail web based. La sorpresa non fu tanto che sembrò che Google stesse allontanandosi dal suo core business (convinzione errata), ma che la casella di posta concessa era di un Gigabyte, cioè – circa 300 – volte più grande di qualunque altro servizio simile gratuito.

Una volta capito che non si trattava di un pesce d'Aprile (Google è famoso per aver annunciato di cercare personale per un ufficio sulla luna e che la sua tecnologia PageRank fosse basata sui piccioni – PigeonRank), molti altri sono corsi ad aumentare la dimensione delle loro offerte o ad annunciare servizi uguali a Google.

Ma il business model di Google si è rivelato innovativo quanto la casella da 1GB. Grazie all'ultra collaudata tecnologia di ricerca e all'esperienza acquisita con AdSense hanno creato un programma che mostra pubblicità rilevante al contenuto dell'e-mail che viene letta (e forse rilevante anche ad un eventuale filo conduttore che lega tutte le e-mail nella casella?).

In molti hanno sollevato dei dubbi sui rischi per la privacy con un sistema simile, ma Google ha promesso di non creare dei profili degli utenti (cioè è tutto calcolato in real time) e che tutto sia eseguito da algoritmi e le critiche si sono chetate.

Quindi voi leggete la posta e accanto ci vedrete della pubblicità, non male che considerando che con un Gigabyte non avrete bisogno di cancellare la posta per molto molto tempo. Unico neo è che Google, ovviamente, non vi dà accesso alla casella utilizzando client – anche se voci di corridoio dicono che la cosa stia cambiando –, altrimenti come farebbe a guadagnare?

Google sta inoltre creando molta aspettativa. Il servizio è in fase beta e loro hanno distribuito solo una manciata d'inviti per provarlo. A queste persone sono poi concessi altri inviti che loro possono distribuire. La diffusione è piuttosto lenta, ma è sicuramente rumorosa!

Nel frattempo si sono scatenati tanti programmatori creativi attorno a quest'iniziativa di Google. È già stato fatto di tutto, ma come accade con i prodotti migliori, c'è ancora tanto da fare e Google ha solo da guadagnarci a seguire queste iniziative.

C'è chi ha creato programmi per avvertire di un nuovo messaggio nella casella, oppure c'è chi ha trovato il modo di farsi inoltrare la posta su account che possono essere controllati via client di posta.

Ma il premio per la creatività va alle seguenti due applicazioni.

La prima è di Jonathan Hernandez, un programmatore messicano, che ha scritto un programma in PHP per gestire il suo blog utilizzando GMail come back-end. Ovvero, ogni volta che lui spedisce un'e-mail al suo account su GMail, viene aggiunta una entry al suo blog personale.

Un altro programmatore, Richard Jones, ha invece esteso il suo desktop trasformando l'account su GMail in una cartella remota dove archiviare i suoi file. Comodo no?

Etichette, non cartelle

Una delle prime cose che si nota con un account GMail è che non è possibile creare delle cartelle come siamo abituati, ma che Google ci incoraggia invece a etichettare la nostra posta e lasciarla dov'è. Questo ha il duplice effetto di facilitare la ricerca per Google e di offrirgli un'indicazione in più per capire quale pubblicità può interessarci.

La posta non viene spostata fisicamente da una cartella ad un'altra, ma gli si dà un'etichetta ("Amici", "Fidanzata", "Lavoro"...). Filtrando la posta per un'etichetta è però come se navigassimo in una cartella.

Quella cosa in più nell'e-mail

Google supporta un tipo di indirizzo abbastanza particolare. Mettiamo ad esempio di avere un indirizzo pcfacile@gmail.com. Google vi permette di ricevere e-mail anche se il destinatario è: pcfacile+newsletter@gmail.com o pcfacile+xyz@gmail.com.

Questo vi permette di avere un numero infinito di indirizzi virtuali che possono essere tutti convogliati sul vostro indirizzo principale. In questo modo, con dei semplici filtri, potete catalogare facilmente la posta. Oppure potete creare indirizzi ad-hoc da dare solo ai siti di cui non vi fidate tanto. Nel caso dovesse cominciare uno spam spropositato è facile aggiungere un filtro che blocca questo particolare indirizzo.

Purtroppo non ci vorrà molto agli spammer a togliere il + e tutto ciò che segue.

Allegati

Per tutti quelli che pensano di poter mandarsi per posta i film scaricati da Kazaa, ripensateci: Google

non permette allegati superiori ai 10MB!

Conversazioni

A differenza della normale concezione delle email in cui ogni messaggio è trattato individualmente, Google lega le varie risposte ad un'e-mail in "conversazioni" in modo che sia facile seguirne lo sviluppo. Selezionando quindi un solo messaggio si vedrà tutto lo storico dello scambio di messaggi al posto di doverne selezionare uno per uno come siamo abituati a fare adesso.

Google, come in molte occasioni, è stato il primo a introdurre questo modello, ma saranno in molti a copiare l'idea e a migliorarla. Speriamo che Google riesca a stare dietro alle migliori idee!

<http://www.gmail.com>

Google Desktop Search

Viene da chiedersi come mai una compagnia come Google abbia prodotto un software per fare qualcosa che una normalissima copia di Windows dovrebbe già fare. Quando ci si rammenta che Microsoft è, di fatto, un monopolio e che è la competizione a spronare l'innovazione, allora ci si rende conto che la risposta è ovvia.

Il Google Desktop Search è l'estensione di Google applicata alla ricerca sul proprio computer. I file vengono letti, indicizzati e categorizzati. Quando si effettua una ricerca, usando il GDS, i risultati saranno velocissimi e rilevanti, proprio come ci ha abituato Google.

L'idea è ottima, ma il software è ancora in fase beta per cui, benché il servizio di base funzioni benissimo, questo non è adeguato ad un uso intenso e complesso come avrei avuto in mente io. Mancano troppe opzioni prima che io possa decidere di utilizzarlo: fermo restando che mi dà fastidio usare un'applicazione esterna per fare qualcosa che Windows dovrebbe già fare!

L'eseguibile è piccolissimo, si scarica velocemente e si installa con facilità. Facendolo partire bisognerà dargli qualche ora (in cui non facciamo uso del computer) per indicizzare i file presenti sul PC – operazione che può richiedere circa 1 GB di spazio su disco. Dopo questa prima fase Google indicizzerà soltanto i nuovi documenti e quelli modificati.

Al momento Google indicizza i seguenti documenti:

- E-mail di Outlook e Outlook Express
- File Microsoft Word, Excel e PowerPoint
- Chat log di AOL Instant Messenger
- Pagine web viste con Internet Explorer
- File HTML
- File TXT

- Nomi dei file GIF e JPEG

Di ogni file aperto viene creata una copia cache del tutto simile alla copia cache di Google quando effettuiamo ricerche su Internet. Il vantaggio qui, però, è che viene creata una copia cache ogni volta che si modifica un file. In questo modo è possibile ricostruire uno storico dei propri documenti anche se li abbiamo già modificati irreparabilmente.

Inoltre, l'anteprima della cache non richiede l'apertura del software che ha creato il file – se abbiamo una copia cache di un'e-mail, non avremo bisogno di far partire il monolitico Outlook per leggere la posta.

Come ho detto, però, il GDS è ancora in fase beta e soffre di alcune lacune piuttosto importanti:

È possibile installare il Google Desktop Search solo sul disco C: (per cui l'indice, che può richiedere fino a 1GB, deve risiedere sul disco C:)

Indicizza solo i contenuti del disco C:

Mancanza di una pagina per le opzioni avanzate

Solo 10 risultati per pagina, non è possibile impostarne di più

Manca l'opzione per aggiungere nuovi tipi di file. Servirebbe un'API, come quelle solitamente rilasciate, per permettere di creare nuovi formati da ricercare. Sono sicuro che la maggior parte dei nostri file siano in formati non contemplati al momento.

L'indice non è criptato. Chiunque può accedervi senza fatica: servirebbe una password

Non è predisposto ad un ambiente multi-user. In una situazione di computer condivisi tutti potrebbero vedere tutti i file di tutti

Qui di seguito alcune query che funzionano per aggirare alcune limitazioni. Questi comandi non sono documentati, sono il frutto di esperimenti personali e potrebbero cambiare senza preavviso. I seguenti operatori vi aiutano a effettuare ricerche soltanto tra certi tipi di file:

- Word: [filetype:word] o [filetype:doc]
- Excel: [filetype:excel] o [filetype:xls]
- PowerPoint: [filetype:powerpoint] o [filetype:ppt]
- Text: [filetype:text] o [filetype:txt]
- E-mail: [filetype:email]
- Chat: [filetype:chat]
- File HTML: [filetype:web] o [filetype:html]
- Immagini: [filetype:jpg] o [filetype:gif]
- Acrobat: [filetype:pdf]
- Windows Media: [filetype:wma] o [filetype:wmv]
- MP3: [filetype:mp3]

È inoltre possibile specificare alcuni parametri quando si fanno ricerche nella posta:

- [to:marco]
- [from:maria]

- [folder:lavoro]
- [phone:123]
- [category:ufficio]

Non fraintendete le mie parole, il Google Desktop Search è uno strumento concettualmente validissimo, anche se continuo a essere infastidito dal fatto che non sia stato proposto dalla Microsoft. Non penso sia però ancora un prodotto pronto all'uso.
<http://desktop.google.com>

Google Images

L'utilità di ricerca di immagini di Google è tra le più complete del Web poiché permette di ricercare e visualizzare oltre 880 milioni di immagini.

Google analizza il nome dell'immagine (il nome del file) e il testo contenuto nella pagina adiacente all'immagine stessa e ne determina la rilevanza rispetto alla ricerca.

Voci di corridoio ci informano che stanno lavorando a tecnologie in grado di “vedere” e riconoscere un'immagine in modo accurato, ma non tratterrei il fiato nell'attesa. Al momento questo genere di tecnologia è la migliore tra quelle disponibili sul mercato.

Le immagini così individuate possono essere protette da copyright. Pertanto, l'utente è autorizzato ad utilizzare il servizio di ricerca al solo di scopo di visualizzare le immagini sul Web. Per poter utilizzare le immagini trovate il servizio di Google, l'utente, dove necessario, dovrà chiedere l'autorizzazione al proprietario dei diritti d'autore

<http://images.google.it/>

Le News

Forse uno dei servizi in cui più si intuisce la potenza della ricerca di Google è nel suo servizio di news la cui beta fu lanciata già a Settembre del 2002. Una serie di algoritmi selezionano notizie tra oltre 4.500 fonti accreditate. I siti vengono visitati, le nuove news vengono individuate e indicizzate. Altri algoritmi le suddividono e decidono quali fonti sono più rilevanti e quali news più recenti e più importanti. Il tutto viene poi presentato in un formato comodo e sempre aggiornato.

Ancora più interessante è la tecnologia di raggruppamento (di cui ne parleremo nel prossimo capitolo dedicato ai Google Labs - i laboratori di Google) che raggruppa le notizie a seconda della categoria a cui appartengono e che raggruppa diverse notizie sullo stesso argomento come riportate dalle diverse testate.

<http://news.google.it/>

Google groups e Usenet

Nel 1979, due studenti della Duke University nel North Carolina, Tom Truscott e Jim Ellis, ebbero l'idea di usare un sistema, allora chiamato UUCP (Unix-to-Unix CoPy), per distribuire informazioni tra gli utenti di UNIX.

All'inizio il sistema era molto spartano e consisteva di un server che raccoglieva i messaggi e di un software che permetteva agli utenti di leggere e scrivere questi messaggi. In altre parole crearono il precursore dei moderni forum in un tempo in cui Internet non esisteva ancora.

L'idea ebbe un gran successo e ben presto nacquero newsgroup di tutti i generi, riguardanti tutti gli argomenti possibili e in tutto il mondo. Dal 1981 Usenet raccoglie sui suoi server tutti questi newsgroup e tutti questi messaggi. Google ha acquistato questo archivio contenente oltre 845 milioni di post e 42.000 newsgroup ufficiali e li ha resi ricercabili attraverso Internet. Se poi avete un account su Google potete anche partecipare alle discussioni.

Ci sono due cose da tenere bene a mente.

Il primo è che non si può accedere ai gruppi con un newsreader, ma si deve per forza passare via web.

Il secondo è che non è possibile inviare messaggi anonimamente. Questo non sarebbe un gran danno se non fosse che i newsgroup sono una delle principali fonti di indirizzi e-mail per gli spammer.

<http://groups.google.it/>

Froogle - lo shopping comparato

Froogle, lanciato in modalità di test a Dicembre del 2002, è un servizio che dà la possibilità agli utenti di compiere ricerche sui siti di e-commerce per prodotto ottenendo e paragonando prezzi, immagini e condizioni tra i milioni di siti che recano informazioni rilevanti.

Sono un convinto utilizzatore di questo servizio anche se ammetto che funziona meglio negli Stati Uniti, dove l'e-commerce è più sviluppato e i siti si sono adattati meglio.

Sono rimasto sorpreso nello scoprire quanti siti di e-commerce esistono e quanto divergono i prezzi tra questi siti. Conoscerli tutti, sapere quali sono specializzati in quale settore e quali sono i migliori è un'impresa impossibile. Un servizio che mi permette di paragonare al volo i prezzi di tutti i siti che offrono un determinato prodotto è per me preziosissimo.

Molto comoda è anche la possibilità di cercare tra siti che hanno un prezzo all'interno di una certa

fascia. Questo serve a scremare le offerte troppo care o che appaiano, tra i risultati, anche gli accessori del prodotto che cerchiamo. Facendo ricerche, ad esempio, per un telefono vengono visualizzate anche le custodie, gli auricolari e altri accessori per lo stesso.

Un piccolo trucco quando decidete di acquistare via Internet.

Una volta individuato il sito da cui volete effettuare l'acquisto aprite un altro browser e (immaginando che volete acquistare da Amazon) digitate query come [amazon coupon] e [amazon discount]. Molti siti esterni ad Amazon, offrono dei codici (buoni) sconto nella speranza di farsi un po' di pubblicità. Perché non approfittarne?

Se stiamo invece acquistando da un sito Italiano possiamo provare con query come [nome_sito sconto], [nome_sito buono sconto] e [nome_sito buono].

Con le poche, semplici, tecniche indicate qui, è possibile risparmiare attorno al 40% del prezzo che si può pagare per lo stesso prodotto in un comune negozio. Acquistare su Internet ha però senso solo quando il prodotto è standardizzato – è, cioè, un prodotto di massa per cui il rivenditore non dà un valore aggiunto, ad esempio i CD, i DVD, i libri e le scarpe (di cui conosciamo esattamente il modello, il numero, il colore...) – e quando le spese di spedizione non incidono che in minima parte sul prezzo totale dell'acquisto. Quest'ultimo punto è particolarmente importante per cui è importante che effettuiamo una spesa grossa, come nel caso di un orologio, oppure tante spese piccole, se invece decidiamo di acquistare tanti libri o CD accorpendo, il più possibile, le spese di spedizione.

<http://froogle.google.com/>

Traduzioni on-line

Non sempre i risultati delle nostre ricerche conducono a pagine scritte in una lingua che capiamo. Per venirci in contro Google offre la possibilità di tradurre intere pagine in varie lingue. Attenzione però che la traduzione è effettuata da un computer (e di conseguenza la traduzione è letterale) e quindi è utile soltanto a darci una vaga idea di cosa tratta una pagina.

È utile anche che Google permetta di tradurre pagine che non abbiamo visitato o anche dei semplici brani di testo.

http://www.google.it/language_tools/

Google Answers

Google è un ottimo strumento per trovare informazioni on-line, ma anche i ricercatori esperti possono trovarsi in difficoltà quando le informazioni sono di difficile reperibilità.

Google Answers mette a disposizione degli utenti un gruppo di 500 esperti nella ricerca che cercheranno di fare ciò che voi non siete riusciti a fare: cioè a trovare le informazioni che vi servono.

Per fare ciò avete bisogno di una Carta di Credito e di creare un Account su Google. Fatto questo vi verranno dedotti \$0,50, come costo fisso, più una cifra che va da \$2 a \$200 a seconda di quanto pensate sia difficile trovare le informazioni che vi servono – più difficile la reperibilità più alto sarà il prezzo.

Tutto qui?

Non proprio. Una volta inserita la domanda e assegnato un valore monetario al lavoro i 500 ricercatori daranno un'occhiata alla vostra richiesta e se qualcuno pensa che il compenso sia adeguato chiuderà al domanda e si metterà al lavoro. Dopo qualche ora vi fornirà una risposta. È ovvio che più alto sarà il compenso e più i ricercatori si lanceranno alla ricerca delle informazioni.

È importante capire che i ricercatori non sono esperti della materia di cui voi chiedete, ma sono esperti nella ricerca di informazioni di difficile reperibilità.

E se la risposta non è quella giusta?

Come prima cosa potete chiedere una chiarificazione. Se poi anche la seconda risposta non vi soddisfa potete chiedere di essere rimborsati e ciò avverrà – tranne per i \$0,50 iniziali.

Un ottimo servizio davvero prezioso per chi ha bisogno di informazioni.

<http://answers.google.com/answers/>

Google Alerts

È un modo comodo per essere aggiornati in tempo reale su nuovi documenti o nuove news indicizzati da Google.

Se volete seguire gli sviluppi di una notizia scegliete un “News Alert” e Google vi invierà un’e-mail ogni volta (oppure con scadenza programmata – a seconda di cos’avete impostato) che un documento raggiunge i primi dieci risultati nella ricerca.

Se invece volete seguire gli sviluppi di un vostro concorrente o della squadra di calcio del cuore allora scegliete un “Web Alert” specifico a una determinata query e Google vi invierà un’e-mail ogni volta che un documento raggiunge i primi venti risultati nella ricerca.

Potete avere quanti alert volete e potete cancellarli ogni volta che volete. Comodo, no?

Se poi avete un account su Google potete gestire comodamente tutti gli alert da un solo pannello di controllo.

<http://www.google.com/alerts/>

AdSense

AdSense è il programma, di Google, di web -agency per webmaster. In altre parole, aderendo a AdSense avrete la possibilità di essere pagati per mostrare pubblicità ai vostri utenti. Non è un sistema che promette ricchezza immediata, ma vi ripagherà adeguatamente dei vostri sforzi se avete lavorato bene.

Appena registrati vi verrà fornito un semplice codice HTML da inserire all'interno delle pagine e poi penserà a tutto Google. I webmaster potranno personalizzare i colori dei banner per meglio integrarsi nelle pagine e scegliere di non mostrare alcuni tipi annunci, e se mostrare gli annunci solo in formato testo o anche i classici banner con immagini; ma poco altro.

Il pagamento viene effettuato con un assegno in dollari Americani spedito solo a determinati indirizzi. Con questo voglio dire che non è ancora possibile farsi spedire i compensi, ad esempio, in l'Italia.

L'assegno viene spedito 30 giorni dopo la fine di un mese in cui si sono totalizzati almeno \$100 di revenue. Se non si raggiunge questa quota minima, il parziale verrà riportato al mese seguente e così via fino ad arrivare a \$100.

Essendo un programma passivo, i consigli che si possono dare sono pochi, nonostante ciò alcuni sono molto importanti:

Personalizzate i colori del banner, o del formato da voi scelto, in modo che sembri parte integrante del sito.

Se possibile non usate il classico banner 468x60, in quanto siamo diventati ciechi a questo tipo di formato. Il migliore è, al momento, lo **skyscraper**, Google ci ha abituati a leggere la pubblicità in verticale.

Sperimentate con i vari formati pubblicitari. Google vi dà la possibilità di monitorare vari **canali** pubblicitari: impostate un canale per ogni formato pubblicitario che vi sembra accettabile e agite a seconda dei risultati; scartate i formati che non portano introiti e tenete quelli che sono più favorevoli.

Mettete la pubblicità in un punto che si vede **appena aperta la pagina** in fondo a una lunga pagina non serve a molto!

AdSense è la controparte di AdWords e paga per ogni click che il banner sul vostro sito riceve. Bisogna, però, ricordare che **i banner non sono tutti remunerativi allo stesso modo**, ma ogni click vale qualcosa in più o in meno a seconda dell'offerta massima dell'inserzionista.

Non create pagine che traggono Google in **inganno** inserendo, ad esempio, parole come valium, texas holdem o poker nel titolo mentre il resto della pagina tratta di giardinaggio. Il segreto di AdSense, se ce n'è uno, è di avere delle pagine molto molto specifiche dove l'utente trova le informazioni che stava cercando e dove trova anche della pubblicità che davvero gli serve.

Se avete il tempo, fate ricerche su quali sono le **keyword più remunerative** – ad esempio potete aprire

un account su AdWords e vedere cosa Google suggerisce di offrire per le varie keyword – e cercate di includere queste nel contenuto delle vostre pagine. Attenzione, però, che queste keyword siano attinenti al resto del contenuto!

Un ultimo consiglio. Internet pullula di offerte per libri che “*per la modica cifra di \$xxx*”, promettono di svelarvi i segreti di AdSense e di farvi guadagnare migliaia di dollari. Per favore non cascateci: non ci sono segreti!

Google Local

Questo servizio esiste solo negli Stati Uniti al momento ma dato l'interesse di questo servizio ho deciso di parlarne lo stesso.

In questo caso Google cerca di darvi un risultato che abbia senso geograficamente.

Immaginate di fare una ricerca per [pizzeria trancio].

Il numero di risultati sarà altissimo. Anche aggiungendo la parola “milano” il numero non calerà di molto. Anzi, vi accorgete che i risultati includeranno molte pagine in cui sono presenti una lista di ristoranti di cui uno solo, magari, è a Milano. Risultati come questi sono solitamente poco utili e vi dicono poco della qualità del ristorante.

Immaginate ora che Google sia in grado di paragonare i risultati ottenuti con al query [pizzeria trancio milano] e le pagine gialle! In questo caso sarebbe capace di capire se un risultato contiene informazioni geograficamente compatibili con la nostra query. Poi ordinerebbe i risultati trovati per rilevanza - magari se una pizzeria ha tanti link vuol dire che è buona?

Google Local non esiste ancora in Italia, ma, se nel frattempo se volete sapere qual'è la miglior pizzeria da asporto a Milano, vi posso dire che è alla fermata della metropolitana di Inganni; proprio all'angolo.
<http://local.google.com/>

Google Catalogs

I cataloghi di Google sono un interessante esperimento di OCR (Optical Character Recognition) e di portare on-line informazioni che sono tradizionalmente off-line.

