

## Castello

### Come e' strutturata la Rete?

La Rete e' come ce la immaginiamo. Ne' potrebbe essere altrimenti essendo la Rete una serie infinita di interpretazioni di oggetti e simboli, materiali e immateriali, collegati fra loro nei modi piu' bizzarri e complessi.

Per Marcos7 "la rete siamo noi tutti che parliamo e ascoltiamo" - ed e' sicuramente una definizione su cui riflettere-, per molti utenti storici della telematica la Rete sono le aree di discussione delle BBS 8 o al massimo i famigerati newsgroup9. Per la maggior parte delle persone la Rete e' il Web10, per altre ancora la posta elettronica. Per chi scrive e' un'interconnessione rizomatica, un qualcosa che non si fa a tempo a descrivere che e' gia' cambiato profondamente come un sistema biologico in un percorso co-evolutivo11.

Abbandonando il terreno delle interpretazioni e dell'immaginario collettivo, la Rete e' innanzi tutto un numero enorme (milioni ma forse anche miliardi) di computer di vario tipo che costantemente, oppure occasionalmente, approfittano di un vettore fisico che puo' essere il comune cavo telefonico, ottico ecc., ma anche l'aria stessa - attraverso cui i radio modem 12 e ponti satellitari si interconnettono per comunicare fra di loro.

Ovviamente non tutti i computer hanno le stesse funzioni, finalita' e possibilita' comunicative. La distinzione principale in una comunicazione di tipo client-server<sup>13</sup> come avviene in Internet e' fra i computer che ospitano i servizi, ad esempio quelli che ospitano i nomi dei domini 14 - e che un recente studio dell'Universita' di Berkely in California stima intorno alle 13.000.000 di macchine - e i computer che ne usufruiscono. Questi sono i computer che possono leggere le pagine web appartenenti a tali domini: un numero elevatissimo e in crescita esponenziale, impossibile da determinare scientificamente!

Una volta fotografata l'attuale situazione di una miriade di computer, tra fornitori di servizi e utenti collegati fra loro fisicamente perlopiu' attraverso una ragnatela di cavi ma anche attraverso l'etere, rimane da sapere come avviene la comunicazione fra macchine. Le regole, in telematica si chiamano protocolli, per cui queste macchine comunicano fra di loro sono costituite dal famoso insieme di protocolli TCP-IP (Trasmission Control

---

Protocol / Internet Protocol). Il TCP/IP permette attualmente a una vera e propria infinita' di macchine di attivare sessione di comunicazione nei modi piu' disparati fra i quali, i piu' noti, sono il Web, la posta elettronica e i newsgroup, e l'IRC e l'FTP che rappresentano rispettivamente le modalita' storiche di chat e trasferimento di file in Rete, cosi' come il Telnet (accesso diretto a computer remoti), senza tralasciare le moderne possibilita' di attivare sessioni audio e video attraverso nuovi algoritmi di compressione come l'MP3 oppure tramite programmi proprietari come il REAL-AUDIO-VIDEO.

Il protocollo di comunicazione attuale IPv4 permette gia' delle funzioni mirabolanti che hanno fatto la fortuna di Internet assicurando la possibilita' di comunicare in tempo reale fra soggetti molto distanti fra loro; ma la sua futura implementazione gia' annunciata da tempo - IPv6 - consentira' un aumento considerevole delle macchine che potenzialmente possono accedere alla Rete insieme a un'ottimizzazione e controllo del traffico della Rete stessa<sup>15</sup>. Tutto cio' si attuera' attraverso una traduzione alfabetica del numero di riconoscimento della macchina in Rete, detto appunto numero IP, delegato a una rete di server speciali noti come DNS (Domain Name Server), e ad un'organizzazione gerarchica della Rete. Una macchina allora accede alla Rete solo attraverso un'altra macchina che e' gia' presente in Rete e che ha gia' acquisito il credito e la configurazione necessaria per essere visibile al resto della Rete. Un network, quindi, fondato su una costruzione piramidale e verticistica, apparentemente tutt'altro che orizzontale e anarchica.

