

Ippolita – copyleft 2007 – Creative Commons 2.0 by-nc-sa

www.ippolita.net

info@ippolita.net

THE DARK SIDE OF GOOGLE

(a.k.a. luci e ombre di google... - versione consegnata all'editore, gennaio 2007)

nota di Ippolita, l'autore:

Questo testo è la versione integrale consegnata all'editore nel gennaio 2007. Molte le modifiche rispetto a quella presentata in occasione di Hackit 2006 – Parma; aggiunte fra l'altro introduzione e conclusioni. Il testo non è conforme alla versione cartacea, la cui pubblicazione è prevista per il venti aprile 2007. Ogni eventuale refuso, errore di impaginazione, ecc. è quindi completamente a carico di info@ippolita.net

contributi e suggerimenti sono sempre bene accetti.

Indice

I. La storia di un successo / Storie di motori

1. Motori e ricerche
2. La nascita di Google: in principio fu il garage, anzi, l'università
3. Google.com: pubblicità diretta all'interno delle pagine
4. Il self-service pubblicitario: sopravvivere alla bolla della new-economy
5. Stile, morfologia e moltiplicazione dei servizi
6. Google, il gigante buono, va in borsa
7. Google, Inc.: il monopolio della ricerca

II. BeGoogle

- La fuga di cervelli verso Google: la guerra per il controllo del web
- Code lunghe nelle reti: dentro l'economia della ricerca, Google vs. Microsoft
- La guerra degli standard
- Arma1: Googleplex, il capitalismo morbido
- Arma2: perfezionare la strategia dell'accumulo
- Arma3: L'immagine è tutto, ma un po' di filosofia non guasta
- Arma4: Google e l'Open Source

III. Google Open Source: teoria e pratiche

- Open non è Free
- Gli hackers di Stanford
- L'era dell'Open Source economy: concorrenza e bontà d'animo
- Sedurre gli hackers: autonomia, soldi facili, strumenti gratuiti
- Ambienti ibridi fra università e azienda

IV. Algoritmi che passione!

- Algoritmi e vita reale
- La Strategia dell'oggettività
- Spider, basi di dati e ricerca
- Dalla brand-identity all'interfaccia partecipativa
- PageRank, o l'autorità assoluta di un mondo chiuso
- PageRank, o la moneta della scienza

V. In aggiunta altre maliziose funzionalità

- **I filtri sugli algoritmi: banche dati preconfezionate e controllo degli utenti**
- **I Cookie di Google, biscotti che lasciano il segno**
- **Onanismo tecnologico: crea, ricerca e consuma i tuoi contenuti**
- **Browser come ambiente di sviluppo**
- **Privacy, paranoie e poteri**

VI. Qualità quantità relazione

- **L'emergenza dell'informazione**
- **Quantità e qualità**
- **Il mito della ricerca istantanea**
- **Dietro il velo del mito**
- **Modelli di ricerca**

VII. Tecnocrazia

- **Tecnocrazia: gli esperti della scienza**
- **Miracoli della tecnologia: dalle opinioni soggettive alla verità oggettiva**
- **Pubblico e privato**
- **Vie di fuga: p2p, media indipendenti, crittografia, blog, foaf**

APPENDICI:

- I. La fine del mondo in un biscotto**
- II. Interzone: influenze e domini nei mondi digitali**

LICENZA

Introduzione

Google è il motore di ricerca più noto e utilizzato dell'intera Internet, tanto da essersi affermato negli ultimi anni come il principale punto di accesso alla Rete. I navigatori si sono adattati progressivamente alla sua interfaccia sobria e rassicurante, alle inserzioni pubblicitarie defilate e onnipresenti; hanno adottato i suoi comodi servizi e l'abitudine al suo utilizzo si è trasformata in comportamento: "Se non lo sai, chiedilo a Google". Si ricorre a Google anche quando si potrebbe ricordare il *post-it* appiccicato al frigorifero, consultare l'agenda, guardare le Pagine Gialle, o sfogliare la collezione di Garzantine che s'impolvera sugli scaffali, insieme alle altre pesanti enciclopedie cartacee, troppo faticose e difficili da consultare.

Google ha saputo sfruttare magistralmente il nostro bisogno di semplicità. Google aspira ad essere il motore di ricerca perfetto, in grado di capire esattamente le richieste degli utenti e restituire, in un batter d'occhio, proprio ciò che desiderano. Le candide interfacce, ormai altamente personalizzabili direttamente sul Web e tuttavia implacabilmente riconoscibili per il loro stile minimale, sono la via di fuga quotidiana dalla claustrofobia delle scrivanie digitali di un numero impressionante di utenti, in costante aumento. Sono una boccata d'aria, sono finestre privilegiate spalancate sull'affascinante mondo della Rete. Quante persone usano Google come pagina iniziale del proprio browser? Eppure dietro una tale semplicità e facilità d'uso si cela un colosso, un sistema incredibilmente complesso e pervasivo per la gestione delle conoscenze del *mare magnum* della Rete. Google offre decine di servizi gratuiti, per soddisfare ogni desiderio di ricerca e comunicazione: e-mail, chat, newsgroup, sistemi di indicizzazione dei file sul proprio computer, archivi di immagini, video, libri e molto altro ancora. Perché? Cosa ci guadagna? Criticare Google attraverso una disamina della sua storia e la decostruzione degli oggetti matematici che lo compongono è un'occasione per disvelare una precisa strategia di dominio culturale. Questa indagine offre l'opportunità di fornire un metodo di studio più generale utile alla scoperta dei retroscena di molti fra gli applicativi che ci siamo abituati ad utilizzare.

Il volume si apre con una breve panoramica sulla storia dei motori di ricerca, passando quindi in rassegna i momenti più significativi dell'ascesa di Google. Sopravvissuto senza danni allo scoppio della bolla della *new economy*, Google ha intrecciato solidi rapporti con diverse multinazionali dell'*Information Technology*. La continua espansione delle sue attività in ogni settore delle comunicazioni digitali sta diffondendo uno stile inconfondibile e modellando un intero universo culturale, quello del Web.

"Don't be evil" (non essere cattivo) è il motto di Sergey Brin e Larry Page, i due fondatori di Google. Gli ex-studenti di Stanford, grazie ad un'oculata gestione della propria immagine, hanno creato un "Gigante Buono", impaziente di archiviare le nostre "intenzioni di ricerca" nei suoi sterminati database. L'alter-ego digitale di milioni di utenti sembra essere in buone mani, affidato al *datacenter* principale di Mountain View, California, noto come Googleplex. Qui, come negli altri centri di archiviazione dati di Google - che stanno spuntando come funghi in tutto il mondo - si mettono a punto vere e proprie armi per combattere la guerra per il controllo delle Reti. In primo luogo, si diffonde la pratica del capitalismo morbido dell'abbondanza: si tratta di una strategia di controllo biopolitico in senso stretto, che propina ambienti di lavoro confortevoli, pacche sulle spalle e gratificazioni ai dipendenti. I lavoratori, soddisfatti e lusingati, sono contenti di farsi sfruttare e diventano i maggiori sostenitori dell'azienda, fieri di propagandare un'immagine vincente e "buona".

Gli obiettivi e i metodi di Google sono buoni per tutti; infatti, la filosofia aziendale, basata sull'eccellenza di stampo accademico, l'impegno per l'innovazione e la ricerca, si trova esposta in dieci rapide verità sul sito stesso del motore di ricerca. Questi dieci comandamenti costituiscono una sorta di buona novella dell'era informatica, il Google-pensiero, propagato con l'aiuto di veri e propri "evangelizzatori" (*evangelist*), ovvero personalità di spicco del mondo informatico. Ultima arma, ma non meno importante, è la cooptazione delle metodologie di sviluppo cooperativo tipiche dell'Open Source e l'uso di software liberi, non protetti da copyright o brevetti, come base per i propri prodotti. In questo modo Google abbatte i costi per l'implementazione dei propri servizi, si

assicura l'appoggio di tecnici, smanettoni e hacker di tutti i tipi e si spaccia per il sostenitore della causa della libera circolazione dei saperi, poiché l'uso del motore di ricerca sembra offrire l'accesso gratuito alla Rete nel modo migliore.

Ma il sogno di Brin e Page di "Google contenitore di tutta Internet", coltivato fin dai tempi dell'università, è solo un'idea demagogica, utile ad affermare un culto quasi positivistico dell'oggettività scientifica: nel caos della Rete solo una tecnica superiore può farsi garante della trasparenza dei processi, della correttezza delle risposte, addirittura della democrazia.

Infatti Google dichiara di essere uno strumento "democratico", basato sul presunto carattere "democratico" del Web. Il suo algoritmo di indicizzazione della Rete, PageRank(TM), si occupa di copiare i dati digitali nei *datacenter*, sfruttando i collegamenti associati a ogni singola pagina per determinarne il valore. In pratica, Google interpreta un collegamento dalla pagina A alla pagina B come un voto espresso dalla prima in merito alla seconda. Ma non si limita a calcolare il numero di voti, o collegamenti, assegnati a una pagina. Google esamina la pagina che ha assegnato il voto: i voti espressi da pagine "importanti" hanno più rilevanza e quindi contribuiscono a rendere "di maggiore valore" anche le pagine collegate. Il PageRank assegna ai siti Web importanti e di alta qualità un voto più elevato, utilizzando filtri e criteri non pubblici, di cui Google tiene conto ogni volta che esegue una ricerca. La "democrazia" di Google ordina perciò la Rete in base al numero di voti ricevuti da ogni pagina, e dell'importanza di questi voti: una democrazia filtrata dalla tecnologia.

Vi sono alcuni segreti attorno al colosso di Mountain View, molti dei quali, come vedrete, sono segreti di Pulcinella. L'alone di leggenda che circonda la tecnologia googoliana è dettato in gran parte dall'assenza di un'istruzione di base, di rudimenti pratici per affrontare culturalmente l'onda lunga della rivoluzione tecnologica. Per esempio, la straordinaria rapidità dei risultati di ricerca è frutto di un'accurata selezione niente affatto trasparente. Infatti, come potrebbero milioni di utenti fogliare contemporaneamente in ogni istante l'intera base dati di Google se non ci fossero opportuni filtri per restringere l'ambito della ricerca, ad esempio limitandolo ai dati nella loro lingua d'origine? E se esistono filtri creati per garantire una migliore navigazione linguistica, non è lecito supporre che ne esistano molti altri, studiati per indirizzare anche le scelte dei navigatori? Il prodigio di Google è in realtà una tecnologia opaca e secretata dal copyright e accordi di non divulgazione dei suoi ritrovati. La ricerca non è trasparente né democratica come viene spacciato: non potrebbe esserlo sia per motivi tecnici, sia per motivi economici.

Il campo bianco di Google in cui si inseriscono le parole chiave per le ricerche è una porta stretta, un filtro niente affatto trasparente, che controlla e indirizza l'accesso alle informazioni. In quanto mediatore informazionale, un semplice motore di ricerca si fa strumento per la gestione del sapere e si trova quindi in grado di esercitare un potere enorme, diventando un'autorità assoluta in un mondo chiuso. Il modello culturale di Google è dunque espressione diretta di un dominio tecnocratico.

Con questo volume Ippolita intende sottolineare il problema, o meglio l'urgenza sociale di alfabetizzazione e orientamento critico del grande pubblico attorno al tema della gestione delle conoscenze (*knowledges management*). Internet offre agli utenti straordinarie opportunità di autoformazione, tanto da surclassare persino la formazione universitaria, in particolare in ambiti come la comunicazione e l'ingegneria informatica. Il movimento del Software Libero, come Ippolita ha mostrato nei suoi precedenti lavori, è l'esempio più lampante della necessità di autoformazione continua e della possibilità di autogestione degli strumenti digitali.

Ma esiste un rovescio di questa medaglia, doppiamente negativo: da una parte, lo svilimento della formazione umanistica, che ha nella Rete pochi e male organizzati ambiti di riferimento; dall'altra, il sostanziale collasso cognitivo dell'utente medio. Disorientati dalla ridondanza dei dati disponibili sulla Rete, ci si affida ai punti di riferimento di maggiore visibilità -di cui Google è solo l'esempio più eclatante- senza domandarsi cosa avvenga dietro le quinte; si inseriscono i propri dati con leggerezza, conquistati dal mero utilizzo di servizi decisamente efficaci e, com'è ancora uso in buona parte della Rete, assolutamente gratuiti.

Ippolita cerca di segnalare il vuoto, tutto italiano, nella divulgazione scientifica dei fenomeni

tecnologici da cui la società intera è investita. La manualistica tecnica abbonda, la sociologia parla con disinvoltura di Società in Rete, la politica si spinge sino ad immaginare una futuribile Open Society, nella quale le Reti saranno il sostrato tecnologico della democrazia globale. Ma quanti navigatori assidui sanno cosa sia un algoritmo? Ben pochi, eppure moltissimi si affidano al responso di PageRank, un algoritmo appunto, che ordina senza sosta i risultati delle loro interrogazioni e indirizza la loro esperienza in Rete. Occorre il coraggio di riportare al centro la divulgazione scientifica, senza chiudersi nella torre d'avorio del sapere accademiche. Bisogna parlare di macroeconomie senza essere economisti, di infomediazione senza essere esperti di comunicazione, di autoformazione senza essere educatori, di autogestione degli strumenti digitali senza essere politicanti. Bisogna provocare dibattiti insistendo su concetti di base come "algoritmo", "dati sensibili", "privacy", "verità scientifica", "reti di comunicazione", troppo spesso discussi da Authority e Garanti che non possono garantire assolutamente nulla.

L'abitudine alla delega provoca un disinteresse generalizzato per i grandi mutamenti in corso nel nostro mondo tecnologico, che avvengono in sordina o coperti dal fumo mediatico, senza essere stati minimamente assimilati dal grande pubblico.

L'atteggiamento più comune oscilla fra l'incantata meraviglia e la frustrazione nei confronti dei continui, incomprensibili "miracoli della tecnologia"; si giunge spesso all'adorazione mistica, come se il digitale ricoprisse il mondo di un'aura esoterica, penetrabile solo da pochi iniziati, coniugata alla frustrazione per la propria incapacità di officiare adeguatamente il culto del nuovo progresso.

Il gruppo di ricerca Ippolita si riunisce proprio attorno alla convinzione che attraverso lo scambio e il dialogo tra competenze e linguaggi diversi si possa trasformare la cosiddetta Rivoluzione Digitale in una materia utile per comprendere la contemporaneità, le sue anomalie e probabilmente anche il tempo a venire. La ricerca scientifica, la tradizione umanistica, le passioni politiche, il metodo femminista sono altrettanti linguaggi da usare in questa esplorazione.

L'attività di Ippolita rivela che "mettere in comune" non basta, perché il livello di riflessione sulle tecnologie è ancora limitato e la cassetta degli attrezzi degli utenti ancora troppo rozza. È necessario assumere un'attitudine critica e curiosa, sviluppare competenze a livello soprattutto individuale, capire in quali modi si può interagire nei mondi digitali, mettere a punto strumenti adeguati ai propri obiettivi. La sfida è quella di moltiplicare gli spazi e le occasioni di autonomia senza cedere a facili entusiasmi, senza soccombere alla paranoia del controllo. *Just for fun*. La pratica comunitaria non è una ricetta capace di trasformare per incanto ogni novità tecnologica in un bene collettivo, non è sufficiente a scongiurare il dominio tecnocratico in nome di una grande democrazia elettronica. Si tratta di una visione fideistica del progresso, dimentica del valore delle scelte individuali. La sinergia fra i soggetti sulle reti, mondi vivi e in perenne mutazione, non è una banale somma delle parti in gioco, richiede passione, fiducia, creatività e una continua rinegoziazione di strumenti, metodi e obiettivi.

Vincolare gli elementi più strettamente tecnici alle loro ricadute sociali è sicuramente il primo e arduo passaggio da compiere. Per questa ragione, il testo che avete tra le mani è integralmente scaricabile sotto una licenza copyleft.

ippolita.net/google/
info@ippolita.net

I. Storie di motori

Motori e ricerche

I motori di ricerca si presentano oggi come siti appartenenti al World Wide Web (web) che permettono il reperimento di informazioni (*information retrieval*). La maggior parte degli utenti ritiene erroneamente che Internet e il web coincidano, perché il web è l'aspetto più semplice e immediato di Internet. In realtà la Rete è assai più complessa, eterogenea e variegata del web: comprende anche i mondi delle chat, dei newsgroup, della posta elettronica, e in generale tutto ciò che gli individui desiderano “mettere in rete”, a prescindere dalla “forma” con la quale queste informazioni si presentano. Inoltre, la Rete non è statica, ma dinamica. Ciò significa che le connessioni fra le risorse si modificano, le risorse stesse sono soggette a cambiamenti (nascita, trasformazione, morte). Allo stesso modo, i vettori di connessione fisica a queste risorse cambiano in continuazione, si evolvono dal doppiino telefonico del modem fino alla “banda larga” e alla fibra ottica. Non ultimi, gli individui che creano la Rete proiettando i loro alter ego digitali su di essa sono in continua mutazione, in quanto esseri viventi. La Rete dunque non è il web, ma una dinamica coevolutiva costituita dalle interazioni complesse fra tre differenti tipi di macchine: macchine meccaniche (computer, cavi, modem, ecc.), macchine biologiche (individui umani), macchine significanti (risorse condivise).

In ogni caso, dal momento che reperire informazioni nel magma della Rete è sempre stata un'esigenza fondamentale, per quanto possa sembrare strano, la storia dei motori di ricerca è assai più antica rispetto a quella del web.

Il web così come noi lo conosciamo è un'invenzione di Tim Bernes-Lee, Robert Caillau¹ e altri scienziati europei e americani. Tra il 1989 e il 1991 al CERN di Ginevra creano i primi browser, il protocollo http e il linguaggio html per la composizione e visualizzazione di documenti ipertestuali, ovvero contenenti link (sia interni a ogni documento, sia fra documenti diversi). Questa nuova tecnologia si presenta come un'implementazione della rete Internet, un progetto americano esito della fusione di diversi progetti universitari e militari.

Mentre il web si diffondeva tra i laboratori e le università di tutto il mondo, i motori di ricerca offrivano già da anni servizi di indicizzazione e ricerca di informazioni su Internet.

I primi motori di ricerca ovviamente non erano consultabili via web: erano veri e propri programmi da installare e configurare sulla propria macchina. Questi strumenti indicizzavano le risorse di Internet attraverso protocolli come ftp (per lo scambio di file), Gopher (un rivale dell'emergente http) e altri sistemi ormai caduti in disuso.

Nel 1994 diventava operativo il primo motore di ricerca basato su web: WebCrawler², un esperimento finanziato dall'università di Washington. Le innovazioni introdotte da questo motore di ricerca erano straordinarie: oltre a funzionare come un sito Internet, e a offrire la possibilità di effettuare ricerche *fulltext*³, era dotato di uno strumento per catalogare le pagine automaticamente, lo *spider*. Lo *spider* è un software che, come vedremo approfonditamente in seguito, assolve a due

¹ Si veda: Tim Bernes-Lee, L'architettura del nuovo Web - Dall'inventore della rete il progetto di una comunicazione democratica, Feltrinelli, Milano, 2001.

² Esiste ancora, ed è un ottimo metamatore di ricerca, combinando i risultati dei migliori motori:

<http://www.webcrawler.com>

³ Grazie alla ricerca *full-text* è possibile trovare parole singole, combinazioni di parole, frasi complete, scorrendo i documenti nella loro completezza.

funzioni: mentre naviga su Internet, memorizza le informazioni contenute sulle pagine web e le rende accessibili agli utenti del servizio di ricerca.

Per quanto WebCrawler fosse incredibilmente innovativo per l'epoca, i risultati che era in grado di fornire erano semplici elenchi di indirizzi web e i rispettivi titoli delle pagine indicizzate.

Negli ultimi mesi del 1994 il motore di ricerca Lycos riusciva a indicizzare in pochissimi mesi oltre il 90% delle pagine web di allora (10 milioni circa in tutto). La principale innovazione di Lycos era quella di escludere sistemi *fulltext* per la catalogazione, analizzando soltanto le prime 20 righe di ogni pagina indicizzata, invece che l'intero contenuto testuale di ogni pagina. Questo permetteva a Lycos di mostrare nei risultati anche una breve descrizione estrapolata da quelle stesse righe.

Con l'avvento di Excite, nel dicembre 1995, i risultati cominciano a essere ordinati in base al criterio dell'importanza delle pagine. L'introduzione di un sistema di valutazione del "peso" di una pagina apriva anche a una forma rudimentale di catalogazione tematica: finalmente venivano accantonate le liste interminabili di risultati non ordinati. Si configurava così una prima consultazione a "directory" dei siti, che possiamo paragonare a un'organizzazione in stile biblioteconomico classico delle risorse web, ovvero ad albero, con una indicizzazione suddivisa per materie, lingue, ecc.

Tuttavia, Excite è passato alla storia per un'altra ragione: è stato il primo motore di ricerca a dotarsi di strumenti finalizzati esplicitamente all'attività commerciale. Dopo aver acquistato WebCrawler, ha offerto personalizzazioni agli utenti e caselle di posta gratuite, trasformandosi in meno di due anni in uno dei primi portali web (1997). Di lì a breve, con la nascita di motori di ricerca sempre più evoluti, Excite mette da parte la propria vocazione originaria, optando per l'utilizzo degli strumenti di ricerca di altre società, tra cui oggi anche Google⁴.

Ma il panorama dei precursori di Google non sarebbe completo senza una menzione di quello che nel 1997 era il miglior motore di ricerca esistente, nonché il più popolare: AltaVista. AltaVista, che vuol dire "vista dall'alto", si avvaleva dei risultati di un gruppo di ricerca della Digital Equipment Corporation (DEC) di Palo Alto, California, che nella primavera del 1995 era riuscito a immagazzinare tutte le parole di una pagina HTML presente su Internet in un indice nel quale fosse possibile effettuare ricerche. Questo aveva consentito ad AltaVista lo sviluppo del primo database aperto alla consultazione diretta del World Wide Web. La mente di AltaVista era Louis Monier⁵ che, usando *cluster* di computer collegati fra loro, hardware all'avanguardia e il meglio delle tecnologie allora disponibili, era riuscito in breve tempo a rendere la propria creatura il motore di ricerca più usato e più apprezzato dagli utenti. AltaVista è stato anche il primo motore di ricerca multi-lingua di Internet e la prima tecnologia di ricerca in grado di supportare testi in cinese, giapponese e coreano. Infine, ha introdotto un sistema di traduzione automatica Babel Fish, tuttora utilizzato.

Prima del suo crollo⁶ nel 1997, AltaVista serviva 25 milioni di richieste al giorno e raccoglieva 50 milioni di dollari l'anno di sponsorizzazioni, fornendo il servizio di ricerca per gli utenti del portale Yahoo!, ancora oggi il maggiore antagonista di Google nel mondo Web.

⁴ L'idea di usare le funzionalità di ricerca di altri siti, e quindi poter offrire fra i risultati anche quelli altrui, non è stata di Excite. Già alla fine del 1995, MetaCrawler restituiva sulle proprie pagine web i servizi di tutti i motori di ricerca presenti. A oggi esistono moltissimi meta-motori di ricerca che compiono ricerche incrociate su diverse tipologie di database

⁵ In seguito Monier ha lavorato per Ebay; è stato assunto da Google nel 2005.

⁶ Nel 1996 il valore di questa ricerca non è ancora monetizzabile, considerato anche lo stato embrionale in cui si trova

La nascita di Google: in principio fu il garage, anzi, l'università

La parola Google deriva da “Googol”, termine matematico che indica un 1 seguito da 100 zeri. Leggenda vuole che questo fosse il numero di pagine web che Larry Page e Sergey Brin sognavano di indicizzare con il loro nuovo strumento di ricerca. Si incontrano nel 1995 a Stanford, quando il 24enne Larry Page, laureato all'Università del Michigan, si reca in visita a Stanford con l'idea di iscriversi al dottorato in informatica. Il 23enne Sergey Brin faceva parte del gruppo di studenti assegnati ad accompagnare i visitatori nel campus. Stanford era il luogo ideale per sviluppare progetti fortemente innovativi dal punto di vista tecnologico. Infatti poteva offrire da una parte laboratori di ricerca all'avanguardia; dall'altra, legami stretti sia con aziende del settore Information Technology (IT), sia con finanziatori attenti e disponibili a investire grossi capitali sulle ricerche accademiche più promettenti. Brin e Page sono entrambi affascinati dalla crescita vertiginosa del web e dai problemi connessi alla ricerca e all'organizzazione delle informazioni; si dedicano al progetto BackRub, chiamato così perché era in grado di analizzare i “back links” che puntavano a un sito e di fornirne una mappatura. In seguito BackRub sarà ribattezzato Google quando verrà dotato di una pagina web, nel 1997.

L'innovazione fondamentale introdotta da Google è quella di ribaltare il processo di indicizzazione delle pagine: non si elencano più i siti in base alla “prossimità” rispetto alla richiesta effettuata, ma si mostrano i risultati esattamente nell'ordine “corretto”, ovvero rispondenti alle aspettative dell'utente. Il primo link reso disponibile deve essere la risposta “esatta” all'interrogazione; i successivi si allontanano via via dal fulcro della richiesta⁷.

In quest'ottica nasce il noto pulsante “mi sento fortunato”: cliccando, si apre direttamente il primo link della ricerca di Google, presentato come il link inequivocabilmente “giusto”.

L'algoritmo che Larry Page si “inventa” per calcolare l'importanza di una pagina web, noto come PageRank, si basa sulla matematica statistica dei primi del Novecento e in particolare sulle formule sviluppate da Andrej Andreevic Markov per calcolare, all'interno di una rete, l'importanza di un nodo rispetto agli altri⁸.

All'inizio, Google è solo un progetto accademico e scientifico, per il quale il sistema di valutazione dipende in larga misura dai giudizi dei recensori (*referees*) che operano secondo il meccanismo della “revisione dei pari” (*peer-review*). Il metodo che teoricamente offre le garanzie maggiori di oggettività è il cosiddetto “doppio cieco” (*double-blind*): prima di essere accolto in una rivista scientifica, un articolo viene sottoposto ad almeno due ricercatori di chiara fama, che non devono conoscere il nome dell'autore (per non essere influenzati nel giudizio). Il secondo passaggio “cieco” riguarda l'estensore dell'articolo, che non deve conoscere il nome dei recensori. In sostanza, più un articolo scientifico viene accolto in maniera positiva da altri scienziati (che si suppone lavorino in maniera indipendente), più l'articolo stesso viene considerato importante e meritevole. Page trasla nel suo campo di ricerca questo approccio e sviluppa la teoria secondo cui il numero di link che conduce a una pagina è un modo per valutare il valore della pagina stessa e in un certo senso la sua qualità. Vedremo meglio più avanti come si realizza questo passaggio dalla

l'economia legata a Internet, ma il suo potenziale è individuato chiaramente dalla DEC, società che si occupava dell'aspetto hardware di AltaVista. Tuttavia il gruppo DEC non riuscirà a sfruttare l'occasione a causa di una serie sfortunata di ristrutturazioni interne che smembrano il team di AltaVista, acquisito nel 2003 da Overture (ceduta successivamente a Yahoo!).

⁷ Riteniamo questo approccio assolutamente irrealistico, oltre che politicamente scorretto e intrinsecamente autoritario, poiché sottintende l'esistenza di un'autorità in grado di esplicitare e fornire una risposta “giusta, rapida e corretta” a un desiderio. Possono esistere al limite molti percorsi individuali per costruire una risposta accettabile e soddisfacente a un bisogno/desiderio (si veda Cap. V).

⁸ Si veda il cap. IV.

“quantità” di informazioni servite alla “qualità” del risultato rispondente alle aspettative dell’utente⁹.

Ma questo criterio non era di per sé sufficiente, perché non tutti i link sono uguali né hanno lo stesso valore: più precisamente, spostando l’attenzione dal valore statico di un collegamento al valore dinamico di un percorso, la Rete è un ambiente (in termini matematici, un *grafo*) in cui non tutti i percorsi hanno lo stesso valore, ma diversi “valori di percorrenza” a seconda del “peso” dei nodi. Del resto, proseguendo nel parallelo con la pratica scientifica, non tutte le recensioni sulle riviste scientifiche hanno la medesima importanza. Il parere positivo di recensori non prestigiosi, o addirittura malvisti dalla comunità scientifica, è nocivo: troppe recensioni di scarso peso inficiano l’autorevolezza di una pubblicazione. Così, secondo Page, i link provenienti da siti particolarmente connessi, con un elevato numero di link, sono di maggiore importanza. Pertanto, un percorso (link) che giunge a una pagina da un sito molto popolare ha maggiore rilevanza rispetto a un percorso proveniente da una pagina poco nota. In questo modo un link da una pagina A a una pagina B viene interpretato come una recensione scientifica (ma va sottolineato che Brin e Page parlano esplicitamente di “voto” e “democrazia”), di maggior peso a seconda della popolarità del recensore/votante che fornisce il link. Il criterio dell’autorevolezza del recensore scientifico diventa più banalmente la misurazione della popolarità di un sito.

La valutazione delle pagine web di Google, detta *PageRanking*, si costruisce dunque in base a un sistema di recensione “pubblica” analogo al sistema della “repubblica scientifica”¹⁰, non limitato agli scienziati ma esteso a tutti i navigatori della Rete.

Oggi, l’organizzazione della repubblica scientifica e la questione del reperimento delle informazioni in genere è diventato un problema di cruciale importanza: in questo contesto di diluvio informazionale¹¹, in particolare sul web, ragionare non solo sull’importanza ma anche sull’attendibilità delle informazioni diventa sempre più complesso. Tanto più dopo che il meccanismo della “revisione dei pari” è stato messo in discussione dagli scienziati stessi¹². Fra le

⁹ Si veda il cap. VI.

¹⁰ Il sistema della pubblicazioni scientifiche nasce nel 1665 con *The Philosophical Transactions*, la prima rivista scientifica fondata da Henry Oldenburg della Royal Society di Londra, una sorta di pubblico registro della proprietà intellettuale; essa funzionava come una specie di ufficio brevetti delle idee scientifiche. La rivista voleva introdurre chiarezza e trasparenza nel processo di fondazione delle pretese innovative e tali pretese venivano soddisfatte col conferimento, da parte di pari, di un titolo di “nobiltà intellettuale”; la quale assicurava un particolare titolo di proprietà: quella intellettuale, appunto. L’importanza delle pubblicazioni contribuiva alla creazione di una “gerarchia tra i pari”, fondata su un ideale di eccellenza; tuttavia, poiché i risultati per essere apprezzati avevano bisogno di pubblicità, l’organizzazione delle regole di pubblicazione assunse un ruolo centrale, poiché per uno scienziato rendersi visibile diveniva fondamentale per acquistare notorietà e prestigio.

¹¹ Il concetto di diluvio informazionale (“*déluge informationnel*”) è stato sviluppato soprattutto da Pierre Lévy, rielaborando l’idea di “secondo diluvio” dell’artista Roy Ascott. Ne deriva un quadro filosofico in cui il movimento di “virtualizzazione” assume un ruolo centrale, si veda: Pierre Lévy, *Il virtuale*, Cortina, Milano, 1997 (online la versione originale francese: <http://hypermedia.univ-paris8.fr/pierre/virtuel/virt0.htm>). In maniera indipendente, Manuel Castells parla di “informazionalismo” ed “economia informazionale”. Gran parte delle letture economiche in merito all’epoca del diluvio informazionale, specie in ambito francese, italiano e statunitense, sono di derivazione marxista e si basano su interpretazioni forzose dei fenomeni di “intelligenza collettiva” delle Reti come ipostasi del *General Intellect* marxiano; si vedano ad esempio: Olivier Blondeau, *Le Capitalisme Informationnel - Genèse et subversion* http://www.vulgum.org/article.php3?id_article=158; la rivista monografica “La società del General Intellect”, *Derive approdi*, Roma; Wark McKenzie, *Un manifesto hacker*, Feltrinelli, Milano, 2005.

¹² SITOGRAFIA: La revisione dei pari. Definizione: http://it.wikipedia.org/wiki/Revisione_paritaria; La topologia di una Rete di citazioni: Stefano Mazzocchi, “Peer Review vs. Citation Network Topology” <http://www.betaversion.org/~stefano/linotype/news/74/>; Recenti articoli critici, da The scientist: Alison McCook. “Is Peer Review Broken?”. *The Scientist*. Volume 20, Issue 2, page 26 <http://www.the-scientist.com/2006/2/1/26/1/>; Richard Gallagher. “Taking on peer review”. *The Scientist*. Volume 20, Issue 2, page 13 <http://www.the-scientist.com/2006/2/1/13/1/>; Google e la revisione dei pari: <http://www.lib.uiowa.edu/hardin/md/notes4.html>; Il fenomeno dei Blog tra informazione e rumore di fondo: Fabio Giglietto,

proposte alternative più interessanti figurano meccanismi di valutazione distribuita delle pubblicazioni, rese disponibili sotto licenze libere (copyleft) mediante progetti di accesso pubblico (Open Access), anche per quanto riguarda le opere umanistiche (si veda ad esempio HyperJournal)¹³.

Così nel marzo 1996 Page lancia il suo spider, un programma per esplorare il web, testandolo con l'algoritmo di ranking che aveva sviluppato.

Il motore di ricerca dei due talenti di Stanford, costruito attorno allo spider, diventa immediatamente di successo tra studenti e ricercatori, acquisendo rapidamente una straordinaria popolarità. Tuttavia per gli amministratori di rete di Stanford il consumo di banda del motore di ricerca comincia a rappresentare un problema significativo. Non solo, ma i proprietari dei siti indicizzati si preoccupavano da un lato dei diritti di copyright dei propri materiali e dall'altro del fatto che il ranking di Google non tenesse in alcuna considerazione i premi ufficiali o gli altri sistemi di valutazione consolidati, fatta eccezione per la quantità e qualità (nel senso di popolarità) delle relazioni che una pagina è in grado di muovere attorno a sé. A Google interessa l'economia relazionale espressa in termini di link e null'altro: "lo spider non si interessa dei contenuti di una pagina".

Una simile affermazione sottintende che il valore di un risultato si costruisce attraverso la valutazione delle relazioni di una pagina con altre, e non imponendo una classificazione arbitraria in base a termini di ricerca. Questa forzatura sarà la chiave dell'efficienza di Google negli anni a venire: i risultati non sono più stabiliti in maniera assoluta una volta per tutte, ma variano dinamicamente a seconda della posizione assunta dalle pagine nel contesto della rete.

Google.com: pubblicità diretta all'interno delle pagine

Page e Brin sviluppano e testano Google per tutto il primo anno e mezzo di vita usando strumenti liberi messi a disposizione delle comunità Free Software e Open Source¹⁴, a partire dal sistema operativo GNU/Linux. Implementano così un strumento fortemente modulare e scalabile, che si può espandere e modificare facilmente anche a pieno regime di utilizzo. Tale struttura modulare è la base del *datacenter*¹⁵ odierno di Google, il Googleplex, che permette la sostituzione o l'aggiunta di hardware e software, la sua manutenzione e aggiornamento senza dover mai bloccare il servizio.

<http://nextmedia.blogspot.com/2004/10/informazione-non-informazione-fra.html>; Citazioni e revisione dei pari, scienze dure e scienze morbide: Francesca di Donato: "Verso uno European Citation Index for the Humanities - Che cosa possono fare i ricercatori per la comunicazione scientifica" <http://bfp.sp.unipi.it/~didonato/ovre/ECIH.html>; L'esaustivo lavoro di Jean-Claude Guédon, "In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing" Université de Montréal, <http://www.arl.org/arl/proceedings/138/guedon.html>
¹³ SITOGRAFIA: L'accesso ai saperi; la libertà dell'informazione e l'accesso aperto alla letteratura scientifica. Segnaliamo in particolare il volume del Gruppo Laser, "Il sapere liberato", Feltrinelli, Milano, 2005, online su www.ippolita.net/laser/. In campo umanistico, il lavoro svolto da Francesca di Donato che riesce a coniugare un approccio qualitativo e non scienziato con la necessità di decentramento degli archivi e delle informazioni e di definizione di un numero minimo di criteri a priori di catalogazione (metadati, ad esempio attraverso XML), da scambiarsi tramite un protocollo comune. Si veda "HyperJournal. Un software open source per l'Open Publishing" (abstract e slides) <http://purl.org/hj/bfp/102/>; Seminario: "Dopo Berlin 3: politiche di accesso aperto alla letteratura scientifica" (Pisa, 16 febbraio 2006) <http://bfp.sp.unipi.it/blog/html/notizie/oapisa.html>. Il sito di HyperJournal: <http://www.hjournal.org/>

¹⁴ Per un'analisi dettagliata di questi rapporti si veda il cap. III

¹⁵ Il datacenter di Google è composto al momento attuale di svariati datacenter sparsi in giro per il mondo che elaborano le richieste effettuate e si occupano di propagare le modifiche delle indicizzazioni. L'aggiornamento dei database, però, così come il calcolo del pageranking, risente della decentralizzazione: i dati vengono di solito allineati nel giro di qualche giorno, o anche più; questo sfasamento dà luogo alla cosiddetta "Google Dance", la danza dei ranking. Per una panoramica di massima, si veda l'articolo <http://www.alverde.net/articoli/stampa.asp?ID=145> e

A metà del 1998 Google serve circa diecimila richieste di ricerca al giorno; il sistema casalingo ammassato nella stanza affittata da Page e Brin è ormai sull'orlo del collasso. Bisognava trovare un sistema di finanziamento per il progetto che andasse al di là dei normali fondi per la ricerca.

L'uscita dall'università viene attribuita all'incontro con Andy Bechtolsheim, fondatore della Sun e consumato *talent scout* dell'IT, che stacca ai due ricercatori un primo assegno da centomila dollari.

La nascita della società coincide con l'assunzione dei primi dipendenti necessari allo sviluppo e alla manutenzione del datacenter, tra cui Craig Silverstein, direttore della tecnologia. Il datacenter di Google si configura sin dall'inizio come un sistema fortemente ridondante in cui le informazioni vengono copiate in più luoghi per minimizzare la possibilità di perdita dei dati (ovvero la carta moneta di un sistema di ricerca) e soprattutto nel quale in ogni momento è possibile aggiungere e togliere i moduli presenti per ottenere una maggiore efficienza. L'approccio tipico della filosofia degli universitari smanettoni di Brin e Page, che recuperano e riadattano hardware a buon mercato e usano software gratuito, si rivela una carta vincente. Le ridotte risorse economiche consentono poi di farne la chiave di volta della loro filosofia aziendale, che sarà basata sull'estrema scalabilità a ogni livello. Il sistema-Google è *scalabile* perché in grado di “crescere” o “decretere” (aumentare o diminuire di *scala*) in funzione delle necessità e delle disponibilità. Non è necessario riprogettare il sistema quando vengono aggiunte nuove risorse (di qualsiasi tipo: macchine, servizi, dipendenti, ecc.): la struttura altamente modulare consente l'acquisizione di nuovi moduli senza soluzione di continuità.

Google Inc. apre le porte il 7 Settembre 1998 a Menlo Park, California. Per la precisione, Larry apre queste porte con un telecomando: gli uffici sono stipati nel garage che un amico subaffitta alla società. Un ufficio-garage spartano, ma dotato di alcuni comfort non indifferenti: una lavatrice, un essiccatore e un idromassaggio. Fin dal principio, la filosofia aziendale di Google è all'insegna della soddisfazione dei lavoratori.