Con questo servizio, Google ha scannerizzato più di un migliaio di cataloghi di prodotti per la vendita per corrispondenza dando accesso ai loro contenuti nel tipico formato ricercabile che ha reso popolare Google.

Si tratta di un altro servizio dedicato allo shopping e, come in Froogle, Google si limita soltanto a

fornire le informazioni.

Pensate se Google ottenesse il permesso di scannerizzare la letteratura mondiale, come vorrebbe fare Amazon.com!

<http://catalogs.google.com>

Account

Un account su Google è il punto di partenza se volete cominciare a utilizzare a pieno il potenziale di questo sito. Come nella tradizione di Google, la vostra privacy sarà garantita e i vostri dati protetti.

Al momento un account vi serve per:

utilizzare Google Answers

scaricare le Google Web APIs

scrivere messaggi su Google Groups

gestire comodamente i Google Alerts

In futuro questa list si estenderà e includerà servizi come:

Google AdWords

Store

<https://www.google.it/accounts/>

Google Sets

In realtà, questo servizio è ancora in fase beta, cioè è ancora nei laboratori di Google, ma i risultati mi hanno stupito a tal punto che voglio mostrarvelo lo stesso.

Si tratta quasi di un esempio di intelligenza artificiale. Dati alcuni elementi di una categoria Google è in grado di trovarne altri appartenenti alla stessa categoria.

Inserite ad esempio: Milan, Inter, Juve, Roma e Fiorentina e Google vi troverà altre squadre di serie A.

Inserite Italia, Francia, Germania, Spagna e Olanda e Google vi indicherà altri paesi.

Inserite Indiana Jones, Guerre Stellari, Frantic, Blade Runner e Apocalypse Now e Google vi troverà altri film di Harrison Ford.

Utilissimo anche per trovare altre keyword quando non sapete più cosa inventarvi.

<http://labs.google.com/sets/>

Google Labs

I laboratori di Google sono quel luogo dove gli ingegneri di Google ci mostrano i loro progetti ancora in fase beta. Si tratta di progetti non ancora perfezionati, che richiedono feedback da parte degli utenti per essere finalizzati, ma che – d'altra parte – rischiano di non essere mai completati.

Si tratta di un luogo interessante per sapere cosa passa per la mente collettiva di Google, anche se i progetti davvero interessanti non vengono mostrati fino alla fine per non essere compromessi.

Servizi come, ad esempio, Google Images sono passati per la fase beta e si sono laureati a pieni voti.

Al momento di scrivere sono in fase beta GMail, Google Groups II e Google Sets, tra gli altri.

Ma facciamo un passo indietro, torniamo al Dicembre del 2001 quando Krishna Bharat, come uno dei 10 addetti alla ricerca nei laboratori di Google, invia un'e-mail interna ai suoi colleghi invitandoli a provare un servizio a cui sta lavorando. È solo un'idea in fase embrionale e ci sta lavorando su da un mesetto, ma questo, del resto, è il suo lavoro. Si tratta di un motore che visita una quarantina di siti di news ogni ora, ne prende le notizie, le categorizza a seconda dei contenuti e li presenta in una pagina facilmente navigabile. Un po' come una rassegna stampa elettronica.

Vi dice niente questa descrizione? Perché all'interno del Googleplex l'idea fu molto apprezzata. Più importante è che attirò l'attenzione di Marissa Mayer, un giovane ingegnere diventato project manager. Mayer assegnò a Bharat un piccolo team di ingegneri che, nel giro di un mese e mezzo, trasformarono la demo testuale in un servizio che visitava 155 fonti continuamente, non ogni ora. Appena pronto, questo servizio – che ormai avrete capito essere le Google News –, fu messo nei laboratori di Google.

Perché? Perché presentare le idee al pubblico li aiuta a identificare facilmente e velocemente le idee che non funzionano. I fallimenti vanno bene. I buoni fallimenti vanno ancora meglio. I fallimenti buoni hanno due caratteristiche. La prima è che sappiamo perché abbiamo fallito e sappiamo come migliorare la seconda volta. Ad esempio, quando Google sperimentò con degli screenshot dei siti nei risultati, vide subito che i tempi di caricamento aumentavano drammaticamente e decisero subito di abbandonare l'idea.

La seconda ragione è che identifica rapidamente i progetti vincenti.

Entro due settimane dalla sua introduzione, le Google News avevano già un pubblico di 70.000 utenti e queste 70.000 persone erano munizioni per spingere lo sviluppo del servizio. “La prova pubblica aiuta a muoversi velocemente,” dice Mayer. “Se funziona, aiuta a creare passione e interesse e tutti noi cominciamo a pensare a come migliorare il servizio.”

E cosa se ne farà Google? Non importa, se l'idea è buona un modo per trarne profitto salterà fuori. Questa è una società di ingegneri che costruiscono quello che loro stessi vorrebbero utilizzare.

Andate a vedere i tipi di persone che cercano per lavorare da loro: voi non lavorereste in una società che cerca esperti in algoritmi genetici, data mining, robotica, intelligenza artificiale e file system design? (Non ci capisco molto di queste cose, ma immaginate di lavorare in una società in cui si sentono queste

parole tutti i giorni – sarei davvero orgoglioso del mio lavoro!)
<http://labs.google.com/>

Google Calculator

La calcolatrice di Google è un interessante esperimento

Non esiste un manuale di riferimento con tutte le operazioni che si possono eseguire, ma sappiamo che le seguenti sono valide. Preghiamo notare che i nomi e le convenzioni sono anglo-sassoni: ad esempio “sqrt” indica la radice quadrata e i decimali sono indicati con il punto, non con la virgola.

Costanti: Avogadro constant, electron mass, googol, molar gas constant, pi, Planck’s constant, Stefan-Boltzmann constant.

Calcoli

- Aritmetica: $(1 - 2^{(3 / 4)}) / (4 + 5) * 6 + (3^2^3)$
- Trigonometria: $\sin(30 \text{ degrees}) + \arctan(2 \text{ radians})$
- Funzioni trascendentali: $e^7 + \ln(1000) + \cosh(3.6)$
- Meccanica: $(1 \text{ kg}) * (9.8 \text{ m/s}^2)$
- Termodinamica: $((1 \text{ mol}) * (8.315 \text{ J/mol} * \text{K}) * (280 \text{ K})) / (1.013 * 10^5 \text{ Pa})$
- Fisica: $\sqrt{(1 / 93 * 10^{11} \text{ Pa}) / (1.26 * 10^3 \text{ kg/m}^3)}$
- Elettromagnetismo: $(4 \text{ farad}) * (1 * 10^{-3} \text{ m}) / (7.43 * 10^{-12} \text{ farads/m})$
- Ottica: $((0.09 \text{ m}) (600 * 10^{-9} \text{ m})) / (2 * 0.021 * 10^{-3} \text{ m})$
- Relatività: $(1.98 * 10^{-6} \text{ s}) / \sqrt{1 - (0.995)^2}$
- Meccanica quantistica: $(4.894 * 10^{-15} \text{ eV} * \text{s}) (3 * 10^8 \text{ m/s}) / 0.0114 \text{ eV}$
- Fisica nucleare: $25 * (1.007825 \text{ u}) + 35 * (1.008665 \text{ u}) - ((548.5 \text{ MeV}) / (931.5 \text{ MeV/u}))$

Conversioni

- Conversioni numeriche: 0b1010 in decimal
- Angoli: 2 radians in degrees
- Unità di misura: inches in feet
- Temperature: kelvin in celsius
- [2 meters + 5 feet]
- [three quarters of a cup in teaspoons]
- [98.6 degrees Fahrenheit in degrees Celsius]
- [130 lbs in kg]
- [130 lbs in stones]
- [65 mph in kph] o [65 mph in km/h]
- [160 pounds * 4000 feet in Calories]

- [1500 in hex] o [1500 in hexadecimal]
- [64 in binary]
- [LVII in decimal]
- [1 a.u./c]
- [56*78]
- [1.21 GW / 88 mph]
- [e^(i pi)+1]
- [100 miles in kilometers]
- [sine(30 degrees)]
- [G*(6e24 kg)/(4000 miles)^2]
- [0x7d3 in roman numerals]
- [0b1100101*0b1001]
- <http://www.google.com/help/calculator.html>

Google Games

Mentre decine di milioni di utenti amano Google c'è una sconcertante minoranza che ne è ossessionata. Aggiungete a questo che Google distribuisce le sue API e capirete perché questi giochi che seguono non mi sorprendono più di tanto. Certo non si tratta dei servizi che offre Google, ma si può dare un'occhiata alla reale potenza di Google.

Googlewhack

Trovate due parole che, quando combinate in una query, danno UN SOLO risultato. A oggi Googlewhack sostiene di aver trovato oltre 400.000 combinazioni. Date un'occhiata al sito per vedere quali sono le ultime:

<http://www.googlewhack.com>

Googlebomb

Un esempio di terrorismo informatico. Quando i tecnofili non sopportano, ma proprio non sopportano, una persona si infilano tra le pieghe del codice di Google e attaccano come solo loro sanno fare. Bombardano Google con una serie di pagine inesistenti ricche di keyword e fanno puntare queste pagine al sito che vogliono colpire. Il risultato? Tempo fa inserendo keyword come “buffone” o “miserabile fallimento” si veniva portati ad una biografia di Silvio Berlusconi.

A Maggio del 2004 fu indetta una competizione per vedere chi riusciva a portare una pagina Internet al primo posto dei risultati per la query priva di senso “nigritude ultramarine”. Il vincitore di questa singolare competizione si portò a casa un iPod, ma molti temettero che le tecniche usate in questa competizione potessero essere usate da società senza scrupoli per scalare le classifiche. Ci fu un sospiro collettivo quando Google annunciò che aveva seguito la competizione con attenzione e aveva imparato le tecniche e modificato i suoi algoritmi.

Googlism

Digitate un nome, una data, un luogo o una qualunque cosa e Googlism vi dirà cosa ne pensa il web - non sempre è piacevole sapere certe cose di se stessi. Ad esempio Bill Gates “l’anti-Cristo,” “un ladro” e “un eroe”.

<http://www.googlism.com>

Google Smackdown

Chi è più popolare su Google? Inserite due query e vedete chi vanta più nomine: guerra o pace, odio o amore, vita o morte?

<http://www.onfocus.com/googlesmack/down.asp>

Quelli che sono stati presentati qui sono solo alcuni dei servizi che offre Google, ma sono quelli più interessanti ad un pubblico italiano. I servizi di Google sono in continua evoluzione; per essere sempre aggiornati seguite il link qui sotto.

<http://www.google.it/options/>

4. Search Engine Optimisation

Questa sezione non è propria di questo volume in quanto le informazioni qui di seguito non servono tanto a chi effettua una ricerca, ma a chi vuole fare in modo che il proprio lavoro venga trovato. Detto questo, sapere come e cosa Google cerca all'interno del suo indice, può aiutare un ricercatore esperto a formulare meglio le query.

Le informazioni contenute in questa sezione sono soggette a cambiamenti repentini senza preavviso. Prima di mettere in pratica questi consigli è bene cercare di verificare quali siano le attuali voci di corridoio su come Google crei il suo indice. In ogni caso, il confine tra il Search Engine Optimization (SEO) e cercare di barare, è molto fine. Visto che la pena è l'esclusione dagli indici di Google, il mio primo consiglio è di cercare di stare sempre dalla parte della ragione.

Una considerazione da fare quando si parla di SEO è che questo, molte volte, contrasta con i nostri sforzi di web -usability. Applicare quindi questi consigli di SEO con il dovuto buon senso.

Le informazioni contenute in questa sezione sono tutte frutto dell'esperienza personale. L'utilizzo di queste tecniche ha portato un quadruplicarsi del traffico di pc-facile.com dopo un solo mese dall'applicazione. Non sono in nessun ordine particolare in quanto si presuppone che chi affronta il problema del SEO abbia già tentato alcune delle tecniche qui elencate.

Alcune delle soluzioni proposte in questa sezione richiedono una conoscenza abbastanza approfondita di argomenti come HTML e web -server; si consiglia di fare dei back-up prima di apportare qualunque modifica.

Il tag <title>

Le informazioni contenute nel tag <title></title> sono molto importanti. Cercate di indicare chiaramente il contenuto della pagina con un titolo appropriato che contenga al suo interno le keyword importanti. Va da sé che i titoli delle pagine dovrebbero essere diversi per ogni pagina anche quando usate contenuti dinamici.

Se offrite una sezione per il download del software, avere un titolo uguale per tutte le schede software non vi aiuterebbe molto. Potrebbe essere più utile invece indicare il titolo del software insieme a qualche keyword.

Su pc-facile.com abbiamo optato per dei titoli molto descrittivi come indicato qui sotto:

DivX [Download software Codec Microsoft]
--

Dove “DivX” è il nome di un codec (un particolare tipo di software), “Download software” dice all’utente cosa si può fare e allo stesso tempo sono due keyword molto ricercate su Internet, “Codec” è la categoria sotto cui è classificato DivX (ed è un’altra keyword importante) e Microsoft è la piattaforma su cui gira DivX (ed è una keyword).

Seguendo la medesima logica è importante che i titoli del contenuto che scrivete siano descrittivi, magari anche a discapito del sensazionalismo che potrebbero provocare. Un titolo come “Importante notizia” potrebbe attrarre molti utenti (ok, in realtà no, ma è l’idea), ma a Google servirebbe a poco. Meglio qualcosa di più specifico e più ricco di keyword utili a Google.

L’HTML dei titoli

Il titolo di una pagina, di una sezione o di un paragrafo (inteso propriamente come titolo non come il tag <title>) viene solitamente indicato usando una combinazione di tag , , <i> e <u>. Questo va benissimo - come avremo modo di vedere più avanti -, ma non è la soluzione ottimale.

L’HTML prevede sei tag specifici per indicare un titolo (heading) e questi sono: <h1>, <h2>, <h3>, <h4>, <h5>, <h6> dove <h1> è il titolo più importante e l’importanza diminuisce nei seguenti tag.

Non posso sottolineare abbastanza l’importanza di usare questi tag, in particolar modo il tag <h1>. Fatelo.

Se i tag <h1...6> non soddisfano il vostro gusto estetico potete modificarne tutti gli aspetti facilmente usando i CSS.

Inserite il seguente codice nell’header (adattando font, dimensione... alle vostre esigenze). Includo qui anche come modificare un link all’interno dei tag <h1...6>, in quanto è stato un problema che mi ha portato via molto tempo.

```
<style type="text/css">
<!--
h1
{
    font-family:Verdana, Arial, Helvetica, sans-serif;
    font-size: 12pt;
    color: #000000;
    text-decoration: none;
    margin-bottom: 0;
}

h1 a
{
```

```
font-family: Verdana, Arial, Helvetica, sans-serif;
font-size: 12pt;
color: #000000;
text-decoration: none;
margin-bottom: 0;
}
-->
</style>
```

Fatto questo usate il tag <h1...6> esattamente come fareste per con gli altri tag.

<h1></h1>

Per ulteriori informazioni su come usare i CSS vi consiglio:

<http://www.w3schools.com>

[Nota: l'istruzione "margin-bottom: 0;" è tecnicamente inutile ma risolve un brutto problema estetico presente solo in alcune versioni di Internet Explorer.]

Il dominio

Il tipo e il nome del dominio sono fattori molto importanti.

Google dà molta importanza ai domini di secondo livello e tende a considerare poco i domini di terzo livello. Se volete una presenza importante su Internet è vitale che il vostro dominio non sia del tipo <http://pc-facile.xoom.com>, ma sia invece <http://www.pc-facile.com>.

Altro fattore importante nella scelta del dominio è il nome. Mentre le regole di usability impongono che il nome sia breve, facilmente ricordabile e possibilmente fonetico, Google pare apprezzare i nomi di domini che contengono delle keyword come ad esempio:

<http://keyword1-keyword2-keyword3.com>

A voi il delicato equilibrio tra usability e la sempre crescente importanza dei motori di ricerca.

Nome della pagina

Il nome del dominio non è l'unico di cui Google tiene conto. I nomi delle pagine individuali sono molto importanti. Ad esempio la pagina [software.php?id=123](#) dice a Google solo che si tratta di software. Invece la pseudo-pagina [software_DivX_codec.php](#) dà molte informazioni sul contenuto della pagina.

Questo tipo di lavoro è facile quando un sito è piccolo, ma quando si lavora con siti dinamici che rappresentano le pagine usando dei template fissi la cosa diventa difficile da gestire.

In nostro aiuto vengono il mod_rewrite su Apache e alcuni filtri su IIS di Microsoft.

Per fare in modo che al posto di questa pseudo-pagina, che non esiste, venga mostrata quella giusta senza tradire il nostro segreto è necessario inserire il seguente codice nel file .haccess di Apache (attenzione che diverse versioni di Apache possono differire leggermente: leggere quindi prima la documentazione).

```
RewriteEngine on
RewriteBase /
RewriteRule ^software_DivX_codec/?$ /software.php?id=25 [L]
```

Se usate invece IIS avete molte soluzioni per ottenere questo stesso risultato. Forse la più semplice, ma la meno potente consiste nell'installare la libreria URL Replacer messa a disposizione gratuitamente dallo sviluppatore cecoslovacco Antonin Foller.

<http://www.motobit.com/help/url-replacer-rewriter/iis-mod-rewrite.asp>

In testa è meglio

Google considera importante il contenuto che si trova in cima ad una pagina, quindi una buona tecnica è posizionare il contenuto importante il più in alto possibile. Il problema e la soluzione sembrano banali: non lo sono.

Prendiamo il caso tipico in cui la pagina è divisa in due colonne; dove la colonna di sinistra rappresenta la barra di navigazione, mentre la colonna di destra rappresenta il contenuto.

Il codice HTML che costruirebbe questa pagina (solo la parte che ci interessa) sarebbe:

```
<table>
  <tr>
    <td>codice per barra di navigazione</td>
    <td>contenuto della pagina</td>
  </tr>
</table>
```

Come vedete il contenuto si trova dopo la navigazione. Ma con un piccolo espediente possiamo far salire il contenuto della nostra pagina al di sopra della navigazione senza influenzare la usability.

```
<table>
  <tr>
    <td>&nbsp;</td>
    <td rowspan="2">contenuto della pagina</td>
```

```
<tr>
  <td>codice per barra di navigazione</td>
</tr>
</table>
```

Nel caso, invece, abbiate la navigazione in alto il codice da usare è il seguente. Questo snippet non funziona perfettamente e, in alcuni casi, può dare dei problemi.

```
<table>
  <tr>
    <td>&nbsp;</td>
    <td rowspan="3" valign="bottom">contenuto della pagina</td>
  </tr>
  <tr>
    <td colspan="2">codice per barra di navigazione</td>
  </tr>
  <tr>
    <td>&nbsp;</td>
  </tr>
</table>
```

Il segreto sta nel contenuto

Sembra una cosa banale da dire, ma un buon sito si basa su di un buon contenuto che viene aggiornato di frequente. La logica che sta dietro a questo ragionamento è semplice: più contenuto avete, più pagine create e ognuna di queste pagine è un potenziale accesso per gli utenti. Se il contenuto è fatto bene (sia dal punto di vista della qualità, sia dal punto di vista del SEO) ogni pagina non solo sarà un potenziale accesso, ma sarà anche ottimizzato rispetto ad alcune keyword.

Ma c'è di più, il contenuto che produceate - articoli, recensioni, tutorial... - dovrebbe essere distribuito su quanti più siti possibili. In cambio potete chiedere che in fondo al contenuto ci sia un link al vostro sito. Questa semplice tecnica, se eseguita bene, vi permetterà di ottenere un buon numero di link che contribuiranno ad aumentare il PR del vostro sito.

Keyword density

Uno dei metodi usati da Google per misurare l'importanza di una keyword all'interno di una determinata pagina è quello di calcolare la densità delle keyword. In altre parole: più volte ripetete una keyword e più Google penserà che quella keyword è rilevante all'interno di quella pagina. Anche qui Google sembra dare maggior peso alle keyword che si trovano in cima al testo.

Un avvertimento: cercate di non sovra-ottimizzare una pagina, il vostro primo obiettivo è che questa venga apprezzata dall'utente. Google viene sempre secondo.

Font

Google riconosce la dimensione del testo e dà importanza alle keyword scritte più in grande oppure in grassetto o sottolineate. Considerando che le keyword sono utili, sia all'utente durante la lettura che a Google, si consiglia di evidenziare le keyword importanti con il tag .

Inoltre, dove non potete usare il tag <h1...6> ma vi serve identificare un titolo o delle keyword, fate in modo che il titolo sia grande, anche a discapito della grafica della pagina. È stato notato che pagine contenente del testo esageratamente grande - e intendo tanto grande da riempire una pagina solo con il titolo - ottengono una rilevanza per quelle keyword ben al di sopra della media. Ovviamente usate il vostro buon senso.

Al massimo 100 link

Non chiedetemi il perché, ma Google, nelle sue FAQ, indica esplicitamente che preferisce pagine che contengano meno di 100 link. Dato che non vogliamo contrariare Google: facciamolo e basta.

Per contare il numero di link in una pagina io uso questo semplice JavaScript. Per azionarlo è sufficiente inserirlo nella barra degli indirizzi e premere invio, oppure potete salvarlo nei vostri bookmark e richiamarlo quando serve.

```
javascript:alert('Questa pagina contiene ' + document.links.length + ' link.')
```

Tanti link, per favore

Come vedrete nella parte dedicata a come Google calcola il PageRank, l'idea originale di Brin e Page era che una pagina era tanto più importante quanti più link puntavano a quella pagina (avevano applicato a Internet il concetto delle citazioni sulle pubblicazioni scientifiche). La matematica alla base dell'applicazione di questo concetto non è semplice, ma come vedrete, non sono importanti solo il numero di link, ma anche il PageRank che questi link hanno.

Nonostante alcune voci di corridoio suggeriscano che il PageRank non abbia più una grande influenza sul posizionamento di un sito all'interno dei risultati, indizi più convincenti indicano il contrario. Ma che sia il PageRank a influenzare il posizionamento oppure sia una variante a farlo, a noi non interessa;

ciò che ci interessa e è che sembra certo che i link che puntano ad un sito conferiscono un punteggio che viene utilizzato nel calcolo del posizionamento.

