

CRONOLOGIA¹

1945

Il mensile americano « The Atlantic Monthly » pubblica “As We May Think”, in cui Vannevar Bush parla del Memex, progetto di una macchina individuale per archiviare e personalizzare la conoscenza.

1957

L'URSS lancia lo Sputnik, primo satellite artificiale terrestre. In risposta, l'anno seguente gli USA creano l'Advanced Research Projects Agency (ARPA) all'interno del Ministero della Difesa, con obiettivi di ricerca nel campo scientifico-tecnologico.

1963

Douglas Engelbart, ricercatore allo Stanford Research Institute, realizza “oNLine System” (NLS), un sistema che permette la navigazione ipertestuale. Inventa il mouse per questo proposito.

1965

L'ARPA finanzia uno studio su una "rete collaborativa di computer time-sharing".
Ted Nelson conia il termine *Hypertext*.

1969

Nasce il primo nodo Arpanet alla UCLA University, in seguito collegato ad altri tre centri universitari tra cui lo Stanford Research Institute (SRI) dove lavorava Douglas Engelbart.

¹ Nel redigere la cronologia ho utilizzato le seguenti fonti: “Hobbes' Internet Timeline v5.4” <http://info.isoc.org/guest/zakon/Internet/History/HIT.html>;
“A little history of the World Wide Web” <http://www.w3.org/History.html>;
“History of Internet and the Web” <http://www.anderbergfamily.net/ant>.

1971

Ray Tomlinson della BBN (Bolt Beranek and Newman, Inc.) scrive un programma e-mail per spedire messaggi su reti distribuite, scegliendo il simbolo @ per il significato di “presso”.

1972

Viene formato l'International Network Working Group (INWG), in cui emerge l'esigenza di uno sforzo congiunto per migliorare le tecnologie di rete. Primo Presidente viene nominato Vinton Cerf, che svilupperà due anni dopo il protocollo TCP (che diventerà poi TCP/IP). Vengono intanto messi a punto il File Transfer Protocol (FTP) e il Telnet.

1979

Viene creata la rete USENET utilizzando l'UUCP (Unix-to-Unix CoPy) fra la Duke Univ. e UNC.
Richard Bartle and Roy Trubshaw della Univ. of Essex creano il primo MUD (Multi-User Dimension).

1980

Tim Berners-Lee, consulente del CERN, scrive Enquire, un programma che permette di creare collegamenti (*link*) tra nodi arbitrari all'interno di documenti.

1981

Ted Nelson lancia “Xanadu”, un sistema a pagamento per la consultazione centralizzata degli ipertesti.
Viene creata la rete BITNET, detta anche "Because It's Time NETwork".

1982

La rete Arpanet si scinde in Arpanet e Milnet, connettendosi ad altre reti. La nuova rete viene chiamata Internet e si basa sul protocollo TCP/IP.

1984

Viene introdotto il Domain Name System (DNS).
Il numero degli host supera le 1.000 unità.
William Gibson conia il termine *cyberspace* nel suo romanzo “Neuromancer”.

1988

Il virus “Internet worm” viene rilasciato, colpendo circa 6.000 dei 60.000 host su Internet.
Jarkko Oikarinen inventa l'Internet Relay Chat (IRC).

1989

La prima proposta di un progetto web per favorire la condivisione delle informazioni tra i ricercatori è avanzata da Tim Berners-Lee al CERN di Ginevra.

1990

ARPANET cessa di esistere.
Tim Berners-Lee mette a punto HTML (HyperText Markup Language), HTTP (HyperText Transfer Protocol) e URL (Uniform Resource Locator).

1991

La Wide Area Information Servers (WAIS) viene creata da Brewster Kahle.
Gopher viene rilasciato da Paul Lindner e Mark P. McCahill della University of Minnesota.
Il primo browser testuale viene rilasciato da Tim Berners-Lee su alt.hypertext.
Philip Zimmerman inventa PGP (Pretty Good Privacy).
Linus Torvald rilascia la versione 0.02 del sistema operativo Linux.

1992

Viene istituita la Internet Society (ISOC).
Il numero di nodi supera il milione.

1993

La NCSA rilascia la prima versione di “Mosaic”, un browser creato da Marc Andressen che in seguito lascerà l’istituto e fonderà insieme a James Clark la “Netscape Communication Corp.”.

1994

Si tiene al CERN di Ginevra la prima “International WWW Conference”.

Tim Berners-Lee e Al Vezza sono a capo del W3C (World-Wide Web Consortium), creato dal CERN e dalla NCSA.

1995

La Sun lancia Java il 23 maggio.

RealAudio, una tecnologia di audio streaming, consente agli utenti della Rete di sentire flussi audio quasi in tempo reale.

A marzo, il WWW sorpassa l'FTP come servizio con maggior traffico su internet per numero di pacchetti e di byte.

1996

La discussa proposta di legge americana Communications Decency Act (CDA) diventa legge degli Stati Uniti e proibisce la distribuzione di materiale indecente su Internet. Alcuni mesi dopo una corte di tre giudici ne contesta la sua applicazione. Nel 1997 la Corte Suprema all'unanimità dichiara che la maggior parte di questa legge è incostituzionale.

1997

La Network Solutions registra il milionesimo nome di dominio Internet.

1998

XML (eXtensible Markup Language) diventa raccomandazione proposta del W3C.

Il software *open source* raggiunge la notorietà.

1999

Il 18 giugno, in coincidenza con il Summit del G8, un gruppo di attivisti in diverse parti del web prende di mira i centri finanziari mondiali.

Viene rilasciata la tecnologia di compressione del suono nota come MP3

Il virus “Melissa” infetta milioni di computer.

2000

All'inizio di febbraio viene sferrato un massiccio attacco (denial of service) contro i maggiori siti web, fra cui Yahoo, Amazon, e eBay.

