

La blogosfera nel periodo dell'affidabilità

1. Introduzione

- L'analisi qui proposta si è sviluppata in seguito all'ultima discussione in linea sul problema dell'affidabilità dei blog. Discussione che, in realtà, non fa altro che riproporsi puntualmente da quando esiste la rete.

Una volta che si è creato il contenitore infatti occorre creare un metodo valido per classificarne i contenuti: in internet questo non si è ancora fatto. La nostra riflessione prende spunto da un avvenimento piccolo, ma che ha trovato grande risonanza all'interno della blogosfera italiana; un post di Beppe Grillo basato su un documento prodotto dalla società che lo assiste nella sua avventura sul web.

Chiariamo dal primo momento come questo saggio non si soffermerà sulle caratteristiche tecniche che un eventuale algoritmo di calcolo del rating dovrebbe sviluppare, quanto piuttosto su quale contributo ad esso può fornire la sempre maggiore rete mondiale di blog, su cosa si intenda effettivamente con il termine Affidabilità e sulla necessità stingente di procedere e sviluppare la riflessione su questo tema al fine di rendere lo sviluppo della rete più democratico e gestibile dagli utenti.

Tutti i contributi riportano sempre nome e il link dell'autore originale. Nel caso qualcuno di questi ci fosse sfuggito verranno aggiunti immediatamente in seguito a richiesta. Rendiamo altresì noto come questo saggio non sia un prodotto definitivo, ma un canovaccio di discussione creato per riassumere i termini della vicenda e far nascere nei lettori il desiderio di migliorare tutte le singole parti del documento stesso.

- La causa scatenante dell'ultima ondata di post sul tema dell'affidabilità delle fonti in rete viene da un post di Beppe Grillo¹ che annuncia la morte dei media tradizionali, sostituiti nel suo pensiero dal libero arbitrio della rete internet e dai blog dedicati all'informazione. Grillo basa le sue affermazioni su un rapporto della Casaleggio Associati², società che lo assiste e consiglia

¹ http://www.beppegrillo.it/2006/01/lalbo_dei_blogg.html

² <http://www.casaleggio.it/>

per l'uso del web e che propone un'equazione dalla semplicità disarmante (link alla fonte = attendibilità della stessa) arrivando a sostenere che il weblog di Grillo è maggiormente degno di fiducia del sito web di Repubblica in quanto semplicemente capace di attirare un numero maggiore di collegamenti dall'esterno. Continuando il ragionamento, il documento arriva a considerare imminente la fine dei media tradizionali, sostituiti dai weblog e dalla libera informazione prodotta in rete.

Ovviamente questa considerazione riveste un doppio ambito, quello scientifico di ricerca e quello economico in quanto dallo stesso documento di Casaleggio si evince una seconda equazione non meno importante e cioè link alla fonte = attendibilità della stessa = capacità virtuale di generare denaro. La posizione espressa da Grillo pare evidentemente insostenibile ed incongruente in sé, ma ha generato una lunga discussione che prosegue tuttora. Unendo a questo l'articolo di Umberto Eco, apparso sull'ultimo numero de L'Espresso³, in cui il noto semiologo parla proprio del problema delle fonti in rete e della problematicità di definirle l'affidabilità (In questo caso Eco si riferisce in particolare a Wikipedia), si può parlare senza mezzi termini dell'esistenza di una problematica sempre più sentita, che esce dai confini della blogosfera per arrivare a interessare tutti coloro che, nel bene e nel male, sono costretti a ricorrere all'uso della rete a fini informativi o scientifici.

Il problema di base non è quindi da riferirsi ad una eventuale lotta per il predominio dell'informazione tra vecchi e nuovi media, né il poter stabilire una classifica migliore del blog più letto o del blogger più bello, quanto il poter cominciare a pensare a come l'uso dei blog possa rendere più comprensibile la rete, a come questa massa di utenti possa, anche inconsciamente, migliorare la disponibilità delle fonti rintracciabili in breve tempo. Il tutto si può riassumere nella seguente affermazione: Le informazioni disponibili sulla rete Internet sono già ampiamente oltre la soglia di guardia. Per cercare di perderne meno è indispensabile stabilire dei criteri capaci di decimarle nel minor tempo possibile mantenendo la scelta per quanto si può accurata.

Il nostro contributo tenderà proprio a proporre delle basi comuni da cui partire.

³ <http://www.espressonline.it/eol/free/jsp/detail.jsp?m1s=o&idCategory=4817&idContent=1252511>

2. **Semiosfera e Blogosfera**

Al fine di comprendere meglio il rapporto tra la cultura, la gestione della stessa, la blogosfera e la rete internet riprenderò qui una distinzione già fatta in precedenza⁴ e cioè l'identità tra il concetto di semiosfera proposto da Jurij Lotman e i reali sviluppi e le tendenze presenti sul web di oggi.