Costruire, quindi, un solido network di link che puntano al proprio sito è molto molto importante anche se si rivelerà altrettanto faticoso e dispendioso in termini di tempo.

I punti da tenere a mente quando si vuole costruire questo network sono:

Creare una struttura di link che cresce gradualmente nel tempo. Pare Google riconosca quando tanti (troppi) link appaiono tutti in una volta sola e penalizzi questa pratica. Se potete aggiungete un nuovo link ogni giorno.

Più i siti sono grandi e importanti e più alto sarà il loro PageRank. Cercate di farvi linkare da questi siti. Meglio ancora se questi siti sono rilevanti al vostro contenuto. Farvi linkare da siti piccoli non vi darà invece alcun beneficio. Se è possibile, e ritenete il costo ragionevole può essere una buona idea quella di pagare per essere inclusi in directory molto considerate.

Cercate di fare in modo che il vostro link si trovi su di una pagina che non ha molti altri link in uscita; questo indebolirebbe il PageRank che vi viene trasmesso.

Verificate che il formato dell'URL sia sempre lo stesso. Gli esempi qui di seguito portano un utente alla stessa pagina, ma per Google si tratta di sei siti diversi:

<http://vostro-sito.com>

<http://vostro-sito.com/>

<http://vostro-sito.com/index.html>

<http://www.vostro-sito.com>

<http://www.vostro-sito.com/>

<http://www.vostro-sito.com/index.html>

Cercate di convincere chi vi linka a includere delle keyword per voi importanti all'interno dell'URL. Fate però in modo che tutti i link siano uguali, cercate di avere un po' di varietà: toglierà quella sensazione che i siti siano stati preparati meccanicamente.

I link reciproci sembra stiano perdendo favore all'interno di Google. Il vostro scopo è quello di ottenere link senza che voi dobbiate linkare a vostra volta. Specialmente non perdetevi troppo tempo a linkare pagine preposte solo allo scambio di link.

Evitate di linkare pagine che non sono negli indici di Google o di Yahoo. Evitate anche di linkare link farm, web-ring o qualunque altro tipo di sito che miri ad approfittarsi delle debolezze negli algoritmi di Google. Verreste black-listati e perdereste qualunque beneficio acquisito.

Anchor text

I link da siti esterni, verso il nostro sito, sono importanti, ma questi link vanno sfruttati meglio. Un link come questo dice poco a Google:

Scaricate DivX da <http://www.pc-facile.com/software.php?id=25> >qui.

Farsi invece linkare in questo modo è molto prezioso:

Cliccate qui per scaricare DivX.

La logica dietro questo ragionamento è che i siti esterni saranno sempre più onesti riguardo al reale contenuto di una pagina di quanto possiate esserlo voi.

Ora immaginate di abbinare quanto detto prima agli anchor text. Otterremmo il seguente risultato e vi posso assicurare che sarebbe molto importante questo messaggio per Google.

Cliccate qui per scaricare DivX.

Il consiglio è quindi non solo di cercare di avere tanti link che puntano al nostro sito, ma di cercare di influenzare il testo data la grande importanza che questo ricopre.

Link verso pagine esterne

Come vedrete nella parte dedicata a come Google calcola il PageRank, ogni link che porta dal vostro sito ad un sito esterno scala una piccola parte del vostro PageRank. Per limitare questo danno necessario cercate di limitare i link verso l'esterno e inserirli in pagine ricche di link verso altre pagine del vostro stesso sito.

Non fraintendetemi, però, i link verso pagine esterne sono importantissimi: un sito privo di link esterni viene cancellato dall'indice di Google! Ma anche non considerando questo fatto, i link sono la base di Internet: un sito privo di questi link non è possibile che fornisca un servizio adeguato ai propri utenti.

Pagine lontane dalla root

Come regola di base considerate che Google dà maggiore importanza a quelle pagine che sono nella root directory - questa è la cartella in cui c'è la home page di un sito. A ogni pagina che si trova in una sotto-cartella della root viene tolto un punto di PR. Questo punto può essere recuperato con un buon cross linking o se è linkata da siti esterni, ma parte comunque con uno svantaggio.

Dato che non c'è un numero massimo di pagine che possono risiedere nella root, se non per un problema di organizzazione, si consiglia di tenere lì quante più pagine possibili.

Cross linking

Come vedrete nella parte dedicata a come Google calcola il PageRank, è possibile influenzare il modo

in cui è distribuito il PageRank all'interno di un sito modificando la struttura dei link interni. La cosa non è facile, ma se avete necessità di dare importanza ad alcune vostre pagine potete provare diverse strutture finché non trovate quella che più fa al vostro caso. Per i vostri esperimenti potete usare gli script che trovate nell'appendice.

I primi link

I primi link, esterni al vostro sito, che potete ottenere facilmente sono anche quelli piuttosto ignorati dai novelli webmaster. Google dà una certa importanza a siti che sono indicizzate in Yahoo e in Open Directory. Seguite i link qui sotto e assicuratevi di essere presenti:

<http://www.dmoz.org/add.html>

<http://it.docs.yahoo.com/info/aggiungere.html>

Aggiungerei a questi anche altri due motori piuttosto snobbati, Altavista e Arianna. Non faranno granché ma vi concederanno un link in più:

<http://it.altavista.com/addurl>

<http://arianna.libero.it/addurl/inserisci.html>

E, ovviamente, non dimenticatevi di dire a Google di venire a cercarvi!

<http://www.google.com/addurl.html>

Meta tag

Le voci sui meta tag sono piuttosto discordanti. Molti pareri però indicano che Google considera "description" e "keyword" abbastanza poco, se addirittura li considera. Se avete molto tempo, e volete coprire tutte le opzioni, lavorate sui meta tag, ma fate in modo di NON utilizzare un meta tag unico per ogni pagina del sito.

Stando alle leggende il tag "description" è quello su cui lavorerei se proprio ne avessi il tempo. Altri motori, pare, diano molta importanza a questi tag.

```
<meta name="description" content="">
<meta name="keywords" content="">
```

Sessioni

Alcuni siti utilizzano un sistema di sessioni principalmente per cercare di garantire l'autenticazione dei propri utenti - ad esempio sui forum. Questo avviene aggiungendo in fondo all'URL una stringa alfanumerica valida per la durata di una visita e che però cambia ad ogni visita dell'utente.

Nell'esempio qui sotto la sessione è indicata da "sid=a1b2c3", anche se normalmente queste sono delle stringhe molto più lunghe.

<http://www.pc-facile.com/software.php?id=25&sid=a1b2c3>

Dato che questo rende l'URL di una pagina unico, Google tende a non indicizzare bene queste pagine. Se potete cercate quindi di usare altri metodi per autenticare i vostri utenti.

Cose da NON fare!

Come già spiegato la pena per chi bara è l'esclusione dall'indice. Considerato quanto lavoro c'è dietro a un sito, il rischio dell'eliminazione dall'indice di Google è troppo elevato. Ecco qui la lista nera delle cose da NON fare:

NON usate del testo nascosto per aumentare la densità delle keyword senza disturbare l'utente

Tecniche comuni includono usare testo dello stesso colore dello sfondo, layer invisibili, frame nascosti...

Questo punto è particolarmente importante. Qui di seguito vi mostrerò come vengono realmente messi in pratica questi trucchi: ve lo mostro solo per farvi capire, NON usate questi mezzi!

I tag HTML usati per inserire commenti all'interno di una pagina sono un buon esempio. Sono nati per permettere l'introduzione di brevi annotazioni. Non essendo visibili agli utenti che visitano la pagina web vengono spesso abusati per far trovare ai motori di ricerca un po' di keyword in più rispetto a quelle mostrate agli utenti:

```
<!-- Questo è un commento in cui potrebbero esserci molte keyword -->
```

Nei siti che fanno uso di frame, il tag <noframes> è stato ideato per mostrare testi a quei browser che non possiedono il supporto dei frame o che hanno temporaneamente disabilitato tale caratteristica. Un chiaro abuso del tag noframes consiste nel riempirlo di keyword:

```
<noframes>
Keyword visibili solo ai browser che non supportano i frame.
</noframes>
```

Il tag <noscript> è a volte usato per fornire contenuti alternativi a quei browser che non supportano il linguaggio indicato nel corrispondente tag <script>. Quello che è nato per meglio supportare gli utenti dei browser testuali o dei browser meno aggiornati viene spesso utilizzato come un ulteriore trucco per fornire ai motori di ricerca testi non visibili alla maggior parte degli utenti:

```
<noscript>  
Testo visibile solo ai browser che non supportano il linguaggio indicato in un  
precedente tag SCRIPT.  
</noscript>
```

Infine vorrei indicarvi un abuso del tag <alt> che viene utilizzato per descrivere le immagini o per mostrare del testo alternativo quando queste non sono visibili.

```
<img src="" alt?="Testo visibile durante un mouseover solitamente riempito di  
keyword">
```

Non riesco a ripetere a sufficienza di NON usare questi sistemi: imparateli, ma non utilizzateli!
Si pensa che Google stia utilizzando strumenti di Latent Semantic Indexing (LSI or LSA - Latent Semantic Analysis) che leggono un testo e permettono di determinare se questo è coerente o meno.

NO a pagine gateway o cloaking

Questa tecnica consiste nel creare delle pagine appositamente ottimizzate per i motori di ricerca che vengono date in pasto solo ai bot e non agli utenti veri.

NO link farm

Tecnica che prevede la creazione di un numero elevato di domini tutti inter-linkati tra di loro in modo che questi possano ottenere un PageRank elevato. Questo PR può servire esclusivamente internamente, oppure, come spesso fanno società di SEO, serve per aumentare il PR dei clienti.

NO shadow domains

Tecnica simile al cloaking, ma al posto di una pagina si ottimizza un intero dominio. Il traffico viene poi convogliato con redirect invisibili al sito finale che deve ricevere gli utenti.

Google mi ha cancellato dall'index

Nel caso dovesse succedere il peggio contattate Google per chiedere di essere ri-inclusi o per vedere se ci sono dei problemi. È inutile che sottolinei l'importanza di essere educati e gentili in questo frangente.
help@google.com

Non sovra-ottimizzare

Il vostro primo obiettivo è che l'utente apprezzi il vostro lavoro, non che lo apprezzi Google. I canoni del posizionamento cambiano di continuo, cercare di stare al passo con Google è impossibile. E anche se ci riusciste, il giorno in cui cambiano questi parametri, vi ritrovereste in guai seri. Lavorate bene, ma tenete bene in mente che il vostro obiettivo finale è l'utente.

A dimostrazione di quanto sto dicendo si pensi al 16 Novembre 2003 quando Google, senza preavviso, modificò il modo di ordinare i risultati. All'aggiornamento, come agli uragani Statunitensi, fu dato un nome: Florida. La conseguenza fu che molti siti sparirono dalle prime posizioni mentre altri scalarono le vette. Molti siti che avevano ottimizzato le pagine scomparvero, così come molti siti di e-commerce, e, a oggi, non si sa ancora cosa sia successo esattamente.

Alcuni hanno speculato che sia stata una mossa per togliere visibilità ai siti di e-commerce e costringendoli a aderire al programma pubblicitario di AdWords. Altre teorie sostenevano che Google filtrasse le pagine troppo ottimizzate. La teoria più accreditata sembra, però, essere quella che Google abbia implementato un sistema detto "Hilltop"; cioè una tecnica in cui alcuni siti vengono indicati come esperti del settore e link da questi siti acquistano un'enorme importanza nel calcolo dell'ordinamento dei risultati.

Il 16 Novembre 2003 fu un duro colpo per molti siti di ecommerce. Chi aveva puntato molto sulla conoscenza dei meccanismi di Google si ritrovò improvvisamente senza la terra sotto i piedi. Non ripetete lo stesso errore.

Contrariamente a quanto sostengono molti, e a quanto possa sembrare da questo capitolo, l'unica fonte di documentazione affidabile è questa:

<http://www.google.it/webmasters/>

5. AdWords

Introduzione

AdWords è un programma pubblicitario appartenente alla categoria dei pay per click (PPC), ovvero dove l'inserzionista paga soltanto quando un utente clicca sul banner, o annuncio, pubblicitario. Benché non sia stato inventato da Google, l'adozione di questa modalità pubblicitaria da parte del motore di ricerca sta creando fervore tra gli esperti del settore. Tanto che sono in molti a immaginarsi un futuro in cui altri canali pubblicitari (televisione, radio, stampa...) dovranno rendere conto di ogni centesimo speso dall'inserzionista.

Nella pubblicità tradizionale non solo l'inserzionista influisce poco su dove la sua pubblicità verrà posizionata (fossi un inserzionista che vuole pubblicizzare una marca di automobili vorrei che il mio spot non fosse mostrato accanto a quello di altre auto, e mi piacerebbe fosse inserito in un programma come "Magnum P.I." o "Supercar"), ma è anche difficilissimo misurare il preciso impatto di un singolo spot o di un singolo annuncio pubblicitario – come fa un inserzionista a misurare il rendimento dell'investimento (ROI) di ogni singolo spot?

Il pay per click advertising risolve entrambi questi problemi: il messaggio pubblicitario è contestualizzato al contenuto di una pagina web, o ai risultati di una ricerca effettuata – e quindi viene mostrato solo a persone che possono avere un reale interesse per il prodotto o servizio offerto –, e, tramite software appositi, l'efficacia dell'annuncio è misurabile in modo molto preciso.

Nel mondo del marketing non esiste nulla di così preciso, tanto che da arte, il marketing che può avvalersi di sistemi simili al PPC, sta diventando una scienza esatta e affidabile.

Procediamo ora a vedere com'è strutturato AdWords e poi vedremo l'importanza di ogni singolo aspetto e come fare per ottenere il massimo da questi. Quanto indicato in questo capitolo è da considerarsi corretto fino al momento della stampa. Prima di mettere in pratica questi consigli è bene verificare che le politiche interne di Google – soggette a cambi repentini senza preavviso – non siano state modificate.

La struttura di AdWords

Una volta aperto il proprio account si dovrà procedere alla creazione della strategia pubblicitaria. Per poter fare ciò è necessario impostare le proprie scelte in vari livelli. È importante ricordarsi che un'impostazione a livello più alto influenzerà i livelli seguenti.

A prima vista questo sistema può sembrare inutilmente complesso, ma una volta che si comincerà ad utilizzarlo con una certa abilità si capirà quanto sia importante questa struttura per la buona riuscita del

proprio lavoro.

Campagna

Il primo livello della nostra strategia è la “campagna”. Qui si definiscono le lingue e i paesi in cui si vuole presentare il proprio annuncio. È possibile utilizzare qualsiasi combinazione di questi fattori: ad esempio si può decidere di mostrare i propri annunci solo a utenti che effettuano ricerche in Inglese, ma che sono fisicamente in Italia, oppure solo in Lombardia, o anche specificare che devono essere utenti che sono a Milano quindi il nostro obiettivo sarà di raggiungere gli stranieri in Italia. Oppure si può decidere di fare il contrario selezionando di mostrare il nostro annuncio a chi compie ricerche in Italiano ma non è in Italia; oppure si può limitare il nostro campo ai semplici Italiani che vivono in Italia. Le possibilità sono moltissime.

Come si può vedere la dislocazione geografica è piuttosto dettagliata permettendo di mostrare la propria pubblicità in zone molto precise (anche se questo dettaglio non è disponibile in tutti i paesi), mentre il sistema riconosce 40 lingue.

A questo livello si sceglie il mercato a cui ci si vuole rivolgere; è possibile impostare un numero illimitato di campagne pubblicitarie.

Gruppo di annunci

Un gruppo di annunci è composto da annunci e keyword utilizzati assieme per pubblicizzare un prodotto o un servizio. Qui si imposta l'annuncio pubblicitario e le keyword utilizzate per attivare l'annuncio pubblicitario. Si hanno a disposizione un'intestazione (massimo 25 caratteri) e due righe di testo con un massimo di 35 caratteri ciascuna.

Dovrete poi scegliere le keyword utilizzate per attivare l'annuncio e la vostra massima offerta per le keyword che avete scelto. Notiamo che, benché sia possibile dare un valore specifico ad ogni keyword o frase (intesa come la somma di più keyword che compongono una sola query), il costo massimo per click indicato qui è da intendersi per tutte le keyword di quel gruppo.

Non esiste limite al numero di annunci che potete utilizzare all'interno di una campagna.

Budget giornaliero

Qui potete definire il massimo budget giornaliero che spenderete per l'intera campagna. In questo modo siete certi che l'intero budget mensile non verrà mai utilizzato in un solo giorno e potete tranquillamente farvi le due settimane di mare che vi meritate senza farvi rovinare da qualche cliccatore maniaco-compulsivo.

Non dimentichiamoci del resto

Prima di addentrarmi nelle strategie vorrei ricordare che AdWords è soltanto una parte del processo di vendita. Certo, si tratta di una parte importante, ma è solo una parte.

Allo stesso modo bisogna ricordarsi che AdWords non può essere riassunto in “scegli le keyword migliori”. No, AdWords è scelta delle keyword, è scrivere un buon annuncio pubblicitario, è strutturare in modo intelligente le proprie campagne, è fare tracking dei risultati, è modificare keyword e annunci a seconda del tracking e, infine, è ricominciare tutto da capo. Non perdiamoci nei dettagli!

Come funziona AdWords

Google mostra un annuncio pubblicitario quando un utente effettua una ricerca su Google e le keyword inserite corrispondono alle keyword scelte dall'inserzionista per un determinato annuncio, oppure il contenuto delle pagine di Google (Google Groups, GMail...) sono rilevanti per l'annuncio pubblicitario. Oppure quando uno dei partner di Google (AOL, Ask Jeeves...), o dei siti affiliati attraverso il programma AdSense, mostrano del contenuto rilevante alle keyword scelte dall'inserzionista.

Quando questo succede Google prende gli annunci rilevanti e usa i seguenti due fattori per decidere se l'annuncio verrà mostrato, o no, e in quale posizione:

Il costo massimo che siamo disposti a pagare per un click; ovvero il cost per click (CPC). Più paghiamo e più saremo visibili.

Il nostro click through ratio (CTR); ovvero che percentuale di utenti clicca sul nostro annuncio quando questo viene mostrato.

Mentre il primo fattore è piuttosto ovvio più paghiamo e più possibilità abbiamo di essere visti le implicazioni del secondo sono meno evidenti. Il CTR rappresenta la nostra abilità nell'abbinare keyword ad un annuncio pubblicitario e di scrivere un messaggio accattivante che invoglia l'utente a cliccare. Più il nostro annuncio sarà rilevante per le keyword scelte e più il testo sarà interessante e più l'utente sarà invogliato a cliccare. Più è alto il nostro CTR e meno avremo bisogno di pagare per ottenere posizioni alte e, quindi, per essere visti (Google premia gli inserzionisti più rilevanti mostrando il loro annuncio in posizioni più alte: Google può mostrare fino a dieci inserzioni pubblicitarie su ogni pagina).

Avere un CTR alto ci permette di risparmiare e di poter competere con grandi società che hanno a disposizione grandi capitali e, magari, poca fantasia.

Il primo consiglio

I vostri primi esperimenti daranno risultati pessimi. Anche se siete inserzionisti esperti e conoscete bene il mercato di riferimento, i primi risultati su AdWords saranno deprimenti. Mi spiace, ma è bene che

siate preparati se volete utilizzare questo canale. Vi ci vorrà tempo, impegno e pazienza, ma se li avrete i risultati arriveranno e, dato che queste sono qualità rare, alla fine farete meglio della concorrenza che solitamente non ha nessuna di queste caratteristiche. Ma preparatevi, non pensate di aprire un account e spopolare. Non succederà.

Tracking

Per poter monitorare e valutare il successo delle vostre campagne è importante che facciate tracking, ovvero che controlliate cosa succede una volta che l'utente ha cliccato sul vostro annuncio pubblicitario. Il motivo principale che determina il successo di una campagna è il tracking e l'abilità di agire in base ai dati ottenuti.

Forse dovrei ripetere questo concetto: il ciclo di tracking delle visite e il conseguente aggiustamento delle campagne, delle keyword e del testo degli annunci è il segreto del successo del PPC advertising. Non è possibile ripetere a sufficienza questo punto.

Fare tracking non è difficile ma implica che usiate un linguaggio di scripting server-side (come il PHP, l'ASP, .NET o JSP) e che abbiate a disposizione dei programmatori discreti. In alternativa potete utilizzare delle società esterne che possono svolgere questo compito per voi. Anche Google offre, gratuitamente, questo servizio di tracking e per cominciare non potrei consigliarvi di meglio. Google è ben strutturato, è gratuito, ed è configurato per operare al meglio con AdWords. Detto questo sono in tanti che preferiscono usare società esterne anche se queste non sono gratuite.

Oltre a Google quindi, esistono svariate possibilità. Qui di seguito ve ne indico due anche se vi consiglierei di fare un po' di shopping prima di scegliere.

Il primo è Conversion Ruler. Si tratta di un ottimo servizio con un ottimo prezzo: il pacchetto base vi costa \$20 e tiene traccia fino a 2.000 click. Una buona soluzione per le piccole imprese.

<http://www.conversionruler.com>

La seconda opzione è Click Tracks. Si tratta di un servizio del tutto simile al primo il cui costo del pacchetto di base è però di \$50.

<http://www.clicktracks.com>

Se siete alle prime armi in questo campo vi consiglio di familiarizzarvi con gli strumenti messi a disposizione da Google. Una volta che siete pratici saprete esattamente cosa volete e se questi pacchetti fanno per voi o se avete bisogno di fare altro shopping.

Approccio a AdWords

Per avere successo con AdWords avrete bisogno di una mentalità molto strutturata. Non è necessario fare delle strategie troppo dettagliate a priori. Prendete un vostro prodotto, o servizio, e create una prima campagna pubblicitaria; si tratta di un modo per rompere il ghiaccio, ma anche di un modo per familiarizzarvi con AdWords e con l'utenza di questo canale.