Secondo le stime del NEC-RI e di Inktomi, in Internet vi sono più di 1 miliardo di pagine.

Nasce Napster, un sistema di condivisione di canzoni codificate in MP3 .

Fa scalpore il virus “Love Letter”.

2001

Napster si trova sempre più coinvolta nella causa legale che alla fine la vede costretta a sospendere la propria attività, trasformandosi in un servizio a pagamento.

Il worm “Code Red” e il virus “Sircam” entrano in migliaia di server e indirizzi email, causando un picco vertiginoso di utilizzo della banda Internet, e intrusioni nei sistemi di sicurezza.

SMIL 2.0 (Synchronized Integration Multimedia Language) diventa una raccomandazione del W3C.

1 ASPETTI STORICI DEL WEB

1.1 INTERNET E LE RETI

Per risalire all'origine delle reti di comunicazione bisogna tornare indietro all'inizio degli anni sessanta. Cominciano infatti in quegli anni negli Stati Uniti le prime ricerche sulle *reti a commutazione di pacchetto*². L'opinione pubblica americana era rimasta alquanto scossa dal lancio da parte dell'USSR del primo satellite geostazionario, lo Sputnik I, nel 1957. Uno degli effetti immediati fu la creazione dell'Advanced Research Projects Agency (ARPA), presso il Ministero della Difesa, con compiti di ricerca nel campo della scienza e della tecnologia. Nel 1962 l'ARPA dá inizio ad un programma di ricerca sui computer e nomina uno scienziato del MIT, John Licklider, alla guida del progetto. Licklider aveva appena pubblicato un memorandum sull'idea di un "Galactic Network", una visione futuristica di una rete di computer collegati fra loro e accessibile a tutti. All'interno dell'ARPA un altro scienziato, Leonard Klienrock, stava già sviluppando un sistema che permettesse lo scambio di informazioni molto vicino all'idea della *commutazione di pacchetto*.

² La rete a commutazione di pacchetto (packet switching) si basa su una tecnologia che permette la frammentazione dei messaggi in sotto-parti chiamate pacchetti che vengono instradati in modo indipendente e riassemblati una volta arrivati a destinazione.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

Negli anni seguenti la ricerca si era sviluppata a tal punto³ da permettere al nuovo direttore, Leonard Roberts, di pubblicare un progetto per una rete di computer chiamata ARPANET, i cui primi nodi sono installati nel 1969 rispettivamente alla UCLA e allo Stanford Institute of Research. Nell'ottobre del 1972 durante la prima "International Conference on Computer and Communication" gli scienziati dell'ARPA danno una dimostrazione operativa della rete ARPANET, collegando fra loro 40 postazioni remote e suscitando l'interesse della comunità scientifica. Il sistema fu migliorato successivamente con l'introduzione di un programma per gestire l'invio di messaggi di posta personali (e-mail), e con lo sviluppo di un protocollo *host-to-host*⁴, regolando le modalità dello scambio tramite il Transmission Control Protocol/Internet Protocol (TCP/IP). Lo sviluppo di questo protocollo è un passo cruciale nello sviluppo della rete, anche perché viene ribadita l'intenzione di configurare il sistema come un'*architettura aperta*⁵, attenendosi ai principi della decentralizzazione (ogni network sarebbe stato in grado di funzionare autonomamente), della ridondanza dell'informazione e della sua libera circolazione.

³ Erano impegnati nella ricerca indipendentemente e ignari l'uno del lavoro dell'altro il MIT, il National Physics Laboratory (UK) e la RAND Corporation. Venuti a conoscenza dell'idea di ARPANET, parteciparono insieme allo sviluppo di un protocollo per lo scambio di informazioni, conosciuto come IMP (*Interface Message Processor*).

⁴ All'inizio il sistema consentiva l'accesso ai diversi nodi solamente da terminali remoti. La nuova tecnologia permette l'accesso e l'effettivo scambio di file tra due computer, per la durata della connessione.

⁵ Per una definizione di architettura aperta si può fare riferimento a quella data dall'Internet Society, www.isoc.org

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

Intanto altre reti di computer si stavano sviluppando separatamente. Nel 1976 i laboratori AT&T Bell mettono a punto lo UUCP (Unix-to-Unix Copy Protocol), che venne distribuito alla comunità di utenti Unix e utilizzato nel 1979 dalla rete USENET per la trasmissione di e-mail e news. Nel 1981 la City University of New York crea la rete BITNET, per favorire la cooperazione scientifica, e lo stesso ruolo è svolto sulla costa occidentale dalla rete CSnet. In questo periodo si registra una situazione abbastanza caotica, che vede competere una serie di tecnologie e protocolli per ottenere la supremazia. ARPANET è ancora la dorsale dell'intero sistema e, quando nel 1982 adotta il TCP/IP come standard per connettersi alle altre reti, si può parlare per la prima volta di Internet.