Lotman considera la cultura umana come formata da una grande quantità di linguaggi differenti, capaci di interagire tra loro tramite meccanismi di derivazione biologica. In questa concezione della cultura, contrariamente al pensiero dello stesso Eco, non abbiamo dunque un insieme unico e definito nell'unità di tempo di nozioni e concetti, quanto piuttosto un numero elevato di sottoinsiemi culturali capaci di convivere, a volte separati, a volte entrando in relazione tra loro, fino a compenetrarsi e fondersi. La costante interazione di questi insiemi di segni forma la cultura all'interno della quale viviamo.

Prendendo per buona questa distinzione possiamo considerare la rete Internet come una rappresentazione abbastanza fedele della stessa e possiamo quindi procedere a dividerla in sottoinsiemi indipendenti dotati di un proprio linguaggio ed interagenti con tutti gli altri sistemi presenti. L'importanza di stabilire qui divisioni nette e precise tra questi sistemi complessi non è rilevante. Non è importante quindi stabilire con certezza matematica se si debba considerare insieme autonomo la blogosfera in sé o se sia opportuno fare ulteriori divisioni e categorizzazione. Questo perché appunto i sistemi sono in costante modificazione, ma ancora di più perché per noi l'importante è stabilire delle differenziazioni di base, chiare ed evidenti. Dire quindi che il fenomeno dei blog utilizza un linguaggio differente per rapporto a quello delle altre componenti Internet e considerarlo un sistema unico a parte è già sufficiente per i nostri obiettivi.

Fissato questo punto di partenza comune e prima di prendere in considerazione i modelli già esistenti di ranking e di categorizzazione crediamo sia opportuno fare risaltare ulteriormente l'importanza della discussione su questo tema, proponendo due risultati positivi che una mobilitazione seria ed efficace potrebbe portarci. Come primo punto a sostegno della nostra idea portiamo la constatazione che fino ad oggi, in mancanza di uno standard comune di classificazione dei documenti in rete, si sono succedute strategie diverse, ogni volta cadute progressivamente sotto il controllo economico di qualche società. Così come Yahoo! È divenuta una società

⁴ <http://fasttrack.splinder.com/post/6886069>

quotata in borsa, lo stesso è stato per Google e Technorati sembra seguire la medesima strada, come fa presagire il sondaggio che ha recentemente proposto ai suoi utenti, parlando di un servizio di base gratuito e futuri servizi a pagamento per gli utenti. L'iniziativa privata è quindi ricaduta ogni volta in considerazioni economiche che portano in qualche modo a modificare sia lo spirito che le considerazioni che erano alla base dei progetti. Puntualmente ci troviamo quindi con il rischio che le nostre ricerche siano inquinate da modificazioni legate al fattore denaro, questo mentre ogni volta di più ci troviamo legati alle stesse. Il pericolo è talmente evidente che non vale la pena spenderci sopra altre parole.

In secondo luogo riteniamo che una corretta indicazione che possa venire da questo tema possa anche aiutarci nello sviluppo del concetto di web semantico, che altro non è se non il progetto più ambizioso di classificazione dei contenuti in rete. Utilizzare la sempre maggiore produzione di documenti pubblicati ogni giorno dai blog e la loro classificazione sistematica e comune potrebbe infatti farci compiere un grande passo verso la gestione razionale del materiale presente in rete. Classificazione che non sarebbe più prodotto esclusivo di un calcolo matematico, ma che vedrebbe convergere tecnologia e saper fare umano; libertà del singolo e matematica applicata.

Per questi due motivi è fondamentale che una discussione come questa, sui criteri di gestione dell'informazione risultante dai blog, abbia tutto lo spazio possibile e veda convergere i contributi di individui differenti per estrazione, conoscenze e capacità.

3. Technorati, Google e Yahoo!

I tre grandi passi che la classificazione di massa sulla rete ha compiuto sono legati a tre società americane: Yahoo! In primis per gli anni novanta, Google alle soglie del nuovo millennio, Technorati in seguito all'esplosione del fenomeno dei blog.

Passeremo ora in rassegna brevemente i tre metodi proposti, cercando di identificarne i punti deboli che ci portano a sostenere come essi non siano un punto di arrivo, ma soltanto una traballante ed insufficiente situazione di partenza di fronte alla nostra necessità di reperire informazione in rete.