Nel gennaio 1999 Google abbandona definitivamente il campus di Stanford, dichiarando: “Il progetto di ricerca Google è diventato Google Inc. Vogliamo dare al mondo ricerche di qualità superiore e migliori di quelle che ci sono ora, e una compagnia sembra il modo migliore per portare a termine questo obiettivo. Abbiamo cominciato ad assumere altre persone e a configurare altri server per rendere il nostro sistema scalabile (abbiamo cominciato a ordinare i server in pacchi da 21). Abbiamo anche cominciato a lanciare il nostro spider più frequentemente e i nostri risultati non solo sono rimasti veloci quanto prima, ma sono anche diventati ancora più aggiornati. Stiamo assumendo persone di grande talento molto rapidamente e questo ci porta le ultime e migliori tecnologie per quanto concerne il Web” ... Brin e Page poi dedicano qualche riga ai dieci buoni motivi per lavorare per Google, incluso tecnologie, azioni, *snacks*, bevande gratis e la possibilità che milioni di persone: “usino e apprezzino i vostri software”.

Tra il 1998 e il 1999 tutti i motori di ricerca e i siti più popolari del mondo sono affetti dalla sindrome del portale, una vera e propria mania per la costruzione di siti per trattenere i visitatori sulla propria pagina a ogni costo, includendo sempre più servizi, pubblicità e personalizzazioni. Google è l'unico strumento privo di pubblicità e servizi aggiuntivi: solo motore di ricerca, per altro il migliore, il più veloce, e senza legami commerciali o sponsor.

In ogni caso la compagnia non può sopravvivere con la modesta somma di Bechtolsheim, senza generare ancora profitti apprezzabili e continuando a fare ricerca sull'organizzazione e reperimento delle informazioni. Mostrando grande capacità di interlocuzione con le alte sfere della finanza, e

<http://www.motoricerca.info/articoli/google-dance.phtml> sulla “Google Dance”.

ribadendo in continuazione il loro interesse per la ricerca, Brin e Page riescono a mettere d'accordo due delle principali società di finanziamenti della California, che a sorpresa accettano di cofinanziare una singola compagnia: caso più unico che raro, due enormi società di capitale di rischio (*venture capitalism*) vengono convinte a dividere i rischi, e i profitti, dell'eventuale affare. Il 7 giugno 1999 Google può annunciare che la Sequoia Capital Sequoia Capital e la Kleiner Perkins Caufield & Byers gli hanno concesso un finanziamento da 25 milioni di dollari¹⁶.

Mentre aumentano gli assunti alla Google Inc., i due presidenti cercano un modo per far fruttare in senso economico la mole di dati indicizzata. Inizialmente tentano di vendere il loro servizio di ricerca come OEM (*Original Equipment Manufacturer* – Prodotti forniti senza i supporti originali)¹⁷ a portali, ma i risultati sono alquanto scarsi. Il modello di business che si dimostra più adatto alla nuova società è invece la pubblicità diretta all'interno del motore di ricerca, attraverso il conteggio del numero di visitatori che accedono ai siti dai link promozionali a pagamento. Questo modello economico, CPM (Costo Per Mille Impressioni – in inglese CPT, *Cost per Thousand Impression*)¹⁸ è strutturato in modo da essere il meno invasivo possibile per l'utente; infatti, invece di puntare sull'estrema visibilità delle inserzioni pubblicitarie, vengono mostrati pochi link attentamente selezionati in cima alla ricerca. Questi collegamenti hanno colori e caratteri diversi da quelli non commerciali, ma tendono a non disturbare in alcun modo l'attività di ricerca dell'utente.

Il self-service pubblicitario: oltre la bolla della new-economy

Un modello economico basato sulla comparsa di semplici link da sponsorizzare a lato dei risultati delle ricerche è poco vantaggioso in termini di profitto; a questo punto la pianificazione aziendale a lungo termine di Google necessita di un salto di qualità. I presidenti si guardano attorno alla ricerca di soluzioni più interessanti e proficue, e inciampano su Goto, società fondata da Bill Gross¹⁹, oggi di proprietà di Overture/Yahoo!.

Il business di Goto consisteva nel mescolare risultati effettivi e quelli sponsorizzati, fatturando agli inserzionisti soltanto i click compiuti dagli utenti sui loro indirizzi: un modello noto come CPC (*Cost Per Click*).

Questo metodo era particolarmente innovativo rispetto ai precedenti. I link sponsorizzati comparivano infatti in base alle ricerche dell'utente, ampliando in questo modo al massimo le possibilità di portare a buon termine la transazione economica, contabilizzata come click verso il sito pagante. Google cerca di avviare una partnership con Goto, ma il deciso rifiuto del presidente li obbliga a implementare internamente uno strumento simile. All'epoca tutti i portali (ad esempio Excite, Lycos, Infoseek, AltaVista e Yahoo!) utilizzavano il modello CPM, e il CPC era quasi un desiderio proibito. Ciò dimostra come un tentativo fallito di comprare la tecnologia ritenuta migliore per il proprio business comporti la necessità di implementarla in maniera autonoma.

¹⁶ Le due società detengono tuttora le quote di maggioranza del capitale di Google.

¹⁷ Un OEM è una società che costruisce componenti utilizzati in sistemi venduti da un'altra società come, ad esempio, un rivenditore. Società che vendono prodotti con componenti provenienti da OEM sono comuni nell'industria elettronica ed informatica. Tipicamente l'OEM si contrappone al RETAIL, ovvero il prodotto viene fornito a terzi senza i supporti originali (dischi d'installazione, manuali, ecc.), integrato nel prodotto del rivenditore.

¹⁸ CPM, CPC, ecc.: i modelli di business pubblicitario online sono sempre più complessi e raffinati. Per una rapida introduzione si veda (in inglese) http://earlstagevc.typepad.com/earlstagevc/2005/08/web_20_needs_bu.html; in italiano, segnaliamo il blog di Mauro Lupi, di AdMaiora: chi meglio di un pubblicitario può parlare di pubblicità (anche se non certo in maniera critica)?

¹⁹ La storia di Goto è emblematica, perché mostra come anche nel panorama del web si tenda sempre più alle megafusioni tipiche del capitalismo informazionale. Una buon riassunto in italiano fino al 2004 si trova a questo indirizzo: <http://www.motoridiricerca.it/goto.htm>; su wikipedia, risorse a proposito del ruolo commerciale giocato da Goto/Overture nel business di Yahoo! http://en.wikipedia.org/wiki/Yahoo!_Search_Marketing.

Nel marzo del 2000 esplode la bolla del Nasdaq, trascinando con sé tutti i sogni di gloria delle dot com. Crolla anche il modello CPM, ovvero l'illusione che tonnellate di banner pubblicitari a “migliaia di impressioni”, spesso decontestualizzati perché presenti su siti del tutto estranei al business di riferimento dell'inserzionista, sarebbero stati una fonte inesauribile di introiti. Google si ritrova proprio allora a fare i conti con la necessità di monetizzare le proprie tecnologie di ricerca.

Nell'ottobre 2000 entra in scena AdWords, una sorta di self-service pubblicitario in cui gli inserzionisti possono scegliere le parole di ricerca a cui associare il loro link commerciale: un'implementazione della “pubblicità per parole-chiave” (*keywords advertising*) ideata da Goto.

Google sopravvive perciò al crollo del Nasdaq; o meglio, grazie alla sua natura di società privata non ancora quotata in borsa, ne approfitta per assumere tutti i talenti in fuga dalle macerie delle altre dot com. A metà del 2000 Google serve 18 milioni di ricerche al giorno e il suo indice di documenti conta un miliardo di unità; sei mesi dopo le ricerche sono 60 milioni.

Stile, morfologia e moltiplicazione dei servizi

Agli inizi del 2000 Google si trova perciò con pochi concorrenti e dà inizio a una nuova fase di innovazione, cominciando a offrire molti nuovi servizi²⁰ agli utenti.

Ogni nuovo servizio è parte di un mosaico complesso, in costante ridefinizione, con ramificazioni in ogni ambito dell'IT. Al momento Google offre: diciassette tipologie di servizi di ricerca su database di immagini, blog, notizie, libri, mappe, video, servizi finanziari, sui documenti presenti sul proprio computer e molto altro ancora. Due servizi sono legati allo sviluppo di applicazioni e alle novità in cantiere nei laboratori di Google. Sei i servizi di comunicazione: Email, VoIP, messaggistica istantanea, Gruppi di discussione, condivisione di fotografie, servizi di traduzione. Tre servizi riguardano dispositivi mobili (telefoni cellulari, PDA, ecc.); infine, un servizio propone software consigliati da Google. Non passa giorno senza che se ne aggiungano di nuovi.

Anche l'utente più sprovveduto si rende conto facilmente della potenza di questi strumenti. Oggi è possibile digitare un indirizzo o un numero di telefono e Google fornisce tutti i dati necessari a contattare e localizzare una persona o una cosa; si possono memorizzare le proprie preferenze di ricerca rendendo l'uso continuativo del motore di ricerca di una facilità sconcertante; se si sbaglia a digitare una parola nella ricerca, Google suggerisce la ricerca giusta da fare con un meccanismo di correzione ortografica molto avanzato in grado di “imparare” da ogni ricerca.

Nel 2001 Google lancia “Google immagini”, un motore di ricerca per immagini che nel giro di poche settimane diventa una delle fonti principali di produzione grafica fai-da-te e rappresenta uno dei più grandi database di immagini distribuite sul web. Intanto la compagnia acquisisce da Deja.com l'archivio Usenet, ovvero i messaggi dei newsgroup a partire dal 1995, che con oltre 650 milioni di post è la “memoria storica” dell'Internet ante-web, quando i gruppi di discussione erano la linfa vitale della Rete. Nell'aprile 2001 Usenet diventa “Google gruppi”, con un'interfaccia piuttosto semplice e accattivante per seguire facilmente complessi argomenti di discussione. Tutti questi servizi sono in continuo ampliamento e miglioramento.

Dal 2001 in avanti i servizi nascono a ripetizione, senza nessuno scopo evidente né ritorni economici immediati, come se Brin e Page si divertissero a mostrare che un datacenter sconfinato è in grado di elaborare quasi qualsiasi sogno tecnologico. L'esempio più eclatante è il potentissimo

²⁰ Accenneremo qui solo ai servizi più interessanti; per un elenco esaustivo dei servizi offerti da Google, si veda <http://www.google.com/intl/en/options/> (i servizi disponibili in italiano e nelle altre lingue sono meno, ma in rapida ascesa: <http://www.google.it/intl/it/options/>).

Google Maps per la Terra, Marte o per la Luna (2005 e 2006), un set di software messo a disposizione gratuitamente per visualizzare con immagini satellitari una piantina dettagliata o almeno una ricostruzione fotografica della superficie dei pianeti. “Google Directory”, che presenta i contenuti Open Directory di Dmoz.com, organizzati da agenti umani con un sistema cooperativo e decentrato di open publishing, è già da tempo consultabile dalla sempre meno graficamente scarna pagina di Google.

Nel 2005 vedono la luce Google News, che mette l’enorme database di informazioni di Google al servizio della produzione giornalistica, e Gmail, che offre 1 Gigabyte di spazio mail per ogni utente. Con un meccanismo a inviti che immediatamente crea una rete di relazioni tutta interna al mondo Google, vengono zittite le proteste circa la privacy della propria casella di posta con un lapidario “è un buon prodotto, i pregi sono maggiori dei dubbi e il tempo lo proverà”. Nessun concorrente è in grado di mettere a disposizione gratuitamente tanto spazio. Con Gmail ogni attore coinvolto è controllato dal punto di vista dell’uso del servizio, dal momento l’enorme disponibilità di spazio incita a lasciare le proprie mail sui server di Google. Inoltre, poiché la propagazione del servizio avviene mediante inviti, che ogni utente può utilizzare a proprio piacimento, vengono offerte a Google informazioni essenziali sui propri conoscenti e amici: un progetto pervasivo e mirato di *data mining*.

Si arriva poi al progetto di Google Scholar (Accademia), un motore di ricerca ancora sperimentale, in fase beta (la cosiddetta fase di *beta-testing*), che permette di individuare testi della letteratura accademica come articoli sottoposti a revisione, tesi di laurea, di dottorato, volumi accademici, *preprint*, sommari, rapporti tecnici, recensioni di tutti i settori della ricerca accademica. Infine Google Library, che ha l’ambizione di mettere a disposizione on line tutti i libri del mondo in formato digitale, stringendo accordi con biblioteche di tutto il mondo per la digitalizzazione dei volumi, ma anche con gli editori interessati. Un sogno che solo il datacenter di Google promette di poter realizzare: la biblioteca digitale globale accessibile dalle pagine di Google, in aperta rottura con gran parte degli editori statunitensi, riuniti nella Association of American Publishers, che temono di veder calare drasticamente i propri profitti; già nel 2005 la AAP aveva richiesto di bloccare per sei mesi la digitalizzazione delle opere protette da copyright in attesa di spiegazioni convincenti su Google Library. Ma l’apparenza inganna e, nonostante l’opposizione dei fautori del copyright, non si tratta di un’iniziativa per la libera circolazione dei saperi, ma solo di un cambio di monopolio dell’informazione, che passerebbe da alcuni grandi editori a Google. Infatti, come per ogni sogno, l’incubo è dietro l’angolo: un singolo soggetto privato, Google, aspira a decidere cosa faccia parte della sfera dell’informazione collettiva, rendendolo disponibile tramite formati proprietari. In reazione a questo progetto è nata la Open Content Alliance, sostenuta da Internet Archive, una fondazione no profit, e da Yahoo!: l’obiettivo è la disponibilità totale del materiale acquisito grazie all’utilizzo di formati aperti.

Parallelamente alla nascita di nuovi servizi, Google si rivela estremamente abile nel mettere pienamente a profitto l’economia relazionale, generata dallo sfruttamento in senso commerciale dei dati indicizzati.

Nel 2004 viene lanciato AdSense, cioè la possibilità per chiunque di ospitare sul proprio sito alcuni link commerciali suggeriti da Google in base agli argomenti trattati e ad alcune parole chiave. Gli introiti derivanti dai click su questi link vengono suddivisi tra Google e i proprietari dei siti ospitanti. La novità di questo approccio consiste nel mettere a frutto la rete di fiducia basata sugli utilizzatori del sito: Google non è più sul sito di Google, ma è su chiunque voglia averne la “finestra”, e questo piccolo spazio sarà sempre riempito con dati validi ed interessanti, come sempre con Google, anche se ora saranno suggerimenti commerciali. In pratica AdSense dà sostanza a una “Rete di Google”, una rete che incrocia i dati degli utenti e le relazioni fra di essi per i suggerimenti

pubblicitari. Nelle parole di Google, è una rete “composta da siti e prodotti connessi a Google da un rapporto di partnership per la pubblicazione di annunci mirati AdWords sul proprio sito o prodotto”²¹; ovviamente, anche il sistema AdSense fa parte della “Rete di Google”!

Naturalmente, fatta la rete, bisogna fare i soldi. Non a caso, sempre nel 2005, Google sperimenta un “ritorno” al modello CPM attraverso l’offerta AdWords con la metodologia “targeting per sito”: gli inserzionisti tornano a pagare per visualizzazioni (*impression*), ma con un meccanismo ad asta, e non per numero di click sul proprio banner. Possono scegliere in maniera estremamente particolareggiate quali caratteristiche devono avere gli utenti che visualizzeranno l’inserzione: nazionalità, area geografica di provenienza, lingua, ecc. Ma soprattutto, le visualizzazioni avvengono solo all’interno della “Rete di Google”. Questa strategia interessa soprattutto coloro che desiderano visibilità e promozione di un marchio, piuttosto che utenti a cui vendere un prodotto, ovvero privilegiano una forma indiretta di marketing rispetto a forme dirette. Nello specifico, si tratta di investire nella cosiddetta “consapevolezza del marchio” (*brand awareness*) invece che su prodotti specifici indicati da parole chiave (come es. il *keyword advertising* del modello CPC).

Vi è una sordida specularità tra il modo di gestire i propri prodotti immateriali e l’organizzazione dei dipendenti, dell’ambiente di sviluppo progettuale, perfettamente adeguato al criterio di scalabilità modulare che fonda la filosofia aziendale di Google: la crescita infinita è necessaria affinché il sistema non crolli. Il numero di utenti che compiono ricerche tramite Google e affidano ai suoi server i propri dati deve crescere continuamente, affinché cresca di pari passo il numero di inserzionisti che pubblicizzano nella “Rete di Google” i propri prodotti. Sono necessari sempre nuovi servizi, nuove macchine per gestirli, nuovi dipendenti per implementarli, nuovi utenti per utilizzarli, nuovi inserzionisti per renderli proficui... Ogni nuovo “pezzo” del sistema verrà introdotto come nuovo modulo, in un ciclo di accumulazione senza fine: accumulazione di dati e accumulazione di cervelli e accumulazione di utenti e dei loro dati, bontà nel trattamento dei dati, bontà nel trattamento dei dipendenti, bontà nel trattamento degli utenti e dei loro dati, archiviati nel *datacenter* di Google. Sempre con l’imperativo della rapidità e dello sviluppo.

Brin e Page non nascondono le loro ambizioni: “Perché lasciare che i nostri dipendenti cerchino di fondare la propria società e poi doverla acquistare quando possiamo finanziarli per restare da noi e fare quello che avrebbero fatto comunque?”. Infatti il GooglePlex²², la sede operativa per i lavoratori di Google a Mountain View, è una specie di campus universitario in cui le persone sono costantemente coccolate e ai dipendenti viene persino lasciato un giorno alla settimana per lavorare ai propri progetti; questi progetti vengono poi presentati al “Google Duo” che premia con denaro e sostegno aziendale i talenti più promettenti.

Google, il gigante buono, va in borsa

“Don’t be evil”, non essere cattivo, si può fare qualsiasi cosa, basta non fare del male. Questo il motto del capitalismo umano di Google²³. Ma già si registrano molti distinguo in questa campagna a favore del Bene: cause legali di vario genere, sospetti di frode, oscuramento di siti web, ecc.²⁴.

²¹ Definizione della Rete di Google secondo Google:

<https://adwords.google.it/support/bin/answer.py?answer=6104&topic=82>

²² Si veda il servizio fotografico del Time su Googleplex: http://www.time.com/time/photoessays/2006/inside_google/

²³ La filosofia della Google Inc.: <http://investor.google.com/conduct.html>

²⁴ Solo tre esempi della mole di informazioni al riguardo: un articolo di Wired, Google vs. Evil, www.wired.com/wired/archive/11.01/google_pr.html; un simpatico raccontino di Paul Ford, The Banality of Google (invece della notoria banalità del male...), <http://www.ftain.com/GoogleIP.html>; e un intero portale sul motto, esteso a un’auspicato ritorno della fiducia negli affari, nella politica, nei media, con tanto di manifestazioni perché Google rimanga fedele ai suoi principi, <http://www.dontbeevil.com/>

Nel dicembre del 2002 Google conta oltre 1000 dipendenti e possiede oltre 10000 computer. Indicizza oltre 4 miliardi di documenti e i ricavi netti (anche se rivelati con molta cautela) sono di circa 185 milioni di dollari. Di fronte a dimensioni simili gli investitori cominciano a pretendere maggiore chiarezza e controllo, imponendo una struttura societaria più credibile, con a capo non due ingegneri geniali ed eccentrici, ma un manager dalle provate capacità imprenditoriali.

Dopo una serie di incontri infelici e di scontri verbali, il ruolo di amministratore delegato (CEO, Chief Executive Officer) della Google Inc. viene affidato a Eric Schmidt (già in forza a Sun e poi a Novell). I due giovani prodigi continuano a prendere decisioni a tutto campo, ma la mossa diplomatica si rivela essere anche una buona scelta economica, dato che l'arrivo di Schmidt coincide con il primo trimestre in attivo della compagnia, che dimostra così di aver finalmente girato con successo la boa della monetizzazione dei propri prodotti.

Page e Brin avevano dilazionato la necessità di diventare una società quotata in borsa il più a lungo possibile, temendo che la necessità di un rendiconto pubblico delle proprie prospettive e dei propri guadagni avrebbe reso il loro lavoro molto più difficile, mettendo fra l'altro Google in condizione di essere monitorato meglio e più facilmente e quindi attaccato dai concorrenti sul mercato.

Ma ormai, dopo AdSense, nel 2004, per quanto nelle parole di Page “Google non è una compagnia convenzionale. E non abbiamo nessuna intenzione di diventarlo”, per ragioni legali il nuovo colosso diventa esattamente come tutte le altre società americane e si quota in borsa.

Proprio a ridosso dell'offerta pubblica, Yahoo! e altri concorrenti rovesciano decine e decine di cause per violazione di brevetto, di copyright e altro su Google, per rovinarne il nome in vista della vendita pubblica delle azioni.

Wall Street è sul punto di abbassare il prezzo minimo delle azioni previsto, considerata la difficoltà dell'operazione, ma Page e Brin risolvono la principale contesa con Yahoo!, pagando al colosso di Filo e Yang un compenso in azioni e sistemando la controversia brevettuale. Il duo quindi, nonostante il parere contrario della Borsa, procede con l'offerta pubblica di azioni nel bel mezzo dell'agosto 2004, anche se con un ribasso del prezzo di oltre 20 dollari ad azione.

In un solo giorno di contrattazioni sul mercato le azioni di Google passano dalla base di 85 dollari a 100 dollari, con un guadagno netto di 1 miliardo e mezzo di dollari. Un anno dopo le azioni raggiungono il valore di 400 dollari con un aumento di oltre il 300 per cento. Per Google Inc. tutto sembra andare a gonfie vele, nel mondo meraviglioso dove nessuno è cattivo, tutti guadagnano, nessuno si fa mai male. Certo, di fronte a simili cifre, qualsiasi minima oscillazione al ribasso significa milioni di dollari in fumo, come si è visto per una flessione di sette punti percentuali nel marzo 2006. Google Inc. è ormai un gigante fra i giganti della Borsa mondiale, e se trema, in molti tremano con lui.

Google Inc.: il monopolio della ricerca

Nell'autunno 2004 Page e Brin, in viaggio sul loro charter privato, vengono a sapere che il colosso AOL (America Online, il principale provider statunitense) ha firmato il contratto per il proprio motore di ricerca con Yahoo!. I neotrentenni cambiano rotta, dirigono su Londra e convincono AOL a strappare il contratto già firmato e passare a Google con un'offerta irresistibile da 50 milioni di dollari. Non è esattamente il comportamento aperto e gentile che ci si aspetterebbe dal gigante buono intento a sconfiggere i cattivi monopolisti, ma del resto, gli affari sono affari, anche per i due bravi ricercatori di Mountain View.

Nel frattempo i guadagni di Google Inc. in un periodo di 5 anni sono cresciuti più del 400 per cento, rendendolo il diretto concorrente di Yahoo! e Microsoft, non solo in termini di movimentazione finanziaria, ma soprattutto per la popolarità e quindi per il dominio culturale sull'immaginario. Milioni di utenti della Rete usano le pagine di Google come punto di partenza per le loro navigazioni, si fidano dei risultati ottenuti con gli strumenti messi a punto a Mountain View. Ormai Google è sinonimo di web, quando non addirittura di Internet. Google è in grado di sfruttare al meglio la rete relazionale dei suoi utenti e riesce a spremere ogni centesimo da milioni di piccoli inserzionisti, cosicché nel 2005 i dati parlano di circa sei miliardi di dollari di ricavi pubblicitari (Yahoo! è fermo a quota 4,6 miliardi di dollari).

Il primo pantano nel quale Google rimane invischiato è il rapporto fra le sue ricerche e la legislazione sui marchi registrati (trademark, TM). Emblematiche le cause intentate contro Google dalla *Geico* e dalla *American Blind & Wallpaper Factory*²⁵. In entrambi i casi le società accusano il servizio AdWords del motore di ricerca di vendere termini sotto il loro trademark a concorrenti, violando le leggi sul marchio registrato. La questione è spinosa, perché le aziende chiedono a Google di impedire che appaiano inserzioni di proprie rivali o comunque non da loro autorizzate quando gli utenti cercano termini come “geico”, “american blind” o “american wallpaper”. Se questo concetto passasse in tribunale, una delle più importanti fonti di reddito di Google e dei suoi partner verrebbe ampiamente messa a rischio, perché consentirebbe a chiunque possedeva un trademark di inibirne l'uso per AdWords o di denunciare Google per il suo utilizzo. Louis Vuitton in Francia intenta una causa simile, e vince. La risposta ufficiale di Google è che i responsabili della violazione, al massimo, sono gli inserzionisti e non certo Google che in questo caso è un medium neutro, e che “limitare la vendita di termini registrati equivale a limitare la libertà di espressione”. Una posizione forte da parte della società e apparentemente più che condivisibile.

Tuttavia proprio la società di Mountain View cade nello stesso tranello teso alle società sulla libertà d'espressione, violando il vincolo di fiducia che la lega a molti utenti e che costituisce una delle sue risorse fondamentali. Google si è sempre trincerata dietro la presunta oggettività degli algoritmi di ricerca e l'obiettività delle macchine nel fornire risultati senza manipolare in alcun modo i risultati. Eppure, poco prima del causa con American Blind, Google aveva deciso di rimuovere alcune AdWords acquistate dal gruppo di attivisti Oceana²⁶. La “colpa” di Oceana era stata la pubblicazione di denunce ecologiste circa l'attività di un grosso investitore di Google, la Royal Caribbean Cruise Lines, usando come keywords dei propri annunci parole come “cruise vacation” o “cruise ship”, normalmente utilizzate per cercare pagine riservate a vacanze in barca o simili. La motivazione ufficiale sostiene che Google è un canale neutro, e non vuole essere veicolo di campagne propagandistiche atte a discreditarne altre aziende. Evidentemente in questo caso la libertà di parola non era così importante.

Inoltre, proprio il giorno dell'udienza per il caso American Blind, nel distretto di San José dove si teneva il processo i risultati di Google si mostravano misteriosamente differenti da ogni altro luogo nel mondo. Per la prima volta Google veniva colto a manipolare i risultati per fini diversi da quelli del “miglior risultato di una ricerca”. Il parere positivo della corte rispetto al caso con Geico (in tutto e per tutto analogo a quello con American Blind) fatica a cancellare questa realtà.

Il caso più recente e noto rimane l'entrata in Cina del colosso della ricerca. Per cercare di conquistare il crescente mercato cinese, Google si è infatti piegato pubblicamente per la prima volta

²⁵ Al momento, solo Geico ha perso, negli altri due casi Google è stato condannato a un'ammenda. Segnaliamo qui alcune fonti autorevoli in italiano su queste vicende giudiziarie: il caso American Blind and Wallpaper Factory: <http://punto-informatico.it/p.asp?i=46751>; Google condannato a risarcire Louis Vuitton in Francia <http://punto-informatico.it/p.asp?i=45643>; Google vince contro Geico <http://webnews.html.it/news/2514.htm>

²⁶ Il gruppo no-profit Oceana si batte per la qualità delle acque: www.oceana.org

alle richieste di censura, rendendo inaccessibili i siti indicati dall'autorità censoria cinese agli utenti che effettuano le ricerche dal territorio cinese. Ma già nel 2002 uno studio di Harvard aveva mostrato che Google censurava 113 siti, nelle versioni locali francese e tedesca (Google.fr e Google.de). Google rispose che era vero, ma che erano stati rimossi solo su segnalazione di enti governativi, autorità di polizia e altri soggetti, e solo dopo attenta analisi dei loro contenuti: molti siti erano di ispirazione razzista, altri ispirati dal fondamentalismo religioso. Qualcuno polemizzò sul fatto che la tanto sbandierata trasparenza di Google cominciava a vacillare: bisognava rendere noto agli utenti l'applicazione di una censura che rischiava altrimenti di rimanere “nascosta”; altri fecero notare come la colpa non fosse certo di Google ma fosse insita negli ordinamenti giuridici che prevedono la possibilità di trascinare in tribunale un'azienda per la presenza di un link e niente più sulle proprie pagine²⁷. In quel caso, è evidente che Google, per evitare possibili conseguenze legali, preferì rimuovere i link valutando caso per caso. Da notare per inciso che la fine del diritto di link è destinata a pesare sulle libertà digitali in misura sempre maggiore: chi deciderà cosa è lecito censurare? L'ennesima super-autorità, magari sovrastatale? Il diritto del più forte, ovvero nel sistema di mercato il diritto di chi paga (o vale) di più? Oppure i ricatti dei fondamentalismi locali, soprattutto di matrice religiosa, che minacciano ritorsioni ogni qual volta viene lesa la loro particolare visione del mondo da qualche sito “sovversivo”? La portata del problema è vasta quanto l'importanza della libertà d'espressione e certo non si può risolvere nelle aule di tribunale. Il clamore sollevato dal caso cinese è però legato a una censura ordinata da un governo: ma nonostante questo cambiamento di scala, Page e Brin non hanno cambiato idea, coscienti dell'importanza di un mercato rivolto a un quarto della popolazione mondiale.

Secondo Google presto il mondo sarà un enorme indice in cui le risorse digitali corrisponderanno completamente e perfettamente alla definizione della realtà circostante; ogni indice potrà essere raccolto e offerto nel modo migliore attraverso un algoritmo nei risultati di una ricerca: e Google è nella posizione perfetta per essere lo strumento per gestire questo indice.

Ma, a parte l'ovvia e banale obiezione che i mondi reali e quelli digitali, per quanto intrecciati, non corrispondono affatto, da un punto di vista tecnico non esiste l'algoritmo perfetto quando si parla di reperimento delle informazioni in Rete; soprattutto, nulla è neutrale nell'ambito tecnologico, e tanto meno quando si gestiscono i dati di individui reali su Internet.

A prescindere dalle alleanze che verranno strette, la convergenza tecnologica che si sta realizzando ci sembra una direzione nuova della visione di Google come unico punto d'accesso, di gestione e di mediazione dei dati digitali. Si delinea insomma la distopia di Google come aspirante grande fratello, pericolosa e affascinante come ogni scontro epocale: il web è il nuovo territorio di competizione selvaggia verso l'affermazione di un nuovo standard di comunicazione. Uno standard paradossalmente “personalizzato” con servizi e proposte differenziate sulla base dei gusti degli utenti: ormai da alcuni anni la parola d'ordine è “personalizzazione di massa”. Un ossimoro, certo, che rivela però l'importanza della posta in gioco, un vero e proprio slittamento di paradigma, dal consumismo seriale dell'industria di massa al consumismo personalizzato, che viene spacciato come frutto della “libertà di scelta”. Al di là delle facili retoriche, la via di fuga può essere solo una nuova direzione nelle nostre scelte: non si tratta di usare o non usare Google o i suoi servizi, ma di scegliere diversamente il modo di mettere informazioni su Internet e di imparare a collegarle in modo nuovo, con percorsi nuovi e più interessanti per ciascuno di noi²⁸.

²⁷ La versione integrale del rapporto, a firma di Jonathan Zittrain e Benjamin Edelman, si trova all'indirizzo <http://cyber.law.harvard.edu/filtering/google/>

²⁸ In tal senso, una fonte ricca di riferimenti e proposte, indispensabile anche per la stesura di questo testo, è Ferry Bite, Claudio Parrini, “I motori di ricerca nel caos della rete - kit di sopravvivenza”, ShaKe, Milano, 2001.

Da alcuni anni Google sta comprendendo a sue spese (e sulla pelle dei suoi utenti, naturalmente) che la purezza non è una qualità del mondo e tanto meno del mondo degli affari, che la bontà totale è una sciocchezza in genere, tanto più per una società che mira a fare utili, e soprattutto che la neutralità è una strada sempre più difficile da percorrere nella guerra dei motori di ricerca. Del resto, sarebbe bene ricordare che le nazioni da sempre neutrali, come la Svizzera, sono anche quelle armate fino ai denti. Vediamo allora quali sono le armi “buone” di Google, che ne hanno fatto un fenomeno di rilievo mondiale.

II. BeGoogle

La fuga dei cervelli verso Google: la guerra per il controllo del web

“Ucciderò quel fottuto di Google! Eric Schmidt è un fottuto e lo seppellirò, come ho già fatto”²⁹: così è esploso nel maggio 2005 Steve Ballmer, amministratore delegato di Microsoft, alla notizia dell’assunzione da parte di Google di Kai-Fu Li, dirigente di punta della casa di Redmond e suo uomo chiave in Cina (Kai-Fu Li aveva infatti realizzato il motore MSN Search per i cento milioni di utenti cinesi). Un’invettiva rivolta evidentemente all’attuale Amministratore Delegato di Google, ex top manager di Sun Microsystems e di Novell, aziende che Microsoft ha in passato affrontato sia sul mercato che in aule di tribunale. Kai-Fu Li è diventato il responsabile di un laboratorio di ricerca aperto nei dintorni di Shanghai.

Microsoft ha immediatamente avviato un procedimento legale nei confronti dell’ex dipendente e soprattutto nei confronti di Google. Questo perché Kai-Fu Li avrebbe violato, grazie all’aiuto determinante del re dei motori di ricerca, dei riservatissimi accordi contrattuali stipulati con il colosso di Redmond. Essendo un direttore esecutivo, Kai-Fu Li conosce segreti commerciali e industriali e, secondo i legati di Microsoft, non esiterà ad utilizzare tecnologie e know-how socioeconomico per massimizzare i guadagni della società concorrente. L’uomo dello scandalo costa caro, circa due milioni e mezzo di dollari l’anno di “salario” e 20.000 azioni dell’azienda. Cifre esorbitanti che danno la misura dell’altissima posta in gioco, non solo nel mercato cinese.

Dopo oltre un anno, la contesa giudiziaria fra i due colossi si conclude con il patteggiamento di un’intesa bilaterale nel dicembre 2005, appena un mese prima di finire davanti al giudice: un accordo privato assolutamente confidenziale. Forse sono state pagate grosse somme di denaro, o forse Microsoft ha vincolato Kai-Fu Li a non rivelare assolutamente qualsiasi tipo di segreto.

Questa storia è solo una delle più curiose ed emblematiche di una tendenza in atto da qualche anno: Kai-Fu Li infatti è solo l’ennesimo dirigente passato a Google ovvero “la società che più assomiglia a Microsoft”, come ha tuonato Bill Gates, ogni giorno più incerto se demonizzare i due studenti prodigio – rafforzando indirettamente la loro nomea di protagonisti “buoni e generosi” del mondo dell’IT – oppure fingere che non esistano affatto, che non siano interlocutori degni di nota.

In realtà, Bill Gates sa bene quanto le trasmissioni di manager, soprattutto nel settore tecnologico, comportino l’assunzione del “core business” della società dalla quale il manager proviene: Microsoft ha usato la stessa strategia contro i propri rivali in diverse occasioni. La tattica commerciale di rubare i segreti industriali, meccanismi di produzione e gestione delle risorse attraverso l’assunzione di elementi di spicco delle società concorrenti è sempre stata nelle corde della competizione industriale; si fa sempre più diffusa e frequente nell’epoca dell’economia informazionale.

La scelta del management da parte di Brin e Page è un chiaro indice degli obiettivi di Google: diventare la piattaforma del web più completa e maggiormente personalizzabile, promuovendo una personalizzazione dei servizi, sostenuta da un immenso bacino di informazioni. In sostanza, Google spinge sull’acceleratore per catalogare ogni genere di informazione digitale, dai siti web ai forum di discussione, dalle immagini ai video, dalle email ai blog, senza che si intraveda alcun limite all’orizzonte; ciò significa fare la guerra a Microsoft, che al momento rimane nonostante tutto il maggior concorrente nel settore, con il suo browser Internet Explorer, il suo portale MSN, il servizio di posta elettronica Hotmail, ecc.

Gli interessi delle due società si stanno sovrapponendo in misura crescente: entrambe mirano a

²⁹ Dichiarazione giurata di Marc Lucovsky, dipendente di Microsoft, durante una seduta del tribunale dello Stato di Washington, per il caso Kai-Fu Li; Lucovsky avrebbe messo in risalto anche una presunta “sete di vendetta” da parte di Microsoft. Le notizie sono tratte da vari articoli di Punto Informatico, <http://punto-informatico.it/>

essere elementi di mediazione indispensabili per ogni attività informatica. Microsoft è riuscita a imporre il suo predominio grazie ai sistemi operativi e ai programmi per ufficio, che attualmente sono uno standard nel mondo del lavoro e nell'home computing; da parte sua, Google si pone ormai da tempo come mediatore globale dei servizi web, quali la ricerca di informazioni (che è il suo *core business*) in tutte le sue declinazioni, ma anche servizi aggiuntivi come la posta elettronica. Schematizzando in un'opposizione netta, Microsoft domina da anni grazie a dei prodotti che portano a dei servizi, Google sta cercando di dominare grazie a dei servizi che necessitano di prodotti di supporto.

La competizione quindi si basa sulle scelte degli utenti e sugli standard futuri che Google cerca di affermare. Utilizzare programmi via web limitando al solo browser i requisiti per accedere ai servizi significa sottrarre mercato a chi ha sempre investito nel prodotto e nella creazione di nuove forme di architetture informatiche basate sul web. Lo stesso vale nei mercati economici: si passa da una economia all'ingrosso (Microsoft), dove si cerca di vendere licenze dello stesso programma/prodotto, a una economia totalmente personalizzata per ogni utente, dove i prodotti sono reperibili sulla rete.

Code lunghe nelle reti: dentro l'economia della ricerca, Google vs. Microsoft

Google sfrutta, secondo l'argomentazione chiave dell'ampia trattazione di John Battelle³⁰, l'ascesa dell' "economia di ricerca". Nel saggio *The Search*, Battelle, giornalista tra i fondatori di *Wired*, sostiene che il futuro del commercio online sarà legato alle ricerche personalizzate pagate dagli utenti. Google, che gestisce il più grande database delle "intenzioni di ricerca" dei navigatori, si trova nella posizione ideale per soddisfarle, grazie alla sua rete capillare, costituita da una piattaforma pubblicitaria eccezionalmente efficace (AdWords) e da un network di inserzionisti che ormai conta milioni di siti in tutto il mondo (AdSense). La scommessa di Google è quella di soddisfare ogni bisogno/desiderio che gli utenti esprimono nelle loro ricerche, fornendo servizi per il nuovo consumismo "a misura d'individuo": ciascun utente/cliente troverà proprio quello che cerca, il prodotto studiato appositamente per lui. L'esempio più noto di "personalizzazione di massa" è quello dei servizi di commercio online offerti da Amazon, in grado di guadagnare molto più dalla vendita di singole copie di un libro o di un cd piuttosto che dai picchi di vendita di centinaia o migliaia di copie dei best-sellers. I tantissimi clienti che comprano online volumi non particolarmente popolari rappresentano una miriade di "eventi" che si verificano raramente, o addirittura una sola volta; riuscire a soddisfare queste "ricerche personalizzate" costituisce il punto di forza della distribuzione di Amazon. Infatti per una libreria tradizionale, basata su magazzini, giacenze, ordinativi minimi, è impossibile ottenere la disponibilità di milioni di titoli come Amazon; i proventi maggiori derivano perciò dalle novità e dai titoli più venduti. La fornitura di una singola copia per soddisfare la richiesta di un solo cliente risulta antieconomico per una libreria tradizionale, ma è possibile e redditizio per Amazon, che sfrutta appunto l'"economia della ricerca" dei mercati online.

