

I Segreti di Google ed il futuro

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.

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.



Il **PageRank** è quel valore tra 0 e 10 che vediamo rappresentato da una barra verde se abbiamo installata la GoogleToolbar

<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.

Il **PageRank** è definito come segue:

Consideriamo che la **pagina A** abbia **pagine B1...Bn** che la linkano

Il **parametro d** è una *costante di smorzamento* che può

essere fissata tra 0 e 1. Io di solito la imposto a 0,85.

Il **parametro C(A)** è definito come il numero di link che escono dalla pagina A.

Il **PageRank** della **pagina A** **PR(A)** è determinato come segue:

$$\mathbf{PR(A) = (1 - d) + d (PR(B1)/C(B1) + \dots + PR(Bn)/C(Bn))}$$

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.

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.

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(B_n) - Ogni pagina indicizzata da Google ha un PageRank associato: **PR(B₁)** per la prima pagina fino a **PR(B_n)** per l'**ultima**.

C(T_n) - Ogni pagina suddivide equamente il suo voto tra tutti i link in uscita. Il numero di link in uscita di una pagina è C(B₁), ...C(B_n).

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

$d^*(...)$ - 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)**.

$(1 - d)$ - 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 almeno uno.

Com'è calcolato il PageRank?

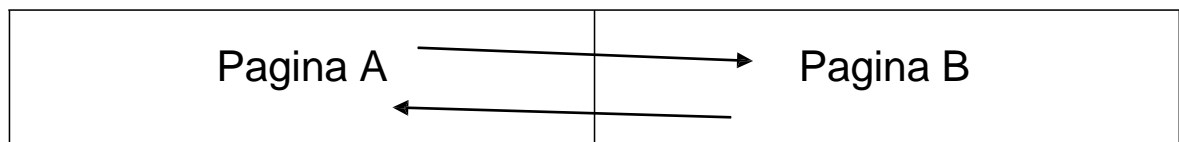
Il PageRank di una pagina dipende dal PageRank di tutte le pagine che puntano a questa pagina. Ma non possiamo sapere quale sia il PR di queste pagine finché non calcoliamo il PR delle pagine che puntano a loro, e via dicendo... Il calcolo sembra impossibile!

In realtà il problema non è così difficile:

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:

$$d = 0.85$$

$$\begin{aligned} PR(A) &= (1 - d) + \\ &d(PR(B)/1) \quad PR(B) = (1 - d) \\ &+ d(PR(A)/1) \end{aligned}$$

cioè:

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

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

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:



$$PR(A) = 0,15 + 0,85 * 0 = 0,15$$

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

Seconda iterazione:

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

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

Terza iterazione:

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

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

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:

$$PR(A) = 0,15 + 0,85 * 10 = 8,65$$

$$PR(B) = 0,15 + 0,85 * 8,65 = 7,5025$$

Seconda iterazione:

$$PR(A) = 0,15 + 0,85 * 7,5025 = 6,5272$$

$$PR(B) = 0,15 + 0,85 * 6,5272 = 5,6982$$

Terza iterazione:

$$PR(A) = 0,15 + 0,85 * 5,6982 = 4,9932$$

$$PR(B) = 0,15 + 0,85 * 4,9932 = 4,3942$$

I numeri tendono decisamente a 1!

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

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.

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.

Usando questo codice 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.

Evitando che questo venga sprecato nei link esterni al sito.

Codice sorgente per calcolare il PageRank

La formula per calcolare il PageRank non è complessa, ma richiede un computer (ed anche veloce) a causa dell'elevato numero di calcoli necessari.

Qui di seguito ti **indico il codice per calcolare il PageRank** usando il **PHP**.

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("%.01.2f", $temp);

}

    return $pagerank;

}

function populate ($webMatrix, $guess)
{

for ($i = 0; $i < count($webMatrix); $i++)
{

        $pagerank[$i] = $guess;

    }

    return $pagerank;

}

?>
```

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?

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. Google sta facendo delle ricerche nel campo dell'AI applicata 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 di google". 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 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** per citare il caso più famoso.

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 legale 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.

Infatti:

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 il movimento open-source, 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.
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 generalmenteresiduale: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, nonne 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 innovative 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, a conti fatti, cerca sempre più collaborazione con Google, piuttosto che la guerra.

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

Tipico dei grandi player 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'accadde 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?

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!

2. Modello di business: Netscape regalava il browser e vendeva software per server. Microsoft lo capì 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.

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 sonomossi immediatamente

- 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.



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 uno dei punti di forza!

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, questoera 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 in fondo mostra come il primo pericolo di Google non siano i competitor, ma siano le eventuali scelte miopi del proprio management.

Conclusioni

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.

Google è molto di più di quanto scritto in questo volume; in realtà Google è tutt'altro. La ricerca, 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 reinventa 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?

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 solo per generare profitti - utili anche ad una compagnia per cui i profitti non sono l'unico obiettivo - 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.

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.

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 e-mail 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.



Un'altra geniale idea alla base di Google è stata quella di motivare le menti che ci lavorano. Con un sistema rivoluzionario che ha generato vantaggi incredibili.

Il sistema del 4+1.

Per 4 giorni si lavora ai progetti Google.

Per 1 giorno si segue un proprio progetto, ci si concentra su una attività diversa, si studia, si sperimenta.

Questo ha portato a "invenzioni geniali". Come ad esempio alla creazione di "Virtual Earth" le famose Mappe di Google.

Chi lavora in questa società si sente libero di esprimere la propria genialità. Non è limitato da rigidi controlli o dal solo scopo commerciale.

Google se lo può permettere!

Pensa a chi ha investito in borsa anni fa in azioni Google.
Che affare!

Però....

ATTENZIONE.

Poi è arrivato **Facebook**.

Ne parleremo in un prossimo eBook.

E se Facebook fagocitasse tutti???

**Copyright Acon Software S.r.l.
@2013**

- <http://www.aconsoftware.com>
- <http://www.10eurooo.com>
- <http://www.adwordsforum.it>
- <http://www.wordpress-blog.it>
- <http://www.un-sito.it>
- <http://www.jobsit.it>
- <http://www.bid7live.it>
- <http://www.bid7live.net>



**Puoi distribuire gratuitamente a
tutti questo Report a patto di
non alterarne il contenuto, il nome
dell'autore ed i links presenti in esso.**