Yahoo! è un motore di ricerca di prima generazione, dispersivo, costoso e soggettivo (Brin & Page 1998)⁵. L'evidenza di questo ragionamento è contenuta nell'articolo che sta alla nascita di Google, pubblicato in rete dall'università di Stanford 8 anni fa. Questo primo modello di motore di ricerca funzionava infatti attraverso una

⁵ <http://www.cis.upenn.edu/~yicn/cse591f05/anatomy.pdf>

scansione umana, che per ragioni di possibilità limitate copriva e copre tutt'ora una limitata parte della rete, quella considerata di maggiore interesse per il pubblico. I risultati di questa scansione manuale sono tuttavia necessariamente soggettivi in quanto prodotto di una scelta umana e sono oltretutto estremamente costosi in termini di manutenzione e di sviluppo. L'inserzione in un secondo tempo di meccanismi di ricerca automatizzati funzionanti a parole chiave non ha migliorato di molto i risultati per via del gran numero di risultati di qualità scadente che essi propongono per via del rumore prodotto dalla rete (è facile infatti ingannare questi meccanismi automatizzati inserendo codici e parole all'interno dei documenti html per sviarli e catturare l'attenzione e la prima pagina del ranking).

Proprio per ovviare a queste difficoltà Brin e Page crearono il sistema Google che resta ancora oggi il più utilizzato motore di ricerca del mondo.

Google funziona secondo un metodo completamente differente se visto in rapporto a quelli precedenti. L'idea alla base del progetto era infatti quella di trovare un modo per crescere con l'indicizzazione al pari della crescita della rete e di utilizzare in maniera più efficiente gli spazi di memoria disponibili ottenendo al contempo un miglioramento significativo dei risultati proposti. Il sistema di Google si basa sul page rank, ovvero sull'utilizzazione da parte del motore di ricerca di tutti quei link prodotti dai singoli utenti della rete al fine di stabilire l'importanza e la rilevanza delle singole pagine web. L'idea si basa dunque su un presupposto di senso comune: *"These maps allow rapid calculation of a web page's PageRank", an objective measure of its citation importance that corresponds well with people's subjective idea of importance*"⁶ (Page & Brin 1998). Google utilizza semplicemente il lavoro degli utenti della rete per dare ordine ai documenti degli stessi e stabilisce l'importanza delle pagine in base all'importanza soggettiva che *già* gli utenti hanno dato ai documenti. Siamo passati in un campo di riflessione diverso; il motore di ricerca ci fornisce un prodotto riflesso. Anche Google ha però incontrato diverse difficoltà negli ultimi tempi, essendo il suo un metodo facilmente influenzabile, non troppo diverso dalla categorizzazione precedente. Sono infatti nate nel tempo pagine web il cui solo scopo è quello di indicizzare il motore di ricerca verso determinati siti, le cosiddette *"Farms links"*. La possibilità di poter modificare i risultati in maniera così semplice ha portato ad un'inflazione di link in rete, rintracciabile oggi anche nel mondo dei blog. Col passare del tempo e, almeno in parte, come

⁶ Trad: Queste mappe permettono un rapido calcolo della classifica delle pagine web, una misura oggettiva dell'importanza tramite le citazioni, che corrisponde bene all'idea soggettiva di importanza degli utenti.

conseguenza diretta di questo fenomeno, Google, ha apportato tre grandi modifiche al proprio metodo di lavoro: una separazione delle competenze, un riequilibrio dei valori dati al page rank e la possibilità per le aziende di acquistare i risultati prodotti dal motore di ricerca.

- Per quanto concerne il primo punto la progressiva differenziazione di Google in ambiti diversi mi pare parlare da sola. Si è proposta prima una divisione in lingue, poi una che separi le immagini dai testi, una che distingue i tipi di file da cercare ed ultimamente fioccano versioni beta di motori di ricerca che derivano dal vecchio Google per dedicarsi però ad un solo ambito (Google Scholar ad esempio, Google Maps, Google Ride Finder o Google Ricerca Libri). Quindi è la stessa società che, resasi conto del fenomeno che abbiamo appena espresso, sta cercando di migliorare e settorializzare il proprio operare.
- In secondo luogo ci sono state più modifiche al progetto originale di utilizzo del page ranking e oggi Google utilizza (non sappiamo ufficialmente in che modo) un misto di parole chiave e dati recuperati dai link. L'intenzione di incrociare due metodi di lavoro così differenti deriva dall'evidente volontà di ridurre l'errore prodotto dall'aumento di link in rete. Dividendo su più piattaforme diverse il calcolo del ranking, l'errore viene drasticamente ridotto.
- Non bisogna inoltre scordare come Google sia un'azienda capace di capitalizzare oltre 100 miliardi di dollari (138 a quando scrivo) e come quindi abbia la necessità di produrre utili attraverso il suo operare. In quest'ottica si leggono alcune delle nuove iniziative proposte dall'azienda e nella stessa ottica la vendita di parole chiave tramite il sistema "*pay for click*". Oggi non è ancora tuttavia chiaro quanto queste iniziative possano incidere sui risultati del motore di ricerca. In Francia ad esempio <http://voilà.fr>, il motore di ricerca di Wanadoo, mette questi link sponsorizzati nella prima pagina dei risultati, falsandoli in maniera palese, in Google non si sa ancora con esattezza quanto essi incidano. Non mi pare tuttavia fuori luogo manifestare il timore che essi incideranno in maniera sempre maggiore sui risultati proposti; presto acquistare il giusto numero di parole chiave potrebbe portare un sito web direttamente ai vertici della classifica portandolo così a nascondere il naturale ranking prodotto dalla rete e dal tempo.