Questo tipo di mercati viene chiamato in gergo economico a "coda lunga"³¹. La teoria statistica

³⁰ John Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*, Portfolio Hardcover, 2005 (trad. it. Google e gli altri, Come hanno trasformato la nostra cultura e riscritto le regole del business, Cortina, Milano, 2006).

³¹ Le code lunghe: Le implicazioni commerciali della Coda Lunga (Long Tail) – nel contesto dell'e-commerce la definizione è diventata d'uso comune – sono state rese famose da Chris Anderson, redattore capo di «Wired», in un articolo dell'ottobre 2004 che ha scritto per la sua rivista. Anderson ha poi pubblicato nel 2005 il volume *The Long Tail*. Come Battelle, Anderson ha tenuto un blog sulla genesi del suo libro a www.thelongtail.com e il termine è ormai diventato di uso corrente. Le code lunghe non sono un'invenzione di Anderson: si tratta di un concetto familiare a chiunque abbia sostenuto un esame di statistica. Infatti vi sono distribuzioni statistiche assai comuni i cui grafici descrivono un modesto numero di eventi che accadono molto spesso e un grande numero di eventi (la coda lunga) che succedono di rado. In complesso, tuttavia, gli eventi rari possono essere più numerosi degli eventi normali. <http://www.technologyreview.it/index.php?p=article&a=551>. La coda lunga in Italia: un articolo sul webmarketing:

delle “code lunghe” risale almeno alla “distribuzione di Pareto”³², in cui “pochi eventi si verificano con estrema frequenza, molti eventi si verificano con scarsità”. Una simile distribuzione di eventi viene descritta, a livello statistico, da un’iperbole nella quale la “coda lunga” è rappresentata dalla miriade di “eventi”, singolarmente pressoché irrilevanti, ma nel complesso numericamente assai elevati. Matematicamente, una distribuzione a coda lunga segue la cosiddetta “legge di potenza”³³.

La strategia vincente sui mercati a “coda lunga” è quindi non l’abbassamento dei prezzi dei prodotti più popolari, ma l’ampliamento dell’offerta, che permette ricavi maggiori attraverso la vendita in poche copie di tanti prodotti diversi.

In termini commerciali, il reddito maggiore sta nelle piccole vendite. In effetti, la maggior parte dei business in Rete si fonda sulla coda lunga: lo stesso Google produce reddito vendendo pubblicità a costi bassi a milioni di persone sotto forma di banner testuali (text ads) piuttosto che vendendo pochi spazi dai costi elevati a inserzionisti importanti.

Battelle è interessato all’applicazione della ricerca a mercati non sfruttati: nel caso di Google, l’enorme quantità di dati messa a disposizione sui quali effettuare ricerche ha permesso, a suo dire, lo sfruttamento commerciale delle “code lunghe”. Nel contesto dell’e-commerce le code lunghe hanno tre implicazioni. In primo luogo, attraverso Internet i prodotti con poca domanda possono, collettivamente, creare un mercato superiore a quello del piccolo gruppo degli articoli assai venduti. In secondo luogo, Internet dà vita a una proliferazione di venditori, proprio come favorisce una proliferazione di mercati (è il caso del sistema di aste online di eBay). Infine, grazie alla ricerca, diviene possibile un passaggio dai mercati di massa a quelli di nicchia.

Quest’ultima tendenza si basa sulla formazione di gruppi spontanei, un fenomeno molto diffuso nelle Reti. Su Internet, anche i gruppi numericamente più rilevanti non sono masse omogenee di persone, ma variegate comunità di utenti uniti da una passione, un interesse o una finalità in comune. L’opposizione fra masse e nicchie non ha quindi grande rilevanza per l’individuazione di segmenti di mercato utili. A livello commerciale questo provoca la nascita di siti che propongono e-commerce su prodotti legati a tipologie molto specifiche di utenti, che al di fuori della distribuzione on-line non costituiscono in alcun modo un mercato interessante: ad esempio, magliette con loghi per smanettoni, orologi che segnano l’ora in binario, computer luminosi e altri oggetti dedicati ai tecnofili. L’ampiezza dell’offerta ripagherà la scarsità della domanda, ripartita su un numero elevatissimo di prodotti altamente personalizzati.

In un simile scenario, è possibile, come evidenzia il puntuale articolo di Charles H. Ferguson³⁴, che Google e Microsoft combatteranno per il controllo della catalogazione, della ricerca e dell’estrazione di tutte le informazioni, su ogni genere di dispositivo digitale. Microsoft sta investendo parecchio sui servizi web. Nel novembre del 2004 ha lanciato una versione beta (sperimentale) di un motore di ricerca progettato per rispondere alle domande poste in linguaggio corrente e per fornire risultati personalizzati in base alla localizzazione geografica dell’utente³⁵; nel febbraio 2005 il motore, MSN search, è stato migliorato. Con MSN search è possibile anche cercare dentro Encarta, l’enciclopedia multimediale della Microsoft; al momento però la consultazione è limitata a due ore e, durante la navigazione, un contaminuti aggiorna sul tempo rimasto. Quindi il gigante Microsoft ha deciso di sviluppare un proprio sistema per le ricerche sul Web e sui personal computer senza passare da Google, che ormai da alcuni anni è l’asso pigliatutto della ricerca (solo Yahoo! può vantare un’influenza paragonabile).

<http://www.marketinginformatico.com/la-teoria-della-coda-lunga.php>

³² Vilfredo Pareto (1848-1923) fu sociologo ed economista. Una buona introduzione si trova su Wikipedia,

http://it.wikipedia.org/wiki/Vilfredo_Pareto

³³ Si veda: A.L. Barabasi, Link, Einaudi, Torino, 2001

³⁴ Si veda l’articolo in italiano di Charles H. Ferguson, pubblicato su Technology Review:

<http://www.technologyreview.it/index.php?p=article&a=340>

³⁵ Il motore di ricerca di Microsoft: <http://search.msn.com/>

Nell'insieme, i mercati legati all'economia della ricerca sembra siano molto più vasti dell'attuale mercato dei servizi di ricerca. Microsoft è in ritardo, certo, ma potrebbe attuare strategie feroci e inarrivabili per Google: ad esempio, potrebbe investire in perdita e fornire accesso gratuito al suo motore di ricerca, dopo averlo integrato nel browser Internet Explorer; oppure vendere pubblicità sottocosto agli inserzionisti per tagliare la liquidità del concorrente. Nel frattempo, il prossimo sistema operativo sviluppato a Redmond, Windows Vista, dovrebbe offrire strumenti di ricerca innovativi. Del resto, Microsoft era in forte ritardo su Netscape (il primo browser scaricabile gratuitamente dalla Rete) nella battaglia per il dominio sui browser, eppure Explorer l'ha spuntata, e non certo perché fosse il software migliore. Se Microsoft ha dalla sua una lunga esperienza, una posizione di predominio sul mercato e un potere finanziario enorme, anche Google ha delle buone carte da giocare. È l'incarnazione della giovane azienda rampante, si è costruita una fama di impegno nella ricerca e dedizione per l'eccellenza tecnica, predica il culto della velocità nel soddisfare le richieste degli utenti con interfacce scarse ed efficaci: si è imposta perché tecnicamente migliore degli altri motori di ricerca. Nello scontro per il controllo del web, Google sembrerebbe leggermente avvantaggiato; tuttavia, non si può dimenticare che l'estensione delle attività di Microsoft non ha paragoni, non si limita al web e copre anzi l'intera gamma dell'IT, dagli strumenti con Windows o MS Office, ai contenuti con Encarta, alla ricerca di alto livello con piattaforme come dotNet, ecc. Vista la posta in gioco – il dominio dell'accesso a qualsiasi informazione digitale, e i profitti che ne deriverebbero – una convivenza pacifica sembra alquanto improbabile. Per ora Google ha rimandato lo scontro, ma non potrà tardare a lungo.

La guerra degli standard

Continuiamo a seguire Ferguson nel suo ragionamento: lo scenario che si apre è quello di una guerra di standard. Al momento ci sono tre attori coinvolti: Google, Yahoo! e Microsoft. L'industria della ricerca, come risulta anche dall'analisi di Battelle, è in forte crescita. Le tecnologie emergenti o in via di consolidamento, come ad esempio lo *streaming* audio e video attraverso la banda larga, il VoIP (telefonia via internet: si pensi a Skype, e all'omologo servizio di Google, GTalk) e la messaggistica istantanea generano enormi quantità e tipologie di dati che ancora attendono di essere indicizzati, resi disponibili per l'accesso da tutta una varietà di nuovi dispositivi elettronici (dai palmari ai cellulari e lettori di documenti audio-video di nuova generazione, ai navigatori satellitari, ecc.), incrociati fra loro per soddisfare le richieste degli utenti, e in ultima analisi sfruttati per pubblicità pervasive di ogni tipo. Per poter cooperare reciprocamente e con molti altri sistemi saranno necessari nuovi standard, difficilmente introducibili in maniera indolore.

Le guerre tra gli standard avvengono perché il mercato della tecnologia dell'informazione ha bisogno di linguaggi comuni per poter gestire l'aumento di complessità. Il valore dell'informazione risiede nella sua diffusione; ma diffondere e scambiare testi è più semplice che diffondere e scambiare audio, o addirittura video: una maggiore quantità di dati richiede reti di comunicazione più potenti e cambiamenti nelle modalità di gestire le informazioni. Storicamente, il controllo proprietario su un importante standard tecnologico ha sempre creato enormi ricchezze³⁶. A volte, come nel caso del web, l'adozione di uno standard aperto e non proprietario, come è il protocollo HTTP, conduce a una situazione vantaggiosa per tutti. Spesso però non si affermano le tecnologie "migliori", ma "è più importante poter contare su di una strategia efficace. I vincitori tendono tuttavia a essere connotati da alcune caratteristiche importanti. Di solito vendono piattaforme universali, indipendenti dall'hardware, come i sistemi operativi Microsoft, invece che soluzioni hardware e software fortemente integrate, come i sistemi Apple e Sun. Le architetture vincenti sono

³⁶ Il predominio architeturale equivale a battere moneta e, se gestito in modo adeguato, può durare per sempre. L'architettura dei computer mainframe IBM è stata introdotta nel 1964; Intel ha sviluppato il suo primo microprocessore nel 1971; il primo sistema operativo Microsoft è stato presentato nel 1981; Cisco Systems ha venduto il suo primo router nel 1986. Nessuna di queste tecnologie sta dando segni di stanchezza e ciascuna di esse ha complessivamente generato centinaia di miliardi di dollari di fatturato.

proprietarie e difficili da imitare, ma sono anche estremamente «aperte»: mettono cioè a disposizione interfacce pubblicamente accessibili per lo sviluppo da parte di programmatori indipendenti e utenti finali. In questo modo l'architettura può raggiungere tutti i mercati e creare un punto di aggancio e di blocco: in altre parole, spinge gli utenti verso quella specifica architettura, rendendo quasi impossibile il passaggio a un sistema concorrente, se non con grosse difficoltà e spese.”³⁷ Si impone uno standard chiuso per sfruttare una situazione di monopolio.

Un esempio che rende lampante la lotta per l'egemonia sugli standard è la sfida già accennata fra Gtalk e Skype. Al momento Skype ha conquistato una posizione di monopolio nella gestione del VoIP per il mercato domestico; ma è possibile che non siano stati valutati attentamente i tempi di assimilazione delle tecnologie aperte all'interno delle comunità di sviluppo. Fino a poco tempo fa, Skype era l'unica soluzione realmente funzionante per chiunque volesse telefonare via internet pur senza avere alcuna competenza tecnica; ma lentamente le cose stanno cambiando. Le tecnologie proprietarie di Skype potrebbero essere surclassate da Gtalk, totalmente basato su tecnologie libere (in particolare sul protocollo di comunicazione “jabber”), con librerie di sviluppo rilasciate sotto licenze copyleft, che favoriscono l'afflusso di energie nel progetto perché stimolano i programmatori a potenziare la rete VoIP di Google. In questo caso la scelta dell'Open Source potrebbe rivelarsi vincente per scalzare il predominio di Skype, la quale naturalmente potrebbe decidere di rilasciare pubblicamente i propri codici per mantenere la situazione invariata. Scegliere di adottare tecnologie e piattaforme proprietarie, che chiudono l'accesso (ma rilasciando magari delle interfacce pubbliche di sviluppo), oppure aperte, è una parte fondamentale della strategia per il controllo del web e dell'economia della ricerca.

L'accesso al mercato della ricerca è già stato bloccato (in gergo economico “*lock-in*”). Infatti, è impensabile che un'azienda *start-up* possa competere ora con Google o Yahoo! per l'indicizzazione di miliardi di pagine web: l'investimento per le macchine necessarie sarebbe proibitivo, anche se si disponesse di un algoritmo migliore per il proprio *spider*. Tuttavia molti aspetti collaterali, e soprattutto l'interfaccia fra i vari sistemi di ricerca, lasciano ampi spazi a possibili innovazioni “critiche” a basso costo. È il caso delle librerie informatiche, piccole porzioni di software che consentono il collegamento fra sistemi eterogenei, ponti di traduzione da un sistema all'altro, da un linguaggio all'altro, da un motore di ricerca all'altro; le quali, insieme alle metodologie di integrazione fra dispositivi diversi e di condivisione di informazioni e risultati di ricerca, potrebbero essere sviluppati da piccoli ricercatori indipendenti, invece che dalle grandi aziende coinvolte³⁸. Approfondiremo successivamente questi aspetti legati alle interfacce e alle librerie.

Al momento, è importante sottolineare che nessuno degli attori coinvolti ha una posizione di assoluto predominio, e questo è naturalmente un bene per tutti: immaginate cosa significherebbe un monopolio totale del mercato della ricerca ottenuto da un soggetto privato qualsiasi attraverso l'imposizione de facto di uno standard. Le prime ovvie considerazioni sarebbero sulla privacy: di chi sono i dati che vengono indicizzati e sui quali si effettuano ricerche e si generano profitti astronomici? Inoltre, poiché già oggi digitando semplicemente il nome di una persona su Google si accede a una mole impressionante di dati, in un futuro non lontano in cui la quantità e la qualità delle informazioni, e soprattutto la possibilità di effettuare ricerche aggregate su dati eterogenei, migliorasse drasticamente, il controllo sugli individui potrebbe diventare infinitamente più soffocante e totale, incrociando dati sensibili, schede mediche, conversazioni telefoniche, email, immagini, video, blog e opinioni, addirittura informazioni sul DNA. Google potrebbe essere il punto d'accesso privilegiato al panottico³⁹ digitale: vediamo allora quali sono le sue armi

³⁷ Charles H. Ferguson, cit., Technology Review

³⁸ Si veda Ippolita, Open non è free, Elèuthera, Milano, 2005, cap. V.

³⁹ Il panopticon auspicato da Bentham e analizzato da Foucault trova nel progetto Echelon una delle realizzazioni più agghiaccianti. Senza lasciare spazio a paranoie complottiste, segnaliamo fra i moltissimi volumi e articoli un breve scritto introduttivo sull'argomento, Alex Jones e Paul Joseph Watson, Il panottico: una prigione di sorveglianza di massa per l'umanità, 20/01/2006 http://www.ariannaeditrice.it/articolo.php?id_articolo=1878, versione originale

migliori in questa vera e propria guerra per il controllo delle reti.

Arma N 1: GooglePlex, il capitalismo morbido

L'agiografia corrente di Google narra con delizia l'impressionante crescita che ha portato Brin e Page dalla stanzetta di Stanford al garage di Menlo Park, California, che un amico aveva affittato alla neonata società Google Inc., all'ufficio sulla University Avenue di Palo Alto, fino al trasloco al Googleplex, gli odierni uffici a Mountain View, California. Tra il 1998 e il 2000 il duo espande la propria cultura universitaria, sviluppando una cultura aziendale fatta di innovazione, di creatività e di sacrificio. La dedizione universitaria per la ricerca applicata al business: ecco la chiave del successo. Googleplex attirò fin dall'inizio schiere di collaboratori: venne ricreata l'atmosfera del campus universitario americano, dove studio, impegno, sport, svago formano un tutt'uno: se un ambiente ricco di comfort e rilassato favorisce la creatività degli studenti, evidentemente poteva favorire anche la produttività dei lavoratori. Lo spirito accademico della confraternita, dell'élite universitaria che lavora alacremente per ottenere sempre il meglio, spira dai racconti su Googleplex: pare che intere sezioni del parcheggio venissero transennate due volte la settimana per far spazio alle partite di hockey su pattini; oggi più che mai negli uffici i gadget si sprecano, e su tutte le lampade-lava con i fluidi colorati; l'informalità viene eletta a sistema, con Larry e Sergey che dirigevano le riunioni settimanali, insieme a tutte le decine di impiegati, chiamandole TGIF (Thank God It's Friday – Grazie a Dio È Venerdì) nell'*open space* tra le scrivanie.

Di certo, una simile atmosfera informale incoraggiò lo spirito di gruppo e accelerò lo scambio d'idee. Googleplex sembra quasi essere il luogo per esprimere le proprie passioni di ricerca, invece che un luogo di lavoro; sicuramente, non è un luogo di lavoro qualsiasi, nonostante le dimensioni ormai enormi. Eppure la dimensione organizzativa stile "campus", negli USA, è diffusa da almeno trent'anni: Microsoft e Apple, tanto per fare due esempi, da sempre hanno adottato un approccio di questo tipo. La mitologia di Silicon Valley è ricolma di storie sulla creatività al primo posto, sull'importanza della cooperazione fra i collaboratori: in sostanza, sulla necessità di avere lavoratori felici di lavorare e di considerare gli obiettivi aziendali al pari di propri obiettivi vitali, invece che lavoratori frustrati, schiacciati da gerarchie soffocanti, legati a orari di lavoro fissi, a regole rigide, ad ambienti opprimenti.

Forse la novità di Googleplex risiede nell'aver promosso consapevolmente, fin dall'inizio, l'immagine di sé come luogo "diverso", "nuovo", "per le menti migliori". Entrare al Googleplex è impossibile se non si conosce qualcuno all'interno. Dentro poi è persino vietato fare fotografie, almeno in teoria. Come per difendersi dal mondo esterno, popolato da feroci squali della finanza e malvagi predatori dell'IT che vogliono solo sfruttare il talento dei Google-boys.

Tutti vogliono lavorare a Googleplex. Un elenco ufficioso delle amenità che nasconde è d'obbligo: palestre aziendali, piscina, cibo gratis nei 4 ristoranti (di cui uno vegetariano), drink e snack gratis ovunque (basta con la schiavitù delle macchinette a pagamento! Google paga tutto!); campi da pallavolo, basket e spazi all'aria aperta per fare sport, monopattini a motore per spostarsi tra i vari edifici del campus-azienda. Ma queste sono inezie, rispetto all'asilo nido, all'asilo e alle scuole elementari aziendali per i bimbi dei dipendenti, completamente gratuiti, o allo studio dentistico, ovvero un camion trasformato in studio dentistico mobile: in un Paese come gli USA, in cui l'istruzione e la sanità sono un lusso per pochi, si tratta di opportunità ancora più incredibili.

Le postazioni di lavoro poi sono spettacolari, il sogno di ogni fanatico delle tecnologie digitali: monitor standard da LCD 21", schermi al plasma ovunque, giochi e giocattoli (modelli a grandezza naturale dei personaggi di Star Wars, gadget tecnologici di qualsiasi tipo). Onnipresenti, le lampade lava dai colori fluorescenti (*lavalamp*).

Googleplex è un luogo da sogno, immerso nel verde, dove gli orari di lavoro sono flessibili, dove

<http://www.prisonplanet.com/articles/january2006/110106thepanopticon.htm>

tutto sembra possibile. In breve, Googleplex diffonde la filosofia di Google, lo stile di vita di Google: c'è naturalmente un negozio dove acquistare tutti i *gadgets* aziendali che si possono immaginare, raggiungibile online su un sito dedicato; la maggior parte sono veri e propri inni all'inutilità e al superfluo, come ogni *gadget* che si rispetti, ma tutti di sicuro effetto per rendere il senso di appartenenza all'azienda una ragione di orgoglio da rivendicare. La maglietta e il cappellino con il logo dell'azienda per cementare lo spirito di gruppo sono davvero il passato: il presente del condizionamento morbido è Googleplex.

Del resto, non solo Google, ma anche aziende come Apple o Yahoo! offrono ormai da tempo un vero e proprio catalogo di oggetti di consumo, che vanno da linee complete di vestiario a ogni sorta di accessori high-tech, lettori MP3 e chiavi USB, fino agli oggetti d'arredamento, tutti opportunamente contrassegnati da marchi, loghi, motti aziendali. Il negozio di Google, assai versatile, offre le tutine per i neonati, ma anche Google Mini, il sistema aziendale per indicizzare “come Google” tutti i propri materiali.

Googleplex è l'incarnazione del capitalismo dell'abbondanza nell'era dell'informazionalismo⁴⁰: tutte le informazioni del mondo offerte a tutti gratuitamente. Non è più tempo di scarsità: l'abbondanza e la disponibilità dei beni (in questo caso delle informazioni) è totale. Non dimentichiamo però che in ultima analisi la fonte di quest'abbondanza è praticamente solo la pubblicità fatta prevalentemente con link testuali. Tutto il resto, è *free*, libero. E non tutto funziona alla perfezione⁴¹.

Due fra i più influenti personaggi di questo capitalismo dell'abbondanza sono Mikie Moritz⁴², gallese, che oltre a Google ha finanziato anche Yahoo; e John Doerr⁴³, che ha avuto partecipazioni importanti anche in Sun e in Netscape.

Arma 2: perfezionare la strategia dell'accumulo

Dopo una visita a Googleplex diventa più chiara la ragione della fuga di cervelli verso Google di cui si parlava in apertura di capitolo. Per un operaio dell'IT, o sarebbe meglio dire bracciante dell'IT, lavorare a Googleplex rappresenta un sogno. In questo settore sono sempre più numerosi i braccianti sfruttati; una figura professionale esemplificativa è quella del *coder* indipendente che sviluppa progetti personali, magari pubblicandoli su Sourceforge.net o Slashdot.org, e offre la propria competenza sul mercato senza avere alcun tipo di protezione sindacale, contratto di categoria o simili, ormai relitti preistorici nella nostra epoca della flessibilità totale. A Googleplex potrà addirittura dedicare l'ormai famoso 20% del tempo lavorativo ai propri progetti, pagato e

⁴⁰ Sul concetto di informazionalismo si veda in particolare: Manuel Castells, *L'Età dell'informazione*, Università Bocconi Editore, Milano, 2004.

⁴¹ Vi sono molte segnalazioni di disservizi di Google; del resto, è ovvio che si moltiplichino i problemi man mano che le dimensioni e i servizi crescono. Inoltre la pratica di rilasciare versioni beta che gli utenti testano in maniera estensiva aumenta la probabilità di malfunzionamenti. Un esempio concreto: Marco Deseri su Google Analytics:

<http://www.deseri.it/articoli/google-analytics-rallenta-il-sito>

⁴² Gallese, giornalista presso il Time Magazine negli anni settanta e ottanta, oggi è considerato un “*veteran venture capitalist*” e sembra avere una ricetta per tutti: ricreare oggi fenomeni come Google? L'importante, dice, “è creare qualcosa che semplicemente si pensa possa essere utile a se stessi”. Artista, se non avesse dato la priorità agli impegni lavorativi, è l'uomo di marketing vecchia scuola, ma sempre incisivo: le idee giuste sono quelle che soddisfano se stessi, quindi i consumatori, anzi, “gli utenti”.

⁴³ John Doerr continua anche dopo Google, i suoi investimenti: ultimamente ha scommesso 16 milioni di dollari in un progetto dal nome “Zazzle”, il cui *core business* è un mercato on line dove si possono comprare e vendere prodotti altamente personalizzabili. In pratica, Doerr non vuole rimanere fuori dalla “personalizzazione di massa”, ormai in fase avanzata di attuazione negli USA. Ispiratori del progetto sono personaggi come Eric Von Hippel, che “dalla sua cattedra della Sloan School of Management teorizza al MIT da tempo l'idea che il prossimo passo sarà proprio quello: trasformare la produzione di massa di stampo fordista e post-fordista in qualcosa di completamente diverso, su misura, creato e gestito attraverso il web”

http://www.bloggers.it/Marcioweb/index.cfm?idForum=News%20dal%20Mondo&blogaction=archive&file=blog_7_2005.xml

incoraggiato a far sempre meglio.

Tra partite di hockey e pallavolo, cani che gironzolano per il campus/azienda e nei corridoi, riunioni attorno a un tavolo da ping pong, certamente è difficile annoiarsi. Poiché non è facile reclutare nuovi collaboratori che siano in grado addirittura di migliorare questo genere di atmosfera, Google si è distinta per le sue trovate ingegnose in tema di assunzioni. La più curiosa è forse il test lanciato nel luglio 2004: in alcune stazioni della metropolitana della cittadina di Cambridge, nello stato del Massachusetts (USA), e lungo l'autostrada 101 in California, sono apparsi enormi cartelli bianchi con la scritta:

{first 10 digit prime in consecutive digits of e}.com

ovvero:

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

Il logaritmo naturale richiesto è il numero 7427466391; all'indirizzo <http://www.7427466391.com/> indicato dal cartellone ci ritroviamo su un indirizzo IP di Google che ci chiede di completare una sequenza di numeri; trovato il numero 5966290435, si seguono le istruzioni, usandolo come password per entrare in una sezione del sito <http://www.linux.org> (passaggio niente affatto scontato) e di qui veniamo rimbalzati ancora su Google, sul sito <http://www.google.com/labjobs/> dove ci chiedono di inviare il nostro curriculum: se siamo riusciti a risolvere gli enigmi, potremmo essere un buon acquisto per Google.

Ma Google non attira solo i migliori tecnici, hacker e smanettoni geniali: ben presto anche i manager più quotati del mondo dell'IT si accorgono delle sue potenzialità e fanno a gara per entrare a farne parte.

La strategia dell'accumulo di dati su cui fare ricerca, e di macchine in rete per contenere una copia dei dati, procede di pari passo all'accumulo di cervelli. Macchine significanti, macchine elettroniche, macchine biologiche vengono accumulate a Googleplex, per dar vita a uno stile di vita, anzi a una specie di religione dell'eccellenza, incarnata da un evangelizzatore.

Colui che incarna al meglio lo stile dell'azienda, il cosiddetto "evangelist" di Google, in mezzo a tanti giovani, è un grande vecchio della Rete, Vinton G. Cerf, inventore insieme a Robert Kahn del protocollo TCP/IP. Le circostanze sono degne di una piccola parentesi: nel febbraio 2005 Google confermava che l'ICANN, l'organismo che supervisiona il sistema dei nomi e dei domini Internet, aveva dato il via libera affinché l'azienda di Mountain View potesse entrare nel business della registrazione dei domini. Nel settembre 2005 Google annuncia che Vinton Cerf è diventato "vicepresidente ed Evangelizzatore Capo di Internet per Google. Il suo compito è quello di individuare nuove tecnologie ed applicazioni strategiche per l'azienda su Internet e su altre piattaforme."⁴⁴ Fino a quel momento, tra le altre cariche, Cerf sedeva nel consiglio di amministrazione dell'ICANN. A differenza dell'acquisto di Erich Schmidt come Amministratore Delegato, e di altri top manager strappati ai concorrenti, questa è senza dubbio un'operazione completamente di immagine (difficile pensare che Cerf lavori veramente a Googleplex, per quanto divertente possa essere...).

Arma 3: l'immagine è tutto, ma un po' di "filosofia" non guasta

L'immagine di Google non è solo il sito e l'interfaccia⁴⁵ che hanno riscosso tanto successo per la loro semplicità e velocità; non è solo Googleplex, il mondo dorato dei prodigi tecnologici; non è solo "essere buoni" e far soldi lo stesso, coniugando strategie commerciali spregiudicate alle strizzate l'occhio alla cultura accademica e alle comunità Free Software e Open Source, con

⁴⁴ Si veda <http://www.google.it/intl/it/corporate/execs.html#vint> ricordiamo che negli USA praticamente chiunque non sia impiegato è Vicepresidente di qualcosa.

⁴⁵ Si veda il cap. V.

finanziamenti e iniziative a loro dedicate.

L'immagine è anche e soprattutto la “filosofia” dell'azienda, la spiegazione chiara e lampante, in pillole, del Google-pensiero. Non si tratta precisamente di indagini svolte per amore del sapere e della chiarezza, perciò forse il termine filosofia è esagerato; in ogni caso, on-line si possono trovare dieci punti, i dieci comandamenti del buon Google. Si inizia con un cappello introduttivo, che enuncia lo stile: “Mai accontentarsi del meglio”; infatti, come afferma Larry Page, l'obiettivo ultimo è “Il motore di ricerca perfetto”, che “dovrebbe capire esattamente la richiesta dell'utente e restituire esattamente ciò che egli desidera”. Quindi Google non vuole essere per molti: vuole essere per tutti, soddisfare i desideri di tutti, portare la felicità nel mondo. Per farlo, si dedica con tenacia all'innovazione e alla ricerca, seguendo quelle che definisce le sue “Dieci verità”:

1. Attenzione all'utente: tutto il resto viene dopo. La crescita di Google è stata frutto del passaparola, che ha attirato utenti estasiati dalle sue prestazioni, non di aggressive campagne pubblicitarie. La pubblicità non deve essere invasiva, ma solo utile all'utente.

2. È meglio dedicarsi veramente bene a una sola cosa.

“Google si dedica alla ricerca. Con uno dei gruppi di ricerca tra i più grandi al mondo dedicato esclusivamente alla risoluzione dei problemi della ricerca, siamo consapevoli di ciò che facciamo bene e di come possiamo farlo meglio.”. Abbiamo visto che Googleplex si configura appunto come un centro di ricerca universitario. I nuovi servizi aggiuntivi rispetto al semplice motore di ricerca sono visti come opportunità offerte agli utenti per orientarsi e reperire le informazioni che desiderano senza perdersi in masse sempre più enormi di dati.

3. Rapidità è meglio di lentezza.

“Google crede nella gratificazione immediata. L'utente vuole delle risposte e le vuole subito. Chi siamo noi per non essere d'accordo? Google è probabilmente l'unica società al mondo il cui obiettivo dichiarato è fare in modo che i propri utenti abbandonino il sito Google al più presto.”. Due sono le principali intuizioni e realizzazioni che hanno consentito a Google di ottenere “velocità”: la messa a punto e il perfezionamento continuo dell'algoritmo PageRank, che indicizza senza posa le reti; l'uso di piattaforme modulari fortemente scalabili collegate fra loro (*cluster*). Sulla velocità come panacea, tuttavia, sarebbe opportuno riflettere più approfonditamente: a volte, anche nell'ambito delle nuove tecnologie, la lentezza è una virtù⁴⁶.

4. La democrazia sul Web funziona.

“Il successo di Google si basa sui milioni di individui che pubblicano siti Web, per determinare quali altri siti offrono contenuti validi.” Sappiamo già che Google usa PageRank per valutare i siti collegati a una pagina Web e assegnare loro un valore, basato in parte sui siti ad essi collegati. L'immagine di questa democrazia elettronica è idilliaca: il risultato del ranking di Google sarebbe un “indice di gradimento popolare” basato su un algoritmo/legge elettorale che consentirebbe agli utenti/cittadini della rete di esprimere le proprie preferenze/voti attraverso i link offerti e accettati da altre pagine, e trovarle puntualmente espresse nel posizionamento dei siti web votati. L'accostamento link/voto è quantomeno forzato e semplicistico, anche perché vengono introdotti continuamente “correttivi” per calcolare il ranking, “pesando” in maniera diversa i link/voti. Di certo il link/”voto” di un sito porno conterà meno del link/”voto” di un sito universitario. Eppure è discutibile che la cultura accademica sia più popolare della cultura pornografica... Di sicuro, con la crescita delle informazioni, questa “democrazia” è in continua espansione.

5. Non occorre essere alla propria scrivania per avere bisogno di informazioni.

“La società attuale è sempre più mobile e meno disposta ad essere vincolata a un luogo fisso. O

⁴⁶ Per una panoramica complessiva sul tema, si veda: Thomas Hylland Eriksen, *Tempo tiranno - Velocità e lentezza nell'era informatica*, Elèuthera, Milano, 2003.

tramite i PDA, i telefoni wireless o le loro automobili, gli utenti desiderano che le informazioni vengano verso di loro.”. La flessibilità dei tempi e degli spazi è un obiettivo importante: la convergenza dei dispositivi elettronici (TV, radio, telefono, Internet...) in dispositivi mobili miniaturizzati è un’opportunità irrinunciabile per il maggior fornitore mondiale di soluzioni di ricerca. La penetrazione in questo mercato del futuro è strategica, l’abbiamo visto discutendo della “guerra tra gli standard”: tanto più per Google, che non produce dispositivi elettronici sui quali proporre/imporre i propri software (Microsoft, Apple), ma interfacce di ricerca. Ogni nuovo dispositivo è un territorio da conquistare.

6. È possibile guadagnare senza fare del male a nessuno.

“Google è un’azienda. Le entrate generate dalla società derivano dall’offerta alle aziende della sua tecnologia di ricerca e dalla vendita di pubblicità visualizzata su Google e su altri siti Web.”. Ma le pubblicità sono testuali e poco invasive, i collegamenti sponsorizzati sono pertinenti alle ricerche fatte (AdWords); e gli utenti possono diventare inserzionisti in maniera semplice e self-service; e addirittura, se è un gestore di siti web, sfruttare la Rete di Google tramite AdSense, pubblicando annunci pertinenti con il contenuto dei propri siti. “Non essere cattivo”, “Non fare del male”, significa anche “non fare agli altri la pubblicità che non hanno fatto a te”, e naturalmente garantire che il PageRank non è oggetto di vendita. La fiducia degli utenti nella correttezza dei risultati è la principale forza di Google, e non può essere persa per guadagni a breve termine. Viene sfruttata per generare introiti indiretti, di secondo livello (generati dalla pubblicità).

7. Vi sono sempre più informazioni di quanto si immagini.

“Dopo aver indicizzato più pagine HTML su Internet di qualsiasi altro servizio di ricerca, i tecnici Google hanno rivolto l’attenzione alle informazioni non immediatamente accessibili.”. Sì, Google accumula basi di dati eterogenee: immagini, messaggi dei newsgroup (Usenet), numeri di telefono e indirizzi, informazioni finanziarie. Per essere il principale mediatore di informazioni al mondo è necessario non porre limiti all’accumulo.

8. La necessità di informazioni oltrepassa ogni confine.

“Nonostante la sede centrale di Google sia in California, abbiamo uffici in tutto il mondo e la nostra missione è facilitare l’accesso alle informazioni alle persone di tutto il mondo.”. Cultura americana, strettamente accademica, per tutti. Pensare in grande sempre e comunque: indicizzare tutte le informazioni, fornire accesso a tutto il mondo. Le traduzioni sono parte essenziale dell’universalismo di Google: dal coreano al gergo hacker al klingoniano di Star Trek, dall’hindy allo xhosa sudafricano fino al pig latin (ovvero l’alfabeto farfallino americano), dallo zulu all’interlingua e al Bork bork bork! (una specie di finto-russo molvano...) tutti devono poter accedere. Oltre cento lingue d’interfaccia. Domini locali di oltre cento stati. Lodevole, ma un po’ accentratore... operazione forse politicamente corretta e formalmente rispettosa delle minoranze, ma in realtà si tratta di un *super-layer*, un livello superiore, un’unica interfaccia che appiattisce, rende omogenee le differenze e diffonde lo stile di Mountain View.

9. È possibile essere seri anche senza giacca e cravatta.

“I fondatori di Google hanno spesso dichiarato che la società è “seria” solo per quanto concerne la ricerca. Essi hanno costruito un’azienda basata sull’idea che il lavoro deve essere una sfida e che tale sfida deve essere divertente.”. In questo comandamento si compendia l’organizzazione di Googleplex come un campus universitario per massimizzare i rendimenti: esplicitamente, si afferma che “L’enfasi è sui risultati conseguiti dal team e sull’orgoglio dei risultati individuali che contribuiscono al successo globale della società”, e che “Questo ambiente estremamente comunicativo favorisce una produttività e un cameratismo alimentati dalla consapevolezza che milioni di persone confidano sui risultati di Google. Date gli strumenti giusti a un gruppo di persone che amano fare la differenza e loro la faranno”. Forse questo è il modo migliore per sfruttare i creativi, rendendoli contemporaneamente entusiasti sostenitori della “Esperienza Google”.

10. Eccellere non basta.

“Dare sempre di più di quanto ci si aspetta. Google non accetta l’idea che essere i migliori sia l’obiettivo finale, ma ritiene che debba essere il punto di partenza. Tramite innovazione e iterazione, Google si dedica a migliorare qualcosa che già funziona bene per ottenere risultati inaspettati.”. Evidentemente, per soddisfare velocemente, sempre meglio, i desideri di tutti gli utenti del mondo, è necessario rimandare un po’, sempre un po’ più in là, la soddisfazione dei propri desideri. Desiderare di desiderare di essere i migliori. In questo contesto, arrivare secondi è peggio che non esistere. Per quanto ci riguarda, slogan per slogan, preferiamo “Fare schifo, in una società che obbliga all’eccellenza, è un preciso dovere morale”⁴⁷.

Arma 4: Google e l’Open Source

L’arma forse più complessa di Google è la sua strategia di collaborazione-sfruttamento del mondo Open Source. L’iniziativa Google Code (marzo 2005) è un omaggio alla comunità Open Source, di cui i fondatori di Google si dicono “amici e riconoscenti”. Il motore più usato del Web spiega che lo scopo di Google Code non è quello di promuovere lo sviluppo di applicazioni che utilizzino le proprie API (Application Programming Interface), per cui esiste già un sito dedicato, ma quello di mettere a disposizione di tutti gli sviluppatori Open Source strumenti di pubblica utilità. I 4 progetti inizialmente pubblicati su Google Code nascono attorno a programmi creati dagli stessi sviluppatori della società per velocizzare la creazione, l’ottimizzazione o il debugging del codice. I progetti linkati su Google Code sono ospitati anche su SourceForge.net e sono distribuiti sotto licenza BSD 2.0 (il loro codice può essere utilizzato indistintamente all’interno di altre applicazioni Open Source o proprietarie). Inoltre Google promette fin da subito di mettere a disposizione della comunità Open Source diversi altri software, in buona parte frutto di quel 20% di tempo che gli sviluppatori alle dipendenze della società sono incoraggiati a dedicare a progetti personali.

Non a caso segue un progetto interamente dedicato al reclutamento di sviluppatori Open Source, con la “Summer of Code”, una vera e propria gara con in palio 4500 dollari. Poi è il turno di “Google Earth” e infine, come ogni potenza giunta a rappresentare uno stile di vita, Google presenta il sogno: “moon.google.com”, Google sulla luna!

“In onore del primo sbarco dell’uomo sulla Luna, avvenuto il 20 luglio del 1969, abbiamo aggiunto alcune immagini NASA all’interfaccia di Google Maps per consentire a tutti di visitare il nostro vicino celestiale. Buona navigazione lunare”.

I movimenti di Google, “pseudo-monopolistici” nei metodi e nelle prospettive, hanno avuto una immediata ricaduta sulla concorrenza: oggi Google si sta trasformando in un gigante che occupa tutti gli spazi di mercato; sempre nuovi servizi soffocano le società hi-tech più piccole, che faticano ad arruolare ingegneri e tecnici e rischiano di vedere imitati i propri nuovi prodotti.

Il lancio ininterrotto di nuovi servizi e il finanziamento in pratica dall’interno a potenziali *spin-off* dei dipendenti rende di fatto il mercato bloccato in termini di innovazione tecnologica. Chi infatti finanzierebbe oggi un progetto web, con la possibilità che da lì a qualche giorno sia proprio Google a lanciarlo?

Google ha saputo rappresentare, agli occhi degli osservatori, ma anche dei semplici utilizzatori di Internet, una sorta di paladino del progresso: a partire dal motore di ricerca, concepito per essere compreso rapidamente e con semplicità dagli utenti, si sono moltiplicate le proposte di servizi e idee.

Con l’opzione Open Source l’economia relazionale messa in atto da Google diventa una “visione del mondo”, recepita addirittura come uno sviluppo sostenibile, un capitalismo buono che diffonde l’abbondanza, una soluzione economicamente “etica” alle ricerche degli individui.

⁴⁷ SantaMariaVideo, la TV che non trasmette niente: <http://www.santamariavideo.tv/video/anonimato.mpg>

III. Google Open Source: teoria e pratiche

Open non è Free

Free Software⁴⁸ (Software Libero) e Open Source⁴⁹ (Sorgente Aperto) sono sintagmi usati spesso come sinonimi per indicare codici o porzioni di codici informatici; tuttavia, per quanto descrivano oggetti spesso identici, rispecchiano prospettive radicalmente differenti. Free Software è un termine nato agli inizi degli anni Ottanta per iniziativa di Richard Stallman: si riferisce alla libertà dell'utente di usare e migliorare il software. Più precisamente, può essere riassunto in quattro libertà fondamentali:

- 1) Libertà di eseguire il programma, per qualsiasi scopo.
- 2) Libertà di modificare il programma secondo i propri bisogni (perché questa libertà abbia qualche effetto in pratica è necessario avere accesso al codice sorgente del programma, poiché apportare modifiche a un programma senza disporre del codice sorgente è estremamente difficile).
- 3) Libertà di distribuire copie del programma gratuitamente o dietro compenso.
- 4) Libertà di distribuire versioni modificate del programma, così che la comunità possa fruire dei miglioramenti apportati.

L'espressione Open Source, invece, nasce alla fine degli anni Novanta, per iniziativa in particolare di Bruce Perens e Eric S. Raymond, che nel 1998 fondano la *Open Source Initiative*⁵⁰ (OSI); si riferisce alla *Open Source Definition*, a sua volta derivata dalle *Debian Free Software Guidelines*, ovvero una serie di 10 punti pratici che definiscono quali criteri legali debba soddisfare una licenza per essere considerata effettivamente “libera”: o meglio, con il nuovo termine, Open Source.

È evidente quindi che da una parte il Free Software pone l'accento sulla libertà: come sottolineato nella definizione, “il Software libero è una questione di libertà, non di prezzo”⁵¹. Dall'altra parte, l'Open Source si occupa esclusivamente di definire, in una prospettiva totalmente interna alle logiche di mercato, quali siano le modalità migliori per diffondere un prodotto secondo criteri open, cioè aperti. Il Free Software ha un senso che va ben oltre il mercato, pur non escludendolo a priori; l'Open Source esiste, come specificato dai suoi stessi promotori, per adattare un modello preesistente (quello “free” nel senso di “libero”) al mercato.

Gli hackers di Stanford

Ormai da alcuni anni il software Open Source viene considerato comunemente affidabile, capace di prestazioni elevate e a costi sostenibili; in pratica è oggi spesso considerato migliore rispetto ai software proprietari, proprio perché in grado di aumentare l'affidabilità di un prodotto soprattutto in virtù di una metodologia differente di sviluppo, aperta e sottoposta a un meccanismo molto ampio di revisione.

Grazie alla realizzazione di software stabili e funzionali come browser, programmi per ufficio,

⁴⁸ La filosofia del Free Software: <http://www.gnu.org/philosophy/free-sw.it.html>

⁴⁹ La definizione di Open Source: <http://www.opensource.org/docs/definition.php>

⁵⁰ Il sito della Open Source Initiative: <http://www.opensource.org/>

⁵¹ Anche in articoli recenti, peraltro di buon livello, bisogna notare questa confusione fra Free Software e Open Source, quasi che il secondo fosse un miglioramento del primo in perfetta sintonia e continuità. Il movimento fondato da Stallman ha senz'altro mostrato molti limiti concreti, non solo nelle relazioni con il mercato ma anche nell'assunzione di posizioni troppo ideologiche; tuttavia non ci sembra corretto tracciare linee evolutive semplici quando il panorama è assai frastagliato e complesso. Si veda ad esempio “Economia delle Reti Open Source: storia e dinamiche del movimento del software libero”, http://www.pluto.it/files/journal/pj0601/economia_reti.html

editor e interi sistemi operativi (GNU/Linux), quasi ogni utente si è accorto o ha sentito parlare dell'esistenza dei programmi detti genericamente copyleft o Open Source.

Open Source era il termine adatto, e necessario, per sostituire la dicitura Free Software. In inglese infatti la parola Free ha il duplice significato di Libero e di Gratuito: questa ambiguità linguistica rendeva il prodotto poco appetibile dal punto di vista economico. La sostituzione con Open fu un modo strategicamente vincente per mantenere le caratteristiche di cooperazione libera senza rinunciare alla possibilità di un uso più strettamente commerciale dei software.

All'epoca, in realtà, si stava assistendo a un cambiamento radicale nell'assetto delle comunità digitali spontanee, cioè quelle comunità a cui si sentono legati tutti coloro che danno una definizione positiva di hacking.

Questi aggregati erano (e lo sono ancora) estremamente compositi. Ci riferiamo a un interstizio culturale fluido nel quale si formano e collaborano studenti, professori, ricercatori, liberi professionisti, polizia e criminali, programmatori stipendiati da società di sviluppo, appassionati e molte altre tipologie di hacker.

Il movimento del Free Software stava cominciando un confronto serrato con l'economia di mercato. La battaglia della *Free Software Foundation* verteva sulla diffusione della licenza GPL (*General Public License*), creata dal fondatore della FSF Richard Stallman⁵²; tale licenza vincola l'artefatto in modo "virale" alle quattro libertà sopra elencate. In sostanza ogni modifica apportata a codice sotto licenza GPL deve mantenere la stessa licenza d'uso, rendendo impossibile la chiusura del codice; questo meccanismo è noto come "permesso d'autore" (*copyleft*, gioco di parole su copyright). Nascevano e si diffondevano allora le prime distribuzioni del sistema operativo GNU/Linux.

La commistione tra metodo di sviluppo libero e net-economy avrebbe determinato negli anni successivi al 2000 l'esplosione dei prodotti Open Source e allo scatenarsi del dibattito politico circa la brevettabilità del software, il copyright e alla gestione etico-politica di tutto ciò che attualmente definiamo opera d'ingegno umano.

L'azienda Google, per quanto non sia direttamente un produttore di software, non è rimasta ai margini dello scossone Open Source: come altre aziende dinamiche e innovative, Google ne ha cooptato le metodologie e le ha poste al servizio della sua "missione".

La contiguità fra Google e Open Source è spaziale e temporale: nel 1998 a Stanford, proprio mentre Brin e Page mettevano a punto la prima versione del loro motore di ricerca, stavano emergendo alcuni importanti progetti Free Software; ricordiamo ad esempio SND e Protégè che, in campi molto differenti, ovvero l'audio e il web semantico, avrebbero riscosso grande successo sulla scena digitale.

A Stanford la cultura hacker, da cui deriva in ultima analisi l'Open Source, si respira come un'aria di famiglia: non è dunque un caso che il nostro duo formatosi in quegli anni abbia sempre manifestato una certa predilezione per lo sviluppo su piattaforma GNU/Linux.

Se esistono differenze sostanziali tra Open Source e Free Software, vi sono però anche elementi in comune e continuità di vedute. Per semplicità e correttezza parleremo perciò di "metodologie e pratiche aperte", in breve di Open Source, per indicare il fenomeno che interseca Free Software, Open Source e competizione di mercato nel mondo dell'IT.

La prima caratteristica di una comunità Open (ma in questo senso anche Free) è quella di mettere in pratica un metodo di lavoro aperto alla collaborazione di tutti, capace cioè di accettare suggerimenti e interazioni spontanee da ogni tipologia di soggetto coinvolto nella costruzione dell'artefatto informatico: programmatore, traduttore, o anche semplice utente. Questo

⁵² La Free Software Foundation in italiano: <http://www.gnu.org/home.it.html>

procedimento è stato definito nel gergo hacker metodo a “bazar” e la sua applicazione su vasta scala si deve allo sviluppo del Kernel Linux nei primi anni Novanta, un progetto nato per iniziativa di Linus Torvalds e base di tutte le distribuzioni GNU/Linux⁵³.

La nuova tecnica cooperativa proposta dall’underground digitale ha rovesciato la legge di Brooks⁵⁴ che regolava le comunità di sviluppo dei progetti informatici fino a quel momento. Secondo la legge di Brooks, con il crescere della complessità aumentano esponenzialmente gli errori e quindi un progetto a cui contribuiscono migliaia di sviluppatori avrebbe dovuto essere un’accozzaglia di codice instabile e pieno di bug. Invece, attraverso la resa pubblica dei codici sorgenti, la circolazione libera su internet della documentazione, la cooperazione e il *feedback* spontaneo di un numero sempre più elevato di soggetti in gioco, le comunità libere hanno dimostrato come fosse possibile un enorme miglioramento nella costruzione di artefatti digitali, sia dal punto di vista del risultato, sia del processo. I software realizzati in questo modo vengono generalmente resi pubblici sotto la licenza GPL, che come abbiamo visto diffonde in maniera virale materiali copyleft.

La licenza GPL non prevede quindi restrizioni dal punto di vista commerciale; tuttavia, esattamente come il termine Free Software risultava eccessivamente radicale nel porre le libertà al primo posto, così anche la GPL è stata sostituita da licenze edulcorate rispetto al portato etico e politico che il movimento originario voleva esprimere. È il caso di licenze come la BSD (Berkeley Software Distribution), che non implica alcuna restrizione rispetto alla chiusura dei codici e quindi inibisce la viralità, perché porzioni di codice non libero possono essere integrate nel codice libero. Dunque una creazione libera può diventare proprietaria. Oppure la MPL (Mozilla Public Licence) e altre licenze sviluppate su misura per i nuovi prodotti Open Source.

L’economia di mercato diventa così sviluppo sostenibile e la Community degli sviluppatori il nucleo di una vera e propria Open Society⁵⁵, la chimerica Società Aperta. Questo immaginario è determinato non solo dall’adesione morale che suscita la pratica di uno sviluppo comunitario, ma soprattutto dalla qualità superiore degli applicativi, in apparente contraddizione con la gratuità delle competenze messe in gioco.

L’era dell’Open Source economy: concorrenza e bontà d’animo

L’ingresso dell’Open Source nel mercato è, secondo alcuni osservatori, una delle conseguenze della cosiddetta «convergenza tecnologica», uno slogan ormai divenuto quasi un paradigma dell’era informazionale: l’avvicinamento e sinergia di varie tecnologie, precedentemente ritenute estranee, studiate e sviluppate in ambiti separati.

Dinanzi a queste trasformazioni spesso estremamente rapide, la creazione di standard aperti ha creato un varco nella “guerra di tutti contro tutti” del cosiddetto “libero mercato”: “*cooperate on standards, compete on solution*”, è il motto della Ibm, una delle principali imprese coinvolte. Se anche Big Blue decide di cooperare, significa che il gioco vale la candela...

Per molte aziende l’Open Source è infatti una delle poche possibilità per contrastare monopoli e oligopoli ormai consolidati, per sfuggire a dinamiche di competizione classiche senza investimenti enormi, per limitare i costi di sviluppo e quindi diminuire “il prezzo” dei propri servizi .

Le imprese conoscono da tempo e apprezzano il valore di una dinamica reticolare di sviluppo e di alleanze: è ben noto che il valore di una rete è proporzionale al quadrato delle persone/nodi che

⁵³ Si veda in proposito: Eric S. Raymond, *La cattedrale e il bazar*, Apogeo, Milano,

<http://www.apogeoonline.com/openpress/cathedral>

⁵⁴ La legge di Brooks secondo Eric S. Raymond: <http://www.apogeoonline.com/openpress/libri/545/raymond.html>

⁵⁵ Per una trattazione approfondita, rimandiamo a Ippolita, *Open non è free*, Elèuthera, Milano, cap. II, online www.ippolita.net/onf/

collega⁵⁶. Una rete sempre più ampia corrisponde dunque a profitti esponenzialmente maggiori.

L'Open Source sembra offrire alcune garanzie rilevanti nello sviluppo di reti ad alto valore aggiunto: da una parte consente al software di rimanere in qualche modo un bene «pubblico» (adotta lo sviluppo aperto e si avvale di comunità di supporto); dall'altra mantiene molto bassi i costi di passaggio da un sistema all'altro (i cosiddetti *switching costs*), in particolare dai modelli proprietari a quelli aperti, anche e soprattutto nel caso dei sistemi *legacy*, cioè obsoleti. I costi maggiori nell'adozione di nuove tecnologie sono dovuti alla formazione degli utenti, non all'investimento per l'acquisizione della tecnologia stessa; a maggior ragione se si tratta di software eccellente dal costo risibile.

Ma il risultato più importante, il cui prezzo è difficilmente calcolabile, è la creazione di un'immagine completamente nuova: per la propria azienda e per i propri prodotti.

Il successo delle logiche del software libero ha prodotto diversi tentativi di applicazione delle medesime pratiche in altri settori. Sono stati inevitabilmente coniate nuove espressioni e formulati auspici spesso esagerati: Open Laws, Open Science, persino Open Society sembrano a portata di mano. Oggi infatti l'idea di società Open Source è divenuto quasi un paradigma della nuova era volta alla ricerca di uno strumento comune per una prospettiva politica “possibile”. Per società Open Source si intende infatti una società il cui codice sia aperto e al cui miglioramento possano partecipare tutti liberamente. Messa in questi termini non si può che essere d'accordo. Sorprende però la leggerezza con cui si mutua un termine da un percorso particolare, tecnico e informatico, per renderlo generale, applicandolo a teorie filosofiche, economiche, sociologiche, senza considerare le possibilità di modificare quel concetto o quantomeno utilizzarlo con le dovute precisazioni.

Open Source infatti, nel campo del software in cui è nato il termine, significa anche concorrenza, gare per accaparrarsi i cervelli migliori al minor prezzo, finanziamenti di capitale di rischio, operazioni miliardarie di acquisto e vendita, un grande business orientato a una forma più “democratica” e morbida di capitalismo. Questa dinamica mira non più ad asservire la forza lavoro, bensì a cooptarla nella realizzazione della missione dell'azienda, che si identifica sempre più con la realizzazione dei propri desideri individuali⁵⁷.

Tra le tante società che oggi sfruttano quest'onda per ottenere i più diversi vantaggi c'è anche Google che, d'altronde, è buono per motto: perciò, *Don't be Evil*, usa il software libero, è gratuito, è

⁵⁶ Si tratta di una legge matematica formulata alla fine degli anni Settanta da Robert Metcalfe, studente della Harvard University e poi fondatore della società 3Com oltre che pioniere del networking (e inventore del protocollo ethernet, ancora oggi fondamentale per le reti intranet). Ecco la traduzione ad hoc del passaggio saliente: “Il valore di un network cresce esponenzialmente del numero dei computer connessi al network stesso. Quindi, ogni computer che si aggiunge al network da una parte utilizza le risorse di questo (o le informazioni o le competenze) e dall'altra porta nuove risorse (o nuove informazioni o competenze) al network, facendone incrementare il valore intrinseco”. Da questo principio generale deriva che: 1) il numero di possibili relazioni, o meta-relazioni, o connessioni all'interno di un network cresce esponenzialmente rispetto alla crescita di computer collegato ad esso (di qui il valore strategico dei link all'interno di una rete); 2) il valore di una comunità cresce esponenzialmente rispetto alla crescita degli iscritti a questa comunità (di qui il valore strategico delle comunità virtuali). Si veda http://it.wikipedia.org/wiki/Legge_di_Metcalfe

⁵⁷ In un contesto arretrato come quello italiano, in cui le aziende, anche nel settore dell'IT, continuano ad applicare obsolete logiche fordiste di produzione di massa, senza valorizzare minimamente le potenzialità degli individui, potrebbe apparire sprezzante l'attacco nei confronti del capitalismo dell'abbondanza alla Google. Ma come non crediamo nello sviluppo sostenibile e nel consumismo responsabile promossi da un'abile propaganda terzomondista, così non possiamo avallare nessuna forma di sfruttamento degli individui, nemmeno quella di Googleplex, sottile e persino piacevole per i lavoratori ormai diventati creativi entusiasti. Non si tratta di rigidità dogmatica, ma di un minimo, comune buon senso (*common decency*): la missione ultima non è la piena realizzazione delle persone, bensì il dispiegamento senza fine del capitale, non importa se morbido o ferocemente repressivo. Richard Sennett l'ha mostrato con lucidità e un'ampia ricognizione storica in “L'uomo flessibile. Le conseguenze del nuovo capitalismo sulla vita personale”, Feltrinelli, Milano, 1999, il cui titolo originale è ben più icastico: “The Corrosion of Character”, La corrosione del carattere. D'altra parte, il consumismo estremo del capitalismo dell'abbondanza potrebbe costituire il primo passo verso un nuovo fascismo, come mostra la lucida visione di J.G. Ballard nei recenti romanzi Millennium People (Feltrinelli, Milano, 2004) e Regno a venire (Feltrinelli, Milano, 2006).

migliore di quello proprietario, gli sviluppatori sono fieri di collaborare. La visita a Googleplex ha mostrato chiaramente come a Mountain View questa strategia di penetrazione nelle vite delle persone sia stata affinata al massimo grado: lavoratori gratificati, incoraggiati a essere creativi e felici producono meglio e di più rispetto a lavoratori frustrati e oppressi.

Sedurre gli hackers: autonomia, soldi facili, strumenti gratuiti

Lo sfruttamento dell'Open Source da parte di Google raggiunge l'apice intorno al 2005, quando la sua immagine è stata appannata dalle azioni dei concorrenti e da vicende giudiziarie non cristalline⁵⁸.

Per quanto il progetto fosse notoriamente radicato nella cultura informatica e nella pratica dell'eccellenza accademica, non era sufficiente usare il sistema operativo GNU/Linux per far funzionare il datacenter di Google: occorreva esplicitare la fiducia verso l'Open Source con un'iniziativa forte, che richiamasse l'attenzione nel magma delle reti a produzione libera.

Non era più sufficiente offrire il supporto per la lingua h4x0r, il linguaggio dei “veri” hackers, (o per il klingoniano di Star Trek), per accattivarsi le simpatie degli sviluppatori. Inoltre l'atteggiamento elitario da cervelloni universitari cominciava a spazientire gli investitori.

Arroganza e culto meritocratico di derivazione accademica, anche se supportati da risultati sempre ottimali, hanno poco riscontro presso gli investitori che reclamano sostanziosi dividendi. Era inevitabile che si chiudesse la fase in cui i due potevano permettersi scherzosamente di quotare in borsa il titolo ipotizzando azioni per un valore di 2.718.281.828 dollari, un numero che ricalca la costante matematica “e” (la base della funzione logaritmo naturale); o anche uscite balzane come l'annuncio dell'agosto 2005, quando dichiarano che avrebbero venduto 14.159.265 azioni per raccogliere 4 miliardi di dollari di liquidità, senza far parola con gli investitori dell'uso che avrebbero fatto di quel denaro.

Per sostenere concretamente il proprio desiderio di investire in ricerca, per dimostrare come attraverso una simile strategia si possa non solo competere, ma eccellere sul mercato, era necessaria una mossa strategica rivolta non tanto agli utenti “normali”, ma piuttosto ai giovani cervelli, al futuro, all'innovazione; tradotto operativamente questo significa creare comunità, offrire strumenti di sviluppo, stipulare accordi con altre società del settore. Ovvero, corteggiare il mondo dell'Open Source.

Google comincia a investire nella creazione di comunità nell'Ottobre 2005, quando finanzia con 350.000 dollari l'Oregon State University e la Portland State University per migliorare la qualità dei progetti Open Source, favorendo la nascita di nuovi software. Successivamente viene avviato in pompa magna il programma mirato del Giugno 2005: *Summer Of Code*, cioè “L'estate del codice”, promosso direttamente su una pagina del sito di Google, e ora reperibile su <http://code.google.com/summerofcode05.html>.

Lo stile comunicativo è chiaro: offrire opportunità ai migliori. Ogni programmatore che avesse creato un progetto Open Source nuovo o apportato una miglioria degna di nota a un progetto già esistente entro l'arco dell'estate, avrebbe avuto un premio di 4500 dollari. L'operazione mirava evidentemente a presentarsi come un'esplicita dichiarazione d'amore verso l'Open Source, ovvero a rimarcare che l'Open Source era il terreno strategico sul quale coltivare l'innovazione. In secondo luogo, puntava ad attirare le simpatie dei giovani sviluppatori con un'iniziativa di sostegno finanziario concreto al loro lavoro. Infine, cercava di creare una vera e propria comunità nello stile “aperto”, quantomeno relativamente alla sponsorizzazione.

⁵⁸ Si veda il capitolo I.

I programmatori premiati, perlopiù studenti, sono stati oltre quattrocento; per la maggiore hanno apportato modifiche e introdotto novità su progetti già esistenti, piuttosto che far conoscere i loro nuovi software, aggiungendo caratteristiche a programmi come Apache, Fedora, Gaim, lo stesso Google, Inkscape, Jabber, KDE, Mozilla, OpenOffice.org, Python, Samba, Gnome, Mono, Ubuntu. Un bel guadagno per tutti, specialmente per le società che stanno dietro a questi progetti: tra le principali, ricordiamo IBM, RedHat, LinSpire, Novell, Mozilla.com, Sun, HP, Ubuntu⁵⁹.

Alcuni di questi progetti, insieme a quelli sviluppati nel famigerato 20% di tempo libero dai dipendenti di Google, hanno permesso il raggiungimento del secondo obiettivo nel percorso di avvicinamento all'Open Source: offrire strumenti di sviluppo. Già dal 2002 Google offriva strumenti di sviluppo scaricabili gratuitamente dal sito code.google.com. Adesso questa pagina contiene i progetti proprietari creati dai team di sviluppo di Google e insieme i progetti vincenti di Summer of Code che sono in qualche modo legati ai suoi servizi.

La sezione “Code” del sito propone alcuni progetti dedicati ai creatori di software nei più diversi linguaggi di programmazione (Java, C++, Python, etc.). Offrire strumenti di sviluppo è un elemento essenziale per chiunque voglia permettere la creazione di software e comunità, perché si investe proprio sui mezzi di lavoro necessari per la loro creazione. I progetti concepiti dai programmatori di Google come strumenti di sviluppo prendono il nome di Google API, librerie proprietarie per interfacciarsi e utilizzare i principali servizi del colosso di Mountain View.

Una libreria è un'insieme di funzioni condivise: si tratta di porzioni di codice compilate che forniscono strumenti ad altri programmi che abbiano la necessità di funzioni semplificate. Un esempio molto intuitivo è costituito dalle librerie grafiche GTK, QT e FLTK, che implementano gli elementi standard di un'applicazione visuale (pulsanti, menù, icone...), semplificando il lavoro dei programmatori: questi ultimi, appoggiandosi alla libreria preferita, scriveranno solo le funzioni reali del programma. Saranno infatti le librerie a disegnare i pulsanti, a gestire i click del mouse, a tracciare le ombre e molto di tutto ciò che siamo abituati a vedere come utenti. Il tempo e le capacità dei programmatori vengono così ottimizzate; poiché chi scrive codice difficilmente sarà entusiasta di creare i bottoncini delle sue applicazioni, le librerie grafiche rivestono un ruolo essenziale di collante fra i vari progetti. Da una parte, le applicazioni risulteranno relativamente omogenee dal punto di vista grafico; dall'altra, i programmatori potranno concentrarsi sul loro lavoro, senza perdere troppo tempo nell'implementazione della interfaccia.

Vi sono comunità di sviluppo che si occupano delle librerie per fornire strumenti generici e trasversali per la risoluzione di problemi complessi (connessioni di rete, comunicazione tra applicativi, gestione del testo, immagini, compressione). Esattamente come un software viene scritto con l'obiettivo di raggiungere più utenti possibile, una libreria è fatta per arrivare a quanti più sviluppatori possibile.

Le librerie quindi permettono ai programmatori di creare i propri software partendo da un set di elementi messi in condivisione, veri e propri standard *de facto* della programmazione. Appoggiarsi alle librerie significa implementare il lavoro avvalendosi di una base di partenza anche molto ampia e complessa, che utilizza al meglio il codice già disponibile e stratifica le competenze. Le librerie hanno quindi un valore strategico sia nella dinamica di cooperazione spontanea del FS sia nell'economia relazionale dell'OS.

Le librerie di Google, cioè le Google API, sono pubblicate con licenza proprietaria, cioè nascondono al programmatore il meccanismo del loro funzionamento. Ma non è tutto: è presente anche un particolare dispositivo di controllo, infatti lo sviluppatore che scarica gratuitamente le librerie deve autenticarsi attraverso un codice. Questo sistema permette a Google di tracciare in maniera pervasiva tutti i movimenti e le modifiche che derivano dall'uso delle sue API. I programmatori che usano queste librerie hanno l'opportunità di inserire la ricerca di Google nel

⁵⁹ *Summer of code* è stato riproposto nell'aprile 2006 <http://code.google.com/soc/>; il successo è garantito.

proprio sito o conoscere in tempo reale il proprio PageRank. Inoltre possono realizzare software capaci di gestire campagne pubblicitarie attraverso AdWords, generare mappe dinamiche dei loro dati con l'interfaccia di Google Maps o anche implementare client VoIP per la telefonia on-line compatibili con GTalk. In breve, possono sviluppare i servizi di Google come meglio credono, nel linguaggio di programmazione a loro più congeniale, sotto l'attenta supervisione di Mountain View.

L'enorme diffusione dei servizi di Google si coniuga con la possibilità di personalizzazione nei minimi dettagli: infatti, mediante la scrittura di appositi documenti XML⁶⁰, è possibile creare "ponti" di collegamento fra i diversi servizi di Google; per esempio programmando pezzo per pezzo l'*homepage* di Google come se fosse un vero e proprio applicativo, rendendola in questo modo del tutto adeguata alle proprie esigenze. Qualcosa di molto simile si può fare con Google Earth: è possibile costruire particolari navigazioni in 3D sulle foto satellitari, evidenziando graficamente sui computer degli utenti ulteriori zone geografiche, edifici, dati climatici, ecc.

Tutti questi strumenti predisposti per chi sa scrivere codice (in almeno un linguaggio) sono essenziali per trovare nuove combinazioni o semplicemente usare ciò che Google rende pubblico, almeno in parte, all'interno di propri applicativi⁶¹. Esiste addirittura un portale, googleearthhacks.com, dove si possono trovare moltissimi trucchi e *hack* per utilizzare nei modi più impensati questo servizio, incrociando le mappe satellitari con qualsiasi altro database.

Tutte le possibilità che le API di Google ci offrono implicano il rispetto di due precise regole: la registrazione e la licenza. Per attivare funzioni di Google API infatti è necessario richiedere una chiave, cioè un codice di accesso e comunicare esattamente dove le si vuole impiegare. Le API si attivano solo dopo aver effettuato questa registrazione. Il secondo punto è la licenza. Non avendo una licenza copyleft, queste API si possono utilizzare soltanto nel rispetto di alcune limitazioni: ad esempio, è necessario avere un account Google, perché l'accumulo di informazioni non si arresta mai; inoltre, le mappe sono di proprietà esclusiva di Google (o di terze parti) e non possono essere modificate in alcun modo. Naturalmente, per usi commerciali è necessario stipulare un contratto. Il codice di attivazione permette a Google di mantenere sempre il controllo totale sui nuovi programmi generati sulle sue API: può bloccare a piacimento gli applicativi o semplicemente controllare sia il modo in cui accedono ai servizi, sia l'uso che ne viene fatto. Tutto questo è possibile perché il codice sorgente non è pubblico e libero e risulta quindi impossibile comprendere il funzionamento interno delle librerie.

Oltre a far sviluppare gratuitamente e a monitorare lo sviluppo dei propri servizi, un altro motivo per cui Google sta creando comunità con questa strana formulazione che potremmo definire pseudo-Open è quello di ottenere un database sul quale fare ricerca e vendita di statistiche.

Ospitare gratuitamente i progetti dei singoli programmatori significa ottenere la loro fiducia. Permettere a chiunque di effettuare ricerche sul database dei progetti ospitati attiva una solida catena di utenti. Una simile incubatrice gratuita di giovani talenti garantisce inoltre la disponibilità di materiale umano fortemente motivato e la cui formazione, ovvero il costo principale nel settore dell'IT, è già stata compiuta in maniera autonoma e del tutto conforme allo stile dell'azienda.

⁶⁰ XML, eXtensible Markup Language, è un linguaggio estensibile realizzato per poter utilizzare in modo semplice i documenti strutturati, studiato per il Web e per superare i limiti di HTML (HyperText Markup Language), ma con possibilità di utilizzo in ambienti differenti. Sviluppato dal W3C, il World Wide Web Consortium, XML è un sottoinsieme di SGML (Standard Generalized Markup Language), uno standard internazionale che definisce le regole per scrivere markup language; volutamente non comprende alcune funzionalità complesse di SGML difficilmente implementabili su Web. La prima bozza di XML risale al novembre 1996, l'attuale specifica di XML è consultabile all'indirizzo <http://www.w3.org/TR/1998/REC-xml-19980210> la traduzione in italiano di questo documento è consultabile al seguente indirizzo: <http://www.xml.it/REC-xml-19980210-it.html>

⁶¹ Colpisce il fatto che il motore di ricerca, dal punto di vista dell'interazione con le API, sia limitato a mille richieste al giorno. Questo vuol dire che i software che si basano sulle API potranno effettuare al massimo mille interrogazioni giornaliere. Probabilmente le limitazioni diminuiranno con il tempo, perché la capacità di Google cresce senza sosta; tuttavia, le limitazioni vengono imposte da Google senza fornire nessuna motivazione.

L'offerta di strumenti di sviluppo è un meccanismo di *talent-scouting* noto da tempo e in particolare è il cavallo di battaglia di alcune solide aziende come la Va Software Corporation, che mette a disposizione gratuita del mondo dell'Open Source computer estremamente potenti e banda larga, spazio disco e assistenza non alla portata di tutti. Due paradisi digitali possono vantare una fama mondiale e un numero di progetti ospitati superiore a ogni altro concorrente: sourceforge.net e freshmeat.net, entrambi di proprietà della Va Software. Questi portali hanno una tale risonanza che anche i progetti più piccoli, una volta comparsi sulle prime pagine, arrivano tranquillamente a contare centinaia di visite al giorno. Tutti i progetti in seno a code.google.com hanno una loro pagina relativa su freshmeat.net e/o sourceforge.net.

In questo modo gli applicativi possono al contempo godere della visibilità di Google e di tutti i servizi messi a disposizione dal colosso Va Software: mailing list, computer dedicati per la soluzione degli errori di programmazione (*debug*), per sistemi di controllo delle versioni, delle revisioni e dei cambiamenti del codice (ad esempio *cvs*, *Concurrent Versioning System*), forum di discussione, ecc.

È semplice immaginare come, con un database utilizzato gratuitamente da migliaia di *coder*, la Va Software può garantire un ottimo servizio di *business to business* per aziende legate al mondo Open Source e non solo. Un *data mining* di particolare interesse per affari miliardari. Tra gli sponsor e inserzionisti di sourceforge.net e freshmeat.net troviamo Red Hat, Microsoft e molti altri.

Vi sono molti modi per mettere in collegamento gli sviluppatori con le aziende Open Source. In Italia, SUN Microsystem offre la possibilità di pubblicare il proprio curriculum su una mappa di Google (utilizzando le Google API), attraverso il portale javaopenbusiness.it. Sono gli sviluppatori stessi a segnalare il proprio profilo, creando così una mappa delle competenze Open Source in Italia attraverso gli strumenti resi disponibili da SUN e da Google.

Google può quindi contare sull'implementazione praticamente gratuita dei propri prodotti da parte di centinaia di utenti; a questo si aggiunge l'investimento mirato di gare come *Summer of Code*, festival dedicati alla promozione e sviluppo dei propri servizi e, ultimo ma non meno importante, sistemi di reclutamento eccezionalmente dinamici. Tra questi, si trova anche il video-reclutamento, direttamente sulle pagine di video.google.com, con interviste a dipendenti entusiasti e a Sergey Brin in persona, tutti concordi nell'illustrare i privilegi del lavoro a Mountain View⁶².

Ambienti ibridi fra università e azienda

Date queste premesse, l'avvicinamento di Google all'Open Source appare quanto mai strategico e interessato, per quanto senz'altro originato da un comune sentire rispetto alle dinamiche cooperative tipiche delle comunità di sviluppo Free Software, nate nell'humus accademico. La strategia dell'accumulo evidenziata in precedenza è all'opera anche in questo ambito: infatti Google si comporta come una sorta di buco nero che usa codici aperti, o addirittura ne favorisce la stesura e li attira, per poi immetterli nel proprio circuito. Ad esempio, nessuna delle modifiche che i programmatori di Google hanno apportato agli strumenti aperti usati è mai stata resa pubblica. In particolare il loro Google Web Server (GWS) è una versione modificata di una versione di Apache, il server web Open Source più diffuso nella Rete. Questo significa senz'altro sfruttare le potenzialità e le realizzazioni del metodo di sviluppo aperto, senza però condividere le proprie implementazioni e miglioramenti.

Un fattore di primaria importanza a proposito delle relazioni con il mondo Open Source è che Google nasce a Stanford, un'università nota per la sua capacità di generare *start-up* aggressive e competitive basandosi su ricerche di elevato profilo. Per quanto Stanford fosse, e continui a essere, un ambiente favorevole allo sviluppo di progetti Open Source, il legame a doppio filo con il capitale di rischio rende difficile e anzi impossibile proseguire sulla strada dell'eccellenza accademica una

⁶² Il video è disponibile all'indirizzo <http://video.google.com/videoplay?docid=-8618166999532839788>.

volta usciti dal campus.

Un breve accenno alla ricerca accademica americana è necessario per comprendere le origini di Google, tra l'Open Source e la ricerca orientata al profitto. Infatti a livello più generale va sottolineato il carattere accentratore dell'università statunitense a proposito di creazione intellettuale: tutti i progetti sviluppati in campo accademico sono tendenzialmente copyright dell'università che ha ospitato il gruppo di ricerca. Stanford non fa eccezione: del resto, negli Stati Uniti le accademie sono storicamente legate al mondo degli affari, e spesso sono vere e proprie imprese. I brevetti universitari sulle invenzioni dei ricercatori fruttano royalties di tutto rispetto; inoltre conferiscono prestigio ai centri di ricerca e agli studenti/ricercatori/imprenditori.

Le università sono ambienti ibridi, tra il pubblico e il privato. Negli USA fino al 2002, almeno in teoria, i luoghi di ricerca pubblici non potevano brevettare le loro invenzioni; lo stesso si dica per i laboratori privati ma finanziati con fondi pubblici (quindi spesso anche le università). Infatti il pagamento dei dazi ostacola la libera circolazione dei saperi nella ricerca scientifica, la possibilità di riprodurre, verificare o falsificare i risultati sperimentali. Questo in base all'*Experimental Use Defense*, “protezione dell'uso sperimentale”, un principio che autorizza l'uso gratuito di tecnologie brevettate nell'ambito della ricerca, introdotto nel 1813, e abolito appunto nel 2002 con la sentenza a favore del ricercatore John Madey. Madey ha citato in giudizio la Duke University, per cui lavorava, perché usava un'apparecchiatura da lui brevettata per ricerche laser su elettroni liberi. La Corte ha ritenuto che l'*Experimental Use Defense* fosse stato concepito per proteggere lo scienziato dedito alla ricerca disinteressata e libera, ma evidentemente nelle università questa attività non è più così innocente e, anche nel caso non sia direttamente commerciale, può essere considerata un “affare lecito” (*legitimate business*), poiché procura finanziamenti e necessita di forza lavoro e di personale in formazione (studenti). Cade così ogni distinzione fra la ricerca privata e quella pubblica⁶³.

Naturalmente, tutti i progetti nati a Stanford sono sottoposti a brevetto da parte dell'università, e questa commistione fra incentivo ai progetti Open Source da una parte, e brevettabilità selvaggia dall'altra, non giova certo all'ideale, né tanto meno alla pratica, della “ricerca” in sé, tanto sbandierata come punto d'orgoglio e di forza di Google.

La questione del brevetto si fa ancora più interessante se ricordiamo che il successo di Google si basa su un algoritmo ideato da Larry Page, a partire dalla collaborazione con Sergey Brin, quando erano ricercatori alla facoltà di Scienze Informatiche presso Stanford. L'algoritmo che ha rivoluzionato l'indicizzazione della Rete è quindi di proprietà di Stanford, sottoposto a regolare brevetto. Andiamo a scoprire nel dettaglio come funziona questo prodigio, come riesce a fornire risultati in tempi più rapidi di qualsiasi concorrente, quasi che davvero potesse dare a ogni utente “ciò che vuole”.

⁶³ Le notizie sono tratte da Laser, <http://www.e-laser.org/htm/news.asp?idNews=233> e <http://www.e-laser.org/htm/news.asp?idNews=151> e dal volume Gruppo Laser, op. cit., [www.ippolita.net/editoria/2/laser - il sapere liberato.pdf](http://www.ippolita.net/editoria/2/laser_-_il_sapere_liberato.pdf); fonti giuridiche dirette sul caso Madey: www.law.nyu.edu/journals/lawreview/issues/vol79/no4/NYU406.pdf

IV. Algoritmi, che passione!

La crescita vertiginosa di Google non ha intaccato la sua fama di motore rapido ed efficiente, affidabile e completo: tutti abbiamo sentito dire che “se non c’è su Google, non esiste!” e che “Google è più veloce”. Alla base di questo successo, oltre agli elementi che abbiamo analizzato finora, si trova l’algoritmo di PageRank, già citato in apertura, che guida lo spider di Google alla scoperta delle Reti. Vediamo in dettaglio di cosa si tratta e come funziona.

Algoritmi e vita reale

Un algoritmo⁶⁴ è un metodo risolutivo applicato a un problema, un procedimento che si compone di passi semplici da eseguire in sequenza per ottenere un dato risultato. Un algoritmo che perviene alla soluzione del problema è detto corretto, e se la soluzione viene ottenuta in tempi brevi è detto efficiente. Esistono molti diversi tipi di algoritmi, impiegati nei campi più disparati delle scienze; non si tratta però di astruse procedure che riguardano un’esigua minoranza di studiosi, bensì di pratiche che influenzano le nostre vite quotidiane molto più di quanto non sembri di primo acchito.

Ad esempio, le tecniche per registrare un programma televisivo utilizzano algoritmi, ma anche i metodi per ordinare un mazzo di carte o per pianificare le soste di un viaggio particolarmente lungo. In un tempo relativamente prevedibile, realizzando una serie di passi semplici e replicabili in maniera identica, scegliamo più o meno implicitamente gli algoritmi adeguati alla soluzione che stiamo cercando. Semplici significa soprattutto specificati in modo non ambiguo, immediatamente evidenti per chi applicherà l’algoritmo, cioè per il suo esecutore. In questo senso, una ricetta è un algoritmo: “fate bollire tre litri d’acqua in una pentola, salate e gettate cinquecento grammi di riso, scolate dopo dodici minuti, aggiungete spezie a volontà” è una descrizione di passi semplici e non ambigui, se il destinatario della ricetta è in grado di disambiguare passaggi come “salare”, oppure “aggiungete spezie a volontà”.