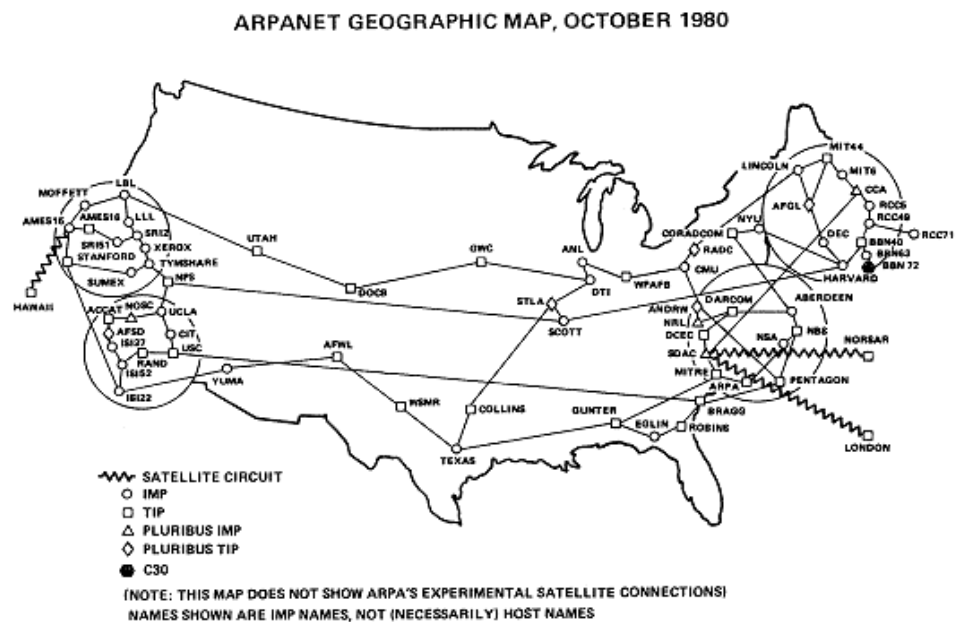


Figura 1.1: Geografia della rete Arpanet nel 1980

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

Fino a questo punto il processo di sviluppo della rete era stato guidato per la quasi totalità dalla comunità scientifica, e il raggio d'azione era volutamente limitato ad alcuni centri universitari.

Le scoperte nel campo dei computer e delle telecomunicazioni facevano però presagire una possibile espansione del sistema. Nel 1984-85 vengono introdotte le reti JANET e NSFNET rispettivamente in Gran Bretagna e negli Stati Uniti, con l'obiettivo di connettere indistintamente tutti i centri di educazione superiore nei rispettivi paesi. Negli stessi anni vede la luce Fidonet, un *Bullettin Board System*⁶ (BBS), che utilizza l'architettura Dos. NSFNET viene scelta come dorsale del servizio Internet negli Stati Uniti in sostituzione di ARPANET, che cesserà di esistere nel 1990.

Nonostante l'ingresso delle realtà commerciali, la crescente espansione della rete continuò ad essere guidata dalle comunità accademiche. Nel 1989 il numero di host supera per la prima volta la soglia dei 100.000, per diventare 300.000 l'anno seguente. La fine degli anni 80 e l'inizio degli anni 90 costituiscono un punto di svolta per diverse ragioni oltre alla già citata cessazione di ARPANET. Nel 1990 alla McGill University di Montreal viene sviluppato il primo motore di ricerca su Internet, Archie; nel 1991 la NSF apre l'accesso alla dorsale al traffico privato; nello stesso anno il World Wide Web viene presentato al pubblico.

⁶ Una BBS è una bacheca elettronica, uno strumento amatoriale per scambiare messaggi e files. Una definizione di Bulletin Board System e più in particolare di Fidonet è disponibile nel libro "Spaghetti Hacker", Apogeo 1997

1.2 IPERTESTI

Parallelamente alla ricerca e alle applicazioni nel campo delle reti e delle infrastrutture di comunicazione altri intellettuali sono impegnati a risolvere il problema dell'organizzazione e del sovraccarico delle informazioni.

Vannevar Bush, personaggio eclettico⁷, è uno dei fondatori del concetto di ipertestualità. Fin dall'inizio degli anni trenta comincia a rendersi conto di come l'eccesso di informazioni rischi di diventare una minaccia al pensiero creativo. Bush si pone il problema specifico della letteratura scientifica, si rende conto che un sistema di catalogazione del sapere deve partire da qui per essere applicato alla società più in generale. La soluzione di questo problema particolare sta nella "meccanizzazione della conoscenza scritta". La sua proposta diventa pubblica nel 1945 in un articolo pubblicato nella rivista "The Atlantic Monthly" intitolato "As we may think", dove Bush descrive in questo modo la sua macchina, il Memex: "*A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory*".

Si tratta dell'idea di un dispositivo per uso individuale, fondato su un principio associativo per cui vengono stabilite connessioni tra i

⁷ Bush fu tra le altre cose consulente scientifico di Roosevelt, coordinatore del progetto Manhattan e docente al MIT, come riportato nel libro "Gli spazi dell'ipertesto", Bettetini, Gasparni, Vittadini, Bompiani 1999

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

documenti, basato sull'analogia tra il cervello umano e la macchina⁸. Bush non è l'unico a fare queste osservazioni. Licklider nel 1960 sostiene che *“La speranza è che, tra non troppi anni, i cervelli umani e i calcolatori siano uniti molto strettamente e che la risultante associazione penserà come il cervello umano non ha mai pensato ed elaborerà i dati in un modo al quale le macchine per l'elaborazione delle informazioni che conosciamo oggi non si avvicinano nemmeno lontanamente”*⁹.

Douglas Englebart, un ricercatore della Stanford University, dimostra nel 1968 durante un convegno a San Francisco la possibilità di uno spazio di lavoro in collaborazione chiamato NLS (oN Line System). Utilizzando un sistema a finestre e muovendosi all'interno di esso con un rudimentale mouse da lui stesso costruito, fa vedere per la prima volta l'uso della posta elettronica e dei collegamenti ipertestuali¹⁰.

L'obiettivo della ricerca di Englebart era: *“usare il meglio di ciò che la tecnologia può offrire per fornire agli esseri umani una maggiore capacità di manipolazione dei simboli, e quindi di esplorare le possibilità risultanti per riprogettare la loro struttura di concetti e metodi per renderli notevolmente più efficienti nella soluzione di problemi concreti”*, così come espresso nel saggio del 1962 *“Augmenting Human Intellect: a Conceptual Framework”*. Il punto decisivo risiede forse nella convinzione di

⁸ Un gruppo di scienziati del MIT era impegnato nella costruzione di modelli computazionali che prendevano a prestito caratteristiche delle reti neurali del cervello. Tra questi vi era Norbert Wiener, padre della cibernetica.