---

Mettere in comune l'informazione

Se per accedere alla Rete bisogna passare attraverso un sistema di accreditamento basato su una gerarchia pluri-piramidale i cui vertici sono praticamente irrintracciabili data la complessità della struttura della Rete e, peraltro difficilmente gestibili, è pur vero che un governo della Rete esiste. Anche se ha delle caratteristiche molto particolari per certi versi di carattere comunitario.

Il carattere di condivisibilità della Rete non è per niente alieno alle sue istituzioni centrali 16 tanto che le regole tecniche di governo della Rete si sviluppano attraverso documenti di proposta detti RFC (Request For Comment) che sono pubblicamente valutati, discussi e nel caso modificati da chiunque si presti al gioco di testare le nuove ipotesi di governo tecnico e di sottoporsi a sua volta al pubblico esame 17. Un network gerarchico ma fondato su una logica "democratica".

---

Per capire come si comunica in Rete ovvero come si mette in comune un'informazione e' importante indagare sulle regole basilari della trasmissione bidirezionale di dati che si svolge, si' attraverso la ragnatela dei cavi, ma anche tramite il sistema ipertestuale dei link.

La ragnatela dei cavi comporta sempre esigenze di banda, in altri termini, di velocita' di connessione (quantita' di dati trasmessi per unita' di tempo) richieste dalle diverse sessioni di comunicazione sulla Rete.

Entrando sempre e comunque come ospiti di macchine gerarchicamente piu' importanti o piu' accreditate di noi, ereditiamo dall'host la nostra legittima capacita' di banda che dovra' bastarci per effettuare - in tempi funzionali ai nostri scopi - comunicazioni di varia natura: dalla navigazione su web alla posta elettronica fino alle videoconferenze. Ogni sessione di comunicazione ha una sua particolare esigenza di banda e quindi, mentre per Telnet, posta elettronica, newsgroup ed IRC sono sufficienti poche risorse, per navigare efficacemente su Web dobbiamo disporre di una buona velocita' di connessione, e per scaricare file tramite FTP, ascoltare un brano o partecipare a una sessione video necessitiamo di una quantita' di banda sempre crescente. Operando in Internet, non e' tuttavia necessario conoscere solo quali sono le nostre risorse di banda ma anche quelle dei nostri interlocutori: essendo la Rete un ambito di scambio comunicativo bidirezionale, pur avendo eventualmente la fortuna di trovarci sotto la confortante presenza di una linea dedicata universitaria, dovremo preoccuparci delle possibilita' tecniche del nostro

---

interlocutore. La reciprocita' di condizioni e' irrinunciabile: se non abbiamo la "banda" necessaria per attivare l'auspicata videoconferenza, forse dovremo piu' realisticamente ripiegare nel piu' rustico ma sempre intrigante dialogo in IRC...

La condivisione dei saperi richiede una "reciprocita' tecnologica" nell'accedere al medium.

Cosa centra tuttavia questo preambolo con l'operativita' dei motori di ricerca?

La struttura della Rete, le sue caratteristiche tecniche e le sue modalita' di interconnessione si riflettono proprio sulle capacita' operative dei motori di ricerca, essendo anch'essi, sotto forma di collettivi redazionali o di spider automatici, degli eterni e insaziabili browser in cerca di informazione in giro per il web. I motori di ricerca risentono infatti non tanto di (improbabili) loro limiti di connettivita' ma della scarsa larghezza di banda - nota anche come bandwidth - della fonte di informazione da indicizzare, tanto che un sito risulta piu' difficilmente indicizzabile ai motori quando e' ospitato su un server con poca banda.

I motori di ricerca

---

I motori di ricerca<sup>20</sup> sono delle complesse macchine basate sia su risorse umane sia su automatismi che hanno il compito di scandagliare un pezzo prescelto della Rete, indicizzarlo e metterlo a disposizione di ipotetici utenti interessati a interrogarlo. Le prestazioni e la geografia della Rete incidono senza ombra di dubbio sulla funzionalità dei motori di ricerca.

Se la velocità di connessione fra le macchine comunicanti in Rete ha un'importanza notevole nell'economia della comunicazione in Internet, e' la geografia dei link a giocare un ruolo di primo piano nel tentare di dare forma e senso alla Rete.

Nel 1999 Albert Re'ka, Hawoong Jeong e Albert-Lazlo' Barabasi pubblicano su Nature<sup>21</sup> uno studio per cui la distanza tra due pagine web qualsiasi era mediamente compresa in un percorso di soli 20 link. Come al solito le medie pero' nascondono una situazione molto sofisticata...

Uno degli studi piu' esaurienti e aggiornati sulla topografia della Rete e' stato svolto dall'IBM Almaden Research Center in collaborazione con il Systems Research Center e la societa' Altavista di San Matteo (California) e presentato all'IX International World Wide Web Conference ~ <http://www.almaden.ibm.com/cs/k53/www9.final> svoltasi ad Amsterdam nel maggio 2000.

Secondo questo studio la Rete si presenta con un corpo centrale di 56 milioni di pagine web strettamente interconnesse ovvero pagine che sono

---

molto linkate le une con le altre seppure secondo criteri e percorsi diversi e di diversa lunghezza. Da questo corpo centrale si staccano due protuberanze di altri 44 milioni di pagine: una protuberanza e' costituita da pagine personali e siti minori i quali sono linkati al corpo centrale - costituito da portali, motori di ricerca, siti d'informazione e pagine delle aziende piu' rinomate - ma che non godono del link di ritorno; l'altra protuberanza e' costituita invece dalla maggioranza delle imprese, delle universita', dei centri di ricerca e di altre istituzioni che beneficiano di link dal corpo centrale ma non contraccambiano con il link di ritorno. Collegamenti di leggera entita' connettono poi le due protuberanze minori, mentre sono numerose le pagine completamente isolate dalla Rete.

Come incide tutto cio' sull'operativita' dei motori di ricerca ?

Moltissimo, ma soprattutto per due aspetti fondamentali: il reperimento dell'informazione e il ranking.

Tanto per cominciare un'informazione per essere presa in considerazione da un motore di ricerca deve essere prima reperita. Nel mare magnum valutato genericamente sulle centinaia di milioni di pagine web, reperire le informazioni diventa un work in progress estremamente complesso. Ad esempio c'e' molta differenza fra un'informazione contenuta in una pagina che e' linkata da molte altre pagine e che facilmente prima o poi intercettera' il laborioso navigare dello spider in questo oceano digitale, e una pagina che non e' linkata da nessun'altra parte. Le pagine linkate infatti hanno il

---

pregio di essere considerate pagine già note o più interessanti e aggiornate della Rete, al contrario di quelle isolate la cui unica speranza resta quella di farsi notare dal motore di ricerca o per un caso fortuito o grazie a una segnalazione extra-ipertestuale che si spera sia presa nella dovuta considerazione.

In mezzo a questi due estremi c'è una scala di importanza della costruzione dei link e delle profondità dei siti <sup>22</sup> che incide, per le ragioni ora espresse, sulla possibilità o meno per una pagina web di essere trovata da un motore di ricerca.

Un altro importante aspetto è rappresentato dal criterio utilizzato dai motori di ricerca per decidere come stilare la classifica dei risultati conseguiti dopo un'interrogazione: il ranking. È considerato un buon risultato (top ranking) un risultato che abbia speranza di visibilità sull'utente in cerca di informazioni. Apparire almeno fra i primi trenta risultati di una ricerca effettuata su un motore è ormai l'aspirazione di ogni sito in cerca di visibilità'.

Al di là delle ipotesi di censura e privilegio di strategie di marketing - non facilmente praticabili ma sicuramente ipotizzabili anche in una classifica come quella della maggior parte dei motori di ricerca basati su automatismi<sup>23</sup> - bisogna per ora mettere in evidenza come la quantità di link che ha un sito abbia influenza sulla possibilità che ha lo stesso sito di scalare le classifiche dei motori di ricerca e apparire fra le

---

sempre piu' ambite trenta posizioni sia da grandi che piccole aziende.

Un'informazione contenuta in una pagina web ha buone probabilita' di essere reperita da un motore di ricerca nell'attuale geografia di Internet, solo se e' posizionata in un server che ha buone prestazioni in termini di larghezza di banda. Ma soprattutto l'informazione godera' di una buona ricaduta di attenzione sul medium se la pagina web e' linkata da molte altre pagine cosicche' non solo e' facilmente reperibile dallo spider, ma essendo considerata "popolare" nel microcosmo delle consultazioni digitali 24, ottiene buone probabilita' di essere inserita nel top ranking dei motori di ricerca.

Il castello dell'informazione...

L'informazione in Rete, seppur malleabile e volatile come un bit digitale, puo' avere caratteristiche fra le piu' differenti. La pagina Web costruita attraverso un linguaggio a marcatura fra quelli accettati dalla filosofia della Rete come il comunissimo HTML fa da padrona fra i possibili file presenti in Rete, e nell'immaginario collettivo la si identifica con la Rete stessa. Tuttavia in Internet sono numerosissimi gli oggetti presenti: pagine testo in formato txt; miriadi di immagini in formato .gif .jpg o

---

magari .png; database dai formati piu' diversi; programmi software (quindi di nuovo file binari); video e audio, i cui formati sono costantemente in evoluzione alla ricerca di quello che meglio si presta ad ottimizzare la qualita' e la velocita' di trasmissione in rete di oggetti di informazione cosi' ricca e cosi' complessa come puo' esserlo un evento audio-visivo.

Il problema che incontrano quindi i motori di ricerca nel catalogare informazione trattata ed elaborata nei formati fra i piu' diversi, e' semplicemente immenso.

E' perlopiu' risolto dal punto di vista dell'interfaccia verso l'utente creando appositi servizi: Altavista ha una sezione specifica <http://image.altavista.com> per le immagini, ad esempio; mentre Lycos ha una corrispondente <http://mp3.lycos.com>, per i file audio di maggiore successo in questo momento. Esistono comunque motori di ricerca specializzati nella ricerca di oggetti specifici 25.

Dal punto di vista, invece, del reperimento dell'informazione, oltre a adottare risorse umane di tipo redazionale e dell'impiego di automatismi 26, alcuni motori di ricerca si avvalgono di algoritmi di analisi dell'immagine in grado di riconoscere segnali, testo e didascalie nascoste nel codice 27. Si tratta di informazioni normalmente usate per la sincronizzazione e la descrizione, per esempio, nelle trasmissioni satellitari oppure per finalita' di titolazione e ricerca interna dai software per i quali poi sono destinati ad essere utilizzati 28.

---

Una volta risolto il problema di catalogare informazione di ogni genere compresa quella sotto forma di codice binario, per l'informazione classica (file di testo, html e comunque decodificabile correttamente<sup>29</sup>) il problema non si pone in termini qualitativi, bensì quantitativi. Accenniamo di nuovo al problema di localizzare l'informazione nella Rete ma anche a quello di acquisire e catalogare razionalmente ed efficacemente l'informazione dopo averla localizzata. Il primo aspetto è stato "risolto" saltando alla ovvia conclusione che è praticamente impossibile localizzare tutta l'informazione presente sulla Rete; si è quindi in un certo senso deciso di lasciare emergere dalla trama magmatica e ridondante della Rete una sorta di classifiche compilate in bell'ordine nello stile pop di un hit parade. Anche la soluzione della seconda problematica risulta tuttavia improbabile e non offre garanzie.

Al di là della bontà e dell'efficacia degli algoritmi di indicizzazione usati dai motori, rimane infatti irrisolta la questione per cui quasi tutti gli spider dei motori di ricerca acquisiscono solo la parte iniziale dei siti che visitano, senza scandagliarne la profondità e i contenuti reali.

Possiamo quindi affermare, trist\* e sconsolat\* che i motori di ricerca che credevamo infallibili ed esaustivi sono in realtà dei laboriosi ma grossolani procacciatori di informazioni, incapaci di districare - senza censure, insabbiature o macchinosi artifici - la trama caotica e ridondante della Rete, questo sofisticato meccanismo-simbolo della complessità

---

postmoderna in cui siamo immersi.

7 Dalla Seconda Dichiarazione della Realidad per l'umanita' e contro il neoliberalismo, Una rete intercontinentale di resistenza per l'umanita', Dalle Montagne del Sudest messicano, Comitato Clandestino Rivoluzionario Indigeno Comando generale dell'EZLN La Realidad, Pianeta Terra, Agosto 1996.

8 BBS sta per Bulletin Board Systems e sono sistemi di comunicazione telematici amatoriali con caratteristiche tecniche particolari e non omogenee fra di loro. Oramai molto rari perlopiu' si riconoscono per la loro attitudine libertaria, localistica e per essere in rete fra loro non in tempo reale.

9 Si sa: quello che non si conosce fa paura ed e' per questo forse che i

---

newsgroups hanno una così brutta fama (tra le persone che ne ignorano la logica). Simili alle aree di discussione delle bbs i newsgroups rappresentano invece per molte persone esperte della Rete la matrice, ovvero la parte viva di Internet, il flusso incontrollato di informazioni e dibattiti dove misurare lo stato di salute della Rete.

10 Il Web o World Wide Web (W3) è la rappresentazione multimediale della Rete attraverso il protocollo http. È lo strumento più immediato ma anche quello meno interattivo per rapportarsi a Internet.

11 Tommaso Tozzi, Cotropia, lifeware e coevoluzione mutuale, La Stanza Rossa, Bologna 1997.

12 Modem o modulatore-demodulatore è uno dei tanti apparecchi di uso più o meno comune ormai che permettono ai nostri pc di collegarsi a Internet attraverso un processo di trasformazione del segnale che permetta di attraversare il vettore scelto (sia esso appunto un cavo oppure l'aria)

13 Si intende per comunicazione client-server un particolare modo di comunicazione nel quale vi sono dei soggetti che fanno da fornitori di servizi ed altri da fruitori di tali servizi. Avviene anche nelle comunicazioni ben note come quella che consiste nello sfogliare delle pagine web dove il nostro programma browser (come communicator di netscape, explorer, opera ecc.) è il fruitore (client) e il computer che ospita le pagine il fornitore (che ha quindi installato un server web adatto allo

---

scopo)

14 Si intende comunemente per dominio il suffisso di primo livello che appare all'estremità destra di un nominativo Internet e che qualifica la nazionalità (ad es. it per Italia) o il tipo di servizio (ad esempio org per organizzazione no-profit) di provenienza.

15 L'avvento dell'Ipv6 per la ragione che tenterà di controllare e ottimizzare il traffico di rete ha suscitato qualche perplessità fra gli amanti della privacy in quanto sono ipotizzabili anche delle procedure che identificano facilmente il mittente della comunicazione, al momento molto difficoltoso con il protocollo in uso.

16 Per un approfondimento del funzionamento e dell'organizzazione della Rete sia in termini tecnici che burocratici si veda anche Luciano Paccagnella, *La comunicazione al computer*, Bologna Il Mulino, 2000, 201 pp.

17 Malgrado queste forme di autogoverno e il senso di libertà che è avvertibile appena si accede alla Rete molte persone e collettivi si sono preoccupati nel corso degli anni di sviluppare o ipotizzare reti autonome che avessero ancor meno strumenti di controllo di quelli presenti in Internet anche per meglio sfuggire a quei casi di censura dell'informazione come quello rappresentato nel 1998 dalla momentanea chiusura del server di movimento di isole nella Rete ~ [www.ecn.org](http://www.ecn.org) che ha fatto discutere della Rete e delle sue potenzialità (anche) repressive mezzo mondo. E', per

---

esempio, opinione comune che prendendo il controllo dei principali nodi di router della rete (non piu' di una cinquantina a livello mondiale) possono essere intercettate praticamente tutte le informazioni della Rete! "Il progetto GAIA ~ [www.ecn.org/loa/terminal/GAIA](http://www.ecn.org/loa/terminal/GAIA) e' nato per risolvere un problema frequente nelle reti ad organizzazione gerarchica: la dipendenza da chi detiene un livello di accesso privilegiato alla struttura piramidale". Paul Garrin, un media artisti di New York ha progettato un network alternativo di DNS in modo che cambiando il settaggio della propria macchina sulle coordinate di uno dei server di Garrin si possa cambiare le regole alfabetiche della Rete. E' di Tommaso Tozzi la proposta di un'INTERFACCIA FLUTTUALE che raccolga al suo interno l'indirizzo di tutti i server nel mondo senza nessun criterio di scelta o censura e con la possibilita' per chiunque di aggiungervi l'indirizzo del proprio server. La proposta piu' forte rimane comunque quella denominata Freenet ~ <http://freenet.sourceforge.net/> consistente in una rete di macchine che pur "appoggiandosi" ad Internet adottano un proprio protocollo di comunicazione che le permette di ricercare, immettere e scambiarsi informazione all'interno di questa Rete senza che ci siano possibilita' tecniche di censura o di rintracciamento dell'autore dell'informazione ovviamente avvalendosi di quello che e' indubbiamente una delle risorse piu' interessanti della telematica attuale ovvero della crittografia cosiddetta pesante.

18 L'etimo di comunicare secondo alcuni filologi deriva dal latino: cum (insieme) e munus (dono). La comunicazione cioe' come scambio paritario e

---

condivisione di saperi.

19 I links sono quelle porzioni di testo riscontrabili in Rete dette zone calde su cui e' possibile semplicemente clickare sopra con il mouse per passare in tempo (quasi) reale da una zona di un documento a un'altra oppure da un sistema informativo a un altro indipendentemente dalla loro localizzazione geografica.

20 Ovviamente sara' oggetto di discussione la struttura e il funzionamento di un motore di ricerca in una sezione apposita

21 Fa ancora piu' scalpore la notizia, riportata sempre su Nature, della scarsa possibilita' dei motori di ricerca di indicizzare la Rete indicando nella cifra ipotetica del 14% il record di Northern Light ~ [www.northernlight.com](http://www.northernlight.com) di indicizzazione della Rete. E il restante 86% (e' venuto da chiedere ai piu'...) ?!?

22 Dicasi profondita' di un sito la quantita' di differenti livelli di link che esso contiene.

23 Fa parte della nostra proposta artistica di aggirare alcuni limiti dei motori di ricerca l'approfondimento di questo argomento che verra' riportato nell'ultima sezione di questo approfondimento.

24 L'indice di "popolarita'" (ovvero di pagine che linkano al sito preso in

---

esame) e' uno dei parametri piu' importanti adottati dai motori di ricerca per considerare di qualche interesse di indicizzazione un sito. Un altro parametro, anch'esso oramai molto conosciuto, e' costituito dall'aggiornamento della pagina: piu' una pagina viene rinnovata in grafica ma soprattutto in contenuto d'informazione e piu' il motore e' invogliato a mantenerla nel proprio elenco di siti da visitare e con frequenza di visita proporzionale alla frequenza di rinnovamento del sito registrato.

25 DEJA ~ [www.dejac.om](http://www.dejac.om) e' il famoso motore di ricerca per i newsgroups mentre TUCOWS ~ [www.tucows.com](http://www.tucows.com) e' la famosa directory per risorse software. Ancora FTP SEARCH ~ [ftpsearch.ntnu.no/ftpsearch](http://ftpsearch.ntnu.no/ftpsearch) interagisce con molte archivi files in FTP e I-SURF ~ <http://isurf.interpix.com> e' specializzato nel recupero di immagini.

26 E' un aspetto sul quale torneremo piu' volte comunque i motori di ricerca sono un risultato variamente bilanciato e derivante sia da un lavoro di tipo redazionale di gruppi di professionisti il cui lavoro e' quello di cercare e filtrare l'informazione sia dallo stesso tipo di lavoro pero' svolto da automatismi di vario tipo.

27 Un esempio di Audio e VideoLogger di nuova generazione e' costituito dalla tecnologia sviluppata da [www.virage.com](http://www.virage.com) ed utilizzata, per esempio, anche da Altavista.

28 Nel caso di un file MP3 viene estrapolata l'informazione normalmente

---

usata dal lettore hardware MP3 oppure dal software di lettura presente sul computer per dare un titolo al brano e verificarne l'autore e la lunghezza.

29 Mentre per un'informazione di tipo testo e' possibile un'indicizzazione che si avvale di dizionari comprendenti vocaboli conosciuti, per un file di tipo binario risulta molto piu' difficile decodificare dal codice binario l'informazione che l'autore ha voluto trasmettere magari sotto forma di prestazione di un software o sequenze di immagini. (Vedi sezione sull'accessibilita' dell'informazione).

30 Una spider e' un programma del motore di ricerca che viene lanciato nella Rete alla ricerca di informazione: lo si puo' immaginare come un browser molto semplice e veloce incaricato di navigare e registrare grosse quantita' di pagine web secondo rotte preordinate dal programmatore che lo ha lanciato.

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