Gli algoritmi non sono necessariamente metodi per raggiungere una soluzione nel minor tempo possibile. Infatti ne esistono alcuni che si occupano di ottenere soluzioni accettabili senza preoccuparsi del fattore tempo; altri ancora permettono di raggiungere un risultato nel minor numero di passaggi, oppure hanno come priorità il risparmio di risorse⁶⁵.

Importa qui sottolineare, al di là di qualsiasi approfondimento specialistico, la natura pratica, applicativa degli algoritmi. Gli algoritmi riguardano tutti noi perché sono pratiche concrete per raggiungere un dato obiettivo. In campo informatico vengono utilizzati per risolvere problemi ricorrenti nella programmazione dei software, nella progettazione delle reti e nella costruzione di

⁶⁴ “Algoritmo: insieme di regole o direttive atte a fornire una risposta specifica a uno o più dati in input”. Per una prima introduzione, <http://it.wikipedia.org/wiki/Algoritmo>. Il termine algoritmo deriva dal nome di “al-Khwarizmi”, importante matematico arabo del nono secolo. Muhammad ibn Musa al-Khwarizmi ha introdotto l’uso dei numeri arabi nella matematica: la sua opera “(Libro) di al-Khwarizmi sui numeri indiani” fu tradotta in latino come “Algorismi de numero Indorum”. Su questo testo l’Europa intera imparò ad usare il sistema di notazione decimale posizionale ancora oggi in vigore; la numerazione romana venne progressivamente abbandonata. Le procedure che permettevano di effettuare calcoli in notazione decimale divennero così note come “Algorismi” o “Algoritmi” e più tardi lo stesso termine fu applicato in generale alle procedure di calcolo necessarie per ottenere un determinato risultato.

⁶⁵ Il metodo migliore per raggiungere Parigi può essere quello di partire con un volo diretto dall’aeroporto più vicino alla propria città, oppure quello di prendere il primo volo in partenza dallo stesso aeroporto, scendere alla stazione d’arrivo, risalire sul primo aereo, scendere e nuovamente ripartire, continuando finché Parigi non sarà raggiunta. È abbastanza certo che entrambi i metodi ci permetteranno di raggiungere la destinazione: con il primo approccio arriveremo alla meta nel minor tempo possibile e probabilmente con il minor spreco di denaro; il secondo, invece, ci permetterà di apprezzare i principali aeroporti delle capitali europee nei diversi periodi dell’anno. Algoritmi differenti descrivono le diverse possibilità.

apparecchiature hardware. Negli ultimi anni, soprattutto a causa della crescente importanza dei modelli reticolari di analisi e interpretazione della realtà, molti ricercatori hanno focalizzato i loro studi sulle metodologie di costruzione e di percorrenza delle reti e dei dati che ne costituiscono la materia viva. L'economia della ricerca di cui parla John Battelle⁶⁶ è resa possibile dal perfezionamento di algoritmi per la ricerca di informazioni, studiati per accrescere le possibilità di reperimento e condivisione dei dati in maniera sempre più efficiente, veloce, affidabile e sicura. Il caso più noto al grande pubblico è il fenomeno del peer-to-peer: invece di creare enormi banche dati a cui è necessario accedere per trovare video, audio, testi, software e ogni genere di informazioni, vengono sviluppati in continuazione algoritmi sempre più ottimizzati per facilitare la creazione di reti altamente decentrate, nelle quali ogni utente si può mettere in contatto direttamente con altri utenti e attuare scambi proficui⁶⁷.

La strategia dell'oggettività

L'aumento vertiginoso della qualità e della quantità di banda dei nostri computer, insieme alla costante diminuzione dei costi, ci ha permesso di navigare in internet meglio, per più tempo e più velocemente. Solo vent'anni fa i modem a pochi *baud* (numero di simboli trasmesso al secondo) erano un lusso per pochi, mentre ora anche in Italia la fibra ottica, attraverso cui viaggiano milioni di *byte* al secondo, è una tecnologia accessibile.

Dieci anni fa erano necessarie elevate competenze informatiche per creare contenuti adatti alle reti digitali; ora invece la maggiore facilità di pubblicazione di contenuti sul web, l'onnipresenza della posta elettronica, il miglioramento dei sistemi di scrittura collettiva online, come blog, wiki, portali, mailing list e parallelamente l'abbassamento dei costi di registrazione e manutenzione dei domini e degli spazi Internet favoriscono la trasformazione degli utenti: da semplici fruitori di informazioni messe a disposizione da specialisti dell'IT, essi divengono sempre più creatori di informazioni.

Il miglioramento della connettività procede dunque di pari passo con una crescita esponenziale dei dati immessi in rete e quindi, come già abbiamo avuto modo di notare, implica la pressante necessità di strumenti di ricerca sempre migliori. L'urgenza diffusa a ogni livello di servizi di ricerca attira forzatamente l'interesse di sociologi, informatici, ergonomisti, designer, studiosi della comunicazione in genere. D'altra parte, il diluvio informativo delle reti globali non è una banale "messa in rete" delle società così come le conosciamo, ma un fenomeno estremamente complesso, che esige interpretazioni non banali. Crediamo pertanto che tale impegno teorico e pratico, non possa essere delegato agli specialisti, ma debba essere frutto di un'elaborazione collettiva.

Infatti se da un lato la costruzione di reti autogestite può essere un'occasione per ampliare e collegare fra loro zone autonome, dall'altro il controllo sociale trova nelle tecnologie dell'informazione uno strumento di repressione formidabile.

La realizzazione di questo secondo scenario, di cui il caso Echelon⁶⁸ è solo la manifestazione più

⁶⁶ Si veda il capitolo II.

⁶⁷ Il peer-to-peer. Generalmente con peer-to-peer (o P2P) si intende una rete di computer o qualsiasi rete che non possieda client o server fissi, ma un numero di nodi equivalenti ("peer" significa, appunto, "pari") che fungono sia da client che da server verso altri nodi della rete. Questo modello di rete è l'antitesi dell'architettura client-server, dove esiste un rapporto gerarchico tra il fornitore di servizio (server) e il ricevente (client). Mediante questa configurazione distribuita, qualsiasi nodo è in grado di avviare o completare una transazione. I nodi equivalenti possono differire nella configurazione locale, nella velocità di elaborazione, nell'ampiezza di banda e nella quantità di dati memorizzati. L'esempio classico di P2P è la rete per la condivisione di file (fonte: http://www2.autistici.org/inventa/doku.php?id=glossario_no-copyright_ecc).

⁶⁸ Echelon è il nome di un sistema di sorveglianza e spionaggio elettronico globale messo in opera dagli USA. La giustificazione ufficiale è oggi la lotta al terrorismo: comunicazioni telefoniche, mail e ogni comunicazione digitale viene intercettata e analizzata per fini politici ed economici. Per un'analisi approfondita, si veda: Duncan Campbell, *Il mondo sotto sorveglianza - Echelon e lo spionaggio elettronico globale*, Elèuthera, Milano, 2003; <http://home.hiwaay.net/~pspoole/echres.html> raccoglie moltissime risorse disponibili.

clamorosa, appare certamente più probabile alla luce del costante aumento del numero di individui che producono informazioni, contrapposto alla diminuzione continua dei fornitori di strumenti di ricerca. L'accesso alle informazioni prodotte da un numero sempre più imponente di individui è gestito da un pugno di monopolisti che riducono una delicata questione sociale e politica a una gara di marketing senza esclusione di colpi, nella quale l'adozione di un algoritmo migliore risulta essere l'elemento vincente.

Infatti un algoritmo di ricerca è uno strumento tecnico che attiva un meccanismo di marketing estremamente sottile: l'utente si fida del fatto che i risultati non siano filtrati e corrispondano alle preferenze di navigazione che la comunità di utenti genera. In sostanza, si propaga un meccanismo di fiducia nell'oggettività della tecnica (nello specifico, la procedura algoritmica che genera il risultato dell'interrogazione) che viene ritenuta "buona" in quanto non influenzata dalle idiosincrasie e dalle preferenze di individui umani. Le macchine "buone", figlie di una scienza "oggettiva" e di una ricerca "disinteressata", non manipoleranno i risultati, non ci diranno bugie perché non possono mentire e comunque non avrebbero alcun interesse a farlo. La realtà è ben diversa e questa credenza si rivela un'ipotesi demagogica, dietro alla quale le macchine del marketing e del controllo accumulano profitti favolosi.

Il caso di Google è l'esempio lampante di questa "strategia dell'oggettività" legata alla tecnica: infatti il motore di ricerca "buono per motto" sfrutta e traccia interamente e in maniera continuativa i comportamenti degli utenti che utilizzano i suoi servizi, al fine di profilare le loro abitudini e inserire nelle loro attività (navigazione, posta, gestione file, ecc.) pubblicità personalizzate, contestuali, leggere, onnipresenti, e possibilmente in grado di generare *feedback*, in modo che gli utenti siano in grado di fornire nel modo più semplice informazioni utili per i venditori e anzi giungano a migliorare essi stessi i "suggerimenti pubblicitari", esprimendo le proprie preferenze. La richiesta continua dell'opinione degli utenti, oltre a lusingare le persone che si sentono partecipi di una vasta "democrazia elettronica", è in effetti il modo più semplice ed efficace per ottenere informazioni preziose dal punto di vista commerciale sui gusti dei consumatori. Sono le preferenze (e l'inconsapevolezza) degli utenti a far vincere un motore di ricerca sugli altri, poiché un sito molto visitato può modificare i suoi contenuti in base a "suggerimenti" commerciali e attivare di conseguenza virtuosi movimenti economici.

Da un punto di vista squisitamente informatico, ai motori di ricerca compete la gestione di quattro elementi: la ricerca di dati nella rete (spider), la memorizzazione delle informazioni in appositi archivi (basi di dati), un valido algoritmo per ordinare i dati secondo le ricerche formulate (interrogazione), e infine lo sviluppo di un'interfaccia capace di soddisfare l'utente; ciascuno dei primi tre aspetti viene curato da un apposito tipo di algoritmo: ricerca, memorizzazione/archiviazione, interrogazione.

La potenza di Google, come di Yahoo! e altri giganti della ricerca in rete, è dunque basata su:

8. "spider", ovvero un software per prelevare contenuti dalle reti;
9. hard-disk di enorme capienza per memorizzare i dati su supporti affidabili e ridondanti, onde evitare qualsiasi perdita accidentale;
10. un sistema rapido per trovare (e ordinare) i risultati di un'interrogazione in base al valore di ranking delle pagine;
11. infine un'interfaccia utente via web (ma non solo: Google Desktop e Google Earth, ad esempio, sono programmi da installare sulla macchina dell'utente) per rispondere alle richieste riguardanti queste informazioni.

Spider, basi di dati e ricerche

Lo spider è un applicativo che, nella maggior parte dei casi, viene sviluppato nei laboratori di

ricerca degli stessi motori di ricerca. Il suo scopo è quello di navigare saltando tra un link e l'altro sulle pagine del web raccogliendo informazioni: formati dei documenti, parole chiave, autori delle pagine, ulteriori links, ecc. Al termine delle sue esplorazioni il software-spider consegna il tutto alla base di dati che archiverà le informazioni. Inoltre lo spider deve preoccuparsi di captare le variazioni di ogni sito e quindi programmare una successiva visita per immagazzinare nuovi dati. In particolare lo spider di Google gestisce due tipologie di scansioni dei siti, una mensile approfondita, *Deep-crawl*, e una giornaliera di aggiornamento, *Fresh-crawl*. In questo modo la base dati di Google viene costantemente aggiornata dallo spider sulle evoluzioni delle reti. Dopo una scansione approfondita Google impiega qualche giorno per aggiornare le varie indicizzazioni e propagare i nuovi risultati in tutti i *datacenter*. Questo lasso di tempo è noto come *Google dance*, (danza di Google): i risultati delle ricerche differiscono anche sensibilmente, poiché fanno riferimento a indici diversi. A partire dal 2003 Google ha modificato le sue metodologie di catalogazione e aggiornamento, limitando drasticamente gli effetti della “danza” e spalmandoli nel tempo; in effetti, ora i risultati delle ricerche variano in modo dinamico e continuativo senza nessuno stravolgimento periodico. In realtà i risultati delle ricerche differiscono anche in base alle precedenti navigazioni degli utenti, che vengono archiviate e utilizzate per “migliorare”, nel senso di “semplificare”, il reperimento delle informazioni⁶⁹.

La sequenza di scelte che l'applicativo compie per indicizzare un sito è la vera potenza dell'algoritmo di Google. Mentre l'algoritmo di base PageRank è depositato sotto brevetto da Stanford, e quindi pubblico, questi ulteriori passaggi algoritmici non sono rilasciati pubblicamente né da Google, né da nessuno dei motori di ricerca attualmente in uso; allo stesso modo non sono pubblici i processi di salvataggio nella base di dati.

In ambito informatico, una base di dati (*database*) è, in sostanza, un archivio digitale; nella sua forma più semplice – e attualmente più diffusa – è rappresentabile sotto forma di una o più tabelle in relazione fra loro che presentano valori in entrata e valori in uscita: si parla allora di database relazionale. Come ogni archivio, una base di dati è organizzata secondo precise regole di stoccaggio, estrazione e continuo miglioramento dei dati stessi (recupero di dati corrotti, correzione di voci duplicate, costante reingegnerizzazione dei processi di acquisizione dei dati, ecc.).

Gli informatici studiano da decenni le metodologie di ricerca, immissione e miglioramento dei dati in database, sperimentando linguaggi di programmazione e approcci differenti (gerarchico, reticolare, relazionale, a oggetti, ecc.). La progettazione di una base di dati è una componente cruciale del processo di sviluppo di un sistema informativo complesso come Google, poiché da essa dipende essenzialmente la sua funzionalità. Per ottenere una rapida estrazione dei dati e, in generale, una gestione efficiente, è quindi fondamentale la corretta individuazione degli scopi del database e, nel caso dei database relazionali, delle tabelle, da definire attraverso i loro campi e le relazioni che le legano. Naturalmente è necessario adottare approssimazioni inevitabili nei passaggi fra le lingue naturali, analogiche, e i dati immessi, digitali, che sono evidentemente discreti: zero o uno, l'informazione è presente oppure no, non esistono vie di mezzo. Il punto dolente è la segretezza di queste metodologie: come avviene in tutti i progetti di sviluppo proprietari, a differenza di quelli liberi, è molto difficile sapere quali strumenti e quali algoritmi siano stati utilizzati.

Attraverso i testi redatti dai centri di ricerca e dalle università è possibile reperire le scarse informazioni rese pubbliche a proposito dei progetti proprietari. Su questi testi si trovano informazioni utili per comprendere la struttura dei computer e la gestione dei dati da parte dei motori di ricerca. Per dare un'idea della potenza di calcolo attualmente disponibile, vengono descritti computer capaci di convertire indirizzi Internet in sequenze univoche di byte utili come indici per i database in 0.5 microsecondi e capaci di eseguire 9000 spider in contemporanea; scendendo nel concreto, si tratta di sistemi in grado di analizzare e immagazzinare circa 50 milioni

⁶⁹ Si veda il cap. V.

di nuove pagine al giorno⁷⁰.

L'ultimo elemento algoritmico che si cela dietro alla "semplice" facciata di Google è il dispositivo di ricerca, ovvero quel sistema che, data una interrogazione utente, è capace di trovare i risultati più congrui, ordinarli per importanza e ranking, infine inviarli all'interfaccia.

Alcune università e laboratori hanno deciso di rendere pubbliche le loro ricerche in tal senso, in particolare le soluzioni raggiunte e i differenti approcci utilizzati per ottimizzare la velocità di accesso alle informazioni, la complessità dell'ordinamento e la selezione dei parametri di input più interessanti.

I motori di ricerca, infatti, devono essere in grado di fornire risultati ottimali quasi istantaneamente, offrendo nel contempo un ventaglio di possibilità di scelta il più ampio possibile. Google rappresenta senz'altro lo stato dell'arte dei motori di ricerca: simili straordinari risultati si possono ottenere solo grazie all'implementazione di opportuni filtri, come vedremo approfonditamente nel prossimo capitolo.

Per ora è importante sapere che l'esito migliore viene assicurato attraverso il giusto bilanciamento tra potenza di calcolo e qualità dell'algoritmo di ricerca. Ricercare un'informazione tra i *terabyte* (1 TB = 1000 GigaByte) o *petabyte* (1 PB = 1000 TB = 1 milione di GigaByte) necessita l'impiego di straordinari supporti di archiviazione e formidabili sistemi di indicizzazione, con il compito di individuare sia in quale punto dell'enorme archivio si trova l'informazione che calcolare il tempo necessario per prelevarla.

La Rete trabocca di leggende non sempre verificate né verificabili a proposito della capacità computazionale di Google, anche perché l'azienda rivela pochi particolari della propria infrastruttura tecnologica. Alcune fonti parlano di centinaia di migliaia di computer collegati fra loro in migliaia di giganteschi *cluster* che montano apposite distribuzioni GNU/Linux; altre di supercomputer, dispositivi la cui estetica rimanda a scenari fantascientifici: enormi silos super refrigerati nei quali uno o più bracci meccanici spostano alla massima velocità migliaia di dischi rigidi. Entrambe le soluzioni sono plausibili, insieme ad altre ancora, e non sono necessariamente in contraddizione. Di certo, l'estrema scalabilità delle macchine di Google consente prestazioni eccezionali, dal momento che il sistema è "aperto" a continui miglioramenti.

Dalla brand-identity all'interfaccia partecipativa

Ricerca, archiviazione e reperimento dei dati sono procedure estremamente complesse e necessitano, per essere comprese a fondo, conoscenze e approfondimenti che esulano dagli intenti di questo testo. Vedremo più avanti alcuni dettagli del loro funzionamento. Un'attenzione particolare va dedicata all'interfaccia perché mentre le performance dell'algoritmo e l'architettura della base di dati sono elementi strutturali del motore di ricerca che rimangono invisibili all'utente, l'interfaccia è progettata e gestita come immagine di Google stesso.

Per interfaccia intendiamo innanzitutto il "blank box"⁷¹, quello spazio vuoto nel quale si immettono le proprie domande o "intenzioni di ricerca" nel quadro della pagina universale di Google, studiata per risultare accogliente, confortevole, familiare.

Si tratta di un'impostazione detta universale perché viene declinata in numerose lingue (al momento, oltre 104 fra lingue e dialetti personalizzabili per oltre 113 paesi) e in ognuna di queste presenta un modello di interazione che rimane invariato e che unifica i comportamenti di ricerca in

⁷⁰ Si veda ad esempio la documentazione resa pubblica da IBM Almaden Research Center:

<http://www.almaden.ibm.com/webfountain/publications/>

⁷¹ Nel gergo informatico, "black box" si riferisce a una "scatola nera" che riceve input, li elabora in maniera non trasparente per l'utente e restituisce un output. Il concetto di "blank box" ricalca questo metodo, ma in maniera implicita, e perciò ambigua, perché pur trattandosi di uno "spazio pulito, vuoto" (*blank*, appunto) è carico di significati e funzioni di ricerca altamente differenziate.

uno schema unico e omogeneo.

Sulla pagina di Google ci troviamo di fronte un'interfaccia lineare composta da elementi essenziali, ciascuno con una funzione ben precisa e universalmente riconosciuta. Essa è in grado di accettare indicazioni di ricerca di diversa natura e complessità, dall'introduzione di semplici parole chiave (es. "ippolita") a parole composte, che vanno poste tra virgolette (es. "comunità scrivente"), fino a ricerche mirate: ad esempio, le ricerche possono essere limitate a un sito particolare, oppure a una lingua specifica, a pagine provenienti solo da un determinato dominio, o ancora a documenti di un certo formato, e così via, a seconda del grado di raffinatezza che si vuole ottenere. Si tratta cioè di un esempio riuscito d'interfaccia che raggiunge il non semplice obiettivo di associare un significato positivo allo spazio bianco della pagina. L'interfaccia si presenta senza orpelli, quasi vuota, o meglio riempita da un unico elemento "vuoto": il *blank box*, che rassicura l'utente e tende a indurre comportamenti attivi, invece di provocare lo smarrimento dovuto all'assenza di punti di riferimento, o viceversa dalla presenza di input visivi sovrabbondanti. Si evita così la confusione generata dalle pagine troppo piene, quasi fossero affette da una sorta di *horror vacui*, da un'ansia comunicativa che, nel tentativo di attirare l'utente con mille banner, effetti grafici, giochini, ottiene spesso l'effetto contrario.

Non esiste una navigazione vera e propria sulla pagina di Google: le diverse componenti della pagina hanno un significato funzionale, servono per accedere a servizi, non per condurre l'utente in un percorso; il loro utilizzo innesca comportamenti che diventano parte molto rapidamente di una routine di ricerca, al punto da apparire istintivi dopo poco tempo. L'interfaccia del motore di ricerca è studiata in modo che l'utilizzo, la dinamica di funzionamento e le aspettative dell'utente, un utente generico, si ripetano; anzi, anche dopo aver immagazzinato e digerito le "personalizzazioni" dell'utente stesso, le pratiche di ricerca rimangono sostanzialmente identiche, tanto che possiamo parlare di uno strumento "universale".

La disposizione di testi e immagini è lineare e si avvale dell'utilizzo di elementi grafici ricorrenti, ad esempio l'impiego dei colori elementari; le immagini usate sono qualitativamente omogenee. Lo stile di progettazione dell'interfaccia è sobrio, quasi scarno e, a dispetto del design di tendenza delle *brand-identity* (e della *corporate-identity*)⁷² orientato alla ricerca di una specificità estetica, fa riferimento a qualità percettive elementari ma molto efficaci nella loro semplicità.

Da questa identificazione visiva immediata deriva una facilità d'uso nettamente superiore rispetto ai motori di ricerca concorrenti. Il livello di ergonomia raggiunto è stupefacente: Google non ha la necessità di mostrarsi come un accentratore di servizi attraverso la propria interfaccia; in altre parole, la sua architettura visiva è quella tipica dei portali multiservizio. Le interfacce dei diversi servizi sono autonome e sostanzialmente indipendenti, caratterizzate tutte dalla presenza della "blank box" e non linkate le une con le altre in maniera diretta. Ad esempio, sono necessari molti passaggi non intuitivi per raggiungere il servizio di code.google.com, pensato per tecnici di vario livello, partendo dal servizio base di ricerca delle immagini, ovvero images.google.com, indirizzato a un pubblico più generico: è necessario scendere "in profondità" nel sito google.com e sapere cosa cercare. Nonostante questa frammentazione, siamo tutti in grado di riconoscere la rete di servizi offerta da Google; inoltre i fruitori sono in grado di utilizzare in maniera combinata e integrata le risorse informative messe a disposizione, sia per coloro che si limitano al semplice uso del browser, sia per Google-dipendenti, i *Google-totally-addicted*⁷³ che si precipitano entusiasti su ogni nuovo servizio.

⁷² L'ideazione di una nuova immagine per un prodotto o un servizio è nota come *brand identity*; quando riguarda una società specifica si parla di *corporate identity*. Ormai il concetto di "brand" ha ampiamente superato l'idea di "marchio distintivo", giungendo a configurarsi piuttosto come una "marca" che ha necessità di espansione psichica, territoriale, commerciale. Per una prima introduzione, si veda <http://it.wikipedia.org/wiki/Marca>

⁷³ La Google-mania dilaga e genera nuovi linguaggi; per una panoramica, si veda: [http://en.wikipedia.org/wiki/Google_\(search_engine\)](http://en.wikipedia.org/wiki/Google_(search_engine)); un elenco di servizi e strumenti correlati a Google, http://en.wikipedia.org/wiki/List_of_Google_services_and_tools

Questa deterritorializzazione dei servizi genera un peculiare meccanismo relazionale: gli utenti non vengono a conoscenza delle nuove sezioni direttamente da Google, ma dalla rete informale degli utilizzatori, da altri siti sui quali i visitatori espongono i loro gusti e discutono delle loro abitudini. La vasta gamma dei servizi offerta da Google viene automaticamente localizzata dal fruitore stesso nel momento in cui si interessa a un nuovo servizio: ad esempio, per quanto riguarda la zona geografica, viene presentata immediatamente l'interfaccia linguistica appropriata all'utente. D'altra parte, è semplice inquadrare la tipologia di utenti a cui un servizio è indirizzato, e valutare il grado di preparazione tecnica richiesto, o il grado di affinità con gli altri utilizzatori. Il meccanismo di passaparola diventa dunque simile a un "PageRank relazionale".

In prima approssimazione, esistono una dimensione relazionale locale, nella quale il passaparola avviene fra amici e conoscenti, e una dimensione relazionale tipologica, nella quale un certo tipo di utenti, identificabili in base a parametri statistici (età, sesso, impiego, ecc.) utilizza un particolare servizio e mette in moto l'economia relazionale.

I dieci problemi relativi all'usabilità dei siti web, discussi da Jakob Nielsen⁷⁴, fra i più noti studiosi di interfacce utente, sembrano non intaccare minimamente il sito di Google che, nonostante sia scritto in linguaggio HTML totalmente fuori standard⁷⁵, riesce ad assicurare la piena visibilità su tutti i browser, grafici o testuali che siano.

La pulizia grafica delle pagine viene esaltata da un'ottima gestione visiva degli aspetti commerciali. Nessun link pubblicitario in homepage o nelle pagine di documentazione e informazione: la pubblicità in Google si trova solo tra i risultati delle ricerche, appositamente separata dai risultati proposti ma non estranea agli argomenti ricercati. Si può dire quindi che Google è capace di esprimere, quantomeno circa la disposizione scenica delle sue interfacce, il giusto compromesso tra rispetto degli utenti e necessità di ritorno economico. La pubblicità, principale fonte di introiti di Google, viene progettata e realizzata in modo da non diventare invasiva e non distrarre gli utenti dal loro utilizzo dei servizi.

I link pubblicitari sono sponsorizzati in modo dinamico per seguire il percorso compiuto da un utente all'interno del motore di ricerca e quindi, in seconda istanza, sui siti Internet.

I collegamenti commerciali dunque non sono statici, ma si modificano e accompagnano le ricerche degli utenti; questo è possibile anche grazie ai feed RSS (acronimo di *RDF Site Summary*, o di *Really Simple Syndication*), uno dei formati più utilizzati per la distribuzione di contenuti Web, e in virtù delle diverse sorgenti informative digitali (quotidiani, riviste, agenzie di stampa, ecc.) in grado di modificare dinamicamente l'homepage di Google. Infatti Google mette la sua homepage a disposizione degli utenti registrati, rendendola totalmente configurabile grazie all'aggiunta di feed RSS: è così possibile impostare le previsioni del tempo automatiche per le città che si desidera monitorare, oppure scandagliare l'archivio storico delle ricerche effettuate. Si possono organizzare i segnalibri e gli ultimi messaggi di posta ricevuti, ma anche tenere sotto controllo i file del proprio computer senza soluzione di continuità rispetto ai contenuti web, grazie all'applicativo Google desktop.

Il meccanismo di promozione pubblicitaria, i servizi e i sofisticati meccanismi di profilazione dell'utente sembrano costituire un tutt'uno a livello estetico e contenutistico; dal canto loro, i link

⁷⁴ Jakob Nielsen, cinquantenne informatico danese, è una delle voci più autorevoli nel campo dell'usabilità del web. Nielsen è noto, fra l'altro, per le sue critiche all'eccesso di grafica e animazioni (ad esempio Flash) che affliggono molti siti popolari a spese dell'usabilità, pratica dannosa soprattutto per i disabili. Il suo ultimo volume pubblicato è Jakob Nielsen, Marie Tahie, *Homepage Usability*, Apogeo, Milano, 2002. I dieci problemi dell'usabilità:

<http://www.shinynews.it/usability/1005-errori.shtml>

⁷⁵ HTML (acronimo per Hyper Text Mark-Up Language) è un linguaggio usato per descrivere i documenti ipertestuali disponibili nel Web. Non è un linguaggio di programmazione, ma un linguaggio di markup, ossia descrive il contenuto, testuale e non, di una pagina web. Si veda: <http://it.wikipedia.org/wiki/HTML>, ma soprattutto il sito della W3 sullo standard HTML: <http://www.w3.org/MarkUp/>

sponsorizzati sarebbero in questo senso nient'altro che semplici suggerimenti, graficamente compatibili e concettualmente allineati con l'operazione di ricerca che si sta compiendo. L'economia di Google è altamente integrata con l'interfaccia, al punto da poter essere esclusa a livello visivo da chi non ne è interessato e sfruttata da chi invece trova interessante i link e i percorsi commerciali proposti.

Anche Yahoo!⁷⁶ e molti altri motori di ricerca e portali mettono a disposizione strumenti analoghi per la personalizzazione della propria homepage; tuttavia la quantità e la qualità delle offerte di Google, al momento, rimane insuperata. Si tratta di configurazioni piuttosto semplici, ma che richiedono in ogni caso una certa dimestichezza con le interfacce web e un po' di tempo per essere realizzate. In ambito web la soglia di attenzione è notoriamente bassissima, le pagine vengono visualizzate e abbandonate in tempi molto rapidi, dell'ordine di pochi secondi; perciò un utente che investe parecchi minuti, o decine di minuti, opera delle scelte che rivelano molto di sé e delle proprie abitudini di consumatore. Queste informazioni, accuratamente archiviate dalla compagnia di turno (Google o Yahoo! che sia) costituiscono la vera ricchezza prodotta dall'utente stesso, e sono fondamentali per proporre beni e servizi mirati da parte delle aziende sponsor.

La personalizzazione delle pagine rende un sito più amichevole, il sito stesso diventa come uno strumento personale in cui l'utente investe tempo, scegliendo colori, aspetto, contenuti. Un visitatore abituale in grado di configurare la propria pagina iniziale viene cooptato e reso partecipe nella costruzione dell'interfaccia web. Concedere il potere e il controllo su alcune pagine all'utente significa promuoverlo da semplice bersaglio di campagne pubblicitarie a consumatore "intelligente" ed è senz'altro il modo migliore e più sottile per creare fidelizzazione promuovendo l'interazione. Si profilano ambienti dotati di interfacce partecipative e non esclusive per ricevere pubblicità sempre più personalizzate, per entrare tutti insieme nel dorato mondo di Google.

PageRank, o l'autorità assoluta di un mondo chiuso

L'algoritmo che permette a Google di assegnare un valore alle pagine indicizzate dallo spider è noto come PageRank.

Sappiamo già che il funzionamento del PageRank si basa sulla popolarità di una pagina web, calcolata in base al numero di siti che hanno almeno un link puntato a essa. A parità di numero link, due pagine web avranno PageRank diversi in base all'importanza di chi li ha linkati: con questo meccanismo si valuta l'aspetto "qualitativo" dei siti. I link presenti nelle pagine web più linkate otterranno importanza superiore rispetto a quelli presenti nelle pagine meno linkate.

Facciamo un esempio concreto: spesso, controllando le statistiche di accesso relative a un sito, si riscontrano un numero enorme di contatti provenienti da siti pornografici. Questo avviene perché Google attribuisce un ranking dipendente dagli accessi, che a loro volta vengono visualizzati nelle pagine di statistiche pubbliche. Esistono perciò programmi che sfruttano la pervasività di questa logica di connessione e valutazione dei nodi di una rete per innalzare il proprio rank; come spesso accade, i primi sperimentatori sono i siti pornografici (come è stato per le gallerie di immagini su web, o per il commercio online).

In pratica vengono utilizzati alcuni programmi che si occupano di cercare i siti con statistiche di accesso pubbliche; viene quindi effettuato un numero molto elevato di richieste, simulando visite provenienti da un finto link presente in un altro sito, che nella maggior parte dei casi è appunto un sito pornografico. Questo meccanismo di bombardamento fa letteralmente impennare il numero di accessi al sito in questione e di conseguenza le statistiche mostrano incrementi evidenti; in questo modo aumenterà sensibilmente il Google-Ranking del sito e in ultima analisi anche quello del sito pornografico da cui i link sono arrivati: insomma, un guadagno per tutti, almeno a livello di visibilità.

⁷⁶ Si veda ad esempio MyYahoo!, my.yahoo.com

Questo tipo di operazione non è illegale: nessuno vieta di fare richieste a un sito Internet; grazie a questa pratica i siti a statistica pubblica ottengono un ranking più elevato. Inoltre, tale meccanismo ci consente di illustrare come anche la magia tecnologica del ranking di Google, ritenuto oggettivo e veritiero, sia legata ai “bassifondi” della rete niente affatto autorevoli e a pratiche di linking piuttosto equivoche.

Altre pratiche non illegali che sfruttano l’approccio all’indicizzazione di Google sono note come SEO (*Search Engine Optimization*); si tratta di un insieme di attività svolte per migliorare il posizionamento di un sito web nei risultati della ricerca. L’offerta di un posto di primo piano avviene spesso attraverso e-mail di spam provenienti da indirizzi improbabili, evidentemente tradotte con programmi automatici, che promettono strabilianti risultati:

“Noi registriamo il Suo sito internet in 910 motori di ricerca, registro e catalogo web. Noi portiamo il Suo sito internet sui primi posti di Google e Yahoo! Provateci! Non si corre nessun rischio. Al posto di 349€ soltanto 299€ (costo unico, senza abbonamento).” Ovviamente Google continua a rivendicare la propria trasparenza: “nessuno può garantire che il vostro sito compaia al primo posto nei risultati di Google”⁷⁷.

Dal punto di vista matematico, una conseguenza del PageRank basato sull’analisi dei link è l’integrità della base di dati; ovvero, la determinazione di uno spazio circoscritto, per quanto ampio, nel quale compiere ricerche. Infatti, se le pagine sono valutate e scoperte solo attraverso link ciò significa che non esistono pagine non linkate o isole di documenti slegati dal resto del web: in sostanza, nel mondo di Google esiste sempre un percorso che porta da una pagina a una qualsiasi altra presente nella base di dati, cioè nelle reti indicizzate.

Le ricerche quindi saranno tendenzialmente funzionali, evitando al massimo la possibilità di link rotti (*broken link*) o di informazioni diverse da quelle precedentemente archiviate, presenti nella memoria nascosta (*cache memory*). Il problema è che in questo modo gli utenti sono indotti a credere erroneamente che Internet sia un mondo chiuso, connesso, completo, privo di strade poco illuminate o di percorsi preferenziali, poiché sembrerebbe che, data un’interrogazione, si giunga sempre al risultato “giusto”.

Ciò dipende dal fatto che la visione googoliana di Internet scaturisce interamente dai percorsi che lo *spider* compie nel suo rimbalzare da un collegamento all’altro. Se una pagina non è citata da nessun altro sito, allora essa non comparirà mai in nessuna interrogazione compiuta da un utente, perché lo spider non ha mai avuto modo di trovarla, pesarla e valutarla. Tuttavia questo non significa affatto che siano assenti isole di dati, tutt’altro.

Ne sono un esempio i siti dinamici, nei quali le funzionalità offerte si basano totalmente sulle scelte dell’utente. Uno di questi siti è trenitalia.com: compilando l’apposita scheda (*form*), il sito è capace di fornire in tempo reale gli orari dei treni, le coincidenze, i percorsi più veloci per raggiungere una destinazione. Google non è in grado di comprendere le richieste di questo *form* e quindi non indicizza gli orari e i percorsi generati dinamicamente da trenitalia.com. Solo l’intervento umano può superare questo scoglio. L’unica soluzione proposta da Google è di inglobare nella sua interfaccia strumenti di reindirizzamento sui siti di compagnie aeree o ferroviarie nel momento in cui l’utente ricerca un percorso, destinazione e arrivo.

L’integrità referenziale proposta dalla base dati di Google deve essere rivista, perché sottintende l’idea di un mondo unico per tutti, chiuso e finito. Al contrario, tracciare un percorso in una rete complessa significa compiere un’esplorazione che determina sempre dei risultati relativi e parziali.

Il sogno di un Google contenitore di tutta Internet è un’idea demagogica particolarmente comoda,

⁷⁷ Testo di una email ricevuta su info@ippolita.net nel maggio 2005. La posizione di Google sui SEO: <http://www.google.it/intl/it/webmaster/seo.html>. Per approfondimenti tecnici, rimandiamo alla guida strategica al posizionamento su Google, <http://www.googlerank.com/>, di cui è disponibile anche una versione italiana.

utile per sostenere la completezza e l'affidabilità delle informazioni disponibili, insomma tutte le caratteristiche che rendono Google un “servizio unico”, un dispensatore di verità. Nelle ricerche della vita quotidiana tale chiusura assoluta è molto utile, perché conduce rapidamente a un risultato; in realtà però ci illude che la libertà consista nell'ottenere una qualità totale. Sappiamo invece che in un sistema reticolare complesso non esistono verità assolute, ma solo autorità distribuite a seconda del percorso che si desidera affrontare, o anche solamente in funzione del tempo che si è disposti a investire nella ricerca. La qualità dipende interamente dalla nostra soggettiva percezione dell'accettabilità del risultato. Le reti che siamo in grado di analizzare, apprezzare e vivere, sono oggetti complessi i cui nodi e collegamenti sono in costante mutamento. Poiché il compito di accettare un elaborato di navigazione relativo a una ricerca dipende in ultima analisi dall'utente, risulta essenziale l'esercizio della capacità critica, la consapevolezza della soggettività del proprio punto di vista. Per generare il percorso che davvero ci interessa analizzare è necessario ipotizzare l'esistenza di una rete finita e limitata, un mondo chiuso solo dalle nostre esigenze personali, sapendo tuttavia che si tratta di una localizzazione soggettiva, non assoluta né costante nel tempo. Esplorare una rete implica la capacità di dividere le reti in sottoreti di analisi e corrisponde alla creazione di piccoli mondi localizzati e temporanei⁷⁸.

In pratica nella navigazione quotidiana i collegamenti casuali sono di primaria importanza: la creazione di collegamenti nuovi e inaspettati non può in alcun modo essere prevista dall'analisi degli elementi della rete suggerita dal ranking Google. Questi collegamenti hanno la funzione di “porte dimensionali” e consentono la diminuzione o addirittura l'annullamento delle distanze fra due nodi della rete.

PageRank, o la moneta della scienza

Inoltre, l'algoritmo del PageRank, a differenza di quanto riporta la vulgata, non è una invenzione originale di Google, ma si fonda sulle scoperte matematico-statistiche di Andrej Andreevic Markov, che nei primi anni del XX secolo analizzò i fenomeni statistici su sistemi chiusi, cioè quei sistemi in cui ogni elemento è causa o effetto solo di altri elementi del sistema stesso⁷⁹.

Sergey Brin e Larry Page sono sicuramente partiti da questa base teorica, ma i miglioramenti che sono stati apportati non sono stati del tutto resi pubblici, al di là del brevetto depositato da Stanford.