Create gruppi di keyword molto specifiche e molto legate tra loro e, per ogni gruppo, create un annuncio pubblicitario (in gergo: copy) unico. Potete creare più gruppi uguali e per ognuno usare un copy diverso. Col tempo capirete quali sono gli annunci funzionano meglio e quali non; capirete quali keyword hanno senso e quali vanno eliminate. Il segreto sta nella sperimentazione. Create più annunci; scartate gli annunci non performanti e createne di nuovi in base a quelli che vedete funzionare. In media, ogni prodotto o servizio dovrebbe avere una campagna dedicata e in ogni campagna dovrebbero esserci una decina di gruppi di keyword con relativo annuncio, di cui almeno un paio con testi e keyword sperimentali.

Cosa funziona meglio per voi? Pubblicizzare un servizio gratuito? Parlare di una promozione? Risolvere un problema specifico del vostro settore? Anche piccole variazioni di punteggiatura, di ordinamento di parole possono apportare grandi differenze in termini di CTR.

Provate vari approcci e vari stili finché non identificate i annunci che funzionano meglio per voi. ma ricordatevi: la struttura!

I gruppi di annunci

Questi sono mattoni su cui è costruita una campagna e su cui è costruito AdWords. Quando vi ritroverete a fare tracking potreste arrivare a calcolare il ROI delle keyword individuali, ma vi renderete conto che, se costruiti con i criteri necessari, i gruppi di annunci è il limite a cui è necessario spingersi. Se la vostra struttura è logica, i dati che potete ottenere a questo livello sono sufficienti a fornirvi il feedback che vi serve.

Con questo non intendo dire che non dovete sostituire le keyword peggiori con keyword migliori e trovare nuove keyword. Voglio dire che il tracking può fermarsi al gruppo; perché bisogna cominciare a pensare al mondo come a un mondo di idee o gruppi di idee che sono espresse dagli utenti sotto forma di query. Le intenzioni degli utenti vanno anticipate dagli inserzionisti sotto forma di gruppi di annunci. Ogni gruppo rappresenterà quindi un'idea, un problema o una soluzione ben precisa che l'inserzionista propone all'utente.

Quante keyword possono esserci in un gruppo? Quanto dev'essere grande una lista per essere una lista grande?

Non lo so, e non importa. Quanti sinonimi possono esserci a una query ben definita? Una lista di 1,000

keyword non può sfociare di certo in un solo messaggio specifico. Ciò che rende una campagna degna di essere chiamata grande è quindi il numero di annunci che contiene. Quanti? Mah, almeno una cinquantina.

Come scrivere un annuncio

Scrivere copy è un'arte e le persone che lavorano nel campo possono confermarvi che non ci si improvvisa in questa professione da un giorno all'altro. L'unico consiglio che sembra migliorare nettamente il CTR è quando l'annuncio e/o il titolo dell'annuncio contengono le keyword digitate dall'utente. Sì, siamo in un campo che richiede pubblicità molto molto specifica, ma, del resto, siamo qui per quello!

Matching

Come nella normale ricerca, Google mette a disposizione un numero di strumenti per aiutarci a individuare gli utenti più rilevanti al nostro annuncio. Questi che seguono sono gli equivalenti degli operatori di cui vi ho parlato nei capitoli precedenti.

Corrispondenza generica (Broad matching)

Digitate semplicemente le keyword, come: *scarpe tennis*.

Il vostro annuncio verrà mostrato quando gli utenti effettuano una ricerca con le parole *scarpe* e *tennis* in qualsiasi ordine e anche se la ricerca comprende altri termini, ad esempio *racchette tennis scarpe*. Riuscire a ottenere dei risultati qualificati con questa tecnica è molto difficile specialmente se usate delle frasi brevi come l'esempio indicato qui sopra. Infatti il vostro annuncio verrebbe mostrato sia nel caso venisse digitato *dove posso comprare delle scarpe da tennis* che *vendo il mio vecchio paio di scarpe da tennis* e anche *posso entrare a teatro con le scarpe da tennis?*. Non potete nemmeno immaginare cosa scrivano le persone quando fanno una query.

Corrispondenza a frase (Phrase matching)

Racchiudete tra virgolette le keyword: *"scarpe tennis"*

Il vostro annuncio verrà mostrato se l'utente ricerca i termini *scarpe tennis* in quest'ordine, anche in presenza di altri termini. Ad esempio, l'annuncio verrà visualizzato per la query *scarpe tennis rosse*, ma non per *scarpe da tennis*.

L'utilizzo di questa opzione crea un effetto carattere jolly che, con un uso sapiente delle keyword negative diventa uno strumento molto preciso. Se dovessi esprimere una preferenza tra questi operatori direi che la corrispondenza a frase è quella che preferisco, anche se di poco.

Corrispondenza esatta (Exact matching)

Racchiudete tra parentesi quadre le keyword: *[scarpe tennis]*

(Preghiamo notare che nella sezione dedicata a AdWords, le query saranno sempre indicate in corsivo; questo perché le parentesi quadre hanno un significato ben preciso per Google).

Il vostro annuncio verrà mostrato quando gli utenti effettuano una ricerca con le keyword *scarpe tennis* in quest'ordine e solo con questi termini. Per esempio, il vostro annuncio non verrebbe visualizzato con query quali *scarpe tennis rosse* o *borse e scarpe tennis*.

Essendo molto precisa, quest'opzione viene poco utilizzata, per cui è possibile trovare delle buone keyword ad un ottimo prezzo.

Corrispondenza inversa (Negative keywords)

Inserite un trattino prima della parola chiave: *-rosse*

Con le keyword *scarpe tennis* e keyword negativa *-rosse*, l'annuncio non verrà visualizzato se l'utente esegue la ricerca di *scarpe tennis rosse*.

L'utilizzo delle keyword negative ci dà molto spazio di manovra e ci permette di ottenere traffico più qualificato. Se, ad esempio, il tracking ci informa che le query da cui provengono i click through sono *vendo scarpe tennis usate*, possiamo pensare di inserire *vendo* e *usate* come parole negative in modo da non dover pagare i click di chi cerca scarpe usate. Il tipo di parole negative da usare varierà molto a seconda del mercato in cui operiamo.

Non esistono delle regole auree che indicano quale di questi tipi di corrispondenze è il migliore. Molto dipende dal mercato in cui operate e dalla vostra abilità nell'individuare le keyword migliori. Come sempre l'unico consiglio è: sperimentare!

Nuove keyword

Fino a qui abbiamo detto molte cose su AdWords. Se qualcuno ne sapeva già qualcosa adesso si starà chiedendo: ma quando parlerà di come trovare nuove keyword?

Eccoci!

Siamo arrivati al punto più critico dell'intera campagna: come inventarsi nuove keyword più specifiche. Qui sotto sono elencati alcuni consigli. Per favore notiamo che questo è un processo che dovrebbe aiutarvi a pensare, non è un elenco completo di idee per generare nuove keyword.

Sinonimi: trovate tutti i sinonimi delle keyword che vi vengono in mente. Vendete un libro? Provate con *manuale*, *volume*, *tomo*, *trattato*...

Usate **variazioni delle keyword:** singolare e plurale, maschile e femminile, coniugate i verbi...

Usate le **corrispondenze** al meglio delle loro potenzialità. Se la keyword che avete in mente è *calligrafia*, ma risulta troppo generica provate ad aggiungere un'altra keyword e usare una corrispondenza generica. Ad esempio *calligrafia bella*, *calligrafia brutta*, *calligrafia gotica*, *calligrafia raffinata*, *calligrafia moderna*, *calligrafia classica*, *calligrafia reale*...

Siate **specifici**: *vendo fiori* o *vendo fiori milano* potrebbe essere un buon inizio; perché non indicare il nome dei fiori: *rose, violette, margherite*? Esiste un intero campionario di ricerche seguendo questa linea di pensiero. Nel PPC bisogna essere il più specifici possibile.

Metodo induttivo: pensate a che interessi può avere il vostro pubblico e sperimentate. Cosa può interessare a una persona che cerca stampa libro? Magari gli serve un servizio *impaginazione*, un *artista* che gli disegni al *copertina* oppure un *editore*..

Metodo deduttivo: quando trovate delle keyword con un CTR alto pensate a tutte le keyword che hanno un legame con queste. Nel caso di libro possiamo pensare a: *autore, manoscritto, editore, stampa, opera, romanzo*..

Matrici di parole. Immaginate di voler vendere *automobili usate* Cercate almeno tre sinonimi per la parola automobile: *vettura, auto, macchina*.... Cercatene altri tre per la parola usate: *vecchie, seconda mano, 20.000km*.. Ora create le nove possibili combinazioni di parole. Ovviamente questo esercizio andrebbe ripetuto con più parole e più sinonimi.

Errori di **ortografia**: pensate a come potrebbe essere sbagliata l'ortografia di una keyword.

Usiamo un linguaggio tecnico e preciso. Se operate in un settore industriale specifico usate le parole che vi qualificerebbero come esperto di fronte ad un vostro collega.

Indichiamo **un problema**. Pensate a come esprimere il problema che il vostro prodotto o servizio risolve. Ad esempio, se vendete fiori: *fiori matrimonio, fiori compleanno*, per una *donna* o per un *funerale*.

Indichiamo **una soluzione**. Descrivete il vostro prodotto o servizio risolve come fosse una soluzione.

Dal **punto di vista dell'utente**. Descrivete il prodotto, il problema o la soluzione come lo descriverebbe un utente, non come lo descrive un operatore del settore.

Aggettivi descrittivi/qualificativi. Provate a aggiungere aggettivi alle query: *gratuito, veloce, professionale, commerciale, raffinata*..

Acronimi e abbreviazioni. Provate con *USA* al posto di *America*, *ROI* al posto di *return on investment*, *USD* al posto di *dollari*..

Diversi **modi di scrivere** la stessa parola: porta cenere, portacenere o porta-cenere? Stati Uniti, USA, America o U.S.A.?

Domini. Per qualche motivo molte persone digitano gli URL nel box di Google. Provate a usare *http://www.keyword.com*, *http://www.keyword.it*, *http://keyword.com* e *http://keyword.it*

Luoghi e eventi famosi. Pensate a battaglie, città, chiese, personaggi di film e del cinema, ritornelli... che possono essere attinenti al vostro prodotto o servizio. Se è appena uscita una nuova canzone potete essere certi che saranno in tanti a fare query con parti del testo.

Avvenimenti. Ogni anno si svolgono migliaia di eventi: concerti, spettacoli, manifestazioni sportive, eventi culturali... Nel periodo che li precede e immediatamente dopo le keyword rilevanti saranno ricercatissime; trovate un collegamento tra questi eventi e il vostro prodotto o servizio.

Siamo ancora nell'infanzia di questo servizio e la grande industria non ha ancora mosso i suoi capitali compilando liste di keyword attinenti e specifiche. L'esperienza è ancora poca e il mercato è ancora relativamente facile. Google risponde a 150 milioni di query al giorno: le possibilità sono infinite!

Le dimensioni contano

Il lavoro con le keyword è fondamentale ed è uno dei segreti del successo di una campagna. Dovrete sforzarvi di trovare keyword, tante keyword. Quante keyword? In alcuni mercati è possibile arrivare facilmente alle **centomila** keyword; non pensate di cavarvela con qualche centinaia di keyword!

Ripeto questo punto molto importante: il numero di keyword che usate è fondamentale!

Keyword illegali?

Nella ricerca frenetica di nuove keyword ci capiterà di pensare ad alcune keyword che potrebbero parerci illegali. Non esiste una legislazione precisa in merito e benché alcune società si stanno muovendo per fare in modo che non si possano usare i loro marchi per effettuare ricerche nulla è ancora stato deciso.

Del resto se il nostro annuncio non è ingannevole, non pretendiamo di spacciarsi per rappresentanti del, o in qualche modo affiliati al, marchio e lo spazio di cui ci avvaliamo è chiaramente riservato alla pubblicità, allora non stiamo infrangendo alcuna legge. Fosse così gli inserzionisti “tradizionali” dovrebbero lamentarsi che il loro spot è stato trasmesso appena dopo quello di una marca competitiva e l’agente di Manuela Arcuri dovrebbe considerare lesivo essere intervistata su di una pagina che mostra la pubblicità di una caciotta.

Ripeto, siamo in un area in cui non esiste ancora una legislazione precisa, per cui chiedete al vostro avvocato prima di procedere.

In questa lista di keyword crepuscolari possiamo pensare a:

Personaggi famosi. vendiamo scarpe da tennis? Usiamo i nomi di tennisti famosi, di tornei e di campi da gioco.

Marchi proprietari. Potrebbe essere difficile giustificare l’utilizzo di un marchio competitivo, ma se non stiamo traendo in inganno l’utente, una keyword famosa potrebbe fare miracoli ed essere relativamente a buon mercato. Vendiamo marmitte per moto? Perché non provare keyword come *marmitta honda*, *marmitta aprilia*, *marmitta yamaha*... anche se non vendiamo i ricambi originali.

Powerposting

Abbiamo detto che la nostra massima offerta per le keyword è definibile a livello di gruppo di annunci. Questo è esatto, ma non è tutto: Google ci dà la possibilità di definire il CPC massimo per ogni singola keyword. Per fare ciò basta aggiungere due asterischi dopo la keyword e poi il valore che vogliamo pagare.

Ad esempio, se abbiamo un gruppo di annunci per cui paghiamo €0,50 possiamo effettuare i seguenti cambiamenti:

scarpe**0,20

scarpe tennis

scarpe tennis rosse**0,75

In questo caso le keyword *scarpe tennis* ci costano ancora €0,50, mentre la keyword *scarpe* – che riteniamo troppo generica abbiamo deciso di pagarla solo €0,20; invece valutiamo di più *scarpe tennis rosse* che è più precisa come query.

Non vi basta? Volete ancora più flessibilità? Allora aggiungete altri due asterischi e specificate un URL diverso a cui deve puntare solo quella query:

scarpe tennis rosse**0,75***<http://www.vostrosito.it/xyz>

Questa pratica viene definita powerposting e, benché molto potente, rischia di essere molto dispersiva e confusionale. È consigliabile utilizzare queste opzioni con parsimonia in modo da mantenere delle campagne ordinate e ben strutturate e facili da gestire. Ripeto quanto detto sopra in termini diversi: cercate di strutturare le vostre campagne e i vostri gruppi in modo che l'utilizzo del powerposting sia minimo.

Il click through ratio

Abbiamo già indicato che il CTR è uno dei fattori più importanti usato da Google per calcolare la rilevanza di un annuncio. Google sostiene che un annuncio cliccato è un annuncio rilevante e un annuncio rilevante è un'informazione preziosa per l'utente. E se l'utente trova informazioni preziose è certo che ritornerà innestando così un circolo virtuoso.

Per questo motivo Google ci tiene molto che il CTR delle campagne degli inserzionisti sia alto (quindi rilevante per l'utente) e farà di tutto per imporre ciò. Per questo, mentre da una parte ricompensa il nostro CTR aumentando la visibilità degli annunci, dall'altra parte penalizza le keyword peggiori disabilitandole. È nel nostro interesse avere un CTR alto.

Teniamo però a mente che:

Google ricorda i CTR passati. Se mettiamo in pausa una campagna l'alto CTR che abbiamo guadagnato non andrà perso: niente paura.

Nonostante venga mostrato un CTR calcolato tenendo presente il click through ratio degli annunci mostrati sui siti partner, il solo CTR che conta verso la disabilitazione delle keyword è quello calcolato sulle pagine dei risultati di Google.com.

Ciò che conta è il CTR complessivo di una campagna. Se la media complessiva di una campagna è al di sopra dello 0,5%, le keyword che non riescono a raggiungere questa soglia sembrano essere ignorate. In altre parole: tenete una buona media e le vostre perle segrete, quelle che pagate davvero poco, ma che non hanno un CTR alto non saranno disabilitate. Pensate, potete avere decine di keyword che costano €0,05 con un CTR molto basso e riuscirete a tenerle attive se la media complessiva della vostra campagna è sufficientemente elevata!

Google calcolerà la performance della vostra campagna dopo le prime 1.000 impression.

In ogni caso, a meno di nuove disposizioni, fate in modo di tenere il vostro CTR al di sopra del 0,5% o le vostre keyword verranno disabilitate. Alcune voci indicano che, per certe posizioni, Google accetta CTR fino a 0,35%, ma qui siamo davvero nel reame del pettegolezzo.

Sindacare o non sindacare

Selezionando la nostra campagna avremo la possibilità di modificarne le impostazioni. Una delle impostazioni principali di cui tenere conto è dove, nel network di Google verranno mostrati i annunci pubblicitari. Non esiste una regola aurea, ogni mercato è diverso e ogni inserzionista, a seconda delle sue caratteristiche troverà più redditizi alcuni canali invece di altri. Il segreto, come in tutto AdWords, è sperimentare. Solo così, solo sperimentando, riusciremo a capire cosa funziona meglio per noi.

Google ci offre due possibilità oltre alle pagine di Google – che non possono essere disabilitate:

La rete di ricerca: ovvero se mostrare i nostri annunci pubblicitari sui siti partner di Google (AOL, Netscape, Libero...).

La rete di contenuto: ovvero se mostrare i nostri annunci pubblicitari sui siti che utilizzano AdSense.

È importante notare che Google offre fino a dieci posizioni, cioè ci possono essere fino a dieci annunci pubblicitari su di una pagina dei risultati delle ricerche. I partner di Google e i siti affiliati (AdSense) ne offrono tipicamente meno della metà. Questo vuol dire che per essere visibili su questi siti per keyword molto competitive potrebbe essere necessario pagare più di quanto siamo disposti.

Il budget giornaliero

Questo serve a Google per calcolare una stima della frequenza con cui mostrare il nostro annuncio per raggiungere il budget giornaliero da noi indicato. Questa stima si basa, oltre al nostro massimo CPC, anche su dati come il CTR medio e il numero previsto di query rilevanti. Si tratta però solo di medie e di stime che spesso non possono essere rispettate; Google, infatti, ci garantisce solo di non farci spendere più del budget stabilito, non di riuscire a spendere esattamente quella cifra.

L'esperienza ci insegna che il budget giornaliero non viene sempre raggiunto. Se decidiamo, ad esempio, di allocare €100 ad una campagna, la nostra spesa media si avvicinerà a €60 o a €70 con punte di €100 e minimi fino a €10. Se vogliamo spendere una media €100 dovremo indicare a Google un budget giornaliero di €200 (notiamo che queste proporzioni possono variare a seconda del mercato, del copy e delle keyword scelte; è importante sperimentare per costruire delle statistiche rilevanti a noi).

In altre parole avere un budget basso significherà spendere ancora di meno e qualunque incapace è capace di spendere meno, il difficile è spendere bene quando si spende di più!

Ma spendere poco implica anche un numero di click molto basso e se abbiamo solo una decina di click a fine giornata ci vorrà molto tempo per costruire un campione affidabile su cui basare le nostre analisi. Come si può dire che un annuncio pubblicitario è scritto male se è stato visto da sole cento persone? Se decidiamo di spendere troppo poco il nostro ciclo di apprendimento e il marketing iterativo che ci siamo proposti di fare non andrà da nessuna parte!

In realtà non esiste nessun buon motivo, una volta che abbiamo affinato per bene la nostra campagna, per non spendere tutto ciò che abbiamo. Se i ritorni sono quelli che vogliamo; se i click si convertono in vendite con una certa regolarità e se abbiamo il capitale per coprire queste spese, allora non c'è motivo per non andare al massimo. L'unico nostro limite, se proprio vogliamo trovarne uno, è la nostra capacità di evadere gli ordini. Perché se spendiamo €1.000 e riceviamo 10 ordini; se spenderemo €100.000 otterremo 1000 ordini e la nostra impresa potrebbe non essere in grado di evadere tutti questi ordini!

Ottimizzare gli annunci

È possibile delegare a Google il compito di ottimizzare la distribuzione degli annunci lasciandogli presentare più spesso gli annunci che hanno un CTR più alto e lentamente ritirando quelli che con un CTR più basso. Benché può sembrare un'ottima scelta, è importante ricordarsi che questo non ci permetterà di sperimentare nuovi testi, e nuove keyword con la meticolosità necessaria.

Cost per acquisition (CPA) e budget

Vorrei che ora mi seguiste un attimo. Forse il concetto che sto per introdurre sembrerà strano e fuori luogo, ma vi chiedo un attimo di attenzione perché le conclusioni potrebbero sorprendere molti di voi.

Nel marketing tradizionale, una delle variabili è il costo dell'acquisizione di un cliente (CPA); cioè quanta pubblicità è necessaria perché un nuovo cliente effettui un acquisto (in realtà è quanta pubblicità è necessaria per acquisire un cliente, ma supponiamo, per semplificare, di essere in un mercato in cui i

clienti effettuano un solo acquisto). In termini tradizionali questo viene calcolato usando il totale della spesa del marketing diviso per il numero di nuovi clienti – non importa se i carissimi spot televisivi non portano clienti mentre il volantaggio è una miniera d'oro: in assenza di strumenti di misura viene tutto calcolato insieme.

L'effetto di una campagna di marketing tradizionale viene misurato a intervalli di 3, 6 o addirittura 12 mesi dall'avvio. Gli strumenti a disposizione non ci danno la possibilità di vedere in tempo reale i risultati del nostro lavoro. Questo vuol dire che, quando si preparano i piani per il nuovo anno, mentre possiamo fissare la spesa di marketing, il numero di nuovi clienti è soltanto una proiezione basata sull'esperienza; non è detto che le nostre previsioni verranno rispettate. Per questo motivo i budget sono contenuti; un risultato al di sotto delle previsioni non deve danneggiarci troppo – come facciamo a predire con 3, 6 o 12 mesi di anticipo l'effetto di un nuovo spot pubblicitario, di un nuovo testimonial oppure come reagirà il mercato se raddoppiamo il budget?

Gli effetti possono essere buoni, anche ottimi, ma possono essere disastrosi. Se vogliamo fare degli esperimenti dovremo farli piccoli e il nostro ciclo iterativo sarà di 3, 6 o 12 mesi.

Con il CPC advertising, invece, è possibile monitorare con molta precisione le nostre campagne pubblicitarie. Possiamo vedere quanto ci costa un'intera campagna fino a quanto ci costa la singola keyword. Possiamo vedere se l'annuncio A attira persone che poi procedono all'acquisto più di quanto non lo faccia l'annuncio B. Possiamo fare piccoli esperimenti ogni giorno e possiamo agire sul feedback ottenuto con cicli anche di 12 o di 24 ore se vogliamo. Possiamo giustificare ogni centesimo speso e possiamo quasi garantire quale sarà il CPA.