Queste ultime tendenze, ci fanno pensare che il metodo proposto da Google potrebbe non rivelarsi più così efficiente con il passare del tempo ed inoltre, il predominio da esso esercitato sulle ricerche sul web, non è in ogni caso un sinonimo di obiettività e una garanzia sulla corretta gestione delle informazioni.

L'ultimo caso che ci apprestiamo ad analizzare è quello di Technorati, il più recente progetto di classificazione, sia come novità concettuale che come diffusione temporale.

Technorati infatti analizza il web grazie all'ausilio dei blogger, grazie ai loro rimandi intertestuali, grazie ai link che essi producono in continuazione pubblicando post, commenti e rimandando a post di blog che sono soliti leggere. Secondo la filosofia di Technorati nel mondo dei blog il link assume un valore ancora maggiore che nella pubblicazione html tradizionale; il motore di ricerca usa allora questi rimandi per ricostruire un senso di temporalità e di connettività propri della conversazione umana. Technorati traccia i link e in base al numero di questi stabilisce una classifica di rilevanza dei blog presenti in rete, proponendo allo stesso tempo un'analisi continua della natura del blog. Il meccanismo di aggiornamento automatico di Technorati mostra in tempo reale quale sia l'argomento di maggior importanza nella blogosfera, mostra su cosa si stia discutendo in rete e cosa potrebbe nascere in prospettiva. Le difficoltà principali di questo sistema risiedono nella sua non-inclusività e in un semplicismo deterministico che deriva dai risultati. Infatti Technorati funziona solamente per inclusione e non traccia spontaneamente i contenuti che non vi giungono per esplicita richiesta dell'autore. Questa scelta se da un lato permette di ottimizzare le risorse e di concentrarsi su quella parte di blogosfera più attiva, esclude del resto tutti coloro che per qualsiasi plausibile ragione non sono giunti a divenire parte del terreno setacciato da Technorati e quindi ogni risultato potrebbe essere falsato in principio. Secondariamente il metodo sviluppato non produce alcuna distinzione tra i link che vengono prodotti, tutti quanti contribuiscono allo stesso modo a formare una classifica generale che in verità è una classifica priva di reali valori da estrarre in maniera diretta; possiamo solo evincerne il numero di link presenti e il rapporto con gli altri blog. Senza azzardare ulteriori ragionamenti giustificati dai dati in sé. A questi due difetti crediamo tuttavia il sistema di Technorati affianchi dati di grande valore, come la possibilità di restare aggiornati su determinati argomenti in tempo quasi reale e l'utilizzo limitato delle tags per gestire la grande quantità di dati giornalmente

prodotta. Technorati sembra dunque prestarsi maggiormente ad un uso esperto e non immediato; capace di fornire notizie estremamente precise e utili, ma di trarre in inganno, come l'ultimo rapporto Casaleggio⁷ fa intendere.

4. **Discriminazione di termini: Affidabilità, Influenza, Visibilità**

Fondamentale diviene a questo punto entrare nel novero della questione, discriminando il termine fondamentale attorno al quale avviene tutta la discussione: Grillo e Casaleggio utilizzano un termine molto generico parlando di *Attendibilità*. Può questo termine rappresentare e risolvere in maniera soddisfacente la richiesta di classificazione di cui esponiamo la necessità?