L'esempio migliore per chiarire la morfologia di questo algoritmo è il passa parola fra amici e conoscenti. Nelle relazioni interpersonali più si parla di un dato evento, più questo assume importanza e contemporaneamente diventa parte di un patrimonio comune. Se si limita la diffusione di quel dato evento a una sfera ristretta la sua popolarità sarà minore. Lo stesso vale per gli uomini di spettacolo: più riescono a far parlare di sé maggiore sarà il loro ranking, più saranno conosciuti e più saranno famosi (è per questo che esistono trasmissioni autoreferenziali come “L'Isola dei Famosi”...). Questa stessa logica viene applicata da Google ai dati online.

Google propaganda questo suo metodo in maniera molto convincente, diffondendo l'immagine di Internet come una grande democrazia, poiché l'algoritmo agisce come se i link fossero voti relativi ai siti: poco importa se si linka per dire male o dire bene: l'importante è che se ne parli. La forzatura di questa “democrazia globale” ottenuta attraverso un algoritmo è evidente a chiunque: come se la democrazia dipendesse dalla tecnologia e non dalle pratiche degli individui.

L'origine culturale di questa pratica, come già accennato⁸⁰, è derivata dal sistema, estremamente elitario, della revisione dei pari (*peer-review*) da parte dei *referees* delle pubblicazioni scientifiche: in questo modo il contributo individuale di ogni ricercatore si inserisce in una rete di rapporti,

⁷⁸ Per approfondimenti sul concetto di mondi chiusi localizzati (LCW, *Localized Closed World*), si veda la dispensa sulle reti a cura di Andrea Marchesini: <http://www2.autistici.org/bakunin/doc/reti/index.xml>

⁷⁹ Per un'introduzione sulle catene di Markov si veda: http://en.wikipedia.org/wiki/Markov_chain.

⁸⁰ Si veda il cap. III.

verifiche e valutazioni che consente la trasmissione e il controllo dei risultati della ricerca. La democrazia globale di Google si configura insomma come l'applicazione del “metodo scientifico” delle pubblicazioni alla Rete, grazie all'algoritmo di PageRank, una sorta di “*referee tecnologico*” in grado di valutare in maniera oggettiva le informazioni del web, tenendo conto delle preferenze espresse dal “popolo dei navigatori” attraverso i link, e proporle nell'ordine giusto.

Il parallelo è stringente: da un lato, le pubblicazioni scientifiche acquistano peso e autorevolezza in base al loro collocamento nel quadro del loro specifico campo di ricerca; tale collocamento viene ottenuto tramite le citazioni, ovvero i riferimenti alla letteratura. In questo modo la ricerca scientifica garantisce la propria continuità, poiché ogni nuovo articolo non nasce nel vuoto, ma si pone come il “presente” del lungo percorso della tradizione scientifica. Dall'altro lato, i link delle pagine web vengono interpretati dallo spider di Google come “citazioni”, che aumentano appunto l'autorevolezza, cioè il ranking, di quella pagina.

L'elitarismo scientifico, base del sentimento di timorato rispetto che incute la “scienza” si basa curiosamente sulla pratica della pubblicazione: del resto, rendere “pubblico” non implica rendere “accessibile” e “comprensibile”⁸¹. Infatti “le scoperte degli scienziati, teoriche o sperimentali che siano, non sono e non possono essere considerate conoscenza scientifica finché non siano state registrate in modo permanente”, come sosteneva negli anni Settanta il sociologo Robert Merton⁸². L'affermazione è forse eccessivamente perentoria (la scienza antica si tramandava in modo tutt'altro che pubblico: si pensi alla scuola pitagorica in Grecia, alla distinzione fra scritti esoterici ed essoterici, ecc.), ma evidenzia correttamente il carattere eminentemente pubblico della conoscenza scientifica moderna. La comunicazione non è quindi un sottoprodotto della ricerca, bensì parte integrante di una forma di sapere i cui presupposti sono il carattere cumulativo e quello cooperativo. La scienza, almeno a partire dal XVI secolo, da una parte è orientata al conseguimento di risultati nuovi, che possano rappresentare un aumento del patrimonio conoscitivo, dall'altra assume come punto di partenza i frutti delle ricerche precedenti. Possiamo abbozzare una storia della comunicazione scientifica che si evolve insieme ai media destinati a supportarla: dalla fitta corrispondenza epistolare fra gli scienziati alla stampa periodica su riviste erudite, fino alla comunicazione digitale. Non a caso i primi nodi di Internet furono centri di ricerca accademica, che avevano la necessità di comunicare e condividere le proprie informazioni.

Tuttavia la mutazione del supporto non ha prodotto un sostanziale cambiamento nel metodo di connessione tipico di questa forma comunicativa, che rimane quello delle citazioni. Descritte come “moneta della scienza”, le citazioni sarebbero una sorta di tributo degli scienziati ai loro maestri e ispiratori. Più concretamente, collegano la ricerca presentata con quelle già svolte dallo stesso autore o da altri. Tuttavia è ragionevole assumere che il numero di citazioni ricevute da un determinato lavoro possa rappresentare un'indicazione della sua importanza o almeno del suo impatto sulla comunità scientifica. Negli anni questo sistema è diventato materia di studio specifica: l'analisi bibliometrica è una disciplina che utilizza tecniche matematiche e statistiche per analizzare i modelli di distribuzione dell'informazione, e in particolare delle pubblicazioni. Attualmente la bibliometria, e in particolare il suo più noto indicatore, l'*impact factor*⁸³, viene comunemente usata come criterio “oggettivo” per valutare la qualità del lavoro scientifico svolto da un singolo ricercatore o da un'istituzione. Un grande archivio per l'analisi bibliometrica è stato messo online

⁸¹ L'impressione che la scienza sia troppo difficile da capire per chiunque non sia uno specialista è socialmente radicata in tutti coloro che, a partire dalla loro vita quotidiana, se ne sentono alieni. Le mura del lavoro tecnico sembrano inviolabili. La comune espressione inglese “non è scienza per razzi” (*it's not rocket science*), di solito sarcastica osservazione fatta a qualcuno che ha insoliti problemi nello svolgimento di compiti facili, è solo un esempio della manifestazione di pubblica riverenza verso l'intensità intellettuale della scienza e la sua separazione dalle comuni attività di ogni giorno. Si veda a tal proposito l'attività di CAE, Critical Art Ensemble, www.critical-art.net

⁸² Robert K. Merton, *Scienza, tecnologia e società nell'Inghilterra del XVII secolo*, Franco Angeli, Milano, 1975.

⁸³ Eugene Garfield, *The Impact Factor*, in “Current Contents”, n. 37 (25) 1994, pp. 3-8; <http://www.isinet.com/isi/hot/essays/journalcitationreports/7.html>

nel 1993 proprio a Stanford, la culla di Google. Il progetto SPIRES (Stanford Public Information REtrieval System)⁸⁴ nacque nel 1974 dalla raccolta di note bibliografiche sugli articoli di fisica delle alte energie curata dalla biblioteca universitaria di Stanford. Limitatamente al ristretto campo d'analisi (la fisica delle alte energie), SPIRES è un database completo e ad accesso gratuito, che consente ricerche complesse anche sulle citazioni, una palestra che Brin e Page hanno saputo sfruttare al meglio per mettere a punto l'algoritmo di PageRank. Accanto all'algoritmo vi sono poi alcuni accorgimenti che contribuiscono a rendere Google un vero e proprio strumento di mediazione globale del web.

⁸⁴ Si veda l'articolo "L'evoluzione delle abitudini di citazione nella comunicazione scientifica primaria. Il caso della fisica delle alte energie", Marco Fabbrichesi, Barbara Montolli; http://jekyll.comm.sissa.it/notizie/lettere02_01.htm

V. In aggiunta altre maliziose funzionalità

I filtri sugli algoritmi: banche dati preconfezionate e controllo degli utenti

La teoria dei grafi⁸⁵ è la base matematica per l'elaborazione di tutti gli algoritmi sulle reti, tra cui figura anche il PageRank. Questo ramo della matematica si occupa tra l'altro di studiare i metodi per gestire, creare, percorrere diverse tipologie di reti descrivibili attraverso grafi e le loro rispettive grandezze. La teoria dei grafi ha avuto un particolare sviluppo a partire dalla metà del XX secolo con l'introduzione dei calcolatori elettronici. Possiamo immaginare geometricamente un grafo come un insieme di punti nello spazio e di curve continue che connettono coppie di punti senza intersecarsi tra loro. Nella teoria dei grafi, si dice *grafo* (da non confondere con grafico) una figura costituita da punti, detti *vertici* o *nodi*, e da linee che li uniscono, dette *lati* o *spigoli* o *archi*⁸⁶.

Una rete è un particolare tipo di grafo, nelle quali è possibile assegnare un valore differente a ogni arco (un “peso”), che potrà essere utilizzato per generare valori di percorrenza diversi. Internet è un grafo e lo stesso si può dire per l'insieme delle pagine Web; il sistema di ricerca di Google è organizzato a partire da questo assunto.

Uno degli elementi fondamentali degli algoritmi sulle reti è il fattore tempo in relazione al numero di nodi esaminati. Il tempo di completamento di un percorso (o di una ricerca) che collega un nodo della rete a un altro dipende dal numero degli elementi della rete e varia sempre tra un valore minimo e un valore massimo, che possono essere anche molto differenti tra loro, a seconda dell'algoritmo che si utilizza per generare il percorso.

Nella rete delle pagine Web ogni pagina è un nodo del grafo complessivo e ogni link è un lato/spigolo/arco. Se consideriamo la premessa sul fattore tempo, appare evidente che i risultati proposti da Google come risultato di una qualsiasi interrogazione (tecnicamente, una query su database) non si possono appoggiare su una consultazione di “tutto” il suo ipotetico Internet.

Lo spider di Google è perennemente impegnato nell'arduo compito di copiare Internet nella sua base di dati; tuttavia, non è credibile che a ogni interrogazione il motore di ricerca sfogli puntualmente tutto il suo database per trovare i risultati più congrui. Il fattore chiave che consente di ottenere risultati quasi immediati dipende da un'oculata serie di limitazioni nella selezione generale, cioè, concretamente, dall'applicazione di particolari filtri. Data un'interrogazione al motore di ricerca, un filtro garantisce la velocità del risultato finale per mezzo di una serie di scarti e scelte studiati appositamente per limitare l'intervallo (*range*) dei possibili blocchi di dati da analizzare.

In questo modo Google può fornire risultati alle interrogazioni in tempi eccezionalmente rapidi. Tuttavia, se la ricerca diventa rapida, contemporaneamente risulta poco trasparente, cioè non del tutto coerente con i dati presenti complessivamente sulla porzione di rete indicizzata. In altre parole, i risultati della ricerca saranno ottenuti velocemente non solo grazie alla potenza di calcolo disponibile, ma anche e soprattutto perché i filtri permettono di ridurre il bacino di informazioni dalle quali attingere.

Il compito di un filtro consiste nell'operare drastiche selezioni sui nodi della rete in modo da escluderli (o valorizzarli) unitamente ad altri eventuali collegamenti associati. L'applicazione di questi dispositivi ha l'obiettivo di escludere (o includere) interi blocchi tra i risultati complessivi.

Tutto questo è possibile mediante l'uso di banche dati di ricerche preconfezionate, per cui a interrogazioni standard vengono fornite risposte standard, ma anche tramite la profilazione

⁸⁵ La letteratura sui grafi è sterminata; per un primo approccio, si veda l'omonima categoria di wikipedia, http://it.wikipedia.org/wiki/Categoria:Teoria_dei_grafi

⁸⁶ Più formalmente, dati un insieme V di nodi e un insieme E di archi, un grafo G è un insieme $G = (V, E)$.

dell'utente basata sulle sue ricerche precedenti, in base alla sua lingua, alla sua posizione geografica ecc. In questo modo, un utente che compie abitualmente ricerche in italiano non interrogherà l'intera banca dati di Google, ma solo la porzione in italiano, con un evidente risparmio di tempo.

Su quantità di dati elevate, è impensabile impiegare algoritmi trasparenti, cioè che vadano a toccare tutti i nodi della rete; è necessario introdurre manipolazioni, semplificazioni o riduzioni delle possibilità di analisi. Questo per ragioni tecniche, di computabilità matematica in senso stretto, e anche, ovviamente, per motivi economici. Inoltre, senza peccare di ingiustificata malizia, in un sistema basato su approssimazioni dettate da filtri si può facilmente concepire l'inserimento di filtri ulteriori per aggiungere o spostare in posizioni visibili i risultati commerciali, a pagamento, oppure di semplice propaganda.

Tuttavia bisogna osservare che dal punto di vista di Google i filtri aggiuntivi non sono legati direttamente a un interesse economico, poiché non servono a vendere un prodotto, ma sono legati al fruitore, alle sue abitudini e ai suoi interessi personali. Google infatti vende pubblicità, non prodotti (o solo in minima parte, ad esempio i prodotti hardware come Google Mini e altri sistemi di indicizzazione per aziende): il suo interesse principale è dunque ottenere dati aggregati in base a parametri sui quali poter effettuare le campagne pubblicitarie con precisione. La personalizzazione dei risultati in base al destinatario è possibile sulla base di informazioni che Google stesso fornisce e che vengono raccolte nel modo meno invasivo possibile. Ad esempio, mailbox, blogs, hardisks virtuali, e altri servizi simili rappresentano altrettante banche dati utili alla profilazione degli utenti molto più di quanto si immagini.

I servizi “extra” messi a disposizione da Google, al di là del semplice motore di ricerca, sono perciò utili all'azienda per sperimentare nuove vie, ma anche e soprattutto perché svolgono un fondamentale ruolo di “accentratori di informazioni personali” degli utenti.

Un esempio eclatante sono state le mailbox di Gmail, una sorta di hard disk virtuale presso Google (al momento, oltre 2GB di spazio online, in costante crescita), offerte attraverso un sistema di propagazione basato sul PageRank stesso. In breve, ogni nodo della rete di Google (utente) ha a disposizione un peso (un certo numero di inviti) e può usarlo per offrire lo stesso servizio (link) ai suoi conoscenti. Grazie a questo metodo ogni soggetto coinvolto viene controllato dal punto di vista dell'uso del servizio; al contempo, mette a disposizione di Google informazioni essenziali sui propri conoscenti e amici.

In un secondo tempo, il meccanismo si propaga alla rete delle persone invitate, che possono effettuare nuovi inviti: in questo modo si va a creare un grafo delle relazioni umane tra gli utenti di enorme valore strategico per la personalizzazione delle inserzioni pubblicitarie.

Se consideriamo tutte le informazioni che si posso estrapolare dall'invio di email (a chi, per quale ragione, in che lingua, in che formati, con quali parole chiave, con quali allegati, ecc.), possiamo immaginare che esista nella banca dati di Google, oltre a una copia forzatamente parziale ma significativa di Internet, anche una copia parziale, ma altrettanto significativa, delle relazioni umane, emotive, professionali dei fruitori del servizio.

In teoria i filtri servono semplicemente per rendere più rapida l'interrogazione e più rispondente alle esigenze degli individui; sono anche assolutamente necessari da un punto di vista tecnico. Tuttavia, il loro utilizzo dimostra come sia estremamente semplice per un soggetto di fatto egemone nel campo dei servizi di ricerca sfruttarne le peculiarità per approfittare dei dati a sua disposizione in senso commerciale e lesivo della privacy.

In breve: attualmente, la base di dati di Google, attraverso alcune parole chiave, è in grado di ordinare i risultati di una ricerca (*query*) in maniera diversificata a seconda del tipo di utente che si trova a servire, ovvero in base ai dati relativi a quell'utente che sono in suo possesso. Lungi dall'essere risultati “oggettivi”, i risultati di ogni ricerca sono quindi calibrati; anzi, l'utilizzo dei servizi di ricerca affina le capacità del sistema di “riconoscere” l'utente e servirgli risultati adeguati.

L'utilizzo di ogni servizio è subordinato all'accettazione di regolamenti e liberatorie da parte degli utenti, e Google in ogni caso si impegna a non divulgare informazioni riservate e personali in genere; è facile però pensare che Google possa commercializzare o sfruttare per i più diversi scopi i dati dei propri utenti. Senza considerare per il momento l'opportunità (sarebbe meglio dire: la certezza) che militari e forze di polizia di vario tipo possano richiedere e avere accesso a simili informazioni per millantate ragioni di “sicurezza nazionale”. L'aggiunta di filtri di ricerca al fine di personalizzare i risultati è in concreto lo scenario più semplice da immaginare.

I cookies di Google, biscotti che lasciano il segno

La profilazione⁸⁷ degli utenti si basa sempre su un sistema di selezione e riconoscimento. Nella rete di Internet distinguiamo fra metodi di profilazione esplicita e implicita. La profilazione esplicita necessita di un'apposita procedura di registrazione, che implica l'invio tramite un modulo (*form*) di dati personali da parte dell'utente. Le informazioni inviate, archiviate nella base dati, vengono analizzate attraverso una serie di parametri utili a segmentare in gruppi omogenei la totalità degli utenti registrati (età, sesso, professione, interessi, ecc.). La profilazione implicita viene invece realizzata tramite il tracciamento di utenti anonimi nel corso delle loro visite a un sito, tramite indirizzo IP oppure tramite *cookies*, ovvero biscotti. I *cookies* sono piccoli file di testo utilizzati dai siti web per immagazzinare alcune informazioni nel computer dell'utente. Nel momento in cui il fruitore ritorna su quei siti, il browser riconsegna le informazioni salvate nel “biscotto”. L'obiettivo è automatizzare alcune operazioni di autenticazione (*login*), ricordare eventuali operazioni in corso, ma soprattutto associare il visitatore a tutte le informazioni memorizzate nelle sue precedenti visite.

La maggior parte dei siti Internet che offrono servizi online utilizzano i cookies e Google non fa assolutamente eccezione⁸⁸. Grazie alla combinazione di cookies e filtri aggiuntivi sugli algoritmi è possibile tenere traccia della navigazione di una persona e accumulare informazioni sulla sua “impronta”.

Facciamo un esempio: l'individuo X possiede un numero di telefono cellulare registrato a proprio nome, con il quale chiama la propria famiglia, alcuni amici e i colleghi di lavoro. Dopo un certo lasso di tempo decide di cancellare quel numero e di utilizzarne un altro non registrato a proprio nome, ritenendo in questo modo di proteggere la propria privacy. Con il nuovo telefono ristabilisce la sua cerchia di conoscenze, contattando la propria famiglia, alcuni amici e i colleghi di lavoro. Quella sequenza di “collegamenti sociali” (famiglia, amici, colleghi), risulta essere una sequenza unica tra tutte le telefonate del mondo, legata indissolubilmente al protagonista di questo esempio. Non è impossibile formalizzare una simile sequenza con un grafo che rappresenta i nodi e gli archi di una rete: i valori (i rispettivi “pesi” degli archi che collegano i nodi) potrebbero essere assegnati associando alla “vicinanza” un proporzionale grado di “amicizia” con il punto di partenza dell'analisi, cioè con il nostro ipotetico individuo X.

La rimozione dei *cookies* è quindi un'ottima pratica di difesa della privacy, ma un ragionamento analogo all'esempio precedente può facilmente essere modulato sull'utilizzo dei motori di ricerca. Tramite i *cookies*, cercando informazioni su alcuni specifici intervalli tematici, è infatti possibile identificare gruppi di persone, se non direttamente individui in base alla loro impronta univoca lasciata sulla rete.

⁸⁷ La profilazione è un settore chiave dell'economia della ricerca, in continua espansione, che comprende concetti come: *Clickstream analysis*, CRM (*Customer relationship management*), *Data mart*, *Data mining*, *Data warehouse*, OLAP (*Online Analytical Processing*). Per un'introduzione elementare all'argomento, si veda: http://www.diodati.org/scritti/2002/g_stat/stat05.asp. Interessante notare che le tecniche di profilazione dei consumatori derivano in ultima analisi dalle attività di *profiling* criminale.

⁸⁸ Per un approfondimento sui cookies volatili, si vedano in particolare: <http://www.google-watch.org/cgi-bin/cookie.htm> e <http://www.scroogle.org/>. Google Watch spiega la gestione dei cookies da parte di Google e metodi per difendersi dalla profilazione; Scroogle si prefigge di anonimizzare le ricerche verso google, facendo in modo che il cookie eterno di Google non possa consentire un tracciamento storico sulle ricerche effettuate dall'utente.

La stessa traccia univoca che scandisce i nostri movimenti, i nostri contatti sociali (o telefonici), è unica quanto le preferenze, i gusti, le idiosincrasie, le passioni che ci distinguono dagli altri. Le nostre passioni sono, in questo caso, i siti che visitiamo, e soprattutto, le ricerche che effettuiamo durante la nostra navigazione. Questa mole di informazioni che noi forniamo a un qualunque motore di ricerca rende possibile la ricostruzione della nostra “impronta”⁸⁹.

Come ogni biscotto che si rispetti, anche quelli Internet hanno una data di scadenza. I siti Internet che consegnano al nostro browser un *cookie* sono obbligati a impostare una data di scadenza, ovvero una data dopo la quale il browser può procedere all'eliminazione dei dati contenuti nel *cookie* stesso. Non è banale utilizzare al meglio i *cookies*; è certamente degno di nota il fatto che Google abbia saputo sfruttare a suo vantaggio un artificio tecnico noto agli sviluppatori del mondo POSIX (lo standard internazionale che permette l'interoperabilità dei sistemi operativi Unix e Unix-like, che comprendono anche GNU/Linux). Infatti, la data di scadenza di Google è il 2038, ovvero pressoché la data massima impostabile in un *cookies* e ciò comporta che il browser dei nostri sistemi operativi non rimuoverà mai quel *cookies* e le informazioni in esso salvate⁹⁰.

Onanismo tecnologico: crea, ricerca e consuma i tuoi contenuti

È impossibile seguire le rapidissime evoluzioni e innovazioni “made in Google”: i rilasci di nuovi servizi si susseguono in maniera quasi convulsa, e diventa molto difficile capire quali siano realmente destinati ad avere un impatto nelle nostre vite e quali invece saranno dimenticati nel giro di qualche mese o settimana. Del resto, è anche poco interessante e utile perdersi in elaborazioni complicate, in tassonomie esaustive che inevitabilmente tralasciano eccezioni significative, se consideriamo l'elevata depauperabilità delle informazioni e delle innovazioni su Internet. La natura fluida e dinamica della rete scoraggia, se mai qualcuno ne sentisse ancora il bisogno, qualsiasi approccio totalizzante e onnicomprensivo, che si rivela velleitario nel momento stesso della sua farraginoso esposizione.

È tuttavia possibile abbozzare, da una prospettiva soggettiva e situata, una visione complessiva del fenomeno Google, senza addentrarci nelle specifiche tecniche né tanto meno in pronostici futuribili di scarsa attendibilità. È senz'altro rilevante l'estremizzazione della figura del “*prosumer*”, attraverso l'accento posto sul concetto di personalizzazione.

Google è noto per l'abitudine consolidata a rilasciare versioni beta, non definitive e ancora in fase di test, dei suoi servizi: questa dinamica, come abbiamo visto nel capitolo precedente, è direttamente mutuata dalle comunità di sviluppo del Free Software. Gli utenti, fornendo feedback, suggerimenti e impressioni d'uso, contribuiscono in maniera determinante all'affermazione o meno di un nuovo servizio: sono al tempo stesso produttori e fruitori dei servizi che utilizzano, figure ibride dette appunto *prosumer*⁹¹.

Ponendosi come soggetto di mediazione globale dei contenuti del web, Google in sostanza vende tecnologie e risultati di ricerca (mediante la pubblicità) agli utenti, che da un lato tendono a essere creatori dei contenuti della rete, dall'altro i consumatori di quegli stessi contenuti attraverso i servizi di Google, che hanno modellato fortemente in prima persona.

Due esempi, apparentemente slegati fra loro, possono chiarire questo ciclo chiuso di produzione e consumo di contenuti: il Google Web Toolkit (GWT)⁹² e la convergenza fra GTalk e Nokia⁹³.

⁸⁹ Per un approfondimento sulla tematica, si veda l'intervento “La privacy e i motori di ricerca”, di Claudio Agosti “vecna” - s0ftpj.org in occasione del convegno “e-privacy - fra Trusted Computing e Data Retention”, tenuto a Firenze il 19-20 maggio 2006 <http://e-privacy.firenze.linux.it/>, disponibile all'indirizzo http://e-privacy.firenze.linux.it/atti/e-privacy_2006_Agosti_Privacy_e_motori_di_ricerca.pdf

⁹⁰ Si veda in proposito l'appendice I: La fine del mondo in un biscotto.

⁹¹ Productor-consumer, produttori-consumatori.

⁹² <http://code.google.com/webtoolkit/>

⁹³ <http://googlesystem.blogspot.com/2006/05/nokia-770-will-include-gtalk.html>

Google ha rilasciato nel maggio 2006 il Google Web Toolkit, un framework che permette di sviluppare applicazioni AJAX attraverso il linguaggio Java. AJAX (Asynchronous JavaScript and XML) è una tecnica per sviluppare applicazioni web interattive e dinamiche usando una combinazione del classico HTML (o XHTML) con CSS per la parte visiva e JavaScript per mostrare dinamicamente le informazioni e interagire con esse. Siti estremamente dinamici, insomma, nei quali non è necessario attendere ogni volta il caricamento di tutte le informazioni della pagina. Gmail, ad esempio, utilizza AJAX. La novità è significativa perché in questo modo cambia l'approccio alla creazione delle applicazioni web, che vengono scritte in un linguaggio a oggetti di alto livello, Java, poi compilate dal GWT e rese compatibili con tutti i browser. Tuttavia, questa innovazione non giustifica l'annuncio in pompa magna dell'avvento di un fantomatico “web 2.0” che rivoluzionerà Internet rendendola adatta alle macchine. Infatti, tecnologie per la creazione di software multiplatforma, per la condivisione dei propri bookmark, il *social networking*, l'aggregazione automatica dei dati, ecc. sono note da anni. Inoltre, l'ipocrisia di multinazionali come SUN che millantano l'ingresso nell'Era della Partecipazione (*Participation Age*), da una parte dimentica l'attitudine alla cooperazione in ambito informatico propria della cultura hacker da decenni, dall'altra spaccia per una grande rivoluzione l'uso degli strati più elementari (standardizzazione attraverso XML/RDF) delle innovazioni proposte da organismi come il W3C a proposito del web semantico⁹⁴.

Di certo AJAX e tecnologie affini risolvono alla radice il problema molto comune della portabilità dei siti web, che al momento sono difficilmente visualizzabili da tutti i browser. Il codice del framework è disponibile sotto licenza Apache, ovvero sostanzialmente Open Source, ma, come accade per le iniziative di code.google.com, alcune parti fondamentali (in questo caso, il compilatore Java-to-JavaScript e l'*hosted web browser*) vengono distribuite solo in binario e occorre sottoscrivere un'apposita licenza, che sostanzialmente vieta di ridistribuirli, di derivarne ulteriori prodotti e di includerli in prodotti commerciali. Inoltre, ogni volta che si usa l'*hosted web browser*, che permette di provare le applicazioni sulla propria macchina prima di pubblicarle su internet, viene contattato un server di Google, ufficialmente per verificare che si stia usando la versione più aggiornata del prodotto. È evidente però che si tratta di un metodo di controllo estremamente efficace, applicato agli sviluppatori invece che agli utenti. Certo, il codice prodotto è distribuibile liberamente, anche con scopi commerciali.

GWT è in pratica uno strumento semplice per creare siti perfettamente compatibili con i sistemi di indicizzazione di Google. Al momento, è necessario conoscere un linguaggio non banale come Java, ma non è avveniristico immaginare lo sviluppo di strumenti che permettano anche a un utente alle prime armi di posizionare sulla propria pagina web oggetti come barre degli strumenti, gallerie di immagini, menù di vario tipo e ogni sorta di oggetti web senza scrivere una riga di codice. Naturalmente esistono già programmi estremamente semplici per creare siti web (WYSIWYG, *What You See Is What You Get*); ma GWT è via web. Questi contenuti sarebbero quindi immediatamente pronti per essere fruiti su dispositivi fissi o mobili di qualsiasi tipo, purché siano in grado di accedere al web.

Immaginiamo ora che Google stringa accordi commerciali per la costruzione di hardware su misura, offrendo a chiunque usa i suoi servizi strumenti semplici per costruire pagine web visualizzabili su PC, palmari, ecc. e indicizzate nella maniera più semplice possibile dai suoi spider. Google infatti non fornisce programmi dietro pagamento di licenza, come Microsoft: ha bisogno, come sappiamo, di diffondere i propri standard per poter gestire più facilmente la sua economia di ricerca.

⁹⁴ A proposito dell'implementazione del web semantico, ben altro rispetto al “web 2.0”, si veda la *roadmap* del 1998 di Tim Berners-Lee, <http://www.w3.org/DesignIssues/Semantic.html>; il coordinamento del W3C, <http://www.w3.org/2001/sw/>; un documento introduttivo alla filosofia del web semantico: <http://www2.autistici.org/bakunin/doc/reti/index.xml>

Ed ecco GTalk e Nokia. GoogleTalk è il servizio VoIP⁹⁵ di Google, integrato da poco nella casella di posta Gmail, in modo che gli appartenenti alla “Google community” possano non solo mandarsi mail, ma anche chattare e parlare in tempo reale. A partire da fine maggio 2006 è disponibile sui nuovi dispositivi mobili di Nokia, chiamati “internet tablet”, sorta di cellulari evoluti pensati per navigare sul web. In questo modo Google entra dalla porta principale nel mondo della telefonia mobile, con la prospettiva di integrarsi ben presto con le reti senza fili pubbliche (*wireless* o *wi-max*) che cominciano a essere installate in diverse città, aeroporti, aree di servizio sulle autostrade, ecc. Una prospettiva di convergenza sul video è altrettanto plausibile: video.google.com è una miniera di video, e la televisione sui cellulari è il prossimo passo.

In altre parole, Google fornisce gli strumenti per creare contenuti secondo i suoi standard. Questa è la personalizzazione estrema, l'equivalente del meccanismo “code lunghe” (che permette di fornire esattamente il prodotto personalizzato per ogni consumatore) nel campo della creazione di contenuti web: l'utente crea “esattamente” quello che vuole nel formato standard di Google. La decentralizzazione totale a livello di creazione di contenuti è parallela alla decentralizzazione totale della pubblicità, e quindi della fornitura di prodotti “personalizzati”.

Un sistema pervasivo per imporre uno standard, ma formalmente “democratico” perché nelle mani degli utenti, a portata di click, a portata di browser. Quella che viene spacciata per democrazia elettronica si trasforma in una standardizzazione che consente di digerire i contenuti creati da una miriade di utenti e associare ad essi la pubblicità più adatta.

Browser come ambienti di sviluppo

L'esplosione di nuovi e sempre più potenti servizi web a partire dal 2005 sta trasformando il browser da strumento di semplice navigazione a strumento di sviluppo. Un ampio ventaglio di tecnologie stravolge gli standard di programmazione web correnti, offrendo agli sviluppatori uno strumento facile, multiplatforma, completo e affidabile: il browser, appunto.

Negli ultimi anni si è lentamente modellato un nuovo stile nella creazione di siti internet, anche grazie alle maggiori attenzioni verso la portabilità e l'accessibilità dei contenuti: ne sono un chiaro indizio la diffusione dei fogli di stile (*Cascading Style Sheet*, standard CSS e CSS2) invece dell'HTML nudo e crudo, dei validatori, dello stesso standard XML. Grafici e web designer trovano nei browser sempre più sofisticati e rispettosi dei vari standard ottimi alleati, riuscendo a realizzare siti conformi alle norme per la visualizzazione sui dispositivi più disparati e contemporaneamente a mantenere, e anzi ampliare, la propria libertà d'espressione.

In particolare, la nascita di browser come Mozilla Firefox e la loro rapida diffusione ha consentito una forte interazione fra gli sviluppatori di siti internet e gli sviluppatori di questo browser, giungendo in breve tempo a sanare quasi tutti i bug e le incompatibilità sugli standard web. L'incompatibilità fra Internet Explorer, Opera, e molti altri browser proprietari e non, è un problema noto a chiunque abbia creato pagine web. La sinergia sviluppata da Mozilla, apparentemente semplice e forse ovvia, costituisce una novità assoluta nella storia dei browser. Un'altra caratteristica interessante dei prodotti Mozilla è la struttura modulare sviluppata intorno al motore di rendering *Gecko*, che consente l'aggiunta di qualsiasi funzionalità immaginabile: tra i più noti e diffusi, strumenti per eliminare la pubblicità dalle pagine web, per avere sempre sott'occhio le previsioni del tempo o l'andamento della borsa.

I browser sono insomma strumenti sempre più affidabili, permettono la creazione di siti internet

⁹⁵ L'acronimo VoIP, Voice over IP (Voce tramite protocollo Internet), indica una tecnologia che rende possibile effettuare una conversazione telefonica sfruttando una connessione internet o un'altra rete dedicata che utilizza il protocollo IP, anziché passare attraverso la normale linea di trasmissione telefonica. Si veda http://it.wikipedia.org/wiki/Voice_over_IP. Non esistono solo Skype e Google che forniscono servizi VoIP, ma naturalmente anche programmi free, come Asterisk, <http://www.asterisk.org/>

complessi e assumono le caratteristiche di veri e propri programmi: tendono addirittura a sostituire gli applicativi di uso comune. L'esempio lampante sono i vari strumenti per ufficio proposti da Google come alternativa web alla suite di produttività Microsoft Office e anche a Openoffice⁹⁶. Al momento, è possibile utilizzare Writely (prodotto da una società acquisita) come elaboratore di testi, Google Spreadsheets come foglio elettronico, Google Page Creator per creare pagine web velocemente. Tutti i servizi sono in fase di *betatesting* a invito: è necessario avere un account Google per utilizzarli, ovviamente sotto stretto controllo.

Gli sviluppatori, da parte loro, spostano il loro interesse sempre più verso il web, anche grazie all'utilizzo di strumenti come GWT. Naturalmente Microsoft non sta a guardare e ha già pronta la versione beta (allineandosi alla strategia di "*betatesting* diffuso" di Google, derivata come sappiamo dalle pratiche del Free Software) di Office System (ovvero Office 2007), che prevede un'integrazione sempre più spinta con gli strumenti web, pur rimanendo un applicativo da installare.

I browser si stanno dunque trasformando in ambienti di sviluppo completi per la creazione di contenuti standard, ovvero SDK (*Standard Development Kit*). Dove sta realmente l'innovazione che ha reso possibile trasformare i browser in SDK? Sicuramente, possiamo parlare di un nuovo paradigma della programmazione: si possono ora creare programmi realmente multiplatforma, distribuiti, lato client, residenti sui server, facilmente aggiornabili, e senza bisogno di complessi *framework* da installare sui computer degli utenti. I contenuti, compresi i dati personali degli utenti, sono sempre più stoccati su macchine remote (ad esempio i server di Google), a cui si accede via web, cioè attraverso browser.

Spesso la scelta di browser Open Source come Mozilla è motivata dalla loro semplicità di configurazione, e dalla disponibilità di numerose estensioni potenti e gratuite. Gli sviluppatori si basano su questi browser per implementare programmi via via sempre più complessi e strutturati. La nascita di programmi basati soltanto sul web provoca due fenomeni entrambi con forti ricadute sul mercato e sugli utenti: i programmi binari da installare diventano obsoleti e i browser diventano i programmi più complessi, spesso fortemente modulari, e più appetibili sul mercato dell'IT. Quindi in futuro possiamo aspettarci meno programmi .exe (Windows), .dmg (Macintosh), o pacchetti Unix installabili, e più strumenti integrati nei browser, dalle estensioni per leggere i feed RSS al podcasting, da Gmail a software completi per ufficio.

Il controllo capillare che può essere effettuato da parte dei fornitori di servizi web rende questa dinamica potenzialmente pericolosa per tutti noi. Chi offre i servizi infatti conosce in maniera esatta l'identità digitale dei propri fruitori, il tempo di utilizzo del software e i contenuti in elaborazione poiché controlla ogni accesso e dettaglio relativo all'uso degli strumenti che mette a disposizione.

Si tratta a livello tecnico di un'interazione continua tra client (il browser) e il server (chi offre i servizi), dove quest'ultimo è in grado di monitorare costantemente le richieste, i tempi e gli elaborati finali. Muta anche lo scenario della sicurezza: con la scusa di fornire un prodotto "al sicuro" dalle grinfie di cracker e hacker di vario tipo, i meccanismi di autenticazione non si trovano più nel programma compilato e rilasciato senza sorgenti, perché avvengono direttamente sui server dei distributori. Quindi per "entrare" in un software non è più sufficiente disassemblare il codice sulla propria macchina per comprendere come funziona, ma diventa necessario "bucare" server remoti.

Privacy, Paranoie, Poteri

La strategia dell'accumulo ha reso Google un soggetto in grado di mettere in difficoltà addirittura

⁹⁶ Esistono anche suite complete per ufficio utilizzabili via web, come gOffice, prodotto dalla società Silveroffice, Inc. di San Francisco, che afferma: "gOFFICE non è affiliata a Google, Inc., ma ci piace pensare che sarà una bella lotta finché la loro missione rimarrà l'indicizzazione di tutte le informazioni del mondo, mentre la nostra è di aiutare la gente a creare quelle informazioni. Abbiamo perciò bisogno di Google per ricercare i contenuti che i nostri utenti creano con gOFFICE! Forse dovremmo aggiungere una scheda alla nostra barra degli strumenti chiamata "ricerca"..."
www.goffice.com

Microsoft, innescando una guerra senza esclusione di colpi per la standardizzazione e il controllo dell'accesso al web e alle reti che usiamo ogni giorno. In quanto soggetto di mediazione globale di informazioni, il fenomeno Google riguarda in maniera diretta tutti i fruitori di informazioni digitali, cioè tutti noi. Percorrere la storia di Google significa perciò ripercorrere la nostra storia di esploratori di internet e del web, che troppo spesso abbiamo delegato la gestione delle nostre informazioni, dei nostri siti, gallerie di immagini, caselle di posta, blog, sms, conversazioni telefoniche, ecc. ad aziende tutt'altro che disinteressate.

La strategia dell'oggettività, che Google attua ponendo l'accento sulla ricerca, sull'eccellenza accademica, sulla superiorità tecnologica, sull'accurato studio delle interfacce, è un velo che nasconde l'inquietante prospettiva di un unico punto d'accesso ai dati prodotti da utenti ingenui.

La strategia dell'Open Source, infine, è necessaria a Google per cooptare il metodo di sviluppo cooperativo proprio delle comunità digitali, rendendolo funzionale alla propria “mission”. Anche in questo caso abbiamo visto come Google si muova con estrema disinvoltura, proponendo metodi “nuovi” per sfruttare dinamiche note: *Summer of Code* ne è l'esempio lampante.

L'attività di Google, insomma, costituisce un pericolo evidente per chiunque abbia minimamente a cuore le tematiche della privacy e, in senso più ampio, della costruzione consapevole del proprio alter ego digitale. Si tratta dell'emersione di un conglomerato di potere che già oggi influenza pesantemente la vita di troppi individui. Google detiene informazioni riservate che analizza senza sosta per promuovere una personalizzazione sempre più accurata del cancro pubblicitario. E poiché l'accumulo di potere generalmente favorisce l'ansia di dominio, è opportuno indagare a fondo questo fenomeno.

Non esistono risposte globali in grado di risolvere una volta per tutte il problema della privacy. Il Grande Fratello non esiste, o quantomeno, come ogni paranoia che esclude orizzonti di liberazione e oscura possibili vie di fuga, è utile e funzionale all'esercizio del potere dominante.