Le implicazioni di quanto abbiamo appena detto portano a una sola conclusione: il concetto di budget di marketing diventa obsoleto – se non nel senso che questo indica il nostro limite di evadere gli ordini.

In altre parole, con un tracking così preciso e con dei cicli di feedback così veloci non esiste motivo per limitare il budget. Infatti, non esiste motivo per parlare proprio di budget: finché riusciamo a restare al di sotto del CPA che ci permette di generare un profitto accettabile, l'amministrazione dovrebbe essere solo felice di darci altri soldi da spendere!

Essere i primi?

Essere il primo annuncio pubblicitario comporta alcuni vantaggi, ma sta diventando opinione comune tra chi opera in questo settore che la miglior posizione è la seconda. Perché? Beh, solitamente il primo posto ha un costo molto elevato e competere per questa può scaturire una guerra al rialzo che è bene evitare sempre. È ovvio che le posizioni più alte sono quelle più in vista e quelle più cliccate, ma alcune ricerche e certi commenti non ufficiali da parte di Google sembrano indicare che non c'è un calo drastico in termini di click tra la prima e la seconda posizione. Inoltre la prima posizione pare attrarre

persone che cliccano in modo compulsivo e questi click si convertono meno in vendite che click su posizioni più basse.

Concludendo, un posizionamento che oscilla tra la seconda e la quinta posizione è ottimale. Vi invito, invece, ad accertarvi che non stiate spendendo troppo se la vostra media è superiore a 1,5 (cioè tende a 1).

Va da sé che se siamo i soli ad aver scelto una determinata keyword, saremo in testa e il nostro CPC sarà di €0,05.

Google inoltre ci fa sapere che il CTR viene normalizzato a seconda della posizione del annuncio pubblicitario. Questo vuol dire che non dobbiamo preoccuparci di raggiungere la soglia minima del 0,5% di CTR se il nostro annuncio viene mostrato nelle posizioni più basse.

Conclusioni

Ottimo, siamo arrivati in fondo: la parte più dura l'abbiamo superata. Ora accendiamo il PC e impostiamo quella dannata prima campagna. Ho un ultimo segreto da svelarvi: i primi mesi saranno i più duri, ma una volta superati quelli è tutto in discesa – in realtà, per alcuni, è sufficiente una sola settimana! Una volta che tutto funziona bene si tratta solo di ritoccare qua e là aggiustando keyword che stanno diventando troppo care e testando nuovi annunci o nuove keyword che ci vengono in mente. Questo è tutto, il resto è sperimentazione!

6. Storia di Google

Perché Google?

Google è un gioco di parole: la parola “googol”, che indica il numero 1 seguito da 100 zeri, fu coniata da Milton Sirota, nipote del grande matematico americano Edward Kasner, e rispecchia l'ambizione di organizzare l'immenso, quasi infinito, volume d'informazioni disponibili sul web.

L'inizio di un lungo cammino

1995, Larry Page, 24enne laureato all'Università del Michigan, va in visita a Stanford con l'idea di iscriversi al dottorato in informatica. Sergey Brin, 23enne, fa parte del gruppo di studenti assegnati a mostrare il campus ai visitatori.

Non si può dire che fu amore a prima vista tra i due. Anzi, la parola amore non passò per la mente di nessuno dei due. Più che una visita sembrò un'aggressiva discussione tra fratelli. Le loro forti opinioni e i punti di vista divergenti si trovarono concordi solo su una cosa: sull'approccio alla soluzione di una delle più grosse sfide nel campo dell'informatica, estrarre informazioni rilevanti da un enorme set di dati.

Ma da queste divergenze nacque una componente importante nei rapporti: il rispetto. E con il rispetto arrivò l'amicizia.

Nel Gennaio del 1996, Larry e Sergey, cominciarono a collaborare alla costruzione di un motore di ricerca denominato BackRub. Si trattava di un progetto universitario, non di un'avventura imprenditoriale: due persone con delle buone idee e le capacità di metterle in pratica, si erano incontrati. Il motore fu chiamato in questo modo perché aveva l'abilità di analizzare i “back links” che puntavano a un sito.

Ma neanche le capienti casse di Stanford non erano sufficienti a sovvenzionare l'esperimento dei due studenti. Afflitti quindi da una perenne mancanza di fondi, male che colpisce gli studenti di tutto il mondo, la coppia cominciò a frequentare i luoghi più improbabili del dipartimento d'Informatica nella speranza di trovare computer da poter prendere “in prestito” per il loro network. Larry, esperto di elettronica – e che aveva ottenuto una certa notorietà per aver costruito una stampante funzionante con i soli pezzi del Lego™ –, si assunse il compito di creare un nuovo tipo di ambiente server che utilizzasse PC comuni invece dei carissimi server, tipici dell'epoca.

Un anno più tardi, voci di una nuova tecnologia di ricerca stavano già diffondendosi per il campus, dando ai due una certa reputazione per l'approccio innovativo all'estrazione dei dati.

In cerca di un acquirente

Larry e Sergey lavorarono per tutta la prima metà del 1998, perfezionando la loro tecnologia. Comprarono un terabyte di hard disk a prezzi stracciati e costruirono il loro computer nella camera di Larry, che divenne il primo data center di Google. La loro fu in parte una scelta ma fu anche in parte dettata dalla necessità di risparmiare. Contrari al modo di pensare dell'epoca i due disegnarono il loro network non basandosi su costosi server, ma su comuni computer che avevano un costo di meno del 10% di un server professionale.

Mentre Larry rivoluzionava il concetto di server, Sergey apriva un ufficio vendite in cerca di acquirenti per la miglior tecnologia di ricerca in circolazione: la loro. Nonostante fossero in pieno dotcom boom, i due, non erano interessati a costruire una società basata sulla tecnologia che stavano sviluppato.

Tra i primi che Sergey contattò ci fu l'amico e fondatore di Yahoo!, David Filo. Filo era d'accordo che la tecnologia fosse valida, ma incoraggiò Larry e Sergey a svilupparla da soli aprendo una società fondata proprio su questa tecnologia. "Quando [la tecnologia] sarà perfezionata e scalabile," disse loro, "allora ne potremo riparlare." Altri dimostrarono ancor meno interesse per Google. L'amministratore di un importante portale disse loro, "Finché siamo all'80% della concorrenza, siamo soddisfatti. I nostri utenti non sono interessati alla ricerca."

Oggi è facile accusare queste persone di miopia, ma bisogna pensare che nel periodo più fervido del dotcom boom, tutti – e intendo davvero tutti – avevano un'idea vincente che richiedeva solo di quel milioncino di Dollari necessario per decollare. Col senno di poi non è sorprendente che in tanti volessero vedere qualcosa in più prima di investire. Se non altro per vedere se c'era davvero un arrosto sotto tutto quel fumo.

Toccati da un angelo

Dato che non riuscivano a suscitare l'interesse dei principali portali del tempo, Larry e Sergey, decisero di provare a fare ciò che tutti consigliavano loro: di provarci da soli. Tutto ciò che serviva loro erano un po' di soldi per potersene andare dai dormitori universitari e per pagare i debiti delle carte di credito che avevano usato per comprare il terabyte di memoria. Scrissero così un business plan, misero in pausa il dottorato e andarono in cerca di un investitore. La prima visita fu a un amico membro della facoltà.

Andy Bechtolsheim, uno dei fondatori della Sun Microsystems, era abituato a una visione a lungo termine, ben lontana dal miope punto di vista di alcuni suoi colleghi. Gli bastò una sola occhiata alla loro demo per capire che Google aveva del potenziale, molto potenziale. Ma anche se avevano stuzzicato il suo interesse, insistette per avere tempo.

Come dice Sergey, “Ci incontrammo una mattina molto presto a Palo Alto sui gradini della casa di un membro della facoltà dell’Università di Stanford. Gli facemmo vedere una demo veloce. Era di fretta, aveva altri appuntamenti, e disse, ‘Invece di discutere tutti i dettagli, perché non vi faccio invece subito un assegno?’ Era intestato a Google Inc. ed era per \$100.000.”

L’investimento creò un piccolo dilemma in quanto non c’era alcuna entità legale che portava il nome di “Google Inc.” e quindi era impossibile depositare l’assegno. Rimase in un cassetto della scrivania di Larry per un paio di settimane mentre lui e Sergey correvano a costituire una società e trovare altri soci tra i familiari, gli amici e altre conoscenze: alla fine riuscirono a portare l’investimento iniziale a un milione di dollari.

Il Googleplex in un garage

Google Inc. aprì le porte il 7 Settembre 1998 a Menlo Park, California. Per la precisione, Larry aprì queste porte con un telecomando: gli uffici erano il garage che un amico subaffittò alla società.

Nonostante l’apparenza spartana, l’ufficio offriva alcuni vantaggi non indifferenti: una lavatrice, un essiccatore e un idromassaggio – penso siano in molti a sognare di lavorare in un idromassaggio. Dava inoltre parcheggio al neo-assunto e primo impiegato della nuova società: Craig Silverstein, direttore della tecnologia.

Già ai tempi del garage-ufficio, Google.com, ancora in fase beta, stava già rispondendo a 10.000 query al giorno. La stampa cominciò a notare questo motore con risultati davvero rilevanti e articoli molto positivi apparvero su USA TODAY e Le Monde. A Dicembre, PC Magazine, nominò Google uno dei 100 principali Siti e Motori di Ricerca del 1998: Google stava cominciando a farsi strada nel mondo.

Di nuovo on the road

Google crebbe in fretta e già nel Febbraio del 1999 gli angusti spazi di Menlo Park dovettero essere abbandonati a favore di un ufficio sulla University Avenue di Palo Alto. Con otto impiegati, lo staff di Google era quasi triplicato e il sito rispondeva ora a quasi 500.000 query al giorno. Anche l’interesse verso la compagnia era cresciuto. Red Hat firmò con Google il suo primo contratto per forniture di servizi di ricerca; attirato in parte dall’impegno di Google di usare server con software open source, in particolare Linux.

Il 7 Giugno, la compagnia annunciò che aveva ottenuto \$25 milioni da due delle principali società di venture capital di Silicon Valley, la Sequoia Capital e la Kleiner Perkins Caufield & Byers. In una

convergenza degli opposti simile a quella che diede vita a Google, le due compagnie, solitamente molto competitive, concordarono sul valore di questo nuovo investimento trovando spazio entrambe nel consiglio d'amministrazione. Mike Moritz di Sequoia e John Doerr di Kleiner Perkins, che insieme avevano aiutato a far crescere Sun Microsystems, Intuit, Amazon, e Yahoo! si unirono a Ram Shriram, amministratore delegato di Jungle, attorno al tavolo da ping pong che serviva formalmente da tavolo del consiglio d'amministrazione.

A breve, personaggi chiave cominciarono a riempire i modesti uffici della compagnia. Omid Kordestani lasciò Netscape per una posizione di vice-presidente del Business Development e Sales, Urs Hölzle fu preso da UC Santa Barbara e assunse il ruolo di vice-presidente dell'Ingegneria (negli Stati Uniti, a chiunque rifiuti di essere considerato un impiegato viene attribuito il titolo di vice-presidente di qualcosa. In una tipica compagnia medio-grande la mano d'opera è costituita da vice-presidenti). Diventò presto ovvio che era necessario più spazio. A un certo punto erano talmente schiacciati che non era possibile alzarsi senza che gli altri non dovessero avvicinare le sedie alle scrivanie.

Google non è più beta

La situazione alla tetris fu alleviata dal trasloco al Googleplex, gli odierni uffici di Google a Mountain View, California. Infilato in un angolo di una palazzina a due piani il kernel (cuore) di Google continuò a crescere attirando a sé nuovo staff, clienti e l'attenzione degli utenti e della stampa – AOL/Netscape scelse Google come tecnologia di ricerca aiutando così a spingere il numero di query giornaliere oltre i 3 milioni. Chiaramente Google si dovette evolvere. Ciò che era nato come un progetto di ricerca universitaria era adesso una compagnia che offriva un servizio molto richiesto.

Il 21 Settembre 1999, fu tolta la scritta “beta” dal sito.

E ancora Google continuava a crescere. Il portale italiano Virgilio divenne cliente così come Virgin Net, la principale guida all'intrattenimento Inglese. La valanga di riconoscimenti che seguirono inclusero un Technical Excellence Award for Innovation in Web Application Development da PC Magazine e l'inclusione in molte liste “The best of..” culminando con la menzione nel “Top Ten Best Cybertech del 1999” del Time magazine.

L'innovazione è nel DNA

Una corporate-culture unica si stava sviluppando al Googleplex: una cultura fatta d'innovazione, di creatività e di sacrificio. Per ottimizzare al massimo la flessibilità degli spazi di lavoro, grosse palle di gomma furono utilizzate come sedie da ufficio estremamente mobili in un ambiente aperto e libero da muri. Mentre i computer erano modernissimi, questi erano posizionati su porte in legno adibite a tavoli

e sorrette da coppie di cavalletti; lampade lava spuntarono come funghi multi-colorati; grossi cani – tra i quali Yoshka, un gigantesco, ma gentilissimo Leonberger – cominciarono a vagare liberi per i corridoi. Addirittura, dopo un rigoroso processo di selezione, fu assunto Charlie Ayers: un cuoco che portò con sé un eclettico repertorio di ricette salutari che aveva sviluppato lavorando per i Grateful Dead.

Interi sezioni del parcheggio venivano transennate due volte la settimana per far spazio alle partite di hockey su pattini, e Larry e Sergey conducevano le settimanali riunioni TGIF (Thank God It's Friday - Grazie a Dio È Venerdì) nell'open space tra le scrivanie che potevano comodamente accomodare i 60 e passa impiegati della compagnia.

L'atmosfera informale incoraggiò lo spirito di gruppo e accelerò lo scambio d'idee. I tecnici di Google apportavano quotidianamente tanti piccoli miglioramenti sia al motore di ricerca, sia attraverso l'aggiunta di nuovi strumenti come la Google Directory (basata sulla Open Directory Project di Netscape) o come la possibilità di effettuare ricerche attraverso apparecchiature wireless. Questa mentalità continuò a crescere dando ben presto a Google una visione globale grazie all'introduzione dell'interfaccia in 10 nuove lingue, per le persone che preferiscono cercare nella propria lingua nativa.

A Maggio del 2000 Google vinse sia i Webby Award che un People's Voice Award per l'innovazione tecnologica. Il discorso di Larry e Sergey fu di sole 5 parole "We love you, Google users!" - "Vi amiamo, utenti di Google!". Il mese successivo Google divenne ufficialmente il motore di ricerca più grande del mondo con oltre un miliardo di documenti indicizzati. Mai nessuno aveva offerto tanti contenuti in un formato in cui era possibile fare ricerche.

Con l'avvento del nuovo millennio ebbe inizio una forte espansione. Furono in tanti a voler utilizzare i servizi di ricerca di Google sui propri portali, ma è grazie all'introduzione di Adwords, il servizio di pubblicità mirata di Google, che la compagnia cominciò a generare profitti. Il rigido programma fiscale aveva permesso a Google di essere una compagnia auto-sufficiente già dalla metà del 2000 e non ebbe quindi bisogno di un secondo round di investimenti.

Il 26 Giugno Google e Yahoo! annunciarono una partnership che non solo ne cementò la reputazione di provider tecnologico, ma la posizionò come serio business che rispondeva a 18 milioni di query giornaliere. Nei mesi a seguire anche i principali portali Cinesi (NetEase) e Giapponesi (BIGLOBE) siglarono accordi con Google.

Per la fine del 2000 Google stava rispondendo a 100 milioni di query al giorno e stava cercando nuovi modi per dare accesso ai propri contenuti dove e quando lo volessero gli utenti. In omaggio alle proprie origini Google tese una mano alla popolazione storicamente affamata d'informazioni: gli studenti e gli insegnanti di scuole e università di tutto il mondo offrendo loro gratuitamente i propri servizi.

Google assume

Partite di hockey nel parcheggio, cuochi, cani che vagano liberi e riunioni attorno a un tavolo da ping pong danno un'idea precisa di come si lavora al Googleplex e del tipo di aria che si respira in quei corridoi. Ma mantenere questo genere di atmosfera non è facile e bisogna sapersi scegliere i nuovi collaboratori.

Google ha sicuramente trovato metodi piuttosto ingegnosi per assicurarsi nuovo personale che condivide il loro modo di lavorare – beh, se non sono ingegnosi loro chi lo è?

A Luglio del 2004, i passeggeri della metropolitana di Cambridge Mass. (Stati Uniti), si sono visti apparire un enorme cartello bianco con questa scritta sopra l'ingresso di alcune stazioni della metropolitana: *{first 10 digit prime in consecutive digits of e}.com*

Si tratta di un'originale offerta di lavoro da parte di Google!

Tradotto, il cartello, dice:

{primo numero primo di 10 cifre consecutive in e}.com

A chi non ha voglia di trovare questo numero posso dire che la risposta è: 7427466391.

Andiamo quindi a dare un'occhiata all'indirizzo <http://www.7427466391.com/> indicato dal cartellone e ci ritroviamo su di un IP di Google che ci chiede di completare la sequenza:

$f(1) = 7182818284$
$f(2) = 8182845904$
$f(3) = 8747135266$
$f(4) = 7427466391$
$f(5) = \underline{\hspace{2cm}}$

Senza dilungarmi in spiegazioni tecniche vi dico che la risposta è: 5966290435.

Seguiamo quindi le indicazioni e andiamo su:

<http://www.linux.org>

Qui usiamo:

Username: bobsyouruncle

Password: 5966290435

E magia magia veniamo portati di nuovo da Google dove ci chiedono di inviare il nostro curriculum con la promessa che questo verrà letto.

<http://www.google.com/labjobs/>

Non ho inviato il mio curriculum, le soluzioni a questi enigmi non le avrei mai trovate da solo. Come

ho fatto allora? Sono bravissimo a cercare con – toh – proprio Google!
Ma non vorreste lavorare per una società del genere voi?

Google si espande

A Febbraio 2001 Google acquista la pietra angolare della cultura Internet, l'archivio di Usenet. Inizia così l'arduo compito di integrare l'enorme volume di dati del più grosso archivio Usenet di Internet in modo che sia possibile effettuare ricerche. In breve Google migliorò il posting, il threading (la concatenazione dei messaggi) e la cancellazione. Il tutto per un totale di oltre 500 milioni di messaggi scambiati, negli anni, su Usenet.

Con l'aumento del suo pubblico globale, emersero gli schemi sepolti nella marea di query dando un'idea di ciò che passava per la mente collettiva. Scavando tra il marasma di keyword, Google catturò i trend delle ricerche e li istituzionalizzò nel Google Zeitgeist, una finestra in tempo reale sulla coscienza collettiva. Il Google Zeitgeist mostra le stelle nascenti e cadenti nel firmamento delle ricerche; nomi e luoghi che passano dall'oscurità al centro del palco per poi scomparire nuovamente. Come un Indice Standard & Poor della cultura popolare, il Google Zeitgeist tiene traccia delle nostre ossessioni e della fama passeggera.

Le capacità di ricerca di Google si moltiplicarono, l'appoggio economico divenne ancor più solido. Già dall'inizio del quarto quarto del 2001, Google avrebbe annunciato di aver trovato ciò che aveva eluso molte società on-line: i profitti.

Nuovo capo, nuova direzione

A Marzo del 2001 Eric Schmidt amministratore delegato di Novell e ex-CTO (Chief Technology Officer) di Sun Microsystems divenne capo del consiglio d'amministrazione di Google. Nei pochi mesi che separarono Eric Schmidt dalla carica di Amministratore Delegato, Google ampliò il numero di lingue a 40, vinse un altro Webby Award, raddoppiò il numero di pagine indicizzate e aggiunse alla sua gamma un numero di protocolli wireless da cui era possibile effettuare ricerche. Nel frattempo Larry Page e Sergey Brin diventarono rispettivamente presidente dei prodotti e presidente della tecnologia.

Eric Schmidt fu la chiave di volta nello sviluppo di Google che fino ad allora stentava a stabilizzare i profitti. Eric chiamò a sé George Reyes, ex-dirigente della Sun Microsystems e gli affidò le finanze della società. Questo, più il fatto di dare una chiara direzione a Google definendola per quello che era ed è – la più grande concessionaria pubblicitaria on-line – diede nuova spinta e nuovi capitali. Google diventò, con lui, maggiorenne.

Informazioni senza barriere

Il cerchio delle amicizie di Google continuò a crescere. Un accordo con Lycos Corea portò Google all'attenzione di un nuovo gruppo di utenti Internet asiatici. A Ottobre, una partnership con Universo On-Line (UOL) fece di Google il principale motore di ricerca in Sud America. Nuovi uffici vendita furono aperti ad Amburgo e a Tokyo per soddisfare la crescente richiesta dei programmi pubblicitari di Google. L'appeal senza frontiere di Google era evidente anche nell'evolversi dell'interfaccia: gli utenti potevano ora limitare i risultati delle ricerche a siti scritti in Arabo, in Turco o una delle altre 40 lingue disponibili.

Nel frattempo il motore di ricerca maturò, imparando a leggere vari tipi di documenti. La ricerca per tipo di file aggiunse una dozzina di formati alla lista supportata da Google. A Dicembre, Google Ricerca Immagini, lanciata quell'estate con 250 milioni di immagini, uscì dalla fase beta con una serie di funzioni avanzate e un indice espanso. Lo shopping on-line fece un balzo in avanti con il lancio di Google Catalog Search, dando la possibilità agli utenti di sfogliare oltre 1.100 cataloghi fino ad allora disponibili solo su carta.

Dicembre portò una nuova pietra miliare: l'indice di Google raggiunse i 3 miliardi di documenti ricercabili, un altro balzo avanti nella missione di Google di rendere l'informazione sempre più accessibile. L'anno si concluse con la retrospettiva di schemi, trend e keyword più ricercate nel Google Zeitgeist del 2001.

La storia di Google viene scritta tutti i giorni e nuovi capitoli vengono pensati quotidianamente dagli ingegneri e dai programmatori. Il resto lo potete trovare sulle pagine di Google da cui sono state tratte molte di queste informazioni.