⁹ Come riportato da Pedemonte in *“Personal media”*, Bollati Boringhieri 1998

¹⁰ Per capire l'importanza del momento basti pensare che il microprocessore non era ancora stato inventato, che i calcolatori dell'epoca erano ingombranti armadi pieni di circuiti integrati a basso livello di integrazione e che l'interazione con i computer avveniva attraverso schede perforate.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

Englebart circa la necessità di trasformare l'immagine del mondo in concetti e simboli che possano essere manipolati fisicamente sullo schermo del computer, intervenendo sul "linguaggio". E' importante sottolineare che fin dall'inizio la sua ricerca è volta a progettare strumenti e tecniche immaginando sistemi che possano lavorare in rete. Grazie agli investimenti dell'ARPA, proprio in quegli anni si sta sviluppando la tecnica del *time-sharing*¹¹, e quando nel 1969 vengono collegati i centri di ricerca americani, creando ARPANET, il laboratorio di Englebart è il secondo a collegarsi al network nascente.



Figura 1.2: Douglas Englebart con il suo gruppo di lavoro

La struttura ipertestuale utilizzata nel NLS era stata teorizzata alcuni anni prima da Ted Nelson, che coniò il termine Hypertext in una pubblicazione inviata ad una conferenza indetta dalla Association for Computing Machinery (ACM) nel 1965: *“Intendo per scrittura non sequenziale un testo che si dirama e che consente al lettore di*

¹¹ La tecnica del time sharing fa sì che un computer potente e di grosse dimensioni (detto “mainframe”) dedichi, in modo sequenziale, una frazione di tempo ad ogni utente; in questo modo ogni lavoratore collegato può andare avanti nel lavoro senza mettersi in coda per attendere il suo turno, come era tradizione.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

scegliere; è una cosa che si fruisce meglio sullo schermo interattivo. (...) Un ipertesto è una serie di grani di testo tra cui sono creati collegamenti che consentono al lettore diversi cammini.”

A differenza di Bush e Englebart, che erano ricercatori legati alla cultura accademica, Nelson è un personaggio piuttosto eccentrico rispetto al contesto della ricerca scientifica, esponente del movimento underground e *hacker*¹² di quegli anni. I suoi due libri, “Computer Lib” e “Literary Machines” sono i manifesti underground della nuova cultura informatica. Nelson in seguito estende il concetto di *Hypertext* a quello più generale di *Hypermedia*¹³, che nel 1967 sfocerà nel progetto della sua vita, *Xanadu*.

Xanadu è l’idea di uno spazio virtuale in cui far confluire tutta l’informazione possibile, dove ciascuno può pubblicare qualunque cosa, collegando il suo testo a qualunque altro documento attraverso collegamenti ipertestuali. Nelson pensa ad un sistema centralizzato ma aperto a tutti, con un sistema automatico per la distribuzione delle royalty. Più in generale si tratta del sogno di una società utopica in cui le persone comunicano su basi egualitarie. Il gruppo *Xanadu* viene formato nel 1979 ma, pur ottenendo alcuni finanziamenti, non riesce a commercializzare alcun prodotto e viene chiuso definitivamente nel 1992.

¹² Si tratta di una cultura tecnica e anticonformista, che aveva un rapporto di puro divertimento con i software. Una descrizione della cultura hacker è contenuta nel “Jargon File”, <http://www.tuxedo.org/~esr/jargon/>.

¹³ Per *Hypermedia* si intende un’estensione dell’ambiente ipertestuale che supporta i contenuti multimediali.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

Una caratteristica accomuna Bush, Englebart e Nelson: l'essere tutti in grande anticipo sui tempi. Bush pensa al Memex quando la tecnologia digitale era agli albori, Englebart parla di finestre e di mouse 15 anni prima della comparsa del primo personal computer¹⁴, Nelson immagina le sue "macchine letterarie" quando il primo software ipertestuale¹⁵ viene commercializzato nel 1987. Questo è il principale motivo per cui nessuna delle innovazioni da loro proposte ha mai visto la luce. Rimane comunque inestimabile il loro ruolo come figure intellettuali e pionieri le cui idee affascineranno intere generazioni.

¹⁴ Il primo personal computer è l'Apple di Steve Jobs e Steve Wozniak, e compare sul mercato nel 1979.

¹⁵ Si tratta di Hypertext, che girava sul Macintosh.

1.3 IL WORLD WIDE WEB

Un gran numero di scienziati informatici d'oltreoceano stava tenendo d'occhio da un paio di decenni la ricerca sugli ipertesti e sui computer network. Ma la tecnologia che consentirà di riunire questi due elementi, il World Wide Web, è stata messa a punto da un giovane ricercatore del CERN di Ginevra (il centro di ricerca europeo sulla fisica delle particelle), Tim Berners-Lee.

Tim Berners-Lee entra al CERN nel 1980 come consulente del sistema informatico per un nuovo acceleratore di particelle. Si trova subito di fronte alla difficoltà di comunicazione fra i vari dipartimenti del centro, dovuta alla presenza di sistemi hardware e software tra loro spesso incompatibili. Nel tempo libero scrive "Enquire", un programma che consente di stabilire link tra due nodi e lo usa per associare i vari moduli software alle persone che lo utilizzano. Nel 1989 fa circolare una proposta ("Information Management: A Proposal") nella quale discute un sistema per facilitare lo scambio di informazioni.