L'intera discussione, per come la si può evincere dal rapporto Casaleggio, si basa su un unico fattore, la credibilità del blog d'informazione e, proprio per dimostrare la tesi dell'obsolescenza dei vecchi media al confronto dei blog, viene utilizzato il metodo del primo Google per mostrare come i blog, avendo un numero maggiore di link rispetto ai siti tradizionali, dispongano di un pubblico maggiore e di una attendibilità informativa superiore rispetto ai siti di molti quotidiani online. L'intera ricerca si basa sui dati di Technorati. Questi dati tuttavia per un sito tradizionale non sono affidabili per i motivi che riportavamo sopra; perché Technorati è una classifica non inclusiva e centrata sui blog e perché al contrario dei blog i siti dei principali quotidiani tendono ad essere testi chiusi in sé stessi e a ridurre i link che inviano, e di conseguenza che ricevono, all'esterno. Il ragionamento quindi è basato su due errori fondamentali; si basa in principio su dei dati che sono incompleti ed allarga la loro settorializzazione all'intera sfera del web. In secondo luogo c'è l'evidente salto logico di presentare come indice di Attendibilità dei link che nella loro natura non contengono alcun giudizio di valore. In base a questi due punti, l'intero rapporto Casaleggio perde un qualsiasi valore scientifico, avendo trascurato anche le variabili più semplici del problema all'interno della propria analisi.

Il termine *Attendibilità* che viene utilizzato ci lascia inoltre molto perplessi e ci ha spinto a riprendere una distinzione fatta da Maurizio Goetz⁸ che preferiva scomporre il termine in tre differenti fattori interagenti: *Affidabilità, Influenza, Visibilità*.

⁷

<http://www.casaleggio.it/rapporto.asp?articleID=286&titolo=Focus:%20La%20Blogosfera%20e%20i%20Media>

⁸ <http://www.bookcafe.net/blog/blog.cfm?id=402>

5. Affidabilità

Il punto più controverso della questione è sicuramente il giungere a stabilire cosa sia e come si possa eventualmente identificare una presunta affidabilità della fonte di informazione in rete. L'affidabilità come la intendiamo noi è un'affidabilità a priori rispetto al singolo post, una sorta di giudizio che la rete ha espresso sulla fonte nel corso del tempo e che è necessario discriminare dall'ultimo post pubblicato, sempre in grado di modificare questa opinione attraverso i propri contenuti. Riguardo la natura e l'affidabilità di un blog cercheremo ora di sviluppare e migliorare, grazie alle critiche ed ai giudizi che sono giunti dalla rete, il documento pubblicato la settimana scorsa⁹.

Proponiamo infatti di partire non dal rapporto Casaleggio e non dal prodotto di page ranking cui Google ci ha abituato, quanto da una considerazione di carattere più generale ripresa dal funzionamento del testo secondo la teoria semiotica di Umberto Eco. Il principio base da cui il nostro ragionamento può partire è proprio anche del senso comune ovvero *“Un testo in principio non è affidabile per nessun lettore”*. Questa considerazione ci permette di svilupparne due assiomi; un testo diviene quindi affidabile quando dimostra di esserlo oppure quando l'autore del testo è da noi riconosciuto come tale e il testo conferma solamente la nostra opinione. Alla prima lettura di un testo non abbiamo nessuna motivazione dunque per riporre una qualche fiducia in quanto leggiamo. Se non conosciamo l'autore stabiliamo un legame fiduciario con il testo grazie alle caratteristiche di quest'ultimo e dunque grazie ai rimandi testuali ed intertestuali che ci possono fare evincere il reale valore del testo; grazie ad una continua non-contraddittorietà. Stabiliamo qui anche una seconda importante base del nostro ragionamento: non possiamo costruire un legame fiduciario con tutti i testi che leggiamo perché questo sarebbe uno spreco di risorse intellettive immane. In base a questa seconda considerazione si spiega il nascere di fonti ritenute affidabili e che, in qualche modo, garantiscono esse stesse per l'affidabilità dei documenti prodotti. In questa categoria possiamo inserire tutti coloro che producono informazione tradizionale; destinata ad essere creduta vera in principio proprio per il ruolo istituzionale che questi organismi hanno assunto. In questo secondo caso c'è solo un controllo a posteriori del testo prodotto e ogni volta l'affidabilità che si era presupposta viene confermata o smentita.

⁹ <http://www.simonemorgagni.it/Blog/Affidabilità%200.1.pdf>

Quella appena proposta è una differenza fondamentale tra vecchi e nuovi media, è la scomparsa del ruolo dell'editore nel mondo della rete e la conseguente crisi di fiducia che ne deriva; chi ci garantisce la verità e il valore di un testo in queste condizioni? L'unica soluzione che per ora pare presentarsi come valida è che sia la stessa rete a creare e garantire attraverso un'opera costante di feedback il valore dei testi che vi vengono pubblicati. Necessità nata proprio perché ogni testo al principio non garantisce valori che, nel bene e nel male, sono ormai standard nel mondo mediatico tradizionale.