<http://www.google.com/corporate/history.html>

7. Il futuro

Poche società suscitano l'interesse, e l'invidia, che riesce a suscitare Google in ugual misura tra gli utenti e gli aspiranti competitori.

Google ha ottenuto un successo mai raggiunto da una compagnia Internet e raramente superato dalle società tradizionali. In qualche modo è riuscito a raggiungere il più alto grado di efficienza tecnica, è facilmente utilizzabile da chiunque, è considerato un innovatore ed ispira la massima fiducia.

Google funziona, Google è gratuito, Google è divertente e Google ci dà sempre qualcosa di nuovo e di utile a cui non avevamo pensato. Se questo non bastasse, Google è adesso diventato anche un investimento economico per alcuni di noi. Date le premesse, è ovvio che l'interesse sul futuro di Google sia un argomento tanto sentito.

Quali saranno le sfide di Google?

L'informazione veicolata attraverso il web sta aumentando esponenzialmente e il rapporto segnale/rumore, che è sempre stato basso su Internet, sta diminuendo ulteriormente. La fortuna di Google è stata di saper scremare le informazioni rilevanti dal mero rumore di fondo con un'efficienza eccezionale. Questo aumento quantitativo dell'informazione disponibile, insieme ai tentativi volontari di influenzare il ranking delle pagine su Google, esercitano una forte pressione negativa sul mantenimento degli standard qualitativi.

Mentre la competizione (Microsoft in testa), cerca di sviluppare prodotti competitivi, Google fa ricerche nel campo della programmazione con linguaggi naturali (come tradurre una domanda scritta in un linguaggio umano in una query comprensibile a un computer), degli algoritmi genetici (usare algoritmi che "imparano" dai risultati precedenti; come gli organismi viventi), delle mappe concettuali (creare dei collegamenti tra diverse aree del sapere in modo da poter proporre solo risultati rilevanti all'utente. Una query per "Franz Ferdinand" darebbe come risultati documenti riguardanti il Duca Austriaco, elemento scatenante della prima Guerra Mondiale, e informazioni su di una band di rock alternativo. Una mappa concettuale sarebbe in grado di separare i due argomenti distinti e presentare solo quello più rilevante all'utente in base alle ricerche precedenti) e dell'intelligenza artificiale - sì, avete capito bene: Google sta facendo delle ricerche nel campo dell'AI applicato alla ricerca su basi dati.

Questa mole di informazioni senza paragoni (mai in passato tanta informazione è stata disponibile in maniera immediata a chiunque abbia quantomeno un PC e un modem) rende l'attività di fare query su Google in maniera mirata una vera e propria scienza. Si aprono così spazi di business per chi fornirà servizi di estrazione qualificata di risultati da Google; questi "compilatori di query di alto livello" potrebbero ricevere domande in linguaggio comune (linguaggio umano), e tradurle in query googliane. Google rimarrebbe come servizio di "basso livello", riservato a tecnici e il suo marchio potrebbe

indebolirsi agli occhi del grande pubblico. Per fare un paragone molto approssimativo: da quando Access ha reso possibile creare tabelle o fare query facendo drag 'n drop, chi ha più bisogno di conoscere SQL? Da quando esistono linguaggi di livello alto-altissimo, chi ha bisogno di conoscere l'assembly?

Perché non open-source?

Una corrente di pensiero vede lo sviluppo di un motore di ricerca open-source che si contrapponga alla posizione eccessivamente dominante di Google. Posizione che, i più paranoici, temono possa essere sfruttata contro gli utenti stessi. A detta di queste persone, esempi come la censura Cinese, sarebbero un chiaro monito e un incentivo a sviluppare un motore di ricerca open-source che risponda al requisito di trasparenza.

La genesi, però, dei prodotti che il movimento open-source crea e gestisce è riconducibile – nella maggior parte dei casi – a uno dei seguenti filoni:

Ideologico-sociale: un prodotto nasce open-source quando il suo analogo commerciale diventa una minaccia per la libera circolazione delle idee, oppure una barriera all'espansione del fenomeno Internet per motivi sostanzialmente economici. È il caso di Linux o dei *BSD, per citare i casi più famosi.

Ideologico-individuale: un progetto ha natura di sfida intellettuale, senza avere le caratteristiche di opportunità commerciale. La mentalità hacker (nel senso originale del termine) è il primo stimolo per lo sviluppo di questo tipo di prodotti.

Opportunistico-individuale: non esiste (per convenienza economica, per estrema specificità del prodotto a contesti molto piccoli o molto poco redditizi) un analogo prodotto commerciale.

Opportunistico-collettivo: un prodotto commerciale cade in disuso, e la comunità open-source eredita una base di codice legacy che ritiene conveniente sviluppare (Mozilla, OpenOffice, solo per citare i due esempi più famosi).

La categoria dei motori di ricerca non ha nessuna di queste caratteristiche. Difatti:

Google non ha mai rappresentato un freno al libero scambio delle idee; anzi, Google rappresenta il canale più efficiente ed efficace per localizzare le idee all'interno della caotica e dispersiva natura del web.

Un motore di ricerca sfrutta algoritmi in larga misura già noti, più o meno ottimizzabili/ottimizzati; prova ne sia che ci sono librerie open-source per l'implementazione di un search-engine (<http://www.htdig.org>, ad esempio). Ciò che fa la differenza è la qualità e l'estensione dei dati da indicizzare, e la velocità di ricerca sulla stessa. È chiaro che questo acquisire e mantenere hardware di eccellente qualità è un problema di risorse finanziarie, non intellettuali – e il movimento open-source ha

soprattutto di queste ultime;
Google esiste, e funziona benone.

Nessun motore in disuso può fornire spunto per l'open-source movement, perché il relativo codice legacy sarebbe poco e poco interessante (un'eredità realmente troppo scarsa qualitativamente e quantitativamente).

Google pubblica?

Un'altra teoria vede la crescente importanza dei motori di ricerca come vero e unico strumento per giungere all'informazione. Considerato il valore che la nostra società pone nelle informazioni, garantire che i meccanismi della ricerca siano trasparenti e non sotto l'influenza di alcun individuo o gruppo, diventa un obbligo pubblico. In molti vedono un futuro in cui la ricerca su Internet - e quindi l'informazione - sarà regolamentata alla stregua dell'energia elettrica e dell'acqua potabile.

Prefigurare un futuro in cui Google possa venire affiancato o sostituito da un ente avente natura pubblica appare poco credibile, ed anche in questo caso per più di un motivo:

1. La natura stessa di Internet non rende probabile una tale ipotesi. Internet è il risultato di una serie di accordi cooperativi tra entità di diversa natura (enti pubblici o semipubblici, enti privati, società commerciali, singoli privati...) per permettere la cooperazione tra hardware di proprietà di queste entità. A livello di protocolli, questo si traduce nel rispetto di standard comuni. A livello di interconnessione e trasporto dell'informazione, questo è il frutto di singoli accordi bi-o-multilaterali tra i singoli agenti. Ognuno di questi negoziati ha caratteristiche proprie: non ci sono regole generali o meccanismi globali. Gli unici enti che possono influenzare la comunità degli agenti sono quelli necessari al funzionamento e al coordinamento del sistema (RIRs, IANA...) o coloro i quali sono deputati alla produzione, alla formalizzazione e alla pubblicazione degli standard (di fatto le RFC).

Tuttavia, questi organismi hanno un mero potere tecnico o consultivo, e le RFC hanno la forza di meri suggerimenti, non di leggi. È la stessa natura cooperativa che dà loro forza: se un player ignora continuamente gli standard, verrà gradualmente isolato (sia per impossibilità di comunicare dovuta alla violazione degli standard, sia per volontà dei vicini di rete, che rifiutano un vicino maleducato che causa loro danni o difficoltà tecniche, con risvolti economici non trascurabili).

Siamo di fronte ad una Internet che si auto-regola, creando gli anticorpi che espellono elementi estranei alla sua natura cooperativa. In questo contesto di autorità con solo potere di moral-suasion, e di singoli player dotati ciascuno di poca forza relativa ma di un buon grado di autonomia, pare improbabile l'affermarsi e il consolidarsi di figure pubbliche in grado di acquisire autorevolezza non grazie ai risultati ottenuti, ma solo per l'autorità derivante dalla loro natura di emanazione statale.

2. Internet non è limitata ad uno stato, ma ha carattere necessariamente internazionale. Nessuno stato è in grado di far pesare il suo ruolo di entità pubblica come contratto sociale tra i suoi cittadini, tutelante interessi comuni, in una tale realtà, proprio perché la sovranità di uno stato - e gli interessi da esso

tutelati – si fermano ai suoi confini.

Un servizio pubblico internazionale dovrebbe essere svolto da un ente sovranazionale. Ora, stando alla mera osservazione dell'attuale stato di salute degli enti sovranazionali non economici (ONU, FAO, UNESCO...), pare potersi affermare che essi sono scarsamente dotati di poteri (sanzionatori, politici) e di mezzi finanziari, e che la loro direzione sia più il frutto di compromessi degli stati associati che di una volontà unitaria a perseguire un obiettivo. Non si vede in quale scenario la costituzione di un'istituzione sovranazionale per la gestione di un motore di ricerca dovrebbe rappresentare un'eccezione.

3 Attualmente, con sfumature, lo stato è sempre più visto come entità che fornisce servizi e prodotti in settori troppo delicati per essere lasciati al libero mercato (giustizia, difesa, pubblica sicurezza) o che verrebbero allocati da quest'ultimo in maniera giudicabile non socialmente equa (istruzione, sanità). La visione dell'intervento dello Stato in economia è generalmente residuale: al di fuori di queste specifiche casistiche, tutto il resto va lasciato all'intervento dei privati.

Al di là di considerazioni personali, questa logica esclude che l'entità-stato possa investire in un settore in cui i privati hanno dimostrato un eccellente grado di efficienza ed efficacia, come quello dei search-engine.

4 Un controllo statale su una tale base di dati, così vicina alla circolazione e al reperimento dell'informazione, configurerebbe concreti rischi di manipolazione e censura per fini politici.

In Italia lamentiamo che il controllo dei media sia di fatto in mano al potere politico, perché ciò rappresenta, o potrebbe rappresentare, una possibilità di influenza nella formazione dell'opinione e del consenso pubblico. Un Google statale sarebbe tutto ciò, elevato all'ennesima potenza.

Competitor

In quello che è il suo core-business, Google, al momento, non ne ha. Il gap con gli altri motori di ricerca è enorme, e si allarga giorno dopo giorno. La realtà è che la necessità di una vasta ed affidabile base di dati viene a costituire una seria barriera all'entrata – prima di potersi mettere in seria competizione con Google è necessario un periodo di accumulazione dei dati, e di conquista di una quota di mercato che venga convinta da un servizio più veloce ed affidabile di Google – difficile, costoso, altamente rischioso, non remunerativo per il periodo di start-up, che può pure essere lungo; viene da dire che non pare esserci spazio per concorrenti.

A questo contribuisce anche la diffusione del marchio Google: gli anglofoni hanno già coniato il verbo to google con il significato di cercare sul web.

Nonostante ciò il mercato è affollato di player minori che cercano di trovare, nelle nuove idee, un loro punto di forza. Purtroppo resta, però, valido quanto detto sopra: è necessaria una seria e competitiva base dati su cui effettuare le ricerche: non basta essere i primi a introdurre tecniche di cluster analysis o delle mappe concettuali visive per guadagnarsi una fetta di mercato.

Anche Yahoo!, dopo essere stato, la prima directory importante su Internet, dopo essere stato un motore di ricerca molto apprezzato e dopo aver usato Google come motore di ricerca adesso sta cercando di rientrare nel mercato sviluppando una propria tecnologia.

Ma la vera sfida viene da un nuovo player in questo mercato: Microsoft.

Tipico dei grandi player, e degli incumbent, Microsoft è stata alla finestra per un lunghissimo periodo prima di entrare. Ha avuto il tempo di studiare il mercato, di valutare le opportunità e di determinare quali siano i requisiti necessari per avere successo.

Il periodo di gestazione è stato lungo – proprio per costruire questa base dati necessaria – e i primi test non sono stati entusiasmanti. Ma a Microsoft non mancano di ingegneri brillanti e i soldi per lo sviluppo, quindi non ho dubbi che il prodotto sarà quantomeno buono.

A sfavore della casa di Redmond possiamo dire che il brand di Google è, al momento, imbattibile, mentre il brand Microsoft gode di una fama pessima tra gli addetti ai lavori. Pochi degli utenti più smaliziati effettueranno uno switch nelle loro preferenze riguardo la ricerca.

A favore di Microsoft dobbiamo invece indicare la possibilità di integrare la ricerca con il suo sistema operativo Windows tanto da sfumare i confini tra la ricerca tra i propri file e la ricerca su Internet. Google sta cercando di anticipare questa mossa tramite il suo Google Desktop, ma Microsoft ha un netto vantaggio in questo campo.

Ricordiamo inoltre cos'è accaduto quando Microsoft entrò nel mercato dei browser.

Nel 1995, Netscape era il browser mentre Microsoft aveva appena acquistato i diritti per utilizzare il codice della SpyGlass, l'unico vero browser concorrente di Netscape. In teoria la Microsoft avrebbe condiviso questa licenza con altre 120 società; in realtà lo strapotere della Microsoft ottenne che il codice divenne, di fatto, solo suo.

Netscape aveva il 90% del mercato, godeva di un'ottima opinione pubblica, ma era economicamente debole. I suoi incassi si basavano su di un solo prodotto e il totale di questi non aveva mai superato gli interessi che Microsoft prendeva sul pronto cassa.

La storia ci insegna che Microsoft ha vinto, no, ha stravinto la battaglia contro Netscape. Come?

Principalmente:

1. Bundling: Microsoft includeva (e include ancora oggi) Internet Explorer con ogni copia di Windows. La pratica è stata dichiarata illegale più di una volta, ma le multe che paga sono inferiori a quanto ne ricava. Questa semplice strategia fa sì che ogni utente abbia, volente o nolente, una copia di Internet Explorer sul proprio PC. Perché scaricare qualcosa quando lo si ha comodamente e gratuitamente?¹

¹ Un passo fondamentale è stato mischiare la shell di Windows con il browser Internet, sia perché ha reso indistinguibile il browser dal sistema operativo (anche concettualmente: Internet Explorer 4 non poteva essere disinstallato), sia perché ha fatto sì che il browser risultasse un sottoprodotto nella produzione del sistema operativo, e quindi con un costo marginale trascurabile. Dalla versione 4 in poi, IE non è più un browser: è un contenitore di ActiveX, che si preoccupano a loro volta di fare il rendering del documento. Ciò che fa parsing e visualizzazione delle pagine HTML non è iexplore.exe, è mshtml.dll

2. Modello di business Netscape regalava il browser e vendeva software per server. Microsoft lo copiò e cominciò a includere Microsoft Internet Information Server insieme ad ogni copia di Windows. In altre parole regalò ciò che Netscape vendeva per vivere.

3 Microsoft fece leva sulla **posizione dominante** per impedire ai costruttori di PC di includere Netscape nei loro prodotti. Anche qui la pratica è stata dichiarata illegale più di una volta, ma il risultato è stata la quasi totale eliminazione della concorrenza.

Dico “quasi totale”, perché è proprio dalle ceneri di Netscape che è nato FireFox, il browser che sta lentamente guadagnando terreno contro il dominio di Internet Explorer e su cui molti vorrebbero scommettere.²

4. Branding: Microsoft diede la possibilità alle compagnie di creare versioni brandizzate di Internet Explorer con il proprio logo. Quando compagnie come AOL approfittarono dell'occasione, milioni di Americani si ritrovarono improvvisamente Internet Explorer proposto da ogni parte.

Microsoft è in possesso di armi molto potenti mentre Google è in una posizione di assoluto dominio. La guerra tra i due sarà dura e il risultato non è scontato. Tutto ciò sarà a nostro vantaggio: mentre i due giganti si sfidano a colpi di innovazioni, di miglioramenti e di nuove idee, noi non potremo che beneficiare da questa situazione di libero mercato.

Al Googleplex i cervelli stanno lavorando a pieno ritmo e, al contrario di molti prima di loro, pare abbiano imparato dalla storia. Da quando è diventato ovvio che il mercato dei motori di ricerca è ricco e che Microsoft sarebbe diventato un concorrente gli ingegneri Californiani di Google si sono spicciati

– che non è parte di Internet Explorer, è parte del sistema operativo.

Microsoft (come Netscape, ma anche come WordPerfect o Lotus per ciò che riguarda l'office automation) ha anche approfittato sostanzialmente del fatto che, essendo anche il produttore del sistema operativo, ha accesso a informazioni che il resto del mercato non ha. Ci sono funzioni non documentate in Windows sfruttate da programmi Microsoft, che risultano inaccessibili (sia proprio perché non documentate, sia perché il codice verifica effettivamente che sia un prodotto Microsoft ad utilizzarle) alle altre società. Questo garantisce a Microsoft prodotti meglio integrati e più efficienti.

Dal momento in cui Microsoft entra in un mercato e vi si afferma in maniera piuttosto solida, il suo schema comportamentale prevede la rilettura degli standard (estendendoli in maniera proprietaria, oppure semplicemente interpretandoli in maniera diversa). Ciò rende di fatto incompatibili gli altri prodotti, cacciandoli gradualmente dal mercato. Un effetto collaterale è che nel fare questo Microsoft trascura più o meno volontariamente che alcune di queste estensioni/interpretazioni possono portare rischi di sicurezza, e che gli standard non le prevedevano proprio per tutelare gli utenti.

² Voci di corridoio sostengono che Google stia contribuendo moltissimo allo sviluppo di FireFox con l'intento ultimo di creare una versione brandizzata FireFox. Versione che aiuti a sfumare i confini tra la ricerca tra i propri file (Google Desktop), la ricerca su Internet (Google vero e proprio) e la posta (GMail). È ironico che questa tattica usata con tanto successo dalla Microsoft stia per esserle usata contro e che per farlo sia utilizzato un browser che ha fatto della brandizzazione (vedi punto 4. sopra) uno dei punti di forza!

nel diversificare le proprie attività, creare nuove fonti di business e cercare in tutti i modi di non trovarsi in uno scontro frontale con Microsoft. Ho fiducia nel loro talento, ho meno fiducia nelle loro finanze!

La reale minaccia al business di Google viene da coloro che cercano di fornire un punto di accesso unico ai servizi della Rete, un portale nel senso pieno del termine.

In questi termini il (nuovo/inesperto) utente userà Internet come un'estensione del suo portale preferito, legandosi fortemente al marchio della società che glielo fornisce, che finirà per essere identificata con la Rete stessa (AOL è Internet, Microsoft produce Internet...). Questo processo è facilitato dalla struttura non lineare del web.

Quasi tutti gli strumenti di informazione hanno una natura seriale: un libro si legge dalla prima pagina all'ultima, un programma TV si segue dal principio alla fine, e così via: il percorso di fruizione è uno solo, con un inizio, un durante ed una fine ben determinati.

In Internet l'informazione è presente tutta nello stesso momento, e ciascuno si trova a dover creare il percorso logico che gli è più consono, per perseguire gli obiettivi che si è posto.

È un grosso cambiamento, rispetto al modello del paragrafo precedente ed è in questo senso si inserisce Google (e i motori di ricerca): esso è uno strumento per crearsi il percorso mentale più corretto per giungere all'informazione richiesta; è fruibile da chi è cosciente del caos esistente oltre il proprio router (o il proprio modem), ed ha imparato a gestirlo.

I portali, invece, serializzano, riportando il nuovo utente su percorsi mentali precostituiti, in una realtà più simile a quella dei media tradizionali.

In sintesi, Google è uno strumento potente per gestire in autonomia una realtà complessa, mentre i portali sono strumenti mirati alla facilità d'uso e tesi alla semplificazione (ma anche all'impoverimento, effetto collaterale inevitabile) della medesima realtà.

Allo stato attuale, i grandi portali non rappresentano una vera sfida concorrenziale per Google, anche perché il mercato è tale da garantire spazi ad entrambi gli approcci.

L'unico esempio concreto di motore di ricerca che perde la leadership e svanisce nel nulla è quello di <http://www.altavista.com>.

Prima di Google, questo era il search-engine giudicato migliore, in termini quantitativi e qualitativi. Una serie di scelte sciagurate – una cattiva gestione della qualità dei dati che ha reso i risultati via via meno affidabili, un'eccessiva inclinazione a cercare risultati commerciali che ha fatto sì che sia la grafica del sito, sia gli stessi risultati fossero profondamente influenzati da elementi pubblicitari – ne ha causato il progressivo abbandono proprio a favore di Google.

In questi termini, Altavista infondo mostra come il primo pericolo di Google non siano i competitor, ma siano scelte miopi del management.

8. Google Matematicamente

Nel progettare Google, Sergey Brin e Larry Page, hanno dovuto risolvere due importanti problemi: come costruire un'architettura veloce e scalabile e in che ordine presentare i risultati ottenuti.

In questo capitolo approfondiremo come Google decide quali pagine sono più rilevanti per la nostra ricerca e vanno quindi posizionate in cima ai risultati.

Nel loro primo articolo³, Brin e Page, citano tre metodi per giudicare la rilevanza di una pagina:

PageRank [2.1]

Il testo dei link che puntano a quella pagina [2.2]

Parole scritte usando una font più grossa (come ad esempio i titoli) devono essere più rilevanti all'interno del contesto della pagina, e quindi più rilevanti [2.3]

Come esattamente Google decida quali sono le pagine più rilevanti è uno dei segreti della compagnia, ma gli esperimenti di alcuni esperti, le congetture e la scarsa documentazione ci lasciano intuire che la seguente, seppure incompleta, è una spiegazione di alcuni dei meccanismi interni che regolano questo motore di ricerca.

È importante notare che i parametri usati da Google per determinare questa rilevanza non sono fissi, ma vengono aggiustati in continuazione per migliorarne i risultati e per stare al passo con quelle persone che vogliono sfruttare debolezze nell'algoritmo corrente.

Due dei tre parametri citati qui sopra sono piuttosto intuitivi; il terzo invece, il PageRank, fu una novità introdotta da Google e merita un'attenzione maggiore.