Tim Berners-Lee racconta così i passi che lo hanno portato a formularla : *"Capivo benissimo che al CERN avevamo bisogno di un programma di ricerca tipo Enquire. [...] Questa meta sembrò irraggiungibile fino a quando non compresi che la diversità dei vari sistemi informatici e reti poteva essere una ricchezza, qualcosa da sottolineare, non un problema da*

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

cancellare. Il modello che scelsi per questo sistema minimalista fu l'ipertesto. [...] Il sistema doveva possedere un'altra proprietà fondamentale, cioè l'assoluto decentramento. Soltanto così si poteva iniziare ad usarlo senza chiedere l'accesso a nessuno. E soltanto così il sistema poteva salire di livello, senza ingolfarsi man mano che aumentava la gente che lo usava. [...] La soluzione consisteva nel comunicare direttamente in rete. Internet era già in funzione negli anni settanta, ma trasferire informazioni restava troppo complesso per quanti non fossero esperti di informatica. [...] Poi hanno inventato la posta elettronica. L'e-mail permette di spedire messaggi da una persona all'altra, tuttavia non creava uno spazio in cui l'informazione potesse esistere in permanenza e dove si poteva andare a cercarla. I messaggi erano volatili. Il World Wide Web, sovrapponendosi a Internet avrebbe appunto regalato all'informazione questo spazio.¹⁶

La stessa proposta ricircola l'anno seguente, mentre Tim Berners-Lee inizia a lavorare su un web client, ossia un programma che avrebbe permesso la creazione, la navigazione e l'editing di ipertesti, utilizzando il sistema di sviluppo Nextstep¹⁷. Il progetto prende il nome di World Wide Web e consiste in un primo momento nella scrittura del codice per il protocollo HTTP (HyperText Transfer Protocol) che regola il trasferimento dei dati, e nella definizione di URI (Uniform Resource Identifiers), lo schema degli indirizzi dei vari documenti. Successivamente viene definito l'HTML (HyperText Markup Language) e messo a punto

¹⁶ Tim Berners-Lee "Weaving the Web", tradotto in italiano con il titolo di "le Architetture del nuovo Web", Feltrinelli 2001

¹⁷ Il personal computer Next, ideato da Steve Jobs, era una macchina innovativa, dotata di interfaccia grafica, posta elettronica vocale, sintetizzatore e, appunto, un editor di ipertesti.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

il primo web server, il software che conserva le pagine web sul proprio spazio di memoria permettendo l'accesso ad altri. Il primo server gira sul computer di Tim Berners-Lee e prende il nome di info.cern.ch. Una stagista, Nicola Pellow, comincia a lavorare su di un browser a linea di comando, mentre Tim si occupa di rendere compatibili con il navigatore web i vari newsgroup presenti su Internet, che utilizzavano FTP e non HTTP come protocollo. Nell'attesa di una risposta dal CERN decide di rendere disponibile su Internet il suo lavoro, pubblicandolo su alt.hypertext e altri newsgroup e aprendo un server telnet¹⁸ pubblico su info.cern.ch, suscitando l'interesse della comunità dell'iperstesto e di quella Next.

```
From: timbl@info .cern.ch (Tim Berners-Lee)
Newsgroups: alt.hypertext
Subject: Re: Qualifiers on Hypertext links...
Message-ID: <6484@cernvax.cern.ch>
Date: 6 Aug 91 14:56:20 GMT
References: <1991Aug2.115241@ardor.enet.dec.com>
Sender: news@cernvax.cern.ch
Lines: 52
[...]

The WorldWideWeb (WWW) project aims to allow links to
be made to any
information anywhere. The address format includes an
access method
(=namespace), and for most name spaces a hostname and
some sort of path.
```

Figura 1.3: l'e-mail con cui Tim Berners-Lee annuncia la nascita del World Wide Web

¹⁸ Telnet è un protocollo che permette di aprire una sessione interattiva a linea di comando su un computer remoto.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

Assieme a Robert Cailliau, un altro ricercatore, partecipa nel 1991 ad un convegno a San Antonio, Hypertext '91, dove adattando il modem svizzero alla rete statunitense e effettuando un collegamento ad internet tramite il numero di un'università vicina dimostrano per la prima volta il Web in funzione.

Da questo momento il Web inizia una lenta ma costante crescita, ma la carenza di browser per macchine che non siano il Next ne limita fortemente la diffusione. Nella prima metà del 1992 compaiono quasi contemporaneamente due browser testuali per Unix, che era il sistema operativo utilizzato da gran parte dei ricercatori e dalle università. Si tratta di Erwise, sviluppato al Politecnico di Helsinki in seguito ad una visita di Tim Berners-Lee stesso, e di Viola, scritto da Pei Wei, studente alla Berkeley University. Robert Cailliau nel frattempo mette a punto un navigatore anche per il Macintosh, Samba.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