Nascondere, crittografare, steganografare sono pratiche utili, ma non sono soluzioni definitive: l'obiettivo e il desiderio rimangono la comunicazione e la condivisione, che solo la “pubblicazione”, ovvero il movimento di “rendere pubblico”, consente. L'ossessione per la privacy scade molto rapidamente nella paranoia del complotto; in questo senso, non è utile indulgere nella costruzione di complicate alternative per creare reti assolutamente sicure e impenetrabili. La tecnologia è un'occasione di apertura e condivisione, poiché usare la macchine significa usare creature ibride, artefatti materiali (appartengono in questo senso al mondo naturale) investiti di valori e significati culturali (appartenenti cioè all'universo della “cultura”). Le reti sono espressione di una dinamica coevolutiva di macchine meccaniche, biologiche e significanti: la tecnologia è meticciasca per nascita. Creare reti significa infatti collegare macchine di vario tipo; significa creare metodi di condivisione, metodi di traduzione, metodi di scambio: non è possibile rimanere chiusi in sé stessi, è necessario mettersi in gioco e cambiare.

C'è bisogno di ricerca e di analisi competenti; la denuncia dei meccanismi del dominio tecnocratico urge più che mai. Abdicare al pensiero critico equivale invece a soccombere alle paranoie del controllo, che pure si fa sempre più pervasivo. È possibile utilizzare in senso paradigmatico la storia di Google per tracciare vie di fuga non ingenui, per immaginare pratiche di autogestione delle tecnologie. Google è infatti il punto d'incontro fra lo stile meritocratico dell'accademia, la spinta all'innovazione a ogni costo e il capitalismo finanziario più avanzato.

È l'occasione giusta per sviluppare reti autonome e decentrate, per contrapporre il desiderio di “esplorare” e “percorrere” le reti alla necessità di “accedere” ai dati, per focalizzare l'attenzione sul percorso invece che sul risultato.

VI. Qualità quantità relazione

L'emergenza dell'informazione

La società dell'informazione presenta un'estrema eterogeneità di manifestazioni: utilizza infatti al contempo e ormai anche con estrema disinvoltura sistemi di comunicazione reticolari (telefonia), versioni digitali di media *broadcast*⁹⁷ tradizionali nati fuori dalla Rete (quotidiani, radio, tv online, ecc.), strumenti di comunicazione universali nati con la Rete (email) e sistemi distribuiti (p2p). Ma tutto questi sistemi a ben vedere si basano su un'unica materia prima: l'informazione. All'interno dello specifico ambito dei motori di ricerca, e quindi del reperimento di informazioni (*information retrieval*), possiamo assumere che l'Informazione sia l'insieme delle pagine che vengono pubblicate sul Web⁹⁸.

La crescita quantitativa, e qualitativa, di queste pagine e dei relativi contenuti è stata smisurata e continua a essere tale. Ciò dipende dalla facilità con cui è possibile attualmente generare nuovi contenuti per la rete. I contenuti non sono isole di dati sconnesse, ma si costituiscono nella varietà delle relazioni e dei link che legano tra di loro le pagine, i siti, gli argomenti, i documenti, i contenuti stessi.

L'accesso diretto, senza mediazione, a questa mole di informazioni è assolutamente impossibile, anche solo in via ipotetica: sarebbe come sostenere di poter sfogliare il Web "a mano". Per questo esistono i motori di ricerca, strumenti studiati per filtrare la complessità della rete e fungere da interfaccia fra noi e l'informazione, risolvendo delle ricerche.

Nei capitoli precedenti abbiamo passato in rassegna gli strumenti fondamentali per il funzionamento di un motore di ricerca, ovvero come Google e gli altri riescano a divorare le pagine Web, come queste vengano analizzate e catalogate in base ad algoritmi di ranking, infine archiviate su appositi supporti hardware e mostrate all'utente in base alle sue ricerche.

La quantità di pagine memorizzate assume quindi un valore centrale per valutare la potenza tecnica ed economica di un motore di ricerca. Maggiore sarà l'ampiezza del patrimonio delle pagine esaminate, maggiore sarà la potenziale completezza e affidabilità delle ricerche compiute, ovviamente nel limite di quel contesto specifico.

Tuttavia, se l'estensione del patrimonio di un motore di ricerca può essere enorme, non potrà mai essere completa e totale, indipendentemente da quanto tempo, denaro e tecnologie si investano. È assurdo pensare di poter conoscere, o più banalmente copiare e catalogare, tutto Internet: sarebbe come pretendere di conoscere l'interezza di un mondo vivo e in costante mutamento.

I dispositivi di stoccaggio delle informazioni dei motori di ricerca come Google sono come dei vasi. Immaginiamo di dover riempire un vaso enormemente capiente con delle minuscole gocce ovvero le pagine Web che costituiscono le informazioni della Rete. Se riteniamo che il vaso sia capace di contenere la totalità delle gocce (la totalità della rete) il nostro compito consisterà nel

⁹⁷ Un sistema trasmittente *broadcast* invia informazioni a molti sistemi riceventi in maniera unidirezionale (comunicazione uno-molti). In una rete di computer, una trasmissione *broadcast* invia informazioni a tutti i computer connessi a quella rete; altri sistemi di trasmissione sono *unicast* (un solo ricevente), *multicast* (molti riceventi), *anycast* (qualsiasi ricevente all'interno di un gruppo). I sistemi p2p, proprio come i sistemi telefonici, sono l'esatto opposto dei sistemi *broadcast*, perché determinano una comunicazione di tipo multi-molti.

⁹⁸ Non è questo il luogo per una disamina esaustiva del concetto di "informazione", ma certo questa definizione è volutamente semplificata. Non tiene conto ad esempio del fatto che motori di ricerca come Google indicizzano anche altri contenuti che non sono Web, come l'archivio dei messaggi Usenet oppure il contenuto dei computer di chi utilizza Google Desktop; tuttavia, poiché questa massa di informazioni viene presentata attraverso un'interfaccia Web di mediazione, riteniamo accettabile la definizione adottata.

cercare di catturarle tutte in modo sistematico e ripetitivo.

Ma se invece pensiamo che le gocce siano troppe rispetto al nostro vaso e che la loro raccolta non sarà prevedibile secondo algoritmi, né rapida, ma potenzialmente potrebbe addirittura rivelarsi un compito illimitato, dovremo cambiare tattica: anche perché le informazioni cambiano nel tempo, le pagine vengono modificate, le risorse cambiano indirizzo....

Potremmo magari decidere di scegliere solo le gocce più grandi, o concentrare i nostri sforzi di raccolta nei punti in cui ne cadono di più, o ancora, semplicemente, potremmo raccogliere solo le gocce che ci interessano maggiormente, e poi cercare di collegarle nella maniera che riteniamo più interessante.

Nonostante i motori di ricerca continuino a porsi l'obiettivo di catalogare ogni cosa, è più corretto applicare al Web un approccio localizzato, ovvero assumere che per ogni "intenzione di ricerca" esistano molte risposte plausibili, fra le quali alcune sono "migliori" perché soddisfano determinati requisiti di rapidità e completezza. Dobbiamo sempre ricordare che la qualità delle risposte dipende dalla nostra soggettiva percezione dell'accettabilità del risultato. Per poter accettare o rifiutare un elaborato di ricerca risulta essenziale l'esercizio della capacità critica, la consapevolezza della soggettività del proprio punto di vista. Per generare il percorso che davvero ci interessa analizzare è necessario ipotizzare l'esistenza di una rete finita e limitata, un mondo chiuso solo dalle nostre esigenze personali, sapendo tuttavia che si tratta di una localizzazione soggettiva, non assoluta né costante nel tempo. Dal punto di vista analitico, esplorare una rete implica la capacità di dividere le reti in sottoreti di analisi e corrisponde alla creazione di piccoli mondi localizzati e temporanei (*LCW, Localized Closed World*), nei quali esiste almeno una risposta alle ricerche eseguite. In caso contrario, molte ricerche sarebbero senza fine, anche perché i dati che possono essere analizzati sono molto superiori alle capacità di elaborazione di un soggetto umano. Tale risposta può non essere soddisfacente: cambiare o specificare la ricerca, arricchire il proprio punto di vista potrà allora generare un percorso più consono alle esigenze. Considerare il Web un mondo chiuso localizzato significa anche riconoscere che la dinamica stessa con cui le informazioni nascono, germinano e si distribuiscono in Rete (e magari vengono abbandonate a sé stesse, non più aggiornate) è un fenomeno "emergente", non causale né casuale. L'emergenza⁹⁹ è un comportamento descrivibile in termini matematici, un'esplosione di complessità inaspettata e imprevedibile, ma soprattutto è un evento in grado di determinare situazioni non descrivibili in maniera esaustiva. Analizzare e percorrere un "universo emergente" come il Web implica dunque un continuo riposizionamento personale che determina un "mondo chiuso localizzato" di competenze e aspettative, ma anche l'apertura a nuove possibilità di esplorazione (altri mondi sono sempre possibili, fuori dal proprio mondo chiuso) e dunque la consapevolezza di poter raggiungere risultati sempre e solo parziali.

Quantità e qualità

L'indicizzazione per accumulo di pagine è un fenomeno quantitativo, ma in sé non determina la qualità delle informazioni del Web: l'obiettivo è raccogliere tutte le pagine e non scegliere fra di esse. Le relazioni fra le pagine, pur essendo generate in base a un criterio semplice, il sistema dei link, generano fenomeni di emergenza. Quindi la qualità delle informazioni, a prescindere dalla loro tipologia, è determinata dalla capacità di tracciare percorsi, senza preoccuparsi di coprire la "totalità" delle informazioni. La qualità dipende soprattutto dall'esplicitazione del proprio punto di vista attraverso un percorso di ricerca: insomma, sono i naviganti, i pirati, gli utenti del Web che

⁹⁹ L'emergenza è un principio che descrive il comportamento dei sistemi complessi. Può essere definita come il processo di formazione ed emersione di comportamenti e schemi complessi a partire da regole più semplici. Le scienze cognitive utilizzano il concetto di emergenza per descrivere "l'esplosione" della complessità tipica dell'intelligenza: a partire da elementi semplici di per sé incoscienti (i neuroni), si generano il pensiero, la coscienza manifesta dei processi cerebrali. <http://it.wikipedia.org/wiki/Emergenza>

determinano e migliorano la qualità delle informazioni, tessendo relazioni fra le pagine. La potenza di accumulo degli algoritmi di Google è utile, ma insufficiente.

La valutazione del contenuto delle pagine e le loro relazioni è delegata ad algoritmi e, soprattutto, alle società che li controllano. Il fenomeno Google è il risultato di questa delega: una potenza apparentemente senza limiti, in grado di offrirci la possibilità di trovare “qualcosa” di utile e interessante tra le molte risorse contenute nel suo patrimonio, che viene spacciato per la “totalità” del Web. Vengono però completamente taciuti i limiti di queste luccicanti offerte: ciò che è assente dal patrimonio, o ciò che è presente solo in parte, e, soprattutto, tutto quello che è stato scartato.

Rimane di stringente attualità uno spinoso problema etico e politico sulla gestione e il controllo delle informazioni: quale soggetto può garantire la correttezza di un’azienda che, per quanto “buona”, ha come obiettivo primario il profitto?

Se da una parte ingenti risorse economiche e un’ottima infrastruttura tecnica possono migliorare continuamente lo stoccaggio e il recupero dei dati, le questioni politiche derivanti dall’accumulo di informazioni non sono in alcun modo superabili, perché Google rappresenta nei fatti un’inedita concentrazione di informazioni, riservate o meno, in ogni caso fonte di un potere enorme senza alcuna trasparenza. Ovviamente nessuna legge sulla privacy potrà migliorare questa situazione, né tantomeno la creazione di istituzioni nazionali o internazionali *ad hoc* per il controllo delle informazioni personali e dei dati sensibili. Una risposta sensata alla questione della riservatezza delle informazioni può venire solo da una maggiore responsabilità e consapevolezza da parte degli individui che creano la Rete, attraverso scelte di formazione individuale, per quanto sia una strada difficile e faticosa da realizzare.

Il mito della ricerca istantanea

Chiarito che l’enorme patrimonio di Google non potrà mai corrispondere alla totalità della Rete, è ingenuo, o furbo, proporsi come interfaccia “istantanea” fra le istanze di ricerca degli utenti e il risultato cosiddetto “esatto”.

Ogni volta che si effettua una navigazione visitando pagine Web si seguono dei link che creano percorsi analizzabili secondo i modelli matematici della teoria dei grafi. Il Web è composto infatti da nodi (le pagine e siti Internet) e da archi (i link che li uniscono).

Per quanto possano crescere le dimensioni del Web, le direttrici proposte dai motori di ricerca saranno sempre efficaci nel portarci all’oggetto “giusto”. Infatti, attraverso l’applicazione di criteri di efficacia/efficienza, il percorso ricavato dall’interrogazione di un motore di ricerca sarà ottimizzato, cioè il numero di nodi visitati sarà molto basso e il tempo per raggiungere il risultato sarà quasi istantaneo. Google si spinge ad auspicare un solo passaggio, tanto da offrire il bottone “mi sento fortunato” nella pagina principale.

Questa ottimizzazione limita la ricerca a uno schema sequenziale articolato in tre punti: *utente-algoritmo-obiettivo*. Alla lunga questa dinamica conduce a un atteggiamento di passività digitale, per cui “aspettiamo” che ci vengano forniti dei risultati per poi “scegliere” fra di essi.

Inoltre, questa efficacia/efficienza si fonda paradossalmente non sull’ampliamento del bacino di dati sui quali effettuare ricerche, bensì sulla riduzione dell’accesso al patrimonio informativo stesso, poiché ogni percorso proposto dal motore di ricerca non sarà mai tracciato sulla rete “in quel momento”, ma calcolato in base al suo patrimonio archiviato, e personalizzato sull’utente in base a filtri algoritmici e *cookies*.

L’accesso alle informazioni proposto da Google è rapido, anzi, nelle sue intenzioni tende a essere immediato, ad annullare la dimensione del tempo, a prescindere dall’enorme estensione delle basi di dati consultate. La mediazione della tecnica (interfacce, algoritmi, ricerche preconfezionate, ecc.)

consente questo “annullamento” temporale e rende l’accesso praticamente istantaneo¹⁰⁰. Ma la qualità della ricerca viene negativamente influenzata dalla rapidità nel raggiungimento del risultato. Infatti, chiunque abbia provato a compiere una ricerca, sa bene che il tempo impiegato è un fattore determinante nella composizione dell’esperienza: creare un proprio percorso, compiere delle scelte differenti in momenti diversi genera senso di partecipazione e soddisfazione. Google ci permette di “localizzare” nello spazio (nel suo spazio multidimensionale) ciò che vogliamo; ma, per quanto breve possa essere il tempo di attesa del risultato, stiamo pur sempre assumendo un atteggiamento passivo nei confronti di un oracolo tecnologico.

Nel quadro dell’esplorazione attiva, l’obiettivo non è invece “accedere” ai dati, ma compiere una navigazione ricca e diversificata, costruire percorsi complessi attraverso la ricerca. Il concetto di efficienza decade. Maggiore sarà il numero dei nodi che visitiamo e la complessità dell’intreccio che costruiamo, maggiore sarà l’opportunità di ampliare i nostri campi di interesse, le possibilità di operare scelte significative, di raffinare le nostre ricerche. Questo approccio consente un arricchimento cognitivo al di là della prestazione immediata. Ad esempio, quando visitiamo i link proposti da un sito di nostro interesse, e proseguiamo nei siti segnalati come amici, creiamo un percorso unico; magari prendendo spunto da un nostro segnalibro. A livello visivo un procedimento simile si oppone alla linearità della sequenza *utente-algoritmo-obiettivo* e crea invece un percorso ricco di biforcazioni, ramificazioni, salti, giri tortuosi, rispondendo a un desiderio cognitivo non lineare¹⁰¹.

In definitiva, i motori di ricerca sono strumenti eccellenti per quanto riguarda la soddisfazione degli aspetti quantitativi della ricerca, quando ci si muove all’interno di risorse già completamente strutturate, come vocabolari, enciclopedie, ecc. La quantità è direttamente proporzionale alla potenza di calcolo e di accumulo: l’estensione di Google è decisamente superiore a quella di qualunque altro concorrente, ma per rimanere tale necessita di una costante espansione, sia essa intesa in risorse, algoritmi, macchine, utenti, ecc.

La macchina dell’accumulo è alimentata dalla ricerca dell’efficienza e dell’eccellenza a livello di prestazioni, a prescindere da una valutazione dei contesti di ricerca: l’interesse principale di un motore di ricerca è l’incremento del proprio patrimonio, e di conseguenza della propria offerta. Una logica esclusivamente quantitativa si limita all’idea che maggiori sono le informazioni disponibili, maggiore sia la possibilità di accedere ad informazioni coerenti con l’intenzione di ricerca; tesi che di nuovo conferma l’approccio lineare sotteso al business dei motori di ricerca.

La qualità invece non risiede necessariamente nella potenza tecnico-economica. Non è scontato che a vastità di bacini informativi disponibili corrisponda la pertinenza dei risultati forniti: l’emergenza del percorso migliore non si prevede, non si calcola, ma si costruisce un passo dopo l’altro.

Dietro il velo del mito

I valori di posizionamento del ranking di Google non rispondono ad alcun criterio trasparente di valutazione; tuttavia, le risposte fornite nella stragrande maggioranza dei casi sono esaustive, o meglio, per noi è impossibile valutare se qualcosa è sfuggito al suo spider, a meno che non siamo esperti di un dato settore e conosciamo una risorsa che non risulta indicizzata.

La diffusione capillare degli strumenti di ricerca di Google lo rendono uno standard *de facto*. Il box bianco (*blank box*) dove inseriamo le parole chiave (*keywords*) delle nostre ricerche è per

¹⁰⁰ La soddisfazione degli utenti per i risultati ottenuti con questo metodo “immediato” è tutt’altro affare: si veda in proposito Jean Véronis, A Comparative study of six search engines, Université de Provence, <http://www.up.univ-mrs.fr/veronis> <http://aixtal.blogspot.com>

¹⁰¹ Il testo base per avvicinarsi alle dinamiche cognitive non-lineari rimane Humberto R. Maturana, Francisco J. Varela, Autopoiesi e Cognizione - La realizzazione del vivente, Marsilio, Venezia, 1985 (ed. or.: Autopoiesis and Cognition. The realization of the Living).

l'utente una sorta di *weltanschauung*, o *vision*, insomma propaganda una ben precisa “visione del mondo”, poiché incarna l'idea di “servizio totale” (*total service*): il motore di ricerca è capace di rispondere a qualsiasi domanda, di soddisfare tutte le esigenze di fronte allo strumento Internet.

In termini epistemologici, il *blank box* rappresenta un modello cognitivo di organizzazione della conoscenza: chiediamo infatti allo spazio bianco la risposta per tutte le nostre intenzioni di ricerca, sia quando ci servono documenti, approfondimenti, informazioni, sia quando vogliamo semplicemente “navigare”. Ormai lo utilizziamo anche quando potremmo consultare le pagine gialle, ricordarci la nota che abbiamo segnato su un libro, chiamare qualcuno, pensarci due minuti, ecc... ma invece Google è più comodo. L'attività di ricerca viene interamente identificata con l'oggetto che fornisce il servizio, Google, del quale abbiamo una percezione sovrabbondante.

L'abitudine all'utilizzo degli strumenti si trasforma in comportamento, in coazione a ripetere: è molto difficile che intenzionalmente gli utenti cerchino di soddisfare la propria “necessità di input” con altre modalità, essendo ormai affezionati e rassicurati dall'efficacia/efficienza del *blank box*.

Essere in movimento sulla rete e avere quindi bisogno di interfacce di accesso, di strumenti per il reperimento delle informazioni e la costruzione di percorsi è un'attività differenziata e profondamente contestuale. L'attività di ricerca non è affatto omogenea e non può dunque essere ridotta all'utilizzo della *blank box*. Ciò che chiediamo e ciò che desideriamo dipende non solo da una necessità esprimibile in termini analitici di informazione quantitativa, ma anche dal modo in cui affrontiamo la ricerca, dal contesto nel quale cerchiamo l'informazione, dal nostro bagaglio culturale e dall'attitudine al confronto con la novità, con nuovi territori, con la diversità. È impossibile esaurire le richieste informazionali attraverso una soluzione univoca.

Poiché l'indicizzazione delle pagine è parziale, ovvero una selezione basata sul ranking, quello che Google ci propone è banalmente la possibilità di trovare “qualcosa” per noi utile e interessante tra le molte cose che sono raccolte nel suo patrimonio di interesse. Un'intenzione di ricerca sottintende però in ultima istanza il desiderio di trovare, anzi scoprire, “tutto ciò che non si sa ma che è possibile sapere”. Il gigante ci appare allora così com'è, enorme, esteso, ramificato, ma non necessariamente adatto per le nostre ricerche.

Modelli di ricerca

L'ambiguità suggerita dai motori di ricerca secondo cui “ricerchiamo in un ambiente infinito”, invece che in un mondo chiuso localizzato in base alle nostre intenzioni di ricerca, dipende dalla sovrapposizione formale tra il livello dell'interfaccia¹⁰² e quello dell'organizzazione. L'interfaccia, in questo specifico contesto, è l'elemento tecnologico attraverso cui si accede alle informazioni e si esegue l'attività di ricerca; l'organizzazione è invece l'architettura, il modello tecnologico con cui le informazioni sono ordinate e archiviate. I due livelli si influenzano a vicenda, infatti le scelte in termini di organizzazione determinano l'utilizzo di specifiche interfacce, e a loro volta le informazioni visualizzate per mezzo delle interfacce rispecchiano la forma con cui vengono conservate.

Il problema di questa sovrapposizione è che le informazioni vengono presentate sotto forma di dati identificabili in maniera univoca, separati tra di loro. L'utente di Google si muove in maniera lineare nella lista dei risultati del ranking; per passare da un risultato all'altro deve tornare all'elenco originario, senza collegamenti trasversali al livello dell'interfaccia.

Infatti con i motori di ricerca si rintracciano informazioni senza che il percorso realizzato venga tenuto in considerazione. L'interfaccia che determina la nostra interazione è il campo delle *keywords*, dove si inseriscono le parole o le formule per cominciare la ricerca; a questo primo livello di accesso tutte le informazioni si trovano sul medesimo piano e omogenee, ma nello stesso

¹⁰² L'interfaccia è il punto, l'area, il territorio che mette in collegamento elementi qualitativamente differenti; nel caso dei motori di ricerca, l'interfaccia permette la relazione fra il soggetto che interroga e il bacino di informazioni.

tempo separate e frammentate, rendendo possibile un elenco di risultati ordinati da un algoritmo in base alla loro pertinenza.

Tuttavia nelle ricerche che si compiono quotidianamente i risultati stessi possono essere collegati fra loro in diversi modi, senza necessariamente ritornare all'elenco ordinato dei risultati, e non vi è un solo risultato corretto; anzi, una ricerca che non sia indirizzata a dati strutturati in enciclopedie, dizionari e simili (che pure mutano nel tempo), potrebbe non avere alcuna soluzione, ma necessitare invece di sforzi creativi, di miscela e ricombinazione.

Quando si verifica identità formale tra il livello dell'interfaccia e quello dell'organizzazione si ottiene un modello riduttivo: nel caso di Google, la rappresentazione di una potenza di ricerca infinita, ovvero l'attività di ricerca che si confonde con uno degli strumenti per realizzarla.

Ad esempio: se cerchiamo “Penna” attraverso l'interfaccia di Google.com troviamo, fra oltre sei milioni settecentomila risultati: “Sandro Penna”, il poeta; “Penna d'autore”, il concorso letterario; il “Comune di Penna San Giovanni”, provincia di Macerata e così via. Non troviamo nei primi risultati (*top ranking*) né risorse relative alle “penne degli uccelli”, né alle “penne per scrivere”, due concetti molto lontani fra loro, di cui volevamo approfondire eventuali legami. Una prospettiva più ampia sul reperimento delle informazioni, che tenga conto criticamente del potenziale cognitivo sotteso a ogni patrimonio informativo, spinge a considerare la funzione accesso-ricerca non come un processo di localizzazione, ma di esplorazione e creazione. L'accento si sposta dunque dall'epistemologia all'ontologia: non solo conoscere delle informazioni, ma essere consapevoli del proprio ruolo di creatori di informazioni¹⁰³. I motori di ricerca agiscono sull'accesso ma non sono di alcun supporto per l'esplorazione, poiché si limitano a intervenire sul primo livello di presentazione delle informazioni.

La navigazione è il momento di reale dinamismo nella messa in relazione fra gli oggetti digitali, che possono così esprimere il massimo delle loro potenzialità comunicative ed euristiche. Si impara facendo, e si cambia mentre si impara, mentre si è presi nell'esplorazione.

Esiste una profonda distinzione tra cercare e trovare: Google ci fa trovare le cose, provoca soddisfazione, sensazione di accumulo. Ma potrebbe anche non essere interessante il “trovare”, quanto piuttosto invece l'atto stesso di “cercare”; forse è interessante anzi non trovare affatto, perché se non si riesce a trovare significa che si è impegnati a cercare¹⁰⁴.

Il motore di ricerca è un modello strumentale attraverso il quale le informazioni si dispongono: è interessante invece immaginare modelli per combinare le informazioni, e creare conoscenza.

¹⁰³ Esplorare le reti significa creare le reti: in questo senso, estendendo l'approccio di Maturana e Varela, possiamo parlare di “autopoiesi delle reti”, reti che si costruiscono attraverso il nostro fare.

¹⁰⁴ :-)

VII. Tecnocrazia

L'analisi del fenomeno Google traccia un panorama variegato, nel quale l'economia della ricerca risulta essere solo una tessera di un mosaico ben più vasto e complesso. Del resto, stando alle dichiarazioni di Eric Schmidt, a Mountain View stanno gettando le basi per la costruzione di un'azienda globale delle tecnologie dell'informazione, "un'azienda da 100 miliardi di dollari": di certo, non un semplice motore di ricerca. Si tratta infatti di un sistema pervasivo di gestione delle conoscenze, di cui abbiamo illustrato le fasi e i metodi più significativi: strategie che coniugano marketing aggressivo e oculata gestione della propria immagine, propagazione di interfacce altamente configurabili eppure sempre riconoscibili, creazione di contenuti standard decentrata presso utenti e sviluppatori, cooptazione di metodologie di sviluppo cooperativo tipiche dell'Open Source e del Free Software, utilizzo di sistemi avanzati di raccolta e stoccaggio dati, sistemi di reperimento delle informazioni correlati a tecniche di profilazione esplicite e implicite, personalizzazione delle pubblicità.

Google è il dominio tecnologico derivato dalla ricerca scientifica che si fa strumento di gestione della conoscenza, espressione diretta della tecnocrazia.

Tecnocrazia: gli esperti della scienza

Gli esperti hanno trovato nel controllo e nella manipolazione della tecnologia un'arma senza rivali per imporre i propri interessi personali, per mantenere il potere o acquisire maggiori privilegi. Il meccanismo è semplice: la tecnologia viene utilizzata come narrazione garante dell'oggettività della ricerca scientifica, ma anche per avallare le decisioni dei politicanti che governano, e in generale di qualsiasi "autorità" che abbia accesso all'oracolo tecnologico.

L'applicazione della ricerca scientifica sotto forma di tecnologia è pervasiva e la realtà viene costantemente interpretata in ossequio a questo paradigma. La curiosità e il desiderio di sapere che muovono la ricerca scientifica vengono castrati dai criteri di bieca redditività che regolano i finanziamenti pubblici e privati. Se la ricerca non produce artefatti tecnologici in grado di generare introiti immediati, non è interessante. Il discorso del potere si fa allora tecnocrazia, l'esatto opposto della condivisione comunitaria, dell'autogestione, della discussione e della mediazione fra individui. Spacciare la tecnocrazia di Google come strumento di democrazia diretta è un gioco linguistico, che ci fa sembrare partecipi di una grande democrazia elettronica, priva di ogni sostanza. Certo, possiamo pubblicare ciò che vogliamo su Internet, e Google ci indicizzerà. Ma non possiamo permetterci da "profani" ed "eretici" di osservare che la strategia dell'accumulo di Google è in perfetta sintonia con il folle sistema dell'economia di mercato basato sulla crescita infinita: infatti, non siamo economisti laureati alla London School, né imprenditori di successo, ovvero non siamo esperti, e non abbiamo dunque alcuna autorevolezza in merito. Eppure il comune buon senso di orwelliana memoria è più che sufficiente per capire che tale crescita senza fine e senza scopo è la manifestazione di una volontà di potenza tecnologica interessata agli individui solo in quanto potenziali consumatori.

Ecco perché il PageRank, che come abbiamo visto non è solo un algoritmo, diventa una veste culturale con la quale Google ci propone di analizzare ogni cosa: in sintesi, si assiste a una estensione forzata del metodo di revisione dei pari, valido in ambito scientifico, all'interezza della conoscenza umana.

Le autorità tradizionali, religiose e politiche, hanno raggiunto il picco minimo di credibilità; ma la loro scarsa presa sulla realtà, invece che favorire la proliferazione di spazi di autonomia, ha condotto a una situazione surreale in cui nessuna asserzione valida può essere compiuta senza il

sostegno di un'autorità tecnologica. L'autorità delle macchine, nella maggior parte dei casi, è una risposta estratta da una base di dati, servita dai sacerdoti della tecnologia, gli “esperti”, al ricco e agiato popolo dei *prosumer*. Il relativismo estremo si realizza nei metodi per estrarre “verità” dai dati disponibili, pressoché senza limiti: corrispondono al numero di algoritmi e filtri che si possono applicare. Una verità adeguata per ogni ricerca corrisponde a un prodotto personalizzato per ogni consumatore evoluto.

Di fronte a questa chiusura nella creazione, gestione e applicazione (sulla nostra pelle) delle conoscenze, sembrerebbero rimanere solo due possibilità: rifiutare la cultura scientifica come fosse la causa di ogni male; oppure, al contrario, accettare con entusiasmo fideistico ogni “innovazione” tecnologica. Eppure, tra i due estremi della tecnofobia e della tecnofilia, è possibile proporre la curiosità dell'etica hacker ovvero la condivisione dei saperi, l'attitudine critica verso le “verità”, il vaglio accurato delle fonti, proseguendo sulla strada della libera circolazione dei saperi.

In questo contesto, la formazione è un nodo centrale, ma non ci sono sbocchi per una istruzione scientifica diffusa. La struttura della formazione sia in Europa che in Nord America ruota intorno alla sola produzione di specialisti. Al momento non c'è nessun modello pedagogico che risponda all'esigenza di un sapere scientifico per così dire “amatoriale”, nemmeno in Paesi di tradizione non occidentale come il Brasile e l'India che producono ricerca scientifica di alto livello e tecnologia di punta a basso costo grazie a una concorrenza spietata. Un sapere scientifico non accademico né aziendale, diffuso e autogestito, non è ammesso nemmeno in fase di discussione: eppure sarebbe indispensabile per la formazione di competenze di base e per la valutazione delle innovazioni tecnologiche che ci riguardano tutti. In particolare, l'intera nozione di “istruzione scientifica” dovrebbe essere ricostruita per rispondere alla necessità diffusa di rudimenti pratici, per affrontare l'onda lunga della rivoluzione tecnologica.

L'affermazione dell'informatica come principale motore dell'innovazione tecnologica apre infatti nuovi scenari: l'informatica non è semplicemente una tecnica per gestire l'informazione in maniera automatica, ma possiede una logica propria, cioè lavora e modifica continuamente le sue stesse fondamenta. Essa è fisica teorica e sperimentale insieme: studia la formalizzazione del linguaggio (dunque formalizza la conoscenza), la applica ai componenti fisici dell'elettronica, ne ricava linguaggi che a loro volta influenzano le teorie della conoscenza, si regge cioè su una dimensione ricorsiva del tutto particolare.

Nelle scienze classiche si osservano fenomeni stabili: la fisica, ad esempio, elabora dati naturali e su questi crea teorie; nell'informatica, invece, i fenomeni elaborati nella teoria sono artificiali, cambiano fisicamente e concettualmente man mano che i progressi teorici e sperimentali permettono di affinarli: un software elaborato su un computer di dieci anni fa sarà strutturalmente diverso da uno recente. Ciò che era vero ieri, sappiamo già oggi che non sarà vero domani, quando avremo macchine più potenti per fare cose nuove, perché parliamo di un mondo vivo e perciò in continuo divenire.

Miracoli della tecnologia: dalle opinioni soggettive alla verità oggettiva

Il buon Google, in quanto gigantesca base di dati, interviene in questo panorama e sostiene che siamo parte di una grande e inedita “democrazia elettronica globale”; sostiene che i risultati del PageRank sono corretti perché espressione della democrazia diretta dei link valutata dai suoi algoritmi, dunque ci restituisce in un certo senso il nostro legittimo diritto a “dire la nostra”.

La popolarità non può essere ritenuta indice di “qualità oggettiva” da un punto di vista epistemologico perché in questo caso il concetto stesso di oggettività si fonda su un fraintendimento, ovvero sulla convinzione che molte idee soggettive (le “opinioni” espresse sotto forma di link/voto) si trasformino per incanto nell'esatto opposto, in oggettiva verità rivelata, nel momento in cui superano un certo numero e diventano la maggioranza. In questo modo il ranking diventa espressione di qualità perché esplicitamente frutto di una tecnologia di manipolazione

dell'informazione.

Ma come è possibile che la quantità diventi qualità? Si assume surrettiziamente che la mediazione tecnica dell'algoritmo sia garante di "oggettività", e si associa all'oggettività la qualifica di "buono", anzi di "migliore", "vero". Il tutto dev'essere rapido anzi immediato, trasparente, grazie all'eliminazione del fattore tempo nella ricerca e allo studio ergonomico sull'interfaccia.

Il meccanismo di creazione del consenso che Google sostiene essere manifestazione della "democrazia diretta" degli utenti/votanti non convince principalmente per due ragioni: da una parte, assume che la maggioranza abbia sempre ragione; dall'altra, implica che le opinioni della maggioranza, per essere fruite dagli utenti, debbano necessariamente passare attraverso una mediazione tecnica che non viene mai esplicitata in quanto tale.

La dicotomia oggettivo/soggettivo, sovrapposta a quella verità/opinione, è del tutto inadeguata nel mondo delle reti. A voler essere precisi, la scienza ha sempre creato ibridi naturali-culturali, ovvero inventato tecniche e promosso tecnologie. Da una parte l'osservazione e la sperimentazione con il metodo sperimentale ha come campo d'azione la "natura", e in questo senso è "oggettiva"; dall'altra parte, il frutto della scienza è altamente soggettivo. La scienza è infatti sottoposta alla volontà e alla percezione politico-sociale, in quanto mediata dal linguaggio (quantomeno, il linguaggio della comunicazione scientifica), e in quanto fonte di potere (dalla bomba atomica in giù).

La tecnologia che implementa le reti è l'applicazione contemporanea del metodo scientifico per creare "ibridi di natura-cultura", ennesimi oggetti scientifici, che vengono portati sempre più come testimoni della realtà al posto degli umani¹⁰⁵. Ormai ha più valore il responso dell'algoritmo di PageRank, un ibrido tecnologico, rispetto a quella di una persona qualsiasi, e spesso persino rispetto all'opinione di un esperto del settore. Se non altro, per il fatto che PageRank è sempre a portata di click, un esperto no.

Come sottolineato in apertura, da un punto di vista materialistico, dopo tutto, Internet è naturale, nel senso di materiale, ovvero costituita da macchine meccaniche ed elettroniche; nello stesso tempo, è culturale, perché non esisterebbe senza il senso che la cultura le attribuisce, costituita dalle interazioni significative degli esseri umani, o meglio, delle macchine biologiche fra loro e con le macchine elettroniche. Il carattere ibrido delle reti è un'ovvia conseguenza del carattere ibrido della tecnologia *tout-court*.

Un altro punto di vista possibile sulla questione soggettività/oggettività riguarda il modello decisionale: come si decide cosa è rilevante. In un contesto relativistico, è facile concepire un'informazione proveniente da un sito (sia esso un blog, Google, o un circuito di informazione ufficiale) come "oggettivo" se questo giudizio di valore proviene da una presa di posizione chiara, da una chiave di lettura resa pubblica, dalla visione del proprio piccolo mondo localizzato. Una rete di fiducia, ovvero un gruppo di persone che si scambia informazioni, pareri e in generale conoscenza, può facilmente comunicare il proprio modo di lavorare, la propria eventuale gerarchia interna e le modalità per aderire a quella specifica rete o progetto. Ogni volta che si consulteranno i risultati di questa rete di fiducia, potranno essere letti come "oggettivi", nel senso di veri per sé, rilevanti per la propria esperienza, proprio perché frutto di tante soggettività diverse e di una forte interazione fra di esse. Se ci si sente in sintonia con questa comunità si potrà ritenere interessante l'informazione trovata oppure scartarla in funzione di altre reti di fiducia.

Secondo questo approccio, se Google comunicasse in modo pubblico il proprio meccanismo decisionale e gli utenti di Internet fossero in grado di comprenderlo si potrebbe facilmente superare

¹⁰⁵ A proposito della visione della tecnologia come "ponte" di ibridazione fra natura e cultura (estremità teoriche della schizofrenia del moderno), si veda l'opera di Bruno Latour, e in particolare *Non siamo mai stati moderni - Saggio di antropologia simmetrica*, Elèuthera, Milano, 1994. Un punto di vista simile è quello di Gilbert Simondon, che elabora un discorso sulla "tecnica transindividuale", uno spazio meticcio costituito di individuale e collettivo al tempo stesso.

il problema oggettivo/soggettivo, arrivando a identificarsi di volta in volta, ricerca dopo ricerca, con la rete che più ci soddisfa, e influenzarla direttamente affinché ci rispecchi, con i nostri gusti, idee, idiosincrasie.

Pubblico e privato

Il PageRank esemplifica anche un'altra dicotomia: quella fra pubblico e privato. Nei fatti, tutto ciò che passa attraverso Google viene pubblicato, reso pubblico: a chi non è accaduto di trovare email personali nei risultati del ranking, magari perché erano state inviate a mailing list pubbliche? Poiché una mole sempre maggiore di informazioni personali passa attraverso Google, e in generale attraverso le tecnologie informatiche, la possibilità di ritrovare un domani le proprie telefonate (magari effettuate via VoIP) archiviate e reperibili attraverso un motore di ricerca appare tutt'altro che remota. Insomma, le tecnologie giungono a ibridare anche le sfere del pubblico e del privato: del resto, connettersi a Internet significa aprire una porta sul mondo, e attraverso questa finestra di certo il mondo entra nella nostra vita.

Esistono già alcune reti le cui pratiche esautorano di fatto la chimera di un'informazione oggettiva, compiendo una scelta precisa e deliberata, e del tutto soggettiva, tra ciò che intendono rendere pubblico e ciò che invece concerne la sfera privata. Questo fenomeno prende corpo nella sua forma più estesa quando il motore di ricerca si rivela incapace di soddisfare un'istanza di ricerca la cui qualità (specifica) oltrepassa la disponibilità quantitativa e la struttura tecnica proposta.