È necessario chiarire che il PageRank è quel valore tra 0 e 10 che vediamo rappresentato da una barra verde se abbiamo installata la Google Toolbar [<http://toolbar.google.com/intl/it/>].

Detto molto semplicemente, il PageRank è il voto che il resto del web esprime attraverso i siti che decide di linkare. Un link a una determinata pagina indica un voto. Nessun link non rappresenta però un voto a sfavore, ma semplicemente un'astensione.

Cominciamo citando il paper originale di Brin e Page:

Le citazioni delle pubblicazioni accademiche sono state applicate al web, principalmente contando le citazioni, o i backlink, a una determinata pagina. Questo fornisce un'approssimazione all'importanza della pagina o alla sua qualità. Il PageRank estende quest'idea contando il peso relativo di ogni link e normalizzando il numero di link in una pagina. Il PageRank è definito come segue:

Consideriamo che la pagina A abbia pagine $T_1 \dots T_n$ che la linkano (i.e. sono citazioni). Il parametro d è una costante di smorzamento che può essere fissata tra 0 e 1. Noi di solito la impostiamo a 0,85. Più

³ <http://www-db.stanford.edu/~backrub/google.html>

dettagli su d nella prossima sezione. Inoltre $C(A)$ è definito come il numero di link che escono dalla pagina A . Il PageRank della pagina A è determinato come segue:

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

Notare come il PageRank è la distribuzione della probabilità, su tutte le pagine e che la somma del PageRank, di tutte le pagine, sia uno.

Il PageRank o $PR(A)$ può essere calcolato usando un semplice algoritmo iterativo, e corrisponde all'autovettore principale di una matrice di link del web normalizzata. Inoltre, il PageRank per 26 milioni di pagine può essere calcolato in poche ore con un computer di media potenza. Ci sono molti dettagli che sono oltre gli obbiettivi di questo paper.

Giustificazione Intuitiva

Il PageRank può essere considerato come un modello di comportamento dell'utente. Immaginiamo un "utente casuale" a cui è data una pagina casuale e che continua a cliccare sui link senza mai premere "back" ma che prima o poi si stufa e ricomincia da un'altra pagina casuale. La probabilità che un utente casuale visiti una pagina è data dal suo PageRank. E, il fattore smorzante d è la probabilità che a ogni pagina "l'utente casuale" si stufi e richieda una nuova pagina. Un'importante variante è dunque l'aggiunta del fattore smorzante d a una singola pagina, o gruppo di pagine. Questo permette una personalizzazione che rende quasi impossibile ingannare il sistema per ottenere un ranking superiore.

Un'altra giustificazione intuitiva è che una pagina può avere un PageRank alto se ci sono tante pagine che la linkano, o se ci sono pagine che la linkano e hanno a loro volta un PageRank alto. Intuitivamente, vale la pena consultare pagine che sono citate in molti posti. Inoltre, anche pagine che sono citate una sola volta, ma da siti come Yahoo! vale generalmente la pena di essere visitate. Se una pagina fosse di bassa qualità è probabile che Yahoo! non le dedicherebbe un link dalla sua homepage. Il PageRank gestisce entrambi i casi e tutti i casi tra questi due ridistribuendo continuamente i relativi pesi delle singole pagine.

Bene, la formula ha una sua giustificazione logica, ma come facciamo a calcolare il Pagerank di una pagina?

Cominciamo a spezzettare la formula in elementi più semplici da gestire:

$PR(Tn)$ - Ogni pagina indicizzata da Google ha un PageRank associato: $PR(T1)$ per la prima pagina fino a $PR(Tn)$ per l'ultima.

$C(Tn)$ - Ogni pagina suddivide equamente il suo voto tra tutti i link in uscita. Il numero di link in uscita di una pagina è $C(T1)$, ... $C(Tn)$.

$PR(Tn)/C(Tn)$ - Il peso del voto che una pagina (Pagina A) riceve da un link di un'altra pagina (Pagina n).

$d(\dots)$ - Tutte queste frazioni di voto sono sommate, ma per evitare che le altre pagine abbiano

un'influenza troppo grande il voto totale è smorzato moltiplicandolo per 0,85 (il fattore d).

$(1 - d)$ - Ri-aggiunge quel 0,15 perso dal fattore di smorzamento. Serve a dare un peso iniziale a una pagina anche se nessuna altra pagina esterna la linka e a fare sì che la media della somma di tutti i PageRank sia uno.

Com'è calcolato il PageRank?

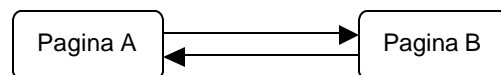
Il PageRank di una pagina dipende dal PageRank di tutte le pagine che puntano a questa pagina. Ma non possiamo sapere quale sia il PR di queste pagine finché non calcoliamo il PR delle pagine che puntano a loro, e via dicendo... Se consideriamo che i link delle pagine possono creare dei cerchi il calcolo sembra impossibile!

In realtà il problema non è così difficile. Riprendiamo cosa c'era scritto sul paper di Google:

Il PageRank o $PR(A)$ può essere calcolato usando un semplice algoritmo iterativo, e corrisponde all'autovettore principale di una matrice di link del web normalizzata.

Questo vuol dire che possiamo calcolare il PR di una pagina senza sapere il PR finale delle altre pagine. Può sembrare strano, ma siccome l'algoritmo è iterativo, a ogni computazione ci avviciniamo sempre di più al valore finale. Dobbiamo solo ricordarci l'ultimo valore calcolato e ripetere la computazione molte volte finché i numeri smettono di variare in modo rilevante.

Prendiamo come esempio due pagine che si linkano a vicenda:



Ognuna ha un solo link in uscita ($C(A) = 1$ and $C(B) = 1$).

Non conosciamo il loro PR quindi proviamo con un numero a caso, diciamo 1 e facciamo qualche calcolo:

```
d = 0.85
PR(A) = (1 - d) + d(PR(B)/1)
PR(B) = (1 - d) + d(PR(A)/1)
```

cioè:

```
PR(A) = 0,15 + 0,85 * 1
        = 1
PR(B) = 0,15 + 0,85 * 1
        = 1
```

Mmmh... Qui, o siamo stati molto fortunati oppure c'è qualcosa di sbagliato. Proviamo con un altro numero di partenza, diciamo 0,5, e vediamo cosa succede:

Prima iterazione:

$$\begin{aligned}PR(A) &= 0,15 + 0,85 * 0 \\ &= 0,575 \\ PR(B) &= 0,15 + 0,85 * 0,575 \\ &= 0,63875\end{aligned}$$

Seconda iterazione:

$$\begin{aligned}PR(A) &= 0,15 + 0,85 * 0,63875 \\ &= 0,69294 \\ PR(B) &= 0,15 + 0,85 * 0,69294 \\ &= 0,738999\end{aligned}$$

Terza iterazione:

$$\begin{aligned}PR(A) &= 0,15 + 0,85 * 0,738999 \\ &= 0,778149 \\ PR(B) &= 0,15 + 0,85 * 0,778149 \\ &= 0,81143\end{aligned}$$

Sembra che ci stiamo muovendo nella direzione giusta, ma ci fermeremo a 1? E cosa sarebbe successo se avessimo scelto un numero più grande di 1?

Proviamo! Scegliamo come partenza il 10

Prima iterazione:

$$\begin{aligned}PR(A) &= 0,15 + 0,85 * 10 \\ &= 8,65 \\ PR(B) &= 0,15 + 0,85 * 8,65 \\ &= 7,5025\end{aligned}$$

Seconda iterazione:

$$\begin{aligned}PR(A) &= 0,15 + 0,85 * 7,5025 \\ &= 6,5272 \\ PR(B) &= 0,15 + 0,85 * 6,5272 \\ &= 5,6982\end{aligned}$$

Terza iterazione:

$$\begin{aligned}PR(A) &= 0,15 + 0,85 * 5,6982 \\ &= 4,9932 \\ PR(B) &= 0,15 + 0,85 * 4,9932 \\ &= 4,3942\end{aligned}$$

I numeri tendono decisamente a 1! Se disegnassimo un grafico lo vedremmo chiaramente.

Non è quindi importante il valore di partenza, l'algoritmo ci porterà inevitabilmente al risultato stabile.

Ottenere il risultato più velocemente

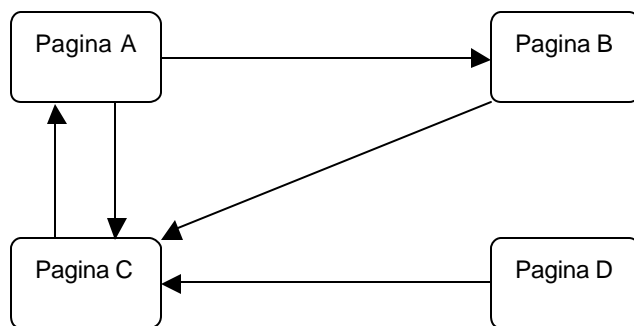
Quante volte è necessario ripetere il calcolo per i grandi network? La domanda è difficile, ma per un network grande quanto il web ci possono volere milioni di iterazioni. Il fattore smorzante ha un'influenza importante. Se è troppo alto servirà troppo tempo per far convergere i numeri. Se è troppo basso si supera ripetutamente la convergenza, come un pendolo impazzito.

Scegliere l'ordine in cui si eseguono i calcoli è anche importante. Si giungerà alla convergenza in ogni caso, ma se si segue un certo ordine ci si arriverà più in fretta.

Sono sicuro che ci siano varie tesi di Laurea che spiegano come rendere questo calcolo il più efficiente possibile, ma per l'esempio usato qui sotto una decina di iterazioni sono state sufficienti.

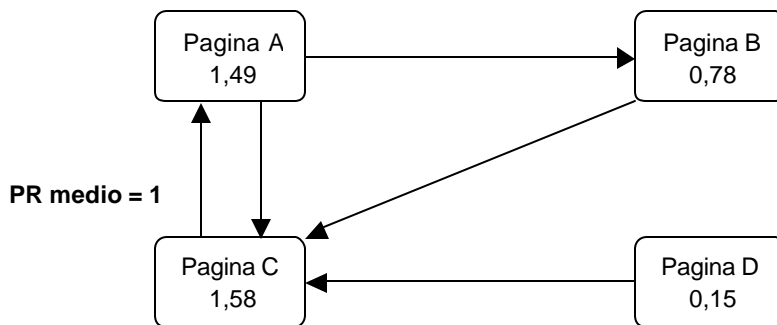
Esempio

Consideriamo le quattro pagine indicate qui sotto:



Il PageRank è stato calcolato usando gli script indicati nell'appendice.

Già dopo 10 iterazioni i valori convergono al PageRank indicato qui sotto



È interessante notare che nonostante la pagina D non abbia link in entrata ha comunque un PR di 0,15. Questo, anche se sembra sbagliato a primo acchito, diventa ovvio quando riguardiamo la formula:

$$\begin{aligned} PR(A) &= (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn)) \\ PR(A) &= (1-d) + d * (0) \\ &= 0.15 \end{aligned}$$

In teoria ogni pagina ha un PR minimo di 0,15. Secondo alcune voci, però, sembra che Google cancelli quelle pagine che non hanno link in entrata.

In questa sezione abbiamo mostrato due esempi semplici per mostrare come Google calcola il PageRank. Usando il codice fornito nell'appendice è possibile per un webmaster costruire complesse architetture per capire dove si concentrerà il PageRank di un sito e in questo modo ridirigerlo dove gli serve o evitare che questo venga sprecato nei link esterni al sito.

Conclusioni

Google è molto di più di quanto scritto in questo volume; in realtà Google è tutt'altro. La ricerca, Usenet, i laboratori, le news e tutto quanto vedete sul sito sono soltanto una facciata. Un'apparenza che i suoi competitori cercano continuamente di afferrare e di emulare. Google re-inventa la ricerca e Altavista cerca di copiarne la semplicità, Yahoo acquista Inktomi e anche Microsoft si butta nell'arena. Google offre una casella di e-mail da un GigaByte (sì, 1GB, che equivale a quanto può contenere un CD e mezzo, cioè più quanto occupa un film scaricato da Internet) e tutti si chiedono come farà a offrire tanto gratuitamente. Come farà ad avere un ritorno economico? E la privacy, come la mettiamo con la privacy? Una volta trovata la risposta a queste domande (risposte ovvie, aggiungerei), il mercato delle caselle da 1GB si riempie immediatamente di piccoli e grandi provider.

Ma ancora nessuna ha capito. Nessuno vede che non possono competere con Google perché Google non ha rivali. Google crea i mercati; gli altri li riempiono con offerte minori.

Ma Google non è un player in nessuno di questi mercati, li crea sob per generare profitti – utili anche ad una compagnia per cui i profitti non sono l'unico obbiettivo – e perché si divertono. Sì, ci sono 600 ingegneri in California che adorano fare ciò che fanno; che pensano giorno e notte come abbassare i tempi di risposta di una query da 0,3 secondi a 0,2 secondi; e siccome 0,2 è comunque più grande di zero, non va bene lo stesso e dunque può essere migliorato. È questo atteggiamento che li ha portati dove sono.

Ma allora cosa sta facendo Google? Google sta costruendo una piattaforma di calcolo distribuito che può gestire dataset su scala web da 100.000 cluster.

Ci siamo persi nella terminologia tecnologica? Allora provate a immaginare San Siro, sì, lo stadio. Immaginate che gli 80.000 posti a sedere siano riempiti non di persone, ma con 80.000 server. Poi riempite gli spogliatoi, i bagni, le gradinate e anche il campo, finché non aggiungete altri 20.000 server (e voci di corridoio dicono che il numero sia ben superiore in realtà). Collegate tutti questi server tra di loro in modo che se uno dovesse rompersi tutti gli altri potrebbero continuare a operare senza problemi – e vi assicuro che quest'ultima semplice frase potrebbe essere oggetto di vari dottorati di ricerca. Ora guardate bene attraverso questo marasma di cavi e server e alzate lo sguardo fino al bar: ci vedrete una manciata di ingegneri il cui unico compito è far sì che tutto questo continui a funzionare. In altre società, se ci sono altre società con un network simile, gli ingegneri che vedreste al bar sarebbero qualche centinaio e correrebbero da un server all'altro senza sosta. A Google ne basta una manciata, e stanno tutti al bar.

Aggiungete a questo un petabyte di file system distribuito a prova di errore, codice sorgente RPC distribuito, memoria condivisa a livello di network e migrazione dei processi. Oltre a, ovviamente, un datacenter che permette agli ingegneri di starsene al bar mentre tutto ciò funziona da solo. Ognuno di questi progetti potrebbe essere il cuore di una compagnia in cui io stesso investirei ad occhi chiusi.

Adesso è facile capire che indicizzare pagine web è un gioco da ragazzi. Al Googleplex esiste un computer con una potenza di calcolo tale da rendere possibile qualunque idea e Google si sta accaparrando i migliori cervelli dalle migliori università e dalle migliori compagnie. Google ha costruito il più grande computer che esista, utilizza un sistema operativo proprietario e un file system proprietario (GFS - Google File System) e rende questo computer più grande e più veloce ogni mese mentre, contemporaneamente, abbassa il costo dei cicli della CPU. Sembra sempre più una piattaforma generica ottimizzata per una sola funzione.

Mentre la competizione prende di mira le singole applicazioni di Google, Google continua nella costruzione di una piattaforma di calcolo distribuito che può gestire dataset su scala web.

Su questo computer gira il miglior motore di ricerca del mondo, un servizio di networking sociale, un motore per il confronto dei prezzi, un nuovo sistema di email e l'intero archivio di Usenet degli ultimi 20 anni. Cosa faranno adesso con il più grande computer al mondo e il più avanzato sistema operativo? Con questi presupposti l'unico limite è solo la fantasia di chi ci lavora.

Glossario

API

Application to Program Interface. Software che consente la comunicazione tra programmi altrimenti incompatibili.

Back

Il bottone del browser che permette di tornare a una pagina precedente della propria navigazione.

Backlink

Se una pagina A ha un link verso la pagina B, definiamo un backlink come il link che dalla pagina B porta alla pagina A.

Bot

Sinonimo di Spider, ma solitamente utilizzato solo dai motori di ricerca. Mentre spider è più utilizzato per indicare un software utilizzato da un utente per crearsi una copia di un intero sito sul proprio PC, un bot è normalmente inteso come il software usato da un motore di ricerca.

Vedi anche Spider.

Carattere jolly

Simbolo che sostituisce un carattere o un insieme di caratteri e che viene generalmente usato per cercare un file o una stringa all'interno di un file di testo.

Le wild card più conosciute sono l'asterisco e il punto di domanda. Ad esempio, cercando nel proprio hard disk i file il cui nome è "*.txt", si troveranno tutti i file con l'estensione ".txt".

Click

Premere un tasto del mouse. Deriva dal rumore che i bottoni del mouse fanno quando vengono premuti.

Click through ratio (CTR)

Il rapporto tra il numero di banner mostrati e il numero di questi cliccati da un utente, ovvero quando la pagina bersaglio viene visualizzata correttamente.

Cost per acquisition (CPA)

Ciò che un inserzionista spende per ottenere da parte di un utente un'azione o una vendita, a seguito della visualizzazione di un banner e del successivo click (ad esempio, l'iscrizione a una mailing list o la vendita di un prodotto o servizio).

Cost per click (CPC)

Rapporto tra costo della pubblicità on line e il numero di click-through ottenuti.

Cracker

Cracker è una persona che penetra all'interno delle reti di computer network elude le password o le licenze all'interno dei programmi; o in altri modi infrange la sicurezza dei computer. Un cracker può farlo per profitto, per cattiveria, per una causa, con motivazioni altruistiche o per sfida. Il cracker è quello che i giornalisti definiscono erroneamente un hacker ed i veri hacker non gradiscono questa confusione.

Crawl

Letteralmente “trascinarsi”, “strisciare” o “camminare a gattoni”. Verbo che serve a indicare l'azione di un software che passa da una pagina all'altra di un sito o tra siti in modo meccanico con scopi esplorativi.

GoogleBot

Il nome del bot, o spider, di Google.

Hacker

Un termine che sta a significare “un abile programmatore” ma che molti, specialmente i giornalisti usano per indicare “qualcuno che cerca di entrare nei sistemi informatici.”

Un “good hack” è una brillante soluzione ad un problema di programmazione e “hacking” è il verbo.

Home page

La prima pagina di un sito. La pagina a cui si accede se non si specifica null'altro di quel sito.

Ad esempio, l'home page di pc-facile.com è: <http://www.pc-facile.com/>

Indicizzare

Il processo per cui una pagina viene analizzata da un bot, a cui viene assegnato un ranking e una rilevanza da un motore di ricerca.

Keyword

Letteralmente “parole chiave”. Solitamente le parole utilizzate per effettuare una ricerca.

Keyword advertising

Pratica pubblicitaria in cui si cerca di legare il prodotto da pubblicizzare al contesto. Nel dettaglio, il promotore di un prodotto sceglie alcune parole che meglio rappresentano il suo prodotto e acquista il diritto a mostrare il proprio messaggio pubblicitario quando queste sono presenti sulle pagine di un editore che accetta di ospitare pubblicità. Saper scegliere le parole che meglio identificano il proprio prodotto è meno banale di quel che sembra.

Questa contestualizzazione del messaggio pubblicitario sembra stia dando degli ottimi risultati su Internet.

Link

Collegamento che permette di passare da una pagina ad un'altra (o ad un altro punto all'interno della stessa pagina) con un semplice click sullo stesso.

PageRank

Posizione di una pagina o di un sito web all'interno di un motore di ricerca. Il valore di PageRank di una pagina, non indica semplicemente il suo grado di "popolarità" sul web ma si spinge oltre fino a indicare un grado di "autorevolezza". Il PageRank è uno dei tanti fattori che contribuisce a determinare la posizione della pagina stessa nei risultati delle ricerche: più il PageRank è alto e più sarà alta la posizione della pagina rispetto alle altre trovate con la stessa rilevanza. Il PageRank è una caratteristica di Google che alcuni altri motori hanno cercato di imitare; alcune voci dicono che il PageRank non ha più alcun'influenza sui risultati di Google.

Pay per click (PPC)

Modalità di acquisto e pagamento della pubblicità on-line: l'inserzionista paga una tariffa unitaria per click-through.

Query

Letteralmente "richiesta", operazione che estrapola dati da un database. Solitamente si tratta di una serie di keyword più degli operatori.

Nel gergo dei database tradizionali è invece una serie di istruzioni che si impartiscono al database e non necessariamente indica una ricerca.

Rank(ing)

Letteralmente "posizione" o "posizionamento". Vedi anche PageRank.

Return on Investment (ROI)

Il ritorno sugli investimenti: il rapporto tra l'utile e il capitale investito. Uno dei tanti metodi usati nel mondo dell'economia per misurare il valore di un investimento.

Search

Parola inglese che significa "ricerca". Nel contesto di Internet può indicare una ricerca (fare un search) o il motore di ricerca interno a un sito (il search del sito). Solitamente non indica un motore di ricerca come Google se non viene utilizzato in abbinamento con la parola "engine" - search engine.

Search Engine Optimization (SEO)

Quella pratica per cui si cerca di adattare un sito agli standard di uno o più motori di ricerca nel tentativo di aumentare il ranking di una pagina o un sito. Sotto questo termine si includono quelle tecniche considerate legittime che quelle tecniche considerate illegali dai motori di ricerca.

Server

Un computer, più o meno ottimizzato, la cui funzione è fornire informazioni e/o documenti su richiesta. Di solito “server” viene usato in riferimento a un “web-server”, ovvero un computer il cui compito è mostrare pagine web quando richieste. In altre parole è un computer che ci mostra un sito Internet.

è importante tenere a mente che un server non è altro che un normalissimo computer a cui, per convenienza, sono state rimosse parti superflue come schede video, schede audio e monitor in quanto non utilizzate.

Spider

Software, utilizzato solitamente dai motori di ricerca o da chi vuole farsi una copia locale di un sito, che copia pagine web, le analizza alla ricerca di link e procede a copiare le pagine a cui puntano questi link. Lo spider si muove “crawl” attraverso quella rete virtuale che è Internet.

Toolbar

Letteralmente “barra degli strumenti”. Solitamente si intende un software che si integra con il browser e utilizzato dai motori di ricerca per facilitare la ricerca.