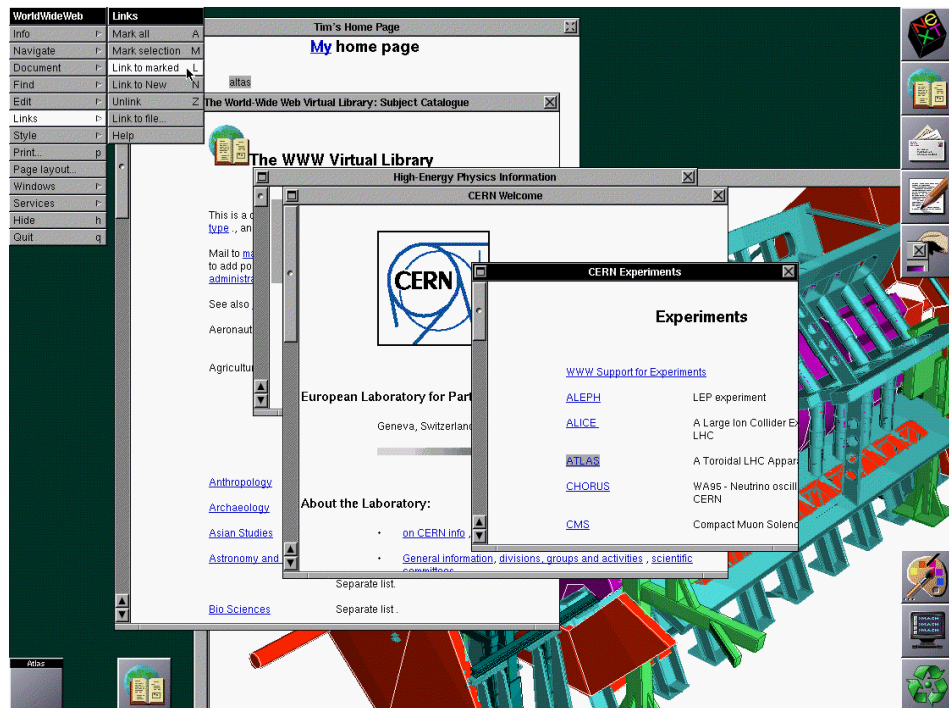


Figura 1.4: una schermata di come il World Wide Web appariva nel 1994

La progettazione di browser era diventata per tanti studenti e tecnici un ottimo pretesto per dimostrare la loro abilità di programmatori. All'interno dello NCSA (National Centre Supercomputer Applications) Marc Andressen, uno studente, e Eric Bina, membro dello staff, iniziano a sviluppare un browser per Unix in grado di girare nell'ambiente grafico X-Window. Eric scrive il codice mentre Marc garantisce una presenza costante all'interno dei gruppi di discussione sul Web, ponendosi in questo modo in assoluto contrasto con il resto degli studenti programmatori. Andressen non era interessato solamente a far funzionare il programma bensì a far utilizzare il suo navigatore a

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

più gente possibile. Il browser viene chiamato Mosaic e viene reso disponibile nel febbraio del 1993. E' facile da installare, consente di visualizzare le immagini e richiede poca preparazione prima di ottenere accesso al web tramite mouse; viene accettato rapidamente dalla comunità internet.

E' questo il periodo durante il quale il Web prende forma. Tim Bernes-Lee si prodiga cercando di orientare gli sviluppatori ad inserire nei loro browser, fra le altre cose, un editor per favorire una collaborazione attiva, ma senza ottenere risposte favorevoli. Preoccupato dal fatto che lo NCSA è impegnato a promuovere con tutte le sue forze Mosaic a discapito del Web, e constatato il rapido abbandono di Gopher dopo che l'Università del Minnesota aveva paventato la possibilità che diventasse un sistema a pagamento, chiede e ottiene dal CERN la diffusione di pubblico dominio della tecnologia del Web senza clausola alcuna.

Intanto il Web comincia a diventare terreno per operazioni commerciali. Marc Andressen, uscito dallo NCSA fonda insieme all'uomo d'affari Jim Clark la Mosaic Communication, che diventerà in seguito Netscape. Bill Gates annuncia in prima persona che la prossima versione del sistema operativo Microsoft, Windows 95, avrebbe compreso un programma per navigare in Internet.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

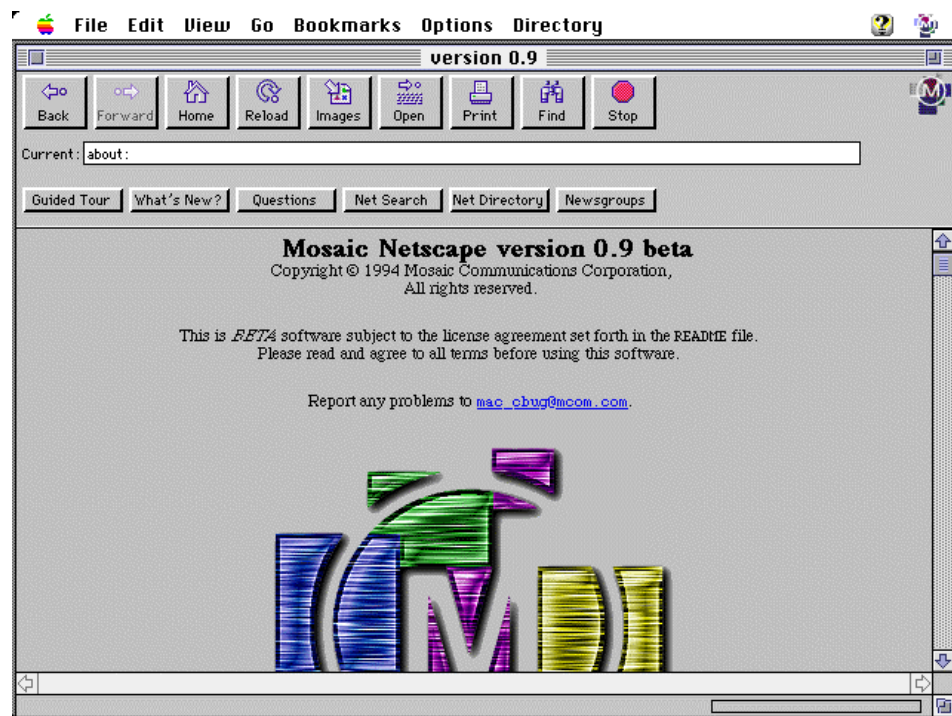


Figura 1.5: la schermata di apertura della versione 0.9 di Mosaic

La compresenza di queste situazioni convinse Tim Bernes-Lee che era necessaria una struttura di supervisione del Web, anche considerando la sua rapida crescita: “ *Non volevo fondare il solito organismo, ma una struttura che aiutasse i programmatori di server e browser a mettersi d'accordo su come doveva operare il web. Con Mosaic verso il traguardo e sempre più utenti gopher che prendevano in considerazione il web, era sempre più chiaro che la rete rischiava di dividersi in diverse fazioni, alcune commerciali, alcune accademiche, alcune gratis, altre no. In questo modo si sarebbe perso il vero intento del Web, essere un unico ipertesto per ottenere informazioni, universale e accessibile.*”

Il 25 maggio 1994 si tiene a Ginevra il primo convegno del World Wide Web, a cui partecipano più di trecento persone, catturando

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

l'attenzione dei giornalisti. Le basi per la formazione di un consorzio vengono gettate invece durante il WWW Wizard Workshop organizzato da Dale Dougherty della O'Really Associates, una casa editrice che si sta interessando al fenomeno. L'organizzazione prende il nome di World Wide Web Consortium, W3C in breve. L'iscrizione è aperta a tutte le organizzazioni commerciali o statali. Il fine del consorzio è "spingere il Web al suo pieno potenziale", soprattutto proponendo protocolli comuni di sviluppo per facilitarne l'interoperabilità e l'evoluzione. Quando viene raggiunto il consenso su una proposta essa diventa "raccomandazione formale." Il W3C ha una base negli Usa, il MIT di Boston, e una in Europa, l'INRIA di Parigi.

Nel maggio 1995 la Sun Microsystem presenta Java, un nuovo linguaggio di programmazione *object oriented*¹⁹ che permette a due computer con sistemi operativi differenti di usare lo stesso applicativo, un *applet java*²⁰, scaricandolo dalla rete. Intanto i provider di rete come Compuserve, America Online e altri che offrono contenuti preconfezionati come news, enciclopedie, e-mail diventano sempre più ansiosi e forniscono controvoiglia l'accesso al Web, che avevano sempre dipinto come una rete "altra" e complessa. La Netscape, capendo che doveva crescere in fretta per competere con colossi del calibro della Microsoft, decide

¹⁹ Un linguaggio object oriented si riferisce ad un particolare tipo di programmazione che combina la struttura dei dati a funzioni che operano su di essi, per creare parti di programma riutilizzabili.

²⁰ In realtà la tecnologia Java consente di operare anche lato server con le *servlet*.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

la quotazione dell'azienda in borsa, in modo da ottenere una grossa trasfusione di liquidi. Le azioni, collocate inizialmente a 28 dollari, un prezzo già alto, salgono rapidamente a 72 dollari. Alla chiusura delle contrattazioni finiscono sul mercato 38 milioni di azioni, e la Netscape si ritrova in un solo giorno con un capitale di 4,4 miliardi di dollari. Due settimane dopo Bill Gates lancia Windows 95 e il browser Microsoft, Internet Explorer 1.0.

Alla fine del 1996 milioni di persone accedono al Web, migliaia di aziende lo forniscono come server e la stampa ne scrive di continuo²¹. Il ruolo del consorzio e la libertà del Web vengono messe sotto esame quando si paventa una possibile reazione censoria del governo degli Stati Uniti di fronte ai timori crescenti sulla pornografia in rete. Il consorzio elabora una piattaforma, PICS (Platform for Internet Content Selection), contenuta in un programma che una volta installato avrebbe permesso ai genitori di bloccare l'accesso ai siti classificati in un certo modo, sulla base della valutazione di organismi diversi. La legge in questione viene comunque emanata nel 1996 con il nome di *Communication Decency Act*, parte della vasta normativa sulle telecomunicazioni per regolare i contenuti di Internet. Contestata da alcuni gruppi per i diritti civili, la legge viene alla fine dichiarata incostituzionale, anche grazie all'esistenza di PICS.

Sempre nel 1996 Netscape rilascia Navigator 2.0, aumentato con un sistema di posta elettronica di facile utilizzo e applicazioni Java.

²¹ Si calcola che nel 1994 vi fossero 3,2 milioni di host e 3.000 siti web. Il numero degli host raddoppia l'anno seguente mentre i siti sono 25.000. Alla fine del 1996 le stime danno 9,5 milioni di utenti e più di 250.000 siti web.

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

Alla metà del 1997 i siti Web contengono belle foto, animazioni, informazioni impaginate in tabelle, audio e moduli per l'interazione, il tutto tenuto insieme dalla struttura ipertestuale. Internet Explorer è ormai sulla via per conquistare il mercato dei browser, ma la Microsoft decide di dare una nuova sferzata con il nuovo sistema operativo, Windows '98. La nuova versione avrebbe infatti compreso un browser avanzato, Internet Explorer 4.0, divenuto parte integrante e inscindibile del sistema operativo. Questa particolarità suscita l'interesse del Ministero americano della Giustizia, che stava già indagando la Microsoft per violazione della normativa antitrust, e viene aperta un'inchiesta²². Nel gennaio 1998 Netscape fa una mossa a sorpresa che ripristina l'ethos primigenio di Internet: rende di pubblico dominio il codice sorgente di Navigator, permettendo a chiunque lo volesse di apportare modifiche²³.

I lavori del consorzio tengono un profilo molto più basso, anche se sempre in contatto con l'evoluzione tecnologica. Alla fine del 1998 il consorzio ha già sfornato una dozzina di Raccomandazioni. Uno dei più importanti progressi tecnici usciti dal consorzio è un linguaggio chiamato XML, da eXtensible Markup Language, che fornisce una base per definire linguaggi e presenta caratteristiche innovative rispetto a HTML.

²² Il processo contro l'azienda di Redmond diventerà un caso che si trascinerà fino ai giorni nostri, fra colpi di scena, condanne e successive assoluzioni.

²³ La politica *open source* prevede che un programma sia gratuito e modificabile apertamente. Anche il codice campione del W3C è open source. Per maggiori informazioni si può fare riferimento a www.opensource.org

1. 4 LE TECNOLOGIE DEL WORLD WIDE WEB

Il Web può essere visto come una singola applicazione che fornisce ai suoi utenti una interfaccia ipertestuale uniforme verso programmi e sistemi operativi differenti. Questa universalità è garantita essenzialmente da due caratteristiche. La prima, e probabilmente la più importante, è la definizione di uno schema

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

uniforme per la localizzazione dei dati su Internet. Combinando il nome di un documento, il nome dell'host sul quale è ospitato il server e il nome del protocollo usato per trasferire un file dal server al client si ottiene il concetto di URL.

La seconda caratteristica è l'utilizzo di una sintassi semplice per la formattazione dei documenti che contengono i link ipertestuali e gli altri elementi che costituiscono una pagina Web. La soluzione pensata da Tim Berners Lee è stata quella di proporre un linguaggio lineare con una sintassi derivata da SGML²⁴ (Standard Generalized Markup Language), chiamato HTML.

Per ovviare ai problemi di lentezza della rete²⁵ è stato inoltre sviluppato un protocollo minimale chiamato HTTP.

Un gran numero di nuovi protocolli ed estensioni sono stati proposti ma queste tre specifiche di base (URL, HTML, HTTP) sono rimaste rimarcabilmente stabili. Prima di analizzare più in profondità i vari aspetti del Semantic Web e considerando il fatto che si fonda a partire dal Web esistente, è perciò necessario discutere brevemente in merito a queste specifiche fondamentali.

²⁴ SGML è un linguaggio di formattazione molto complesso utilizzato prevalentemente nell'ambito della ricerca ipertestuale.

²⁵ Il protocollo di comunicazione usato maggiormente, FTP, era progettato per lunghe sessioni interattive con utenti umani ed era ragionevolmente lento. FTP deve infatti tenere traccia dello stato del programma client e attuare un controllo estensivo degli errori.



Figura 1.6: Tim Berners-Lee

1.4.1 URL (Uniform Resource Locator)

In base alla terminologia in uso nel Web, ogni componente che può essere indirizzato e a cui si può avere accesso come unità singola viene definito *resource* (*risorsa*). Esempi di risorse sono le pagine HTML, immagini, file, fogli di stile, applet, etc. Allo stato attuale, tutte le risorse presenti sul Web sono indirizzate mediante un Uniform Resource Locator (URL).

L'idea che sta dietro ad un URL è molto semplice, si può pensare ad esso come ad un'estensione su scala Internet del percorso (*pathname*) di un file su un computer in locale. Un URL consiste in una prima parte che specifica il protocollo utilizzato per accedere ad una risorsa, mentre la seconda parte ne definisce l'indirizzo. Il protocollo più comunemente utilizzato è di solito HTTP, e l'indirizzo indica il nome della macchina host e il percorso del file. I vari URL formano un sottoinsieme di una categoria più generale di indirizzi detti URI (Uniform Resource Identifier), e sono

organizzati gerarchicamente tramite il DNS (Domain Name System), il sistema utilizzato per nominare gli host. Essendo il collante che tiene insieme HTTP, HTML e molti altri protocolli Web, il concetto di URL è da considerarsi fondamentale nella infrastruttura del Web.

1.4.2 HTTP (HyperText Transfer Protocol)

Il protocollo di comunicazione usato tra i client e i server Web è l'HTTP (HyperText Transfer Protocol). Si tratta di un protocollo veloce, leggero, che non tiene traccia dello stato dell'informazione e privo di una funzione per il controllo degli errori, costruito sul livello immediatamente superiore a quello del TCP/IP. Un client HTTP semplicemente effettua una connessione al server e richiede un singolo documento. La connessione viene terminata quando il documento richiesto è stato ricevuto. La versione 1.1 del protocollo permette in realtà di effettuare più richieste contemporaneamente, questo per ovviare al fatto che documenti di grosse dimensioni o contenenti immagini implicavano tempi di attesa eccessivi. Il punto di forza di HTTP, oltre alla sua velocità, risiede nel fatto che fornisce un supporto per altri protocolli quali Gopher, WAIS e FTP, che erano molto diffusi prima dell'adozione del Web.

1.4.3 HTML (HyperText Markup Language)

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

HTML (HyperText Markup Language) è stato sviluppato originariamente come un semplice linguaggio di formattazione per i documenti presenti sul Web. E' un linguaggio che definisce formalmente come una pagina deve essere visualizzata in un browser. Contiene inoltre informazioni riguardanti il percorso di eventuali link presenti in un documento.

Tutte queste informazioni nel codice HTML sono contenute all'interno di *tag*, il più importante dei quali è sicuramente il tag `<a>`, che consente di stabilire collegamenti tra i documento. Un esempio di codice HTML potrebbe essere il seguente:

```
Autore:<a href:="http://www.sri.edu/doug.htm">Doug  
Englebart</a>
```

che consente di raggiungere la pagina “doug.htm” sul server identificato dal nome logico “http://www.sri.edu”.

Oltre alle informazioni sui link, HTML offre numerosi elementi utili alla formattazione del testo e della struttura di un documento. Mentre le versioni iniziali contenevano un numero sostanzialmente limitato di elementi, il loro numero è cresciuto considerevolmente con le versioni successive. L'aggiunta di nuovi *tag* ha creato difficoltà per quanto riguarda la compatibilità delle pagine web su browser diversi. Il problema fondamentale consisteva nel fatto che la formattazione non era indipendente dalla presentazione. Questo ha provocato le critiche della comunità SGML, critiche accolte dal W3C e che hanno portato alla stesura nel Dicembre 1997 della versione 4.0, che paradossalmente rappresenta un ritorno alle origini. Vengono

Dalla tesi sul Semantic Web di Daniele Salvatico, laureando in Scienze della Comunicazione, Torino, luglio 2002

infatti tenuti separati dal codice HTML tutti gli elementi orientati alla presentazione, che devono essere contenuti in un foglio di stile CSS²⁶ indipendente dal documento HTML.

²⁶ Un foglio di stile *Cascading Style Sheet* (CSS) definisce le caratteristiche relative alla visualizzazione di un documento.