L'esempio più noto è la ricerca tramite network Peer to peer (p2p)¹⁰⁶ come eMule o simili. La mole di dati ricercabili in queste reti corrisponde ai dati condivisi dagli utenti, e varia nel tempo in maniera non periodica; per questa ragione vengono definite reti "transienti"¹⁰⁷. Infatti in questi casi l'utente compie liberamente una scelta tra il materiale che desidera rendere pubblico e ciò che invece rimarrà sigillato nella sfera privata; anche perché una regola basilare dello scambio "alla pari" è l'offerta e condivisione dei propri materiali, ricercabili nel network, per avere in cambio altri materiali. In sé il p2p è una pratica legale; al limite, i dati scambiati possono essere protetti da copyright (audio, video, testi, ecc.): solo in quel caso si entra nell'illegalità. La diffusione planetaria di queste pratiche evidenzia la natura liberticida della vigente legislazione in materia di proprietà intellettuale e brevettabilità della conoscenza: al momento infatti è reato penale diffondere informazioni protette da copyright, ovvero un buon quarto della popolazione italiana dovrebbe finire in galera per aver scaricato almeno un file mp3 illegalmente. Inoltre appare sempre più urgente la scelta individuale sul confine tra pubblico e privato. Si apre un ampio ventaglio di possibilità che spaziano tra la scelta ponderata di spalleggiare e favorire lo scambio di informazioni al di là di qualsiasi considerazione legale e il semplice abbeverarsi alla fonte della pirateria, approfittando di quel che si trova in Rete.

Pirateria diffusa non significa certo essere alle soglie della rivoluzione popolare, anche perché spesso si tratta di pirateria semi inconsapevole, non certo di un'assunzione di responsabilità da parte di individui che scelgono di opporsi al vigente sistema di protezione delle conoscenze¹⁰⁸. Piuttosto,

¹⁰⁶ Letteralmente, p2p significa "da pari a pari"; in generale, è un modello di comunicazione nel quale ciascuna delle parti ha le stesse funzionalità e ognuna delle parti può iniziare la sessione di comunicazione, in contrasto con altri modelli come il server/client o il master/slave. Spesso la comunicazione P2P viene implementata offrendo a ognuno dei nodi di comunicazione le funzionalità di server e client. Nel linguaggio corrente il termine Peer to Peer viene usato per descrivere le applicazioni con le quali gli utenti possono, attraverso Internet, scambiarsi direttamente files con altri utenti. In particolare, per quanto riguarda Internet, p2p è un tipo di network "transiente" che permette ad un gruppo di persone con lo stesso programma, di connettersi e accedere direttamente alle risorse condivise. Napster fino a qualche anno fa, e ora Emule, Lopster, ecc. sono esempi di tali software.

¹⁰⁷ Il termine "transiente", sinonimo tecnico di "transitorio", deriva dall'astrofisica e dall'acustica e si riferisce a una sorgente la cui radiazione varia nel tempo e in maniera non periodica. Le reti "transienti" sono transitorie dal punto di vista dei flussi informativi che le compongono; nel caso delle reti p2p, la "transienza" dipende dalla variazione della quantità di informazioni condivise da ciascun individuo.

¹⁰⁸ Eppure, violare le leggi sul copyright è incredibilmente un reato tanto diffuso quanto grave: un reato penale! Anche

le tecnologie digitali rendono la riproducibilità tecnica un problema obsoleto, e la cultura del consumismo spinge a desiderare molto più di quanto si possa utilizzare. È quasi ovvio desiderare tutto ciò che riusciamo a immaginare, almeno a livello di “informazioni”, per quanto sia evidente che non saremo mai in grado di ascoltare nemmeno una frazione della musica che scarichiamo dalla rete, o di vedere anche solo una porzione significativa dei film che abbiamo sul nostro hard disk. Ma l’istanza del desiderio, per sua stessa natura illimitato, coniugata alla possibilità tecnica, pone seri interrogativi sulla distribuzione e l’accesso alla conoscenza. Il carattere tendenzialmente gratuito e libero di questi scambi costituisce una sfida formidabile al primato dell’economia produttivista, ma la diffusione delle opinioni, ad esempio sotto forma di blog, mette in crisi anche la concezione tradizionale di media di massa.

Il p2p rappresenta infatti solo il livello superficiale, e più diffuso, dello scambio non mediato da un autorità genericamente sovra-comunitaria. Vi sono molte altre istanze di ricerca qualitativa, tutte soddisfatte da network di fiducia e in grado di ricombinare la prospettiva del nostro orientamento su internet. In alcuni casi questi percorsi soggettivi convergono su sistemi di affinità culturale o professionale, ad esempio forum, newsgroup, blog specializzati, oppure sul dissenso esplicito dalle fonti ufficiali; non solo, spesso costituiscono un esempio sufficientemente articolato da rappresentare un modello alternativo per l’organizzazione della conoscenza (*knowledge management*).

È quindi facile ipotizzare una (forse lenta) evoluzione del circuito di conoscenza legato allo strumento del blog verso strumenti di p2p. Scambio di conoscenza attraverso programmi slegati dalla pubblicazione web e basati su network dinamici ove sia possibile condividere flussi di informazioni e file tra gli utenti connessi, quindi realmente appartenenti al proprio circuito di amici. Questo permetterebbe di rendere la divisione pubblico/privato non un qualcosa di statico, ma dinamico e sfumato, nel quale ogni persona è libera di condividere informazioni con un livello di accessibilità scelto in valore intermedio tra pubblico e privato.

Vie di fuga: media indipendenti, crittografia, blog, foaf

Il fenomeno dei blog¹⁰⁹, siti web che contengono riflessioni personali e link dell’autore e commenti dei lettori, ha dato luogo a quella che viene ormai definita blogosfera: si stimano al momento circa 60 milioni di blog con quasi 4 miliardi di link, e oltre 1,3 milioni di post al giorno. La blogosfera aumenta in ragione di oltre centomila blog al giorno e raddoppia ogni sei-sette mesi¹¹⁰.

Dal punto di vista matematico i blog rispettano la legge di potenza e si distribuiscono secondo la caratteristica “coda lunga”: poche centinaia di blog hanno un numero assai elevato di link (circa quattromila costituiscono il gotha dei blog), la maggior parte (milioni) ne ha invece pochissimi. In tal senso, come abbiamo visto nel Capitolo 2, la blogosfera appartiene al quadro dell’economia della ricerca. Il blog consente di condividere il proprio punto di vista, dichiaratamente soggettivo e personale, senza alcun filtro; il numero di link a un blog rappresenta la sua popolarità e quindi la sua autorevolezza, che può raggiungere e superare quella dei quotidiani e di altri media tradizionali, generando fenomeni di influenza sull’opinione pubblica. Affidabilità, fiducia e reputazione sono legate al singolo estensore del blog: al crescere della popolarità diventa difficile mentire senza essere immediatamente penalizzati, ovvero non più linkati e presto finiti nel dimenticatoio¹¹¹.

in Italia è stata recepita la normativa europea nota come EUCD, modellata sull’americano DMCA, nel 2003; si veda:

<http://www.softwarelibero.it/progetti/eucd/>

¹⁰⁹ <http://technorati.com> (Novembre 2006)

¹¹⁰ Si veda il rapporto La blogosfera e i media, www.casaleggioassociati.it

¹¹¹ Vi sono molti metodi per truccare i dati relativi ai link ad un blog, ad esempio le link farm (gruppi estesi di pagine web create per aumentare la popolarità di una pagina), o i fake blog, ovvero blog creati in maniera automatica da programmi, magari sfruttando aggregatori di notizie tipo RSS. Google cerca da tempo di risolvere questo problema, che genera il fenomeno dei click fraudolenti inficiandone, almeno in parte, il meccanismo pubblicitario.

L'autorevolezza del blog di Beppe Grillo, l'unico italiano a comparire fra i primi 100 blog del mondo per numero di link, è maggiore di quella dei blog di Repubblica e del Corriere della Sera. Il soggetto Beppe Grillo si espone a livello personale e non pretende di spacciare verità: espone il suo punto di vista. In questo senso, i blog creano reti condivise autogestite; non solo, a volte possono diventare le uniche fonti di informazione non allineate. È il caso del blogger iracheno Salam Pax (alias Salam al-Janabi) durante la seconda guerra del golfo¹¹².

La novità più importante legata all'informazione sui blog è quella della aggregazione automatica di fonti differenti attraverso i *feed* RSS (letteralmente: "provvista di RSS"), lo standard *de facto* per l'esportazione di contenuti Web. In poche parole si tratta di un metodo automatico per scorrere velocemente siti web, trovare ed esportare le informazioni che interessano. La popolarità dei blog è forse una delle ragioni principali del successo di RSS: migliaia di weblog iniziano a produrre contenuti in RSS e in questo modo proliferano siti che raccolgono una selezione di post dai blog più seguiti (i cosiddetti *blog aggregator*) e programmi per fruire i contenuti di un blog direttamente sul proprio terminale (*RSS reader*). In questo modo è possibile ricercare contenuti nella blogosfera senza passare attraverso Google.

La possibilità di poter comodamente ricevere in modo automatico gli ultimi articoli scritti degli argomenti che l'utente ritiene più interessanti sul proprio computer è chiaramente un'innovazione destinata ad avere enormi ripercussioni sull'utilizzo del Web. Più in generale, poiché l'RSS è un primo passo per avere ogni informazione in un formato che per propria natura è facilmente condivisibile, trasformabile e ampliabile, rende accessibile un'informazione su ogni supporto informatico, sia esso un sito, un programma installato sul proprio computer, un cellulare o qualsiasi altro strumento tecnologico.

Vi sono però molte altre possibilità per cercare informazioni. Se Google si presenta come un mezzo oggettivo e pubblico, Indymedia.org¹¹³ per esempio si dichiara un "network di media gestiti collettivamente per una narrazione radicale, obiettiva e appassionata della verità", dove però esiste un gruppo di persone che agisce secondo una policy pubblica molto specifica: nella colonna a destra delle pagine, chiamata newswire, chiunque può pubblicare. Non viene censurato nulla, tuttavia "I messaggi palesemente razzisti, sessisti, fascisti vengono occultati, ma non eliminati". Indymedia è dunque essenzialmente uno strumento per diventare utenti attivi nella creazione delle informazioni e contribuire alla formazione della conoscenza e di verità condivise: l'autorità che "crea" le verità è diffusa e partecipata, è un'autorità etica (collettivo) e non matematica (algoritmo).

Mettere in discussione le fonti ufficiali dimostrando che è possibile fare informazione in maniera credibile, decentrata ed indipendente è uno degli obiettivi di Indymedia.org e di decine di network nati attorno a temi specifici¹¹⁴.

Rendere pubbliche le informazioni o mantenerle private? Una risposta interessante viene dal progetto Freenet.org¹¹⁵, una rete p2p decentralizzata, creata per resistere alla censura, che sfrutta le risorse (banda passante, spazio su disco) dei suoi utenti per permettere la pubblicazione e la fruizione di qualsiasi tipo di informazione. Freenet.org è stata costruita pensando ad anonimato e sicurezza, non alla velocità di trasmissione. Le nazioni e autorità di varia natura che censurano le comunicazioni hanno tutte una caratteristica comune: esistono dei controllori che decidono cosa tagliare e cosa mantenere, cosa considerare offensivo e cosa innocuo. Freenet.org è una rete che

¹¹² La vita a Baghdad sotto i bombardamenti nel blog di Salam Pax: http://dear_raed.blogspot.com/

¹¹³ www.indymedia.org; sezione italiana: italy.indymedia.org: "Indymedia è un network di media gestiti collettivamente per una narrazione radicale, obiettiva e appassionata della verità. Ci impegniamo con amore e ispirazione per tutte quelle persone che lavorano per un mondo migliore, a dispetto delle distorsioni dei media che con riluttanza si impegnano a raccontare gli sforzi dell'umanità libera" (dalla presentazione americana).

¹¹⁴ Ad esempio: <http://www.cyberjournalist.net/> <http://alternet.org/> www.warnews.it <http://zmag.org/> <http://slashdot.org> <http://wikipedia.org>, <http://digg.com>, ecc.

¹¹⁵ Per entrare a far parte della rete Freenet: <http://freenetproject.org/>

elimina per chiunque la possibilità di imporre la sua scala di valori: a nessuno è permesso decidere cosa sia accettabile, la tecnologia viene messa al servizio della libertà assoluta di espressione.

Nelle reti non anonime, se vogliamo che le nostre informazioni giungano esclusivamente a destinatari prestabiliti, la crittografia pesante fa al caso nostro – e del resto, è una pratica fondamentale per garantire un minimo di privacy nelle comunicazioni digitali. Ad esempio i nostri messaggi email crittati, pur essendo ricercabili attraverso gli strumenti di ricerca, non sono leggibili¹¹⁶. La crittografia si propone di utilizzare le reti in maniera soggettiva e filtrata, creando spazi privati di scambio fra singoli invece che spazi pubblici indicizzabili e fruibili da chiunque. Non si tratta di avere qualcosa da nascondere, ma di preservare spazi privati e di poter decidere autonomamente e liberamente cosa, come e quando rendere pubblico; le nostre comunicazioni personali, i nostri dati sensibili non verranno resi pubblici e fatti oggetti di marketing per il solo fatto di essere transitati sulla Rete.

È possibile anche rendere tutto pubblico, o meglio: decidere di rendere totalmente pubblica, che continua a far rima con “passibile di ricerca e leggibile”, la propria soggettiva visione di sé. La propria identità virtuale (alter-ego digitale) può essere definita nei minimi dettagli e non lasciata in mano alle tecniche di profilazione. La proposta forse più interessante è rappresentata dalla comunità di FOAF, acronimo di *Friend-of-a-friend*¹¹⁷, ovvero “Amico di un amico”. Lo scopo di questo progetto è creare un insieme di pagine web leggibili automaticamente dalle macchine (*machine-readable*) che descrivono le persone, i loro interessi, le cose che fanno, e i loro collegamenti reciproci¹¹⁸. In questo modo il concetto stesso di verità, facilmente preda di passa in secondo piano, sostituito da quello di “fiducia”: tendo a fidarmi degli amici degli amici, formando “reti di fiducia” basate su gusti, affinità, passioni in comune. FOAF è uno strumento utile per creare network sociali, promuovendo un atteggiamento attivo: è necessario “dichiararsi”, “descriversi”, “pubblicarsi”¹¹⁹.

Dal punto di vista delle reti di fiducia (*trusting networks*, base del Web Trust) FOAF può essere una garanzia di trasparenza circa i percorsi che legano i nostri avatar virtuali, favorendo una crescita di credibilità e fiducia nelle relazioni interpersonali sviluppate sulla Rete. Il Web viene utilizzato per creare spazi in cui le persone sono interconnesse in una grande mappatura relazionale, percorribile a partire da uno qualsiasi dei suoi nodi e senza la necessità di uno strumento di ricerca centralizzato.

È possibile usare strumenti come FOAF anche per scopi dichiaratamente politici: ad esempio, il network indyvoter, che si dichiara “un sito web di networking sociale per attivisti politici”, con l’obiettivo esplicito di formare una classe di governo alternativa¹²⁰.

Le false alternative binarie, natura/cultura, soggettivo/oggettivo, pubblico/privato, vero/falso si rivelano nella loro pochezza. L’informatica è costitutivamente duplice e impura: teoria e pratica al tempo stesso, crea oggetti che modificano il nostro modo di pensare e, utilizzati, modificano

¹¹⁶ La crittografia fa paura ai governi, perché genera spazi privati non immediatamente accessibili al controllo poliziesco: non a caso la legislazione statunitense considerava, almeno fino a pochi anni fa, “armi” gli strumenti di crittazione pesante. Al momento, la legislazione in merito è quanto mai confusa. Naturalmente la possibilità di decrittare un documento dipende dalla potenza di calcolo che viene impiegata. Non si tratta quindi di sicurezza assoluta, ma ragionevolmente commisurata all’evoluzione tecnologica. Così utilizzare chiavi di crittazione asimmetrica pesante come il GPG (*Gnu Privacy Guard*) a 1024 kb offre, nel 2006, una protezione molto elevata, anche se non totale. The Gnu privacy guard: <http://www.gnupg.org/>

¹¹⁷ www.foaf-project.org

¹¹⁸ Più tecnicamente, FOAF è un vocabolario RDF/XML per il web semantico con lo scopo di tradurre in un linguaggio comprensibile alle macchine (ad esempio agli spider) quello che solitamente è scritto in una pagina personale circa l’autore di un sito o di un progetto.

¹¹⁹ Scrivere il proprio documento FOAF non è difficile utilizzando strumenti come Foaf-a-matic, <http://www.ldodds.com/foaf/foaf-a-matic.it.html>

¹²⁰ www.indyvoter.org: “The League of Young Voters* is doing national voter-organizing out of New York City. Our mission is to engage pissed off 17-35 year olds in the democratic process to build a progressive governing majority in our lifetime.”

l'informatica stessa. I mondi digitali sono reali, e nella realtà le visioni in bianco e nero lasciano spazio al grigio, e alle infinite varietà di colori, alle sfumature delle opinioni, alle differenze che sono ricchezza, all'eterogeneo che non può essere ricondotto all'unico. Gli strumenti per modificare, espandere e ramificare questi spazi di libertà e queste linee di confine sono a nostra disposizione: le possibilità sono limitate solo dalla nostra superficialità e mancanza di immaginazione.

Conclusione

Siamo giunti al termine della nostra esplorazione. Ecco svelati molti, più o meno notevoli, segreti di Google.

Come abbiamo visto si tratta di un dispositivo che millanta di poter gestire interamente il complesso sistema delle conoscenze disponibili su Internet. Infatti spaccia per verità oggettive risposte che sono solamente percorsi soggettivi filtrati dalle sue tecnologie di ricerca. Non dobbiamo cadere nella trappola di una lettura esoterica del fenomeno, stupefatti dalla rapidità nel servire le nostre richieste. Questi portenti spesso impregnati di misticismo sono in realtà strategie che coniugano l'uso di sistemi avanzati di raccolta e stoccaggio dati con sistemi di reperimento delle informazioni correlati a tecniche di profilazione dirette e indirette con la personalizzazione delle pubblicità.

Non solo, marketing d'avanguardia e un'opportuna gestione della comunicazione sono la summa dell'evangelizzazione googoliana: l'impiego dei colori elementari dello spettro visivo del logo ne sia un esempio per tutti. Aggiungete la diffusione di interfacce altamente configurabili, capaci di conservare la propria *corporate-identity* in ogni ambito, e il piatto è pronto: lo sfruttamento ad ogni livello dell'economia relazionale messa in moto nei confronti degli utenti. In ultimo Google coopta sistematicamente metodologie di sviluppo cooperativo tipiche dell'Open Source e del Free Software, abbattendo i costi per l'implementazione dei propri servizi e spacciandosi per il sostenitore della causa della libera circolazione dei saperi.

Rifiutare l'agiografia corrente del "motore di ricerca perfetto" non significa avviare una campagna di boicottaggio; del resto, il gruppo di Ippolita si è avvalso spesso di Google, anche per scrivere questo libro.

Allo stesso modo, il largo ricorso a risorse diffuse e accessibili come Wikipedia si è rivelato molto utile a livello enciclopedico. Infatti, se si conosce già l'argomento in discussione, si possono verificare la correttezza delle informazioni in maniera indipendente, ricombinando le parti della Rete in maniera laica e disincantata, per parlare della Rete stessa. L'uso critico delle fonti dipende dalla capacità dei soggetti di accettare valutare l'attendibilità delle informazioni, non dalla bontà intrinseca delle tecnologie digitali.

L'informatica non è semplicemente una tecnica per gestire l'informazione in maniera automatica, ma possiede una logica intrinseca, cioè lavora e modifica continuamente le sue stesse fondamenta. Essa è fisica, teorica e sperimentale insieme: studia la formalizzazione del linguaggio (dunque formalizza la conoscenza), la applica ai componenti fisici dell'elettronica, ne ottiene linguaggi che a loro volta influenzano le teorie della conoscenza, si regge cioè su una dimensione ricorsiva del tutto particolare.

Google sviluppa a fondo tale logica ricorsiva: è una straordinaria macchina che si costituisce attraverso il suo stesso utilizzo da parte degli utenti. In questo senso, è una macchina "autopoietica", che accumula tutte le informazioni di base immesse ogni giorno sulla Rete da milioni di utenti (nomi, fotografie, mail, preferenze di ricerca, partecipazione a forum, uso di blog, compilazione di questionari e moduli, percorsi di navigazione, ecc.) per vendere pubblicità in maniera capillare. I dati degli utenti sono diventati un enorme patrimonio economico, sociale e umano. Un patrimonio da proteggere, certamente, ma anche un territorio in cui curiosare, indagare, sperimentare.

Google fornisce una risposta fluida alle intenzioni di ricerca degli utenti, una risposta articolata in servizi sempre più sofisticati e personalizzabili. Ma si tratta di una molteplicità solo apparente, che mira semplicemente a diffondere una forma di consumismo adatta all'economia informazionale: la personalizzazione di massa delle pubblicità e dei prodotti. Il capitalismo dell'abbondanza di Google procede a un'accurata schedatura dell'immaginario dei produttori-consumatori (*prosumer*), a tutti i livelli. Infatti, gli utenti forniscono gratuitamente i propri dati personali, ma anche suggerimenti e impressioni d'uso dei servizi; gli sviluppatori collaborano all'affermazione degli strumenti "aperti" messi a disposizione per diffondere gli standard di Google, che rimangono sotto il vigilante controllo

di Mountain View; i dipendenti di Googleplex e degli altri *datacenter* si riconoscono pienamente nella filosofia aziendale dell'eccellenza.

La profilazione dell'immaginario non è che l'ultima tappa del processo di colonizzazione capitalistica delle Reti che abbiamo chiamato onanismo tecnologico. La mentalità del profitto si ammanta di dichiarazioni a favore della "libera espressione degli individui", salvo poi sfruttare quelle "espressioni" per vendere luccicanti e inutili prodotti personalizzati.

Google promuove la sua "economia della ricerca" come se fosse una nuova ciberdemocrazia in cui centinaia di milioni di persone possono comunicare direttamente e autorganizzarsi, sottraendosi al controllo statale e istituzionale in genere, attraverso i suoi strumenti tecnologici. Questo semplicistico messaggio trova il sostegno convinto dei media e degli intellettuali "democratici" di tutto il globo. Secondo questa visione, la natura di Internet sarebbe intrinsecamente democratica: in Rete, non solo gli individui sarebbero spinti a prendere il posto delle istituzioni, ma persino le stesse istituzioni sarebbero migliori. L'entusiasmo tecnocratico si spinge fino al punto di presentare l'informatizzazione della pubblica amministrazione, un processo noto come *e-government*, come una sorta di governo senza ideologie, che chiama in causa la responsabilità del "cibercittadino". Questa nuova identità è finalmente partecipe in prima persona (digitalmente) dell'emersione di un'opinione pubblica diffusa. Come se fosse possibile fare fronte alla crisi delle forme classiche di rappresentanza attraverso una democrazia locale connessa su scala globale. Abbiamo constatato le lacune principali di questo assunto, riconducibili a un approccio ideologico. L'idea è che le tecnologie siano "neutre" per definizione, e che questa presunta neutralità sia sinonimo di bontà, in quanto frutto di una pratica di ricerca scientifica oggettiva, in grado di offrire a ogni individuo ciò che desidera, senza sforzo e in tutta fretta.

La complessa mediazione informazionale messa in atto da Google viene fatta apparire come una membrana trasparente di alta tecnologia, garante delle libere scelte degli utenti/cittadini/consumatori, che usano/votano/comprano immersi nella libera Rete gestita da Google per il loro bene.

A dispetto di ogni sogno partecipativo, abbagliato dalla favola della ciberdemocrazia ma privo di sostanza concreta, forme realmente autonome di democrazia diretta non si possono realizzare tramite l'accentramento delle informazioni, dei saperi, dei poteri, nelle mani di un'azienda privata (es. Google), nè tantomeno di un soggetto pubblico (es. l'Authority per le telecomunicazioni).

Nemmeno le punte più avanzate dei movimenti alter-mondialisti sfuggono alla trappola identitaria. Auspicano infatti una ricomposizione dell'identità di classe attraverso una rinnovata centralità del lavoro, in questo caso telematico. Ma rimangono lontani dalla sfera del desiderio individuale anche quando promuovono il *social networking* come soluzione taumaturgica di tutte le frustrazioni personali, in un rito di autoaiuto tecnologico globale.

Solo una scelta di autoformazione può costituire una via di fuga percorribile per sottrarsi al dominio tecnocratico. Molti passi vanno compiuti prima di poter "mettere in comune" qualcosa di sè per generare sinergie. Senza un'adeguata preparazione tecnica, la sedicente manna comunitaria si rivela presto un esercizio solipsistico.

Gli individui che creano le Reti devono invece imparare a gestire i propri dati sensibili e cominciare a stabilire cosa rendere pubblico e cosa mantenere privato. Inoltre, devono essere in grado di decidere quali informazioni sono "vere" e quali "false", valutandole in base al proprio soggettivo punto di vista di quel momento. Infatti, mentre percorrono il Web, devono essere consci di modificare il paesaggio informazionale e di cambiare essi stessi nel corso della navigazione. In questo modo gli individui possono sviluppare la propria autonomia, cioè elaborare delle regole per attraversare i territori virtuali e per acquisire una posizione personale.

Al pari di ogni altra tecnologia, le tecnologie digitali non sono né buone né cattive in sé, e come

abbiamo visto nemmeno neutre: dipende dall'uso che se ne fa e dal metodo con cui vengono sviluppate. Di certo, poiché sono ibridi in grado di amplificare alcuni aspetti della vita reale, permettono di svelare le contraddizioni fra la "Natura" e la "Cultura".

Un altro pericolo va allora scongiurato: l'immagine del Web come esperienza smaterializzata, priva di concretezza fisica, che conduce spesso a un rifiuto luddista e reattivo dell'innovazione. In questa prospettiva, la realtà virtuale del cibernazio brulicherebbe di interazioni poco significative, operate da personalità online dimentiche delle disparità materiali della vita concreta: sesso, razza, ricchezza, posizione sociale verrebbero accantonate nella navigazione fluida e senza attrito di identità fittizie.

Questa idea -solo apparentemente materialista- viene solitamente sostenuta da intellettuali e osservatori che, dall'alto delle loro cattedre, discettano senza aver mai avuto l'umiltà di chiedere lumi ai "ragazzini" nati insieme alla tecnologie digitali.

Al contrario, la realtà virtuale è fisica al punto che non potrebbe esistere senza le macchine meccaniche, di silicio e circuiti, e senza le macchine biologiche, ovvero gli utenti: connessa, multipla e deterritorializzata non sono sinonimi di "immateriale". Inoltre, un simile atteggiamento nasconde una paura profonda di veder cambiare il mondo senza riuscire a ritrovare il proprio posto e una sfiducia radicale nelle possibilità di trasformazione e crescita degli individui.

Le tecnologie digitali rappresentano perciò una possibilità di liberazione solo se vincolate allo sviluppo di alter-ego digitali complessi, consapevoli, in grado di interagire in maniera imprevedibile. È possibile utilizzare molteplici linguaggi per costruire un terreno d'incontro: tra gli altri, Ippolita ha trovato indispensabile il metodo sperimentale della ricerca scientifica, la ricchezza inesauribile della tradizione umanistica, la forza dialogica delle passioni politiche, la coralità della narrazione del metodo femminista, la curiosità senza limiti tipica dell'attitudine hacker. La fiducia nella possibilità di modellare tecnologie in base ai desideri degli individui è indispensabile per creare Reti davvero libere, digitali e non solo.

Il caos dei messaggi contraddittori, il rumore di fondo a volte insopportabile e l'ampiezza quasi inconcepibile della Rete possono senz'altro incutere timore, ma l'esplorazione è appena cominciata.

BOX / APPENDICI

Appendice I - La fine del mondo in un biscotto

La data di scadenza di Google è il 17 gennaio 2038. Questa data è interessante perché è giusto due giorni prima della fine del mondo. Il 19 gennaio 2038 è una data data essenziale per tutto il mondo POSIX, che comprende tutti i sistemi operativi derivati dal sistema UNIX (GNU/Linux, BSD, Solaris, ecc.) e tutti i protocolli di rete UNIX style (http, ftp, etc). Sui sistemi POSIX il tempo è regolato per ragioni storiche dal numero di secondi intercorsi dal primo gennaio 1970. L'unità di misura per questo numero è un "intero segnato", valore che nell'informatica occupa in memoria 32 bits.

Se un programmatore crea una variabile di tipologia intero segnato per memorizzare un valore numerico, questo può essere come minimo -2147483648 e come massimo 2147483647. Un numero molto grande, ma che diventa un valore piccolissimo se lo trasformiamo in secondi. In 32bits infatti ci stanno appena 136 anni.

Quando Ken Thompson, quasi per gioco, inventò lo Unix, mai avrebbe immaginato di rivoluzionare a tal punto l'informatica e soprattutto mai avrebbe pensato che una serie sue scelte (il filesystem, l'ora, la codifica caratteri, le chiamate a funzione, il linguaggio di programmazione, ecc.) sarebbero diventate legge per ogni programmatore del globo da lì a brevissimo. Thompson decise nel 1972 che la data 0 dal suo sistema operativo sarebbe stata il 1 gennaio 1970.

2147483647 secondi dopo il 1 gennaio 1970, sarà esattamente il 19 gennaio 2038 alle 3:14 di mattina, ultimo secondo del mondo Unix, di internet (che funziona grazie a protocolli Unix) e dei principali server del globo (che utilizzano sistemi operativi derivati da Unix). Dopo quel secondo saremo proiettati nel 13 dicembre 1901 alle 20:45. Sicuramente questo sarà un problema da gestire da qui ai prossimi anni e richiederà un cambio epocale nella gestione del tempo e di tutto il resto nei sistemi Unix. Sicuramente un problema un po' più serio e fondato del tanto decantato "Millennium bug" nel passaggio dal 1999 al 2000.

La data di scadenza di Google è quindi la data massima impostabile in un *cookie* comportando che il browser dei nostri sistemi operativi, non rimuoverà mai quel *cookie* e le informazioni in esso salvate. Per un approfondimento del problema delle date, si vedano:

- <http://www.2038bug.com/index.html>
- <http://www.iee.org/Policy/Areas/SCS/problemdates.cfm>
- http://en.wikipedia.org/wiki/Unix_epoch~~

Appendice II - Interzone: influenze e domini nei mondi digitali

I progetti Open Source di Google sono ospitati anche sul portale sourceforge.net; non vi sono cambiamenti sostanziali, anzi nemmeno cambiamenti grafici di rilievo, nel passaggio tra il sito Google.com e quello di sourceforge.net. I diversi progetti rimangono inalterati quasi fin dei minimi

dettagli, sia nel codice, sia nell'interfaccia. L'unico elemento degno di nota è il cambio del server ospitante: infatti si passa da una sezione con il nome del progetto sul sito code.google.com a una pagina ospitata da sourceforge.net in un dominio di terzo livello, ad esempio nomeprogetto.sourceforge.net. Esistono quindi due accessi separati e indipendenti per raggiungere la stessa Interzona. Tuttavia queste zone appartengono a domini differenti: Google.com e sourceforge.net.

Il concetto di dominio è alla base della struttura delle Reti. In ambito informatico, con il termine “dominio” si intende il nome di un server all'interno di una scala gerarchica estremamente complessa gestita da una rete di computer chiamati DNS (Domain Name Server). Questi computer hanno il compito di gestire l'associazione server - dominio e permettere a tutti di raggiungere gli indirizzi sparsi nella rete grazie al loro nome.

Ad esempio, quando si vuole raggiungere www.google.com, uno specifico programma all'interno del nostro computer chiede a uno dei DNS quale server risponda al nome “www.google.com”. In sostanza, la rete dei DNS si occupa di cercare all'interno di gigantesche mappe e database il server da noi richiesto, associando l'indirizzo alfabetico a noi comprensibile a un indirizzo numerico (il numero IP) comprensibile per le macchine. Ci comunica poi il risultato, e noi veniamo collegati a Google o a qualunque altro indirizzo richiesto.

Semplificando molto questo complesso meccanismo gerarchico, possiamo dire che i domini sono strutturati a livelli. In un tipico indirizzo web, i livelli si leggono da destra a sinistra. I primi livelli sono quindi i generici .com, .org, .net e quelli nazionali come .it, .fr, .de, .tv. Al secondo livello troviamo i nomi reali dei server, associati a uno solo dei domini di primo livello: google.com, eleuthera.it, somasuite.net, alekoslab.org ecc.. Nulla vieta a un server di avere più domini di secondo livello legati a differenti domini di primo livello: google.it, google.de, google.com sono tutti domini di secondo livello validi. Non è obbligatorio che questi domini di secondo livello si riferiscano a una stessa entità: ad esempio, ippolita.net è il sito del progetto Ippolita, ippolita.it è il sito della poetessa Ippolita Aвали, ippolita.com è il sito di Ippolita che fabbrica e vende gioielli.

Dal secondo livello in poi la gestione spetta a chi gestisce il dominio di secondo livello. Ad esempio, code.google.com è gestito direttamente da Google, così come www.google.com, e anche moon.google.com.

Non esiste una regola precisa che impone il modo in cui si danno i nomi di terzo livello: basti pensare a un indirizzo evocativo, ma assolutamente corretto, come del.icio.us: il suffisso .us indica l'appartenenza al dominio nazionale USA, il nome “icio” indica il server, il prefisso “del” di terzo livello completa il gioco di parole, in modo che l'intero indirizzo significhi “delizioso” nella lingua inglese. Tuttavia è uso associare alcuni domini a servizi specifici. Un indirizzo del tipo www.server.net ospiterà tendenzialmente la prima pagina del servizio di server.net, mentre mail.server.net ospiterà dei servizi di posta elettronica, e così via.

L'immagine forse più adatta per descrivere questa gerarchia è quella di un casa: i servizi hanno una porta di accesso diretto (mail.server.net) mentre il contenuto dei diversi servizi può essere ospitato dentro una delle stanze interne (www.server.net/fotografie/), cioè in una delle directory del server principale.

La gestione dei domini di terzo livello di Google è incongruente, e per questo estremamente aperta e fluida. È curioso infatti notare che alla Luna è stato concesso un dominio di terzo livello (moon.google.com) mentre a Marte no, relegato in una directory secondaria (www.google.com/mars): entrambi questi servizi permettono di esplorare mappe, senza differenziazioni significative nel servizio proposto.

La struttura di Sourceforge.net è invece più rigida, come del resto quella di Freshmeat.net (che pure Google al momento non utilizza) e di altri progetti analoghi. Sourceforge.net offre i progetti

ospitati come servizi per gli utenti a causa della sua stessa natura di portale; per questa ragione tutti i progetti di Google si trovano in un dominio di terzo livello, esattamente come tutti gli altri progetti. Si tratta di uno strumento per le comunità di sviluppo che dedica a ognuna di esse un ambiente *ad hoc* ricco di servizi. Le diverse comunità fanno richiesta dei servizi di cui necessitano e Sourceforge.net dà a loro uno spazio completo e personalizzato. Quindi Google si comporta come la comunità di Gaim o di qualsiasi altro software: eppure, Google non è precisamente una qualsiasi comunità di sviluppatori indipendenti...

Altre interzone vengono create per gestire i campi di influenza delle rispettive comunità e l'eventuale interazione fra i progetti che ne dipendono. Il progetto Open Source Mozilla e la pagina predefinita per il suo browser Firefox è un piccolo esempio di grande visibilità pubblica. Ogni volta che un utente installerà il browser Mozilla-Firefox la prima pagina online che vedrà all'avvio sarà quella di Google modificata graficamente in modo che comprenda anche il logo di Firefox.

Google fornisce anche toolbar da integrare in Firefox. Le Toolbars (Barre degli strumenti) sono blocchi grafici (per lo più a forma di riga o colonna) che, grazie ad una serie di pulsanti e altri oggetti grafici in esse contenute, attivano funzioni particolari del programma che le ospita. Le estensioni per Firefox permettono l'aggiunta di toolbar gratuite all'interno del browser. Quello che il pubblico difficilmente sa è che tali prodotti non sono affatto Open Source. Infatti la toolbar di Google non è un prodotto scaricabile dai classici siti di estensioni per Mozilla, perché rilasciato sotto una licenza proprietaria. Il funzionamento della toolbar è nascosto dentro ad una piccola libreria utilizzata solo da questo strumento e rilasciata senza codice sorgente. Esistono diversi studi illegali di *reverse engineering* che mostrano ogni funzionamento di questa estensione e il modo in cui vengono richiesti dati dal sito di Google.

L'integrazione fra il proprio terminale (computer, palmare, ecc.) e i Google-server è sempre più stretta; infatti, una delle novità fra le estensioni per Mozilla Firefox si chiama Google Browser Sync: bookmarks, cronologia, cookies, password salvate e persino le schede e finestre lasciate aperte durante l'ultima sessione vengono salvate online sui server di Google e poi sincronizzate automaticamente sulle diverse postazioni utilizzate e nelle nuove sessioni di navigazione. Il controllo sull'utente si estende a qualsiasi dispositivo in grado di accedere al web.

Infine, possiamo individuare un'ulteriore interzona sul fronte delle licenze d'uso. Google ha infatti cercato di adeguarsi allo standard delle altre aziende Open Source che tendono a utilizzare versioni più o meno ritoccate della BSD 2.0 (fatta eccezione per la Novell). Questa licenza permette la condivisione del codice ma anche un saldo controllo sul rilascio dei programmi e delle loro versioni. Da una parte ciò consente la creazione di comunità, mentre dall'altra parte lascia tutti i margini per un controllo generale rispetto all'intera filiera produttiva. Qualsiasi programmatore indipendente che sviluppa gratuitamente utilizzando questa licenza potrà essere legalmente espropriato in ogni istante della propria creazione.

Per quanto alcuni responsabili di Google abbiamo comunicato un possibile cambio di licenza in favore della licenza Apache, MIT, o persino GPL, la BSD rimane quella standard per tutti i progetti ospitati e promossi dal motore di ricerca. BSD costituisce dunque l'interzona di licenza utile per attirare a sé i codici liberi del mondo Open Source e farli diventare proprietà esclusiva della propria azienda.

Questo libro è distribuito sotto licenza Creative Commons 2.0 (Attribuzione, Non Commerciale, Condividi allo stesso modo), una licenza di tipo **copyleft** che abbiamo scelto per consentirne la libera diffusione. Ne riportiamo il testo in linguaggio accessibile. Puoi trovare una copia del testo integrale della licenza all'indirizzo web <http://creativecommons.org/licenses/by-nc-sa/2.5/it/legalcode> o richiederlo a info@ippolita.net



Attribuzione - Non commerciale - Condividi allo stesso modo 2.5

Tu sei libero:

di riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare quest'opera
di modificare quest'opera

Alle seguenti condizioni:



Attribuzione. Devi riconoscere il contributo dell'autore originario nei modi indicati dall'autore o da chi ti ha dato l'opera in licenza.



Non commerciale. Non puoi usare quest'opera per fini commerciali.



Condividi allo stesso modo. Se alteri o trasformi quest'opera, o se la usi per crearne un'altra, puoi distribuire l'opera risultante solo con una licenza identica a questa.

12. Ogni volta che usi o distribuisi quest'opera, devi farlo secondo i termini di questa licenza, che va comunicata con chiarezza.
13. In ogni caso, puoi concordare col titolare dei diritti d'autore utilizzi di quest'opera non consentiti da questa licenza.

Le utilizzazioni consentite dalla legge sul diritto d'autore e gli altri diritti non sono in alcun modo limitati da quanto sopra.

Puoi consultare e scaricare la versione digitale del libro dal sito

<http://www.ippolita.net/google/>