Usability

L’usability (usabilità) è una misura della qualità dell’interazione tra un utente e un prodotto (sia che si tratti di un sito web o di un software o di un qualsiasi oggetto tecnologico). Si suddivide in alcuni fattori principali:

Facilità di apprendimento: quanto velocemente un utente, che non è mai entrato in contatto con il prodotto, può apprendere le funzioni base?

Efficienza d’uso: una volta che l’utente ha acquisito una certa esperienza, quanto facilmente riesce a svolgere un compito?

Facilità di memorizzazione: se un utente usa il prodotto e poi se ne allontana, quanto gli ci vuole per ri-usarlo con efficacia?

Frequenza e pericolosità degli errori: quanto spesso un utente fa errori? e quanto questi errori sono seri? quanto ci vuole per correggere l’errore?

Soddisfazione personale: quanto piace all’utente utilizzare il prodotto?)

URL

Acronimo che sta per “Uniform Resource Locator”. Formato standard per identificare una risorsa Internet accessibile tramite World Wide Web. Ad esempio: <http://www.pc-facile.com>.

Appendice I. Google Suggest - Google dalla A alla Z

Londra ha una famosa guida la “London AZ”, il tutto città di Londra per eccellenza. Google cerca di replicare.

Google sa tutto, Google tiene traccia di tutto e Google è abbastanza sicuro di se da pensare di saper cosa volete cercare.

Seguite il link in fondo a questa pagina, digitate una sola lettera e Google vi dirà quali sono le query più comuni con quella lettera.

A... Amazon

Amazon è il e-commerce per eccellenza, forse il primo sito ad aver tentato seriamente di conquistare il web. C'è riuscito e adesso viene premiato: è forse il brand più conosciuto sul web... insieme a Google ovviamente.

<http://www.amazon.com>

B... Best buy

È ancora il e-commerce a fare da padrone: a dimostrazione che su Internet si cercano i prezzi migliori e si trovano i prezzi migliori.

<http://www.bestbuy.com>

C... CNN

“Si en en”: il network mediatico che ci informa sugli eventi di tutto il mondo è la “de facto” fonte d'informazione per chi vive nel web. Quando c'è stato l'attacco alle torri gemelle di New York, la stampa e la televisione erano in netto ritardo sugli avvenimenti, ma Internet e la CNN in versione “lite”, per non soffocare sotto il numero di accessi, sono riusciti a informare il mondo sui fatti in tempo reale.

<http://www.cnn.com>

D... Dictionary

Abbiamo tutti almeno un dizionario in casa, eppure ci è più facile fare una ricerca su Internet per trovare il significato di una parola che aprire un libro.

E... eBay

eBay è l'ennesimo emblema del web; l'idea che pareva non poter funzionare su Internet (un'asta online, ma a chi può interessare!?) è oggi una multinazionale. Per dare un'idea della penetrazione di questo servizio basti pensare che metà della posta Tedesca, tra privati, è costituita da oggetti comprati e venduti tramite eBay.

<http://www.ebay.com>

F... Firefox

Internet Explorer non è un browser sicuro e in questi ultimi anni sono stati in tanti a cercare di competere contro il predominio Microsoft; in testa sicuramente c'è Opera. Ma è stato l'open source, il buon senso, l'ottima programmazione e l'adeguamento agli standard a creare l'unico browser che sembra in grado di competere contro il bundling di Bill Gates.

<http://www.mozilla.org/products/firefox/>

G... Games

"Games", ovvero "giochi". Qualcuno forse ha pensato che la G sarebbe stata di Google, ma chi cerca [Google] su Google?

La verità è che Internet è ancora giovane e così lo sono anche molti dei suoi utenti.

H... Hotmail

Benché non sia stato il primo provider di account di posta gratuita è stato uno dei più amati. Adesso, tra GMail e spam, il servizio di Microsoft è in declino, ma non bisogna dimenticare che è stato uno dei pilastri che hanno aiutato a diffondere la cultura dell'e-mail.

<http://www.hotmail.com>

I... Ikea

Se l'Ikea è tanto presente in rete è perché incarna quel principio di low-cost che tende a sposarsi così bene con il web..

<http://www.ikea.com>

J... Jokes

J come "jokes" – "barzellette" o "scherzi". Su Internet viene sempre prima il piacere e poi il dovere.

K... Kazaa

Adorato dalle folle, silurato dalle major, il software P2P è passato con triste velocità dalle stelle alle stalle: da software più scaricato del web a software più denunciato nei tribunali, Kazaa vive ora un momento difficile. La R non è R di RIAA, per cui questa piccola rivincita gliela si concede volentieri.

L... Lyrics

Mai cercato testi e spartiti musicali on-line? "Lyrics" è la parola giusta.

M... Mapquest

Chi pensava che la M sarebbe stata di Microsoft si sbagliava; la M è per Mapquest, il servizio di mappe che ancora una volta dimostra quanto Internet serva a una sola cosa: lo scambio di informazioni. Oggi le indicazioni non si chiedono più a un passante, si cercano su Internet!

<http://www.mapquest.com/>

N... News

L'avanzata dell'informazione sul web è un fenomeno inevitabile. Aggiungete la parola "news" ad una vostra query e sarete informati sugli ultimi avvenimenti rilevanti. L'informazione è inarrestabile su web.

O... On-line dictionary

"Se in molti cercano [on-line dictionary], qual'è il motivo che li spinge ad aggiungere alla query la parola 'on-line'? Non sanno di essere già on-line? Perché partono dal presupposto che il risultato sarà pescato 'off-line'?"

Kieren McCarthy, The Register.

<http://www.theregister.co.uk>

P... Paris Hilton

Anch'io pensavo che la P sarebbe stata per "porn" o "pr0n", come piace ai "1337", ma anche se la P è per Paris Hilton poco ci discostiamo dalla mia idea.

Paris Hilton, figlia del padrone dell'omonima catena di hotel, ha deciso di riprendere le sue avventure sessuali – un po' come Pamela Anderson qualche anno fa. Non si sa come, questi video sono finiti su Internet e su tutti i principali network P2P. Vi chiedete ancora il perché della P?

Q... Quotes

"Quotes", ovvero "citazioni". Il bello di Internet sta nel suo supporto alla memoria umana, e l'intelligenza umana sta nell'avvalersi di questo supporto.

R... Recipes

"Recipes", "ricette". Volete sapere come si fa il goulash o il tom yam goong Tailandese? Dalle pagine di Internet troverete tutti i segreti di tutte le ricette.

S... Spybot

Teconologia che mira a raccogliere informazioni su di una persona senza che questa ne sia a conoscenza. Di solito si tratta di software che viene installato sul PC ad insaputa della persona, ad esempio nascosto in alcuni programmi shareware come i Morpheus o Kazaa.

Alcune persone considerano i cookies degli spyware ma questi, a differenza, sono largamente conosciuti ed i browser sono solitamente equipaggiati per poterli evitare.

T... Tara Reid

Attrice che in Italia è conosciuta per essere stata la Vicky di "American Pie" o Bunny, la moglie pornstar de "Il Grande Lebowski". Quando alla bionda Ms Reid, è scivolata la spallina del vestito, lei era ignara del fatto che il suo seno sinistro fosse uscito per la felicità di tutti i giornalisti. I fotografi si sono scatenati, ma quei gran bigotti della censura Americana sono corsi a coprire tutto. La soluzione a ogni curiosità? Internet, naturalmente.

U... UPS

Chi acquista via Internet conoscerà sicuramente questo acronimo “United Parcel Service”, uno dei principali corrieri privati utilizzati per il trasporto di merce acquistata via web. I siti che fanno uso di questo corriere inviano sempre un’e-mail per avvisare dell’avvenuta spedizione. All’interno di questa e-mail c’è un link e seguendo questo link è possibile verificare dove si trova il nostro pacco in ogni momento in cui è in transito. Grandioso!

<http://www.ups.com/>

V... Verizon

Un buon indicatore che una delle cose più importanti riguardo a Internet è accedervi. Ed è qui che entrano in gioco gli ISP come l’Americana Verizon. È anche un ottimo indicatore della crescente importanza dell’Internet wireless in cui Verizon sta investendo molto: Internet ovunque, non legato a nessun filo, ma legato a voi, ovunque voi siate.

<http://www.verizon.com/>

W... Weather

“Weather”, il meteo. Perché aspettare il telegiornale per sapere che tempo farà domani? E se vogliamo sapere che tempo fa nel nostro luogo villeggiatura?

X... Xbox

Una seconda entry per la Microsoft dopo Hotmail, è la console per videogiochi che sta facendo spopolando nella sua battaglia contro gli avversari Playstation e Nintendo.

<http://www.xbox.com>

Y... Yahoo

Google: non suggerisce se stesso sotto la G, per poi suggerire la concorrenza sotto altra lettera.

Ma ricordiamo che David Filo, il fondatore Yahoo, è un amico di Brin e Page ed è anche stato uno dei primi a credere e a utilizzare Google.

<http://www.yahoo.it>

Z... Zip codes

Zip code, ovvero l’influenza Americana sul web. L’equivalente del nostro CAP, ha una maggiore importanza oltre oceano tanto che conoscerlo apre molte porte nella ricerca. La domanda che nasce spontanea è: “Ma, allora, siamo tutti americani?”

Ricordiamo che quanto indicato qui non è che una fotografia di ciò che il villaggio globale cerca. Questa fotografia cambia ogni giorno seguendo le mode, i pettegolezzi e i capricci delle 150 milioni di query giornaliere. Cosa si dice su Internet oggi? Google Suggest ne è la risposta.

Le menti più tecniche possono essere interessate a sapere che per realizzare questo servizio Google si avvale di JavaScript, Javascript compresso per la precisione. Ma la cosa davvero interessante è che

questo codice si interfaccia con i server di Google ogni volta che un tasto viene premuto o rilasciato SENZA che la pagina venga ricaricata. Ovviamente, come tutte le informazioni, anche questo codice può essere trovato su – toh – proprio Google.

<http://www.google.com/webhp?complete=1&hl=it>

Appendice II. Codice sorgente per calcolare il PageRank

La formula per calcolare il PageRank non è complessa, ma richiede un computer a causa dell'elevato numero di calcoli necessari. Qui di seguito indichiamo il codice per calcolare il PageRank usando due linguaggi: il PHP e l'ANSI C.

Il codice qui sotto non è ottimizzato per i grandi calcoli e, specialmente nel caso del PHP, non sarebbe efficiente per emulare i calcoli fatti a livello di Google. Ma se volete vedere come modifiche nella struttura del vostro sito possono influenzare la distribuzione del PageRank, all'ora l'implementazione di questo codice vi sarà utile.

Codice PHP

```
<?php
// fattore smorzante
$damping = 0.85;

// numero di iterazioni da compiere
$iterations = 10;

// valori iniziali del page rank
$guess = 1;

// matrice che descrive quali pagine hanno link a quali altre pagine
// in questo esempio vediamo che la pagina "1" ha un link verso la pagina 2 e un
link verso la pagina 3
// la pagina due ha un solo link verso la pagina 3...
$webMatrix = array
(
    array ("0", "1", "1", "0"),
    array ("0", "0", "1", "0"),
    array ("1", "0", "0", "0"),
    array ("0", "0", "1", "0")
);

//fine config utente
/*////////////////////////////////////*/

// popola l'array del pagerank iniziale
$pagerank = populate ($webMatrix, $guess);

// calcola il page rank di tutte le pagine nella $webMatrix
for ($k = 0; $k < $iterations; $k++)
```

```

{
    $pagerank = iteration ($webMatrix, $pagerank, $damping);
}

// vediamo cos'è uscito
print_r ($pagerank);

function iteration ($webMatrix, $pagerank, $damping)
{
    // aggiorna il pagerank di ogni pagina uno alla volta
    for ($i = 0; $i < count($webMatrix); $i++)
    {
        $sum = 0;
        // calcola il pr di ogni pagina
        for ($j = 0; $j < count($webMatrix); $j++)
        {
            // trova quali pagine hanno un link alla pagina di cui stiamo
            // calcolando il pr e calcola il pr che esce da queste pagine
            ($webMatrix[$j][$i] != "0") ? $sum += $pagerank[$j] /
array_sum($webMatrix[$j]) : 0 ;
        }

        $temp = 1 - $damping + $damping * $sum;
        $pagerank[$i] = sprintf("%.2f", $temp);
    }

    return $pagerank;
}

function populate ($webMatrix, $guess)
{
    for ($i = 0; $i < count($webMatrix); $i++)
    {
        $pagerank[$i] = $guess;
    }

    return $pagerank;
}
?>

```

Codice ANSI C

```
/*
```

```

* - Nella configurazione iniziale, bisogna specificare almeno il numero delle
colonne. Infatti, la
* definizione è:
* BOOL webMatrix[][4]=...
* (il numero delle righe viene dedotto dalla definizione)
* il C, infatti, alloca 16 BOOL consecutivi, ma se non gli si dice quante
colonne ci sono, non
* sa come trovare i vari elementi. Infatti, un elemento come webMatrix[2][2]
viene traslato
* in webMatrix[2*numcols+2], e quindi il numero delle colonne deve essere noto
a compiletime
*/

// per printf
#include <stdio.h>

/*
* il c non ha un tipo "BOOL", mentre il c++ ha bool
* Siccome è scomodo scrivere unsigned char, definisco un
* alias
*/
#ifdef __cplusplus
typedef unsigned char BOOL;
#else
typedef bool BOOL;
#endif

double* iteration(BOOL* matrix,double* pagerank,double damping,int sizex,int
sizey,int* totalsPerLines)
{
    int i,y;
    for(i=0;i<sizex;++i)
    {
        double sum=0.0;
        for(y=0;y<sizey;++y)
        {
            sum+=(matrix[y*sizex+i]?(pagerank[y]/totalsPerLines[y]):0);
        }
        //"con qualche semplice passaggio algebrico..."
        pagerank[i]=1+damping*(sum-1);
    }
    return pagerank;
}

int* calculateTotalsPerLines(BOOL* matrix,int sizex,int sizey)
{
    int* totalsPerLines=malloc(sizeof(BOOL)*sizex);
    int i,n;

```



```

        for (n=0;n<sizeX;++n)
        {
            totalsPerLines[n]=0;
            for(i=0;i<sizeY;++i)
                totalsPerLines[n]+=matrix[sizeX*n+i];
        }
        return totalsPerLines;
    }

int main()
{
    /*
     * configurazione
     */
    const double damping=0.85;
    int iterations=10;
    const int guess=1;
    BOOL webMatrix[][4]=
    {
        {0,1,1,0},
        {0,0,1,0},
        {1,0,0,0},
        {0,0,1,0}
    };
    /*
     * fine configurazione
     */
    //calcolo la dimensione della matrice
    int SIZEX=sizeof(webMatrix[0]);
    int SIZEY=sizeof(webMatrix)/SIZEX;
    //alloco la memoria per i risultati
    //ricordarsi di liberarla prima di uscire!
    double* pagerank=malloc(sizeof(double)*SIZEY);
    /*
     * Se voglio riempire pagerank di "guess", tantovale
     * che ce lo copi direttamente dentro...
     */
    int i=0;
    for(i=0;i<SIZEY;++i)
        pagerank[i]=guess;
    /*
     * calculateTotalsPerLines alloca memoria, ricordarsi di liberarla prima
     * di uscire
     */
    int* totalsPerLines=calculateTotalsPerLines(&webMatrix[0][0],SIZEX,SIZEY);
    //ok, iterazione principale
    while(iterations--)
    {

```

```
    pagerank=iteration(&webMatrix[0][0],pagerank,damping,SIZEX,SIZEY,totalsPer
Lines);
    }
    //stampo i risultati
    for(i=0;i<SIZEY;++i)
        printf("pagerank[%d]->%01.2f\n",i,pagerank[i]);
    //libero la memoria allocata
    free(pagerank);
    free (totalsPerLines);
    //esco
    return 0;
}
```

Appendice III. Risorse on-line

Google è in continua evoluzione e io vorrei che tu fossi sempre aggiornato con le ultime informazioni perché essere disinformato vuol dire fallire in questo campo.

Ti indicherò qui sotto una lista di link che potrai consultare per tu stesso. Ora che hai appreso le basi seguire uno spizzico qui e un altro là non sarà un problema.

Ti lascio un ultimo consiglio: se operi in questo settore segui attentamente i siti indicati qui, non riposare sugli allori, non possiamo farlo se vogliamo rimanere competitivi!

Una nota prima di cominciare. Con l'esclusione del primo link, pc-facile.com che è in Italiano, tutti gli altri siti sono in lingua Inglese.

pc-facile.com

Il primo link che ti lascio è quello per il la pagina ufficiale del libro.

Qui ci troverai un forum dedicato per parlare delle strategie di SEO o di AdWords, consigli e trucchi su come cercare al meglio, una newsletter per essere sempre aggiornato, tutti questi link e molto altro.

<http://www.pc-facile.com/google/>

Google Groups

Come prima cosa direi che un'occhiata ai gruppi ufficiali di Google; qui le conversazioni tendono a non essere delle migliori, ma sono sempre i gruppi ufficiali

<http://groups.google.com/groups?q=google.public.translators>

<http://groups.google.com/groups?q=google.public.labs.glossary>

<http://groups.google.com/groups?q=google.public.programming-contest>

<http://groups.google.com/groups?q=google.public.web-apis>

<http://groups.google.com/groups?q=google.public.labs.newsalerts>

<http://groups.google.com/groups?q=google.public.labs.voice-search>

<http://groups.google.com/groups?q=google.public.labs.webquotes>

<http://groups.google.com/groups?q=google.public.support.general>

<http://groups.google.com/groups?q=google.public.labs.google-viewer>

<http://groups.google.com/groups?q=google.public.labs.keyboard-shortcuts>

<http://groups.google.com/groups?q=google.public.labs.sets>

<http://groups.google.com/groups?q=google.public.labs.compute>

Ovviamente questi gruppi non sono stati scolpiti nella pietra; altri nasceranno a seconda delle esigenze.

Per sapere quali sono i nuovi gruppi, qui sotto il link che vi indica la lista dei gruppi ufficiali:
<http://groups.beta.google.com/groups/dir?sel=33604954>

Search

Google Guide

Ottimo sito per principianti che vogliono migliorare le loro abilità nella ricerca e vogliono imparare a usare Google.

<http://www.googleguide.com>

Search Engine Optimization

Site Point Forums

SitePoint è forse il miglior forum che conosca e il loro forum dedicato a Google (focalizzato sul SEO) è spesso informativo.

<http://www.sitepoint.com/forums/forumdisplay.php?f=177>

Search Engine Forums

Altri forum molto visitati per quanto riguarda il SEO. Come tutti i forum ha dei periodi ricchi di utenti esperti e periodi meno buoni. Una fermata obbligatoria per chi gestisce un sito.

<http://searchengineforums.com/>

Market Leap

Sito che si specializza nel SEO, ma di cui io ritengo interessanti tre suoi strumenti più che la documentazione che offre.

Link Popularity Check: Strumento che vi dice quanti siti vi linkano e vi permette di paragonarvi ad altri tre siti da voi indicati

Search Engine Saturation: Quante pagine del vostro sito ha immagazzinato Google? Anche qui potete paragonarvi ad altri cinque siti.

Keyword Verification: Verificate in che posizione si sono piazzate le vostre pagine per le keyword da voi indicate.

<http://www.marketleap.com>

Search Engine Watch

Un sito generico di cui vi consiglio però la documentazione dedicata al SEO. Parte del sito è gratuito,

per gli approfondimenti è necessaria una registrazione di \$99 che dura un anno. Consigliabile solo se questo è il vostro impiego primario, altrimenti consultate solo la parte gratuita.

<http://www.searchenginewatch.com>

SearchEngines.com

Consigli sull'ottimizzazione, submission ai motori di ricerca e la promozione dei siti.

<http://www.searchengines.com>

AdWords

Google AdWords Handbook

Il miglior libro per quanto riguarda AdWords. Si tratta di un PDF che costa \$69: caro, ma è oro se operate nel settore.

<http://www.page-zero.com>

Link generici

Watching Google like a hawk

Aggiornatissimo e fornitissimo sito che raccoglie le ultime notizie e gli ultimi pettegolezzi su Google. NON è tecnico, ma vi terrà aggiornati di ogni passo fatto al Googleplex.

<http://www.watchinggooglelikeahawk.com>

About AdSense

Come suggerisce il nome, questo è un sito dedicato ad AdSense: utile se volete cominciare a guadagnare dal vostro sito.

<http://about-adsense.com>

Google FAQ

Le Frequently Asked Questions (FAQ) di google.public.support.general FAQ. Una pagina lunga fitta di consigli utilissimi. Una delle prime pagine da consultare.

<http://www.geocities.com/googlepubsupgenfaq/>

Fagan Finder

Sito che ha costruito un'interfaccia comoda che racchiude moltissime delle opzioni di ricerca di Google

su di una sola pagina. Davvero utile per quelle ricerche molto complesse e molto specifiche.
<http://www.faganfinder.com/google.html>

Search Engine Showdown

Sito generico con ultime notizie e un po' di documentazione su tutti i motori di ricerca.

<http://www.searchengineshowdown.com>

Search Engine Guide

Informazioni sul SEO e sul search advertising.

<http://www.searchengineguide.com/>

ClickZ

Consigli, notizie, eventi e risorse del marketing elettronico.

<http://www.clickz.com>

eMarketer

newsletter, statistiche di mercato, ricerche e analisi riguardanti l'eBusiness.

<http://www.emarketer.com/>

Nota finale

Questa lista non vuole in alcun modo essere esaustiva. Se pensate che un sito meriterebbe di essere incluso tra questi seguite il link qui sotto e proponetelo; sarà incluso sul sito e potrebbe essere incluso in una futura edizione del libro:

<http://www.pc-facile.com/google/>

Ringraziamenti

Come molte delle cose che facciamo, queste non sarebbero possibili senza l'aiuto di molte persone. In questo caso devo ringraziare (in nessun ordine in particolare):

David De Giacomi

Davide Roselli

Marcello Dall'Aglio (autore del capitolo: Il futuro di Google)

... E tutto lo staff di pc-facile.com, senza il quale questo sito e questo volume non esisterebbero.

Davvero, è grazie a loro se su questo sito funziona tutto così bene!