Partendo dal presupposto che il blog, come ogni altro documento in linea sia un testo a tutti gli effetti possiamo cercare di identificare quali siano i fattori su cui essi maggiormente si basano rispetto ai testi informativi tradizionali. Quali sono quindi le differenze sostanziali tra un post ed un articolo cartaceo trattanti lo stesso argomento?

All'interno della nostra ottica questi fattori sono due che chiameremo fattore **Intertestualità pura** e fattore **Tempo**, volendo intendere con il primo il rimando costante che il testo fa alla rete e la rete fa al testo (molto maggiore che nei media tradizionali) e con il secondo lo sviluppo che il blog ha avuto e la storia che è stato capace di creare¹⁰.

Potremmo quindi cercare di stabilire dei criteri base di affidabilità attraverso l'analisi comparata di questi due fattori.

Il nostro lavoro è partito con la proposta di quattro criteri base che possano essere la base per un nuovo tipo di determinazione dell'affidabilità capace di superare il concetto di ranking puro che viene utilizzato oggi.

Ricordiamo come le quattro caratteristiche sotto riportate debbano essere interagenti tra loro per poter dimostrare la loro utilità; non è possibile fare un calcolo efficace considerandone solo una parte.

- Inserire la connotazione temporale all'interno del calcolo dei link esterni che provengono al blog oggetto dell'analisi. Con questo intendiamo come un singolo link non vada contato, ma come sia più corretto contare in scala esponenziale il numero di rimandi che giungono al nostro blog da una pagina esterna nell'unità di tempo. Un link ripetuto puntualmente all'interno di un periodo dato (ad esempio un mese) ha sicuramente un valore diverso da un

¹⁰ Facciamo notare come nei testi prodotti dai media tradizionali questi due fattori siano secondari, perché i messaggi prodotti paiono essere sempre scorporati dal media in sé e dal contesto, come fuori dal tempo e secondariamente perché i media tradizionali, per ovvi motivi economici tendono a non fare rimandi a concorrenti diretti.

link effettuato una volta soltanto a scopo di critica. Proponiamo questa prima modifica perché oltre una soglia minima all'interno della quale ognuno di noi cita una fonte da lui non condivisa al solo scopo di criticarla, si ha tendenza a continuare a citare solamente le fonti ritenute più interessanti ed attendibili. Proporre quindi una equazione di calcolo che aggiunga il fattore temporale avrebbe effetti simili all'esempio che ora proponiamo.

All'interno di questa discussione noi, come credo tutti gli altri blogger, abbiamo citato il rapporto Casaleggio per criticarlo, aumentandone al contempo Technorati e Google Rank, ma se il calcolo fosse stato ripetuto e avesse avuto come base i 30 giorni, non avremmo dato segno di considerare nemmeno il rapporto; non gli avremmo trasferito affidabilità. Proporre quindi la base mensile e il numero di link al blog superiori ad 1, pena l'esclusione dal conteggio, ci permetterebbe di raggiungere una realtà molto più vicina al pensiero degli autori dei blog riguardo il giudizio espresso sul testo linkato. Abbiamo poi proposto la scala esponenziale con esclusione del primo collegamento perché se una fonte viene da noi citata con allarmante regolarità significa che la riteniamo fondamentalmente affidabile e avremmo risultati non dissimili dalla simulazione effettuata qui sotto (esclusivamente esemplificativa):

1 link in 30 giorni (nessuna modifica del ranking)

2 link in 30 giorni (valore 1 del ranking)

5 link in 30 giorni (valore 10 del ranking)

10 link in 30 giorni (valore 50 del ranking)

30 link in 30 giorni (valore 150 del ranking)

- Stabilire un criterio di importanza relativa tra i differenti tipi di link. I link di un blog sono infatti di quattro tipi diversi: link da blogroll, link in corpo di post, link di commento e link di scambio. Proporrei di escludere dal conto tutti i link di scambio più o meno obbligatorio creati da molti utenti e da molte piattaforme, perché essi sono link che non possono e non devono in alcun modo incidere sul nostro calcolo, perché non farebbero altro che falsarlo. le categorie dovrebbero essere poste in ordine di importanza decrescente:

link di blogroll (significa che l'autore del blog consiglia esplicitamente la lettura del testo linkato, dandogli un imprimatur chiaro ed evidente, quasi a prescindere dai contenuti che, di volta in volta possono essere esposti).

link in corpo di testo (significa che l'autore del blog consiglia la lettura del testo linkato con preciso riferimento ad un post).

link di commento (il link presente in ogni commento manifesta la capacità dell'autore a confrontarsi con il resto della blogosfera attraverso strumenti collettivi, adeguati al mezzo e nel rispetto della netiquette).

Questo accorgimento ci permetterebbe di dare valore diverso a link che in effetti hanno un diverso valore, mostrando tre capacità diverse richieste ad un blogger, l'affidabilità a lungo termine, quella e breve e la capacità di confronto. Inoltre questo metodo ridurrebbe l'errore relativo prodotto nel tempo dagli attuali metodi di rating, mostrando maggiormente l'effettiva volontà dei singoli autori nel donare parte della propria acquisita affidabilità ai collegamenti effettuati.

Ricordiamo sempre come il punto uno sia sempre compresente e quindi come il calcolo sia complesso, inserito in un contesto temporale e con l'esclusione del primo link creato.

- Riduzione immediata e netta di tutti i link automatici prodotti dalle reti e dagli utenti non rispondenti ad effettivi criteri di valore e per questo esclusi dal punto precedente. Tutti i link che sono il prodotto obbligatorio di molte aggregazioni, tutti i link che le piattaforme obbligano ad avere, sono destinati a falsare il rating e quindi devono essere, per quanto possibile eliminati, in quanto non risultanti da un'effettiva volontà dell'autore o risultanti da una sua volontà che nulla ha a che vedere con l'affidabilità dei testi proposti.

Questo è il punto di più difficile realizzazione, come d'altronde possiamo vedere anche nella classica difficoltà di stabilire l'importanza dei testi accademici per via delle amicizie e delle citazioni fasulle, ma resta un obiettivo da perseguire senza sosta fino a dove possibile.

- Inserire un meccanismo di democratizzazione del rating che impedisca che le fonti maggiori oscurino completamente quelle più giovani storicamente. Un secondo elemento temporale va inserito per evitare che i blog più frequentati e quindi dotati di maggiori link e commenti possano oscurare completamente i blog nati da un tempo minore, ma non per questo meno importanti in termini di propositività o di capacità di innovazione. Sarebbe opportuno

quindi poter considerare nel calcolo anche un rapporto tra visite e link riportati per i blog nati entro un determinato lasso di tempo (ad esempio gli ultimi sei mesi). Un numero alto di link guadagnati in rapporto alle poche visite, unito ai tre punti precedenti, potrebbe lasciare spazio aperto verso i piani alti della classifica anche ai nuovi utenti nel caso essi siano particolarmente dotati ed apprezzati. Questo perché, anche in rete, non sempre l'esperienza è sinonimo di competenza.

In breve la nostra proposta non fa altro che apportare modifiche all'attuale concetto di ranking, diminuendo l'importanza in valore assoluto del numero di link creati e trasferendo una parte del controllo a valori soggettivi, ma stabiliti a priori e comuni per tutta la blogosfera. Un grande numero di link non sarebbe più garanzia di pertinenza, un buon numero di link, ripetuti nel tempo e posizionati nel posto giusto invece farebbe la differenza. Dividendo inoltre il calcolo su quattro fattori invece che su uno solo l'errore relativo verrebbe chiaramente a ridursi e le risposte che ci verrebbero proposte sarebbero più rispondenti ai criteri della nostra ricerca. Partire dall'attuale schema di calcolo di Google (rating & parole chiave, i tags) e modificandolo con questi accorgimenti potrebbe avere un risultato più rispondente al criterio di affidabilità della fonte che cerchiamo.

6. **Influenza**

All'interno della divisione del termine originario *attendibilità*, l'influenza di un blog altro non è che il risultato potenziale che un post di un determinato autore potrebbe avere sull'insieme della rete. Quante modifiche potrebbe apportare un singolo post? E in base a quali criteri avvengono queste modifiche?

In linea di massima l'influenza di un blog dipende più dalla rete di rapporti che l'autore è stato in grado di creare nel corso del tempo che dal reale valore dei contenuti che potrebbe proporre. Un blogger dotato di una solida rete di relazioni sarà sicuramente in grado di garantire ai propri post una risonanza maggiore all'interno della blogosfera e, ovviamente, avrà più possibilità di scatenare discussioni e apportare modifiche sostanziali all'ambiente. Come proponevamo qualche giorno addietro¹¹, riprendiamo l'esempio del post di Grillo che ha scatenato quest'ultima discussione. A prescindere dal reale valore dei contenuti espressi Grillo ha ottenuto più link e commenti, più audience se vogliamo, di tutto il resto della

¹¹ <http://fasttrack.splinder.com/post/6848153>

discussione; Grillo ha un'influenza maggiore di tutti i blogger che in seguito hanno trattato il problema. Per dare un'idea indicativa dell'influenza di un blog proponiamo di calcolare il numero di blogger attivi che seguono un determinato autore. Se io avessi 10.000 lettori giornalieri di cui uno solo attivo nel riproporre sul suo blog un ragionamento o un post derivanti dal mio, avrei un'influenza minore rispetto ad un altro blogger che potesse avere solo 1.000 lettori, ma tutti desiderosi di riprendere e rilanciare i suoi post sulle loro pagine personali. Questo avviene perché la rete opera in regime di syndication e questa agisce su fattori esponenziali di riproduzione; un numero maggiore di post, anche in pagine minori, garantisce al 99,9% un numero maggiore di contatti rispetto ad un numero minore di post su pagine di media grandezza. Cercando di esprimere in maniera empirica questo valore, potremmo tentare di incrociare due valori distinti: il numero di collegamenti in corpo di testo verso un determinato blog e l'eventuale presenza all'interno dei commenti nel blog d'origine del creatore del link. Si tratterebbe quindi di cercare un mio link nei commenti di Grillo e poi cercare un mio post in cui lo riprendo e lo linko; se ciò avviene più volte nell'unità di tempo, significa che io sono un lettore attivo del suo blog e gli garantisco quindi anche tutto il mio pubblico. Maggiore è il numero di lettori attivi, maggiore sarà l'influenza del blog in questione

7. **Visibilità**

L'ultimo parametro che deriva dall'esplosione del termine *Attendibilità* è secondo noi la visibilità di un blog all'interno della blogosfera. Termine non troppo distante dal precedente di *Influenza*, di cui è in realtà l'attualizzazione della potenza.

Con visibilità intendiamo cercare di identificare quale parte di blogosfera possa virtualmente giungere a conoscenza dei post pubblicati su un determinato blog. La visibilità è un termine estremamente aleatorio e per questo potrebbe risultare utile calcolarla sui singoli post come somma delle visite ottenute e come somma delle visite ottenute in seconda battuta dai blog che hanno linkato un determinato post. La somma di questi due valori ci dà il valore attuale delle persone che hanno seguito un determinato ragionamento. Una volta ottenuti questi valori si potrebbe metterli in relazione con il numero di post pubblicati. Otterremmo un grafico con il numero di visitatori per rapporto al post e avremmo anche un'idea dello sviluppo del blog nel corso del tempo. Ad un aumento di visitatori e a un aumento di visitatori per argomento (incrociando i risultati con le tags pubblicate e confrontandoli con quelli del passato) corrisponderebbe un secondo parametro atto a garantire l'affidabilità

della fonte, ad una diminuzione potrebbe stare una non affidabilità su certi argomenti o nel peggiore dei casi, una perdita di reputazione e di valore dell'intero blog.

8. Conclusioni: valutazioni incrociate

Si tratta in linea di massima quindi di mettere in rapporto tra loro il numero maggiore di variabili possibile al fine di ottenere i migliori risultati in termini di chiarezza e perspicuità.

Tutto ciò è sempre più indispensabile nell'ottica di un web sempre più fuori dal controllo del singolo, sempre più ricco di documenti che tendono a perdere la loro rilevanza in mezzo alla massa di rumore che viene giornalmente prodotto.

Se i link da soli evidentemente non bastano, potrebbe essere il caso di inserire nel calcolo del rating il feedback che la rete sa dare, le correzioni che essa sa apportare, gestendosi in maniera autonoma.

Un primo passo verso il web semantico?

Bibliografia

Bertrand Denis (2000) *Précis de sémiotique littéraire*, Nathan, Paris

Brin Sergey & Page Larry (1998) *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, in <http://www-db.stanford.edu/~backrub/google.htm>

Cameron Marlow (2004) *Audience, Structure and Authority in the Weblog community*, in <http://web.media.mit.edu/%7Ecameron/cv/pubs/04-01.pdf>

Eco Umberto (1975) *Trattato di semiotica generale*, Bompiani, Milano

Eco Umberto (1979) *Lector in fabula*, Bompiani, Milano

Eco Umberto (1990) *I limiti dell'interpretazione*, Bompiani, Milano

Eco Umberto (1994) *Sei passeggiate nei boschi narrativi*, Bompiani, Milano

Lotman, Jurij (1985) *La semiosfera*, Marsilio, Venezia

Lotman, Jurij (1993) *Kul'tura i Vzryv, Gnosis*, Moskva (trad. It. (1993) *La cultura e l'esplosione. Prevedibilità e imprevedibilità*, Feltrinelli, Milano

Lotman, Jurij (1998) *Il girotondo delle muse. Saggi sulla semiotica delle arti e della rappresentazione*, Moretti & Vitali, Bergamo

Van House Nancy (2004) *Weblogs: Credibility and Collaboration in an Online World*, in <http://www.sims.berkeley.edu/%7Eevanhouse/Van%2520House%2520trust%2520workshop.pdf>