

## Protocolli

### Internet: le basi

Una rete telematica è un insieme di nodi e di connessioni.

- I nodi sono unità di elaborazione. Con il termine unità di elaborazione intendiamo una vasta gamma di dispositivi tra i quali troviamo *personal computer*, *computer* portatili, palmari, telefoni cellulari, eccetera.
- Le connessioni di una rete telematica sono mezzi fisici tramite i quali i nodi possono comunicare ovvero scambiarsi informazione. Esempi di connessioni: cavi elettrici e telefonici, campi radio, fibre ottiche.

Le reti telematiche possono essere classificate in base alla loro dimensione in:

- Reti personali (*Personal Area Network* PAN). Reti che collegano un numero ridotto di unità. Esempio: una rete casalinga che collega due *personal computer* e una stampante.
- Reti locali (*Local area network* LAN). Sono reti di dimensioni ridotte. Esempio: rete di *personal computer* di un qualsiasi ufficio.
- Reti metropolitane (*Metropolitan Area Network* MAN). Sono reti progettate per coprire un'area geografica limitata. Esempio: rete di calcolatori del comune di Bologna.
- Reti geografiche (*Wide area network* WAN). Sono reti progettate per collegare siti o reti geograficamente distanti.
- Inter-reti. Sono collezioni di LAN collegate da WAN. Esempio: **Internet**.

**Internet** è quindi in prima approssimazione una rete di reti. Ovvero è una rete che collega altre reti. Le dimensioni di **Internet** sono enormi. Basti pensare che **Internet** è in grado di collegare milioni di calcolatori e di altri dispositivi dislocati sulla superficie del pianeta in più di cento paesi.

Date le dimensioni di **Internet** e la varietà dei dispositivi ad essa collegati il primo problema da risolvere per rendere possibile la comunicazione è quello di definire un linguaggio comune.

*Computer* diversi usano sistemi operativi, codici di caratteri, strutture di dati, che possono essere anche molto diversi. Per permettere loro di comunicare è necessario definire un insieme di regole comuni di comportamento. Queste regole nell'ambito della telematica prendono il nome di protocolli.

### TCP/IP

Un protocollo di comunicazione stabilisce regole comuni che un *computer* deve conoscere e alle quali deve attenersi per inviare e ricevere informazioni attraverso la rete. Un protocollo deve specificare ad esempio in che modo va codificato il segnale, in che modo far viaggiare i dati da un nodo all'altro, in che modo assicurarsi che la trasmissione sia andata a buon fine.

Nel caso di **Internet**, che interconnette milioni di *computer* e di sottoreti, basati su ambienti operativi e architetture *hardware* diverse, tali protocolli debbono rispondere ad esigenze particolarmente complesse.

L'insieme di protocolli su cui si basa il funzionamento di **Internet** si chiama *TCP/IP* che è un acronimo di *Transmission Control Protocol/Internet Protocol*. *TCP* e *IP* sono due tra i più importanti protocolli di *TCP/IP*, per questa ragione l'intera *suite* di protocolli *TCP/IP* ha questo nome.

*TCP/IP* è la chiave del successo di **Internet**.

*TCP/IP* è indipendente dal modo in cui la rete è fisicamente realizzata: una rete basata su *TCP/IP* può appoggiarsi indifferentemente su una rete locale *Ethernet*, su una linea telefonica, su un cavo in fibra ottica, su una rete di trasmissione satellitare. *TCP/IP* nasconde all'utente la realizzazione fisica della rete che sta usando.

*TCP/IP* è stato progettato per sfruttare al meglio le risorse di comunicazione disponibili; permette un indirizzamento efficiente e sicuro dei *computer* collegati, anche se questi sono diversi milioni; garantisce con la massima sicurezza il buon fine della comunicazione; infine permettere lo sviluppo di risorse e servizi di rete evoluti e facilmente utilizzabili dall'utente.

*TCP/IP* è un *open standard*, le cui specifiche sono liberamente utilizzabili da chiunque. Questo ha permesso il rapido diffondersi di implementazioni per ogni sistema operativo e piattaforma esistente, implementazioni spesso distribuite gratuitamente o integrate in modo nativo nel sistema stesso.

Ciò che viene comunemente indicato come *TCP/IP*, in realtà, è costituito da un vero e proprio insieme di protocolli di comunicazione, ognuno con un compito specifico e organizzati in maniera gerarchica per livelli.

- Livello delle applicazioni. Gestisce i servizi di rete per l'utente e dunque è la fonte e la destinazione finale di ogni transazione di rete.
- Livello di trasporto. Gestisce l'organizzazione dei dati ai fini della trasmissione e ha il compito di controllare che la comunicazione di un certo blocco di dati sia andata a buon fine, e di ritrasmettere quello che eventualmente è andato perso.
- Livello di rete. Gestisce l'indirizzamento dei *computer* e l'instradamento dei dati.
- Livello fisico e di collegamento dati. Gestiscono l'uso dei cavi e l'invio dei segnali fisici sulla rete.

In questo modulo ci concentreremo sull'insieme di protocolli di livello applicazione ovvero su quell'insieme di regole che permettono all'utente di usufruire di servizi telematici quali la posta elettronica e la navigazione del *Web*.

A questo punto possiamo dare una definizione di **Internet** più specifica e dettagliata di quella data in precedenza.

**Internet** è un sistema informativo globale:

- logicamente collegato da un unico spazio di indirizzi, cioè ogni macchina connessa a **Internet** è individuata logicamente da un indirizzo assegnato seguendo le specifiche del protocollo *IP* di *TCP/IP* (Indirizzo *IP*);
- che supporta comunicazioni usando i protocolli *TCP/IP*;
- che fornisce, usa e rende accessibile, in modo pubblico e/o privato, servizi ad alto livello (Esempio: posta elettronica, *chat*, *Web*) basati su *TCP/IP*.

Risulta evidente come *TCP/IP* sia l'elemento cardine per la definizione ed il funzionamento di **Internet**.

Qualunque dispositivo per essere connesso a **Internet** deve aver installato e correttamente configurato il *software* relativo ai (che implementa i) protocolli *TCP/IP*.

Nel seguito prenderemo in esame alcuni dei principali protocolli di livello applicazione nella gerarchia *TCP/IP* ed i servizi a loro associati.

Servizi di Internet - Telnet, Ftp, Gopher

**Telnet**, ftp e **Gopher** sono tre protocolli di livello applicazione nella gerarchia di protocolli *TCP/IP* (ricordiamo che *TCP/IP* è l'insieme dei protocolli che regolano il funzionamento di **Internet**).

**Telnet** e ftp conservano ancora una grande utilità nonostante il loro utilizzo sia stato ridimensionato nel corso degli ultimi anni. **Gopher** ha quasi totalmente perso la sua utilità dopo l'avvento del *Web*.

- **Ftp** è ancora oggi il metodo più efficace e veloce per trasferire *file* da un *computer* ad un altro.
- **Telnet** è ancora oggi utilizzato per accedere (da una qualunque macchina collegata a *Internet*) a numerosi *computer* che accettano solo connessioni in modalità terminale.
- **Gopher** è stato il primo tentativo di organizzare, attraverso un'unica interfaccia (a caratteri o testuale), l'enorme quantità di informazioni disponibili su **Internet**, poi superato dal *Web*.

Ftp

**Ftp** (acronimo di *File Transfer Protocol*) è il protocollo che permette di trasferire *file* da un *host* ad un altro attraverso **Internet**.

Memorizzati negli *hard disk* degli *host* connessi a **Internet** c'è una enorme quantità di programmi, immagini digitali, suoni, eccetera, molti dei quali di pubblico dominio.

Il sistema più efficace per trasferire questi *file* da un *computer* ad un altro è usare un'applicazione (*client*) basata sul protocollo FTP.

Le macchine che entrano in gioco in una sessione ftp sono due: la macchina locale (quella sulla nostra scrivania) e la macchina remota (quella sulla quale o dalla quale vogliamo trasferire *file*).

In una sessione ftp è possibile effettuare due operazioni principali, quella di *download* e quella di *upload*.

L'operazione di *download* è l'operazione con la quale si trasferiscono *file* dalla macchina remota alla macchina locale.

L'operazione di *upload* è l'operazione con la quale si trasferiscono *file* dalla macchina locale alla macchina remota.

Sia l'operazione di *download* che quella di *upload* possono essere eseguite solo sulle macchine *remote* che accettano sessioni ftp.

Non tutti i *file* (programmi, immagini, filmati, eccetera) memorizzati su *host* abilitati a sessioni ftp sono prelevabili o prelevabili gratuitamente. Spesso le università, i centri di ricerca, e anche i singoli utenti privati mettono a disposizione *file* di pubblico dominio e trasferibili con diverse modalità di pagamento.

Programmi *shareware*: è richiesto il pagamento di una modica somma di denaro.

Programmi *freeware*: *software* completamente gratuito.

Oltre al *software shareware* e *freeware* esistono altre categorie di *software*. Per esempio *software* di tipo *giftware* (richiede come pagamento un versamento volontario anche non necessariamente in denaro) e di tipo *cardware* (richiede come pagamento la spedizione di una cartolina!)

Per trasferire un *file* usando il protocollo ftp occorre avere a disposizione un'applicazione ftp chiamata *client* ftp (ce ne sono diverse anche di tipo *freeware*) che viene attivata sulla macchina locale e occorre conoscere l'indirizzo *IP* o il nome logico della macchina remota.

Ci sono due modalità di collegamento ad una macchina remota: ftp anonimo, e ftp con *account*.

Il trasferimento di *file* tramite ftp anonimo è quello tradizionalmente utilizzato per il prelievo di *file* ad accesso pubblico presso università, enti, società. Consiste in un *login*, ovvero nell'ingresso in un *computer* remoto, effettuato senza disporre presso di esso di un proprio codice utente e di una propria *password*, quindi anonimamente. In questa modalità non avremo, per ovvi motivi di sicurezza, pieno accesso al *computer* remoto; potremo quindi entrare solo in determinate *directory* - tipicamente nella *directory* chiamata *pub* (ovvero *public*) e nelle sue *sottodirectory* e potremo solo leggere alcuni *file*, ma non cancellarli, spostarli o modificarli.

L'utilizzazione di ftp con *account*, invece, dà pieno accesso ad una determinata *directory* del sistema remoto, nella quale potremo inserire, modificare e cancellare *file*, proprio come se fosse una *directory* del nostro *hard disk*. Di norma è riservata ai dipendenti dell'università, dell'ente o della società che ospita il *server* ftp, oppure ai loro collaboratori, oppure ancora ai loro clienti. Se, ad esempio, decidete di pubblicare su **Internet** una vostra pagina *Web* acquistando lo spazio presso un *Internet provider*, quest'ultimo con ogni probabilità vi concederà un *account* ftp e una *password* personale.

Esistono *client* ftp dotati di interfaccia a caratteri e *client* ftp dotati di interfaccia grafica.

Programmi con interfaccia a caratteri

I *client* FTP con interfaccia a caratteri possono sembrare di difficile utilizzo, ma sono molto efficienti e versatili.

Vediamo adesso una sessione ftp con *account* eseguita dall'utente margara utilizzando un *client* ftp con interfaccia a caratteri.

Ci colleghiamo ad una macchina del dipartimento di Scienze dell'Informazione dell'Università di Bologna il cui nome logico è `alice.cs.unibo.it`.

Dopo aver impartito il comando:

```
ftp alice.cs.unibo.it
```

otteniamo la seguente schermata:

```
ftp alice.cs.unibo.it
Connected to alice.cs.unibo.it.
220 alice.cs.unibo.it FTP server (Version 6.2/OpenBSD/Linux-0.10) ready.
Name (alice.cs.unibo.it:margara): margara
331 Password required for margara.
Password:
230- Linux alice 2.2.19pre17 #5 Wed Apr 4 15:05:17 CEST 2001 i686 unknown
230- SCIENZE DELL'INFORMAZIONE - UNIVERSITA' DI BOLOGNA - CS.UNIBO.IT
230- -----
230- cluster: Linux CS
230- -----
230- (lm 10.01.2003)
230- -----
230 User margara logged in.
Remote system type is UNIX.
```

```
Using binary mode to transfer files.
ftp>
```

A questo punto il *computer* alice.cs.unibo.it è pronto per l'operazione ftp vera e propria.

Vediamo adesso la schermata corrispondente al *download* del *file* pippo.txt da alice.cs.unibo.it.

```
ftp> get pippo.txt local: pippo.txt remote: pippo.txt 227 Entering Passive Mode
(130,136,2,13,17,206) 150 Opening BINARY mode data connection for 'pippo.txt' (0
bytes). 226 Transfer complete.
```

L'operazione inversa (*upload*) viene eseguita usando il comando *put* come segue.

```
ftp> put pippo.txt local: pippo.txt remote: pippo.txt 227 Entering Passive Mode
(130,136,2,13,17,100) 150 FILE: pippo.txt 226 Transfer complete. ftp>
```

### Programmi con interfaccia grafica

Le interfacce grafiche hanno definitivamente avuto la meglio nel mondo del *software* di rete, così anche per il prelievo di *file* sono ormai disponibili programmi a base di icone e *mouse*. Ce ne sono diversi, per tutti i sistemi operativi. Due *client* ftp molto usati sono quelli inclusi nei *browser* *Internet Explorer* e *Netscape Navigator*.

Il *client* ftp di *Netscape*, è fortemente integrato con le altre funzionalità del programma. Per collegarsi con un sito ftp tramite *Netscape* è sufficiente inserire l'indirizzo (in gergo tecnico **url**) del sito che vogliamo raggiungere (se la macchina remota è ad esempio alice.cs.unibo.it, la relativa url sarà ftp://alice.cs.unibo.it dopo aver fatto click con il *mouse* sul comando *Open Page* presente all'interno del menu *File*.

Notare che è possibile inserire la url completa di *path* (ad esempio ftp://alice.cs.unibo.it/pippo.txt), così da saltare direttamente alla *subdirectory* che ci interessa.

Per trasferire il *file* sul proprio *hard disk*, a questo punto, non si deve fare altro che un click su *Save File* e attendere che il trasferimento sia completo.

### Protocolli, servizi e applicazioni

A questo punto è necessario spendere due parole per cercare di chiarire il legame che esiste tra i concetti di protocollo, servizio ed applicazione.

I termini protocollo, servizio ed applicazione sono ricorrenti in ambito informatico e rivestono una notevole importanza. Quando si parla di **Internet** il significato di questi termini può essere sinteticamente descritto come segue.

- Protocollo: insieme di regole e di convenzioni da rispettare perché due calcolatori possano comunicare.
- Servizio: prestazione fornita da una macchina servente (*server*) a una macchina cliente (*client*) che ne fa richiesta.
- Applicazione: programma.

Spesso lo stesso termine può, a seconda del contesto in cui è usato, rivestire sia il significato di protocollo che di servizio o di applicazione. Facciamo un esempio: ftp.

ftp è un protocollo di livello applicazione in *TCP/IP*. In particolare è il protocollo che regola il trasferimento di *file* da un *computer* ad un altro.

ftp è il servizio fornito da alcuni *host* di rete chiamati *server* ftp. Collegandosi a un *server* ftp è possibile fare *upload* e *download* di *file*.

**Ftp** è un'applicazione. Ovvero è il programma cliente che permette ad una macchina locale di usufruire del servizio ftp messo a disposizione da un *server* ftp seguendo il protocollo ftp.

## Telnet

**Telnet** è un protocollo di livello applicazione nella gerarchia *TCP/IP* che permette all'utente di visualizzare sulla macchina locale davanti alla quale è seduto un terminale agganciato ad una qualsiasi macchina remota sulla quale (in genere) l'utente possiede un *account*. Utilizzando il terminale così attivato, l'utente può accedere alle risorse della macchina remota come ci fosse seduto davanti. Con telnet, attraverso una interfaccia a caratteri è possibile agire sulla macchina remota proprio come se fosse il nostro *computer* locale, sfruttando tutte le risorse che il sistema mette a disposizione. Ad esempio l'utente può mandare in stampa un *file* sulla stampante collegata alla macchina remota oppure leggere la propria posta elettronica.

L'utilizzazione di telnet è molto semplice. Disponendo di un *client* telnet è sufficiente aprire una finestra di testo e come in una sessione ftp digitare:

```
telnet alice.cs.unibo.it
```

dove, come al solito, *alice.cs.unibo.it* è la macchina remota sulla quale vogliamo attivare una sessione telnet. L'indirizzo del *computer* cui ci si vuole collegare può essere un indirizzo numerico (ad esempio: 130.136.1.110) o un nome simbolico (ad esempio: *alice.cs.unibo.it*).

Dopo pochi secondi appare una schermata di benvenuto, e viene avviata la procedura che consente di farci riconoscere e di accedere alle risorse che ci interessano (procedura di *login*).

Molto spesso per portare a termine una connessione telnet è necessario avere il permesso di operare sul *computer* remoto. Appena stabilita la connessione, il *computer* remoto richiede l'identificativo dell'utente e la *password*.

Una volta attuata la fase di *login* si può finalmente disporre delle risorse presenti sul sistema remoto.

L'unica funzionalità importante di **Internet** che i principali *browser* non integrano nella loro interfaccia è proprio il telnet. Ciò significa che non è possibile visualizzare direttamente dall'interno di questi *browser* schermate in emulazione terminale.

## Gopher

Il *gopher* (ormai quasi totalmente inutilizzato) è essenzialmente uno strumento di organizzazione testuale dell'informazione presente su **Internet**. Il principio su cui si fonda questa organizzazione è quello della struttura gerarchica.

La struttura gerarchica è un sistema di organizzazione delle risorse molto efficiente. È molto diffuso in ambito informatico. Esempi tipici di strutture gerarchiche sono l'organigramma di un'azienda, l'albero genealogico, i sistemi di classificazione bibliotecaria.

Per consultare le informazioni organizzate nel *gopherspace* (o almeno ciò che ne resta) bisogna naturalmente avere accesso ad un *client* *gopher*. Ne sono stati sviluppati per tutti i sistemi operativi, sia in ambienti a caratteri sia in ambienti dotati di interfaccia grafica, ma poiché i *browser* *Web* sin

dall'inizio hanno integrato il supporto al protocollo *gopher*, la maggior parte è ormai stata abbandonata.

Un modo alternativo (ed assai più comodo) per accedere ai residui *gopher server* presenti su **Internet**, come detto, è costituito dall'uso di un comune *browser Web*, come *Netscape* o *Internet Explorer*.

L'aspetto di un *gopher server* nella finestra di un *browser* è sostanzialmente uguale a quello di un *ftp server*: un elenco di voci affiancate da piccole icone che ne denotano la tipologia.

### Servizi Multimediali

I documenti multimediali interattivi integrano diverse modalità comunicative e consentono l'interazione con l'utente.

I documenti multimediali interattivi più completi contengono audio, video e ambienti interattivi in grafica 3D (tridimensionale).

Il formato principale dei documenti del servizio *Web* è html, che però ha limitatissime capacità grafiche, video, audio ed interattive.

Per avere documenti interattivi un'alternativa importante che non approfondiamo è l'uso di un linguaggio di programmazione: *Java* e *JavaScript* sono due esempi di linguaggi capaci di animare i documenti.

Diversamente da html, i linguaggi di programmazione non sono alla portata di tutti gli autori di documenti **Internet**.

Al momento esistono pochi *standard de-jure* per servizi multimediali; quindi, le applicazioni che per prime hanno portato sulla rete un certo tipo di formato di documento multimediale sono diventate *standard di fatto*.

Ecco alcuni esempi di *software* per documenti multimediali:

- *Quicktime*: [www.apple.com/quicktime](http://www.apple.com/quicktime)
- *Flash* e *Shockwave*: <http://www.macromedia.com/>
- *RealPlayer*: <http://www.realplayer.com/>

### Formati audio

L'informazione audio viene codificata in digitale in moltissimi modi: i formati più comuni sono .wav, .mid, .mp3, .ra

- .wav, creato da *IBM* e *Microsoft*, rappresenta onde sonore in formato digitale di qualità variabile; infatti, la dimensione di un .wav dipende dalla qualità di digitalizzazione (1 minuto in qualità CD = 10 MB circa);
- .mid denota il formato MIDI, che contiene istruzioni eseguibili da strumenti musicali elettronici; quindi i *file* .mid sono molto compatti (un minuto = circa 10 KB);
- .mp3 è un formato audio digitale compresso di qualità quasi CD (un minuto in qualità quasi CD = circa 1 MB);
- .ra è un formato audio (<http://www.realaudio.com/>) per trasmissione **Internet** a banda stretta, fornito in varie versioni in funzione della connessione modem disponibile.
- .aif, creato da *Apple*, sta per *Audio Interchange File format*.

Il programma riproduttore audio più comune su *Windows* si chiama *Winamp* (<http://www.winamp.com/>).

Un programma su *Windows* molto versatile per effettuare conversioni audio è *AudioCatalyst* (<http://www.xingtech.com/>).

È ad esempio molto facile convertire un brano in versione CD musicale nel formato mp3.

Servizi audio di rete

Distinguiamo due principali classi di servizi: *broadcast* (radio digitale) e punto a punto (*unicast*: telefono su *IP*)

- *Broadcast: Digital radio*. È un servizio che permette di ascoltare la radio usando un *computer* collegato a **Internet**. Ad esempio si veda il sito <http://www.harmony-central.com/> per alcuni programmi audio.
- *Unicast: telephone VoIP (Voice over Internet telephone)*. Con *VoIP* la voce di chi utilizza il servizio viene digitalizzata e spedita via *TCP/IP* (attraverso **Internet**). Al momento esistono tre tipi di *VoIP*: *Pc to Pc*, *Pc to Phone*, *Phone to Phone*, che assicurano rispettivamente la comunicazione tra due *computer*, tra un *computer* e un telefono di rete fissa o mobile e tra due telefoni.

Esempi:

- <http://www.net2phone.com/>
- <http://www.voispring.it/>
- <http://www.picus.it/>
- <http://www.chocophone.com/>

Formati grafici

Anche i formati grafici sono tantissimi e permettono di rappresentare immagini in formato digitale. I principali in uso su **Internet** sono .gif e .jpg, entrambi basati su grafica bitmap (l'immagine è rappresentata come una matrice di *pixel* = *picture element*).

- GIF: formato grafico su 256 colori, compresso senza perdita di informazione, eccellente per icone o logo ovvero immagini che non hanno bisogno di alta definizione.
- JPEG: formato grafico su milioni di colori, compresso con perdita di informazione, eccellente per foto (o immagini che necessitano di un certo grado di definizione).
- Animazione GIF: visualizzazione di immagini GIF in rapida successione che serve per rappresentare semplici animazioni con bassa definizione.

Le immagini GIF usano la compressione *lossless* (senza perdita di informazione) LZW, che cerca le sequenze di *pixel* ripetute: dunque questo formato è adatto ad immagini in bianco e nero, specie se contrastate, ma non è adatto a comprimere fotografie.

**Nota:** il formato JPEG per molte immagini è più compatto di GIF. Esempio: foto di persona, .jpg 156KB, .gif 416KB.

Più importante è il fatto che siccome la compressione JPEG è *lossy* (con perdita di informazione/definizione) si può decidere quanta informazione verrà eliminata (a seconda della definizione desiderata).

## Comprimere documenti video

Problema: un'immagine video digitale in risoluzione 640x480 con un milione di colori occupa circa 1MB. La qualità video cinematografica vuole 24 fotogrammi al secondo, quindi un film di due ore richiederebbe 172.800 MB, pari a circa 260 CD audio.

La grande diffusione di connessioni **Internet** a banda stretta ha fatto nascere formati video speciali: il formato *RealVideo* ad esempio garantisce da 1 a 6 fotogrammi al secondo, risoluzione 80x120, con audio buono (stereo, 16 *bit*, 44KHz).

Nel 1988 nasce il gruppo di lavoro MPEG (*Moving Picture Expert Group*) per mettere a punto un sistema di codifica (CODEC) di video e audio compressi *lossy*.

- Mpeg-1 (ISO 11172, 1991) riguarda sia audio stereo che video (qualità simile a VHS); è uno *standard* che prevede tre livelli di compressione, per 30 fotogrammi/sec, risoluzione tipica 352x240 (max 512x324), audio CD.
- Mpeg-2 (estensione .mpg, ISO 13818, 1997) riguarda audio multicanale per *Home Cinema* (quindi è usato per DVD) campionato a 24KHz. Ha 60 fotogrammi/sec, risoluzione 1280x720, audio CD. **Nota:** Mpeg-2 layer 3 è oggi noto come MP3 (estensione .mp3)
- Mpeg-4 (.avi) è la specifica di *standard* più moderna, ed al momento si incarna in una tecnologia chiamata DivX: un intero film (110 minuti, in media) sta su un CDROM

## Formati di streaming

La trasmissione audio-video su **Internet** a banda stretta si basa su *streaming*, ovvero la riproduzione di audio e/o video contemporaneamente al loro *download*; le principali piattaforme *streaming* sono al momento *RealNetworks*, *Microsoft* e *Apple*.

*RealNetworks* (<http://www.real.com/>) pubblicò il *RealPlayer* (estensione .rm) nel 1995, nel 2000 è alla versione 8.

*Microsoft* include il *Media Player* (versione 7 nel 2000) con tutte le versioni di *Windows* (*WindowsMedia.com*) ed ha circa 2/3 degli utenti di *RealMedia*.

*Apple* ([www.apple.com/quicktime](http://www.apple.com/quicktime)) ha sviluppato il formato *QuickTime* (estensione .mov) che nel 1999 è alla versione 4. La sua forza sta nel fatto di essere davvero multiplatforma; inoltre supporta parecchi tipi di contenuto: testo, immagini ferme, 3-D, e *Virtual Reality*.

## Servizi basati su streaming (WebCasting)

Il *WebCasting* è la diffusione sistematica di informazioni video in formati *streaming*.

Un sito indice dei canali *WebCasting* è <http://www.mediachannel.com/>.

Le principali TV del mondo hanno siti *Web* in cui propongono filmati da vedere via *streaming*.



### Formati per documenti interattivi

I documenti interattivi sono documenti che includono animazioni e suoni e sono capaci di interagire con l'utente (come i videogiochi). Si usano tipicamente per presentazioni graficamente complesse, specie nel campo dell'*e-learning*.

*Macromedia Director* è il più famoso programma di *authoring* di documenti di questo tipo; *Shockwave* è un *plug-in* per *browser* che consente di riprodurre documenti realizzati con *Macromedia Director*. A partire dalla versione 6.0 di *Director*, *Shockwave* permette lo *streaming*.

Nel 1997 è apparso *Flash 2.0*, allo scopo di portare su *Web* alcune caratteristiche importanti di *Director*.

La forza di *Flash* è nel formato vettoriale con cui opera. Consente di ottenere filmati (*movie*) compatti e di ottima qualità grafica, che possono zoomare a tutto schermo senza perdita di qualità. Un altro vantaggio delle immagini vettoriali è che il filmato può racchiudere informazioni aggiuntive. Ad esempio, agli oggetti possono essere associate delle proprietà particolari, come l'attivarsi ad un click del *mouse*.

Perché dunque i *browser* non visualizzano immagini vettoriali in luogo delle GIF e delle JPG (che sono bitmap)? La spiegazione è che non esiste uno *standard* industriale per la trasmissione di immagini vettoriali.

Al momento (2002) la partita degli *standard* si gioca tra *Adobe* e *Macromedia*: la prima ha sottoposto al W3C un formato vettoriale basato sul PDF chiamato PGML, mentre la seconda, con *Flash*, ha imposto de facto il proprio *standard*.

### World Wide Web (WWW)

Il *World Wide Web* (WWW o più semplicemente *Web*), che letteralmente significa ragnatela intorno

(grande quanto il) al mondo, è uno dei tanti servizi disponibili su **Internet**. Anche se il *Web* senza dubbio è uno dei servizi più usati dagli utenti (forse il più usato in assoluto) desideriamo puntualizzare che il *Web* non coincide con **Internet**, ma è uno dei tanti servizi a cui è possibile accedere collegandosi a **Internet**.

In estrema sintesi il *Web* è una collezione di documenti ipertestuali e multimediali contenuti su alcune macchine (*Web server*) collegate a **Internet**. Tramite applicazioni specifiche chiamate *browser* (clienti *Web*) è possibile navigare in questo mare di documenti, visualizzarli e memorizzarli sul proprio calcolatore.

Il *Web* è un vero e proprio sistema di editoria elettronica *on-line*.

Uno degli aspetti del *Web* che ha maggiormente contribuito alla sua diffusione è senza dubbio la capacità di gestione e controllo dei documenti multimediali, e dunque dei linguaggi utilizzati per la loro creazione. Un ruolo importante nel processo di espansione e diffusione del *Web* è rivestito dalle grandi aziende produttrici di *browser*. Nel corso degli anni tanto *Microsoft* quanto *Netscape*, man mano che nuove versioni dei loro *browser* venivano sviluppate, hanno introdotto innovazioni ed estensioni. Questa corsa all'ultima innovazione, se molto ha migliorato l'aspetto e la fruibilità delle pagine pubblicate su *Web*, ha avuto degli effetti deleteri sul piano della portabilità dei documenti. Per ovviare al rischio di una babele telematica, ed evitare che le tensioni indotte dal mercato limitassero l'universalità di accesso all'informazione *on-line*, è stato costituito il *World Wide Web Consortium* (W3C). Si tratta di una organizzazione no profit ufficialmente deputata allo sviluppo degli *standard* tecnologici per il *Web* che raccoglie centinaia di aziende, enti e centri di ricerca coinvolti più o meno direttamente nel settore. In questi ultimi anni il W3C ha prodotto una serie di specifiche divenute, o in procinto di divenire, *standard* ufficiali su **Internet**. Tutti i materiali prodotti dal W3C sono di pubblico dominio, e vengono pubblicati sul sito *Web* del consorzio.

Brevi cenni storici sul Web

- Nel 1945, *Vannevar Bush* (tecnologo del MIT e consulente del governo USA) scrive un articolo in cui descrive la nozione di ipertesto.  
([www.theatlantic.com/unbound/flashbks/computer/bushf.htm](http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm))
- Nel 1965 *Ted Nelson* inventa la parola **Ipertesto** nell'articolo: *A File Structure for the Complex, the Changing, and the Indeterminate 20th National ACM Conference, New York*. Vedi [www.sfc.keio.ac.jp/~ted/TN/PUBS/LM/LMpage.html](http://www.sfc.keio.ac.jp/~ted/TN/PUBS/LM/LMpage.html)
- Nel 1968 *Doug Engelbart* costruisce il prototipo di un *oNLine System* (NLS) con funzioni di *editing* ipertestuale, e inventa il *mouse* per interagire con tale sistema. Vedi [stanford.edu/MouseSite/](http://stanford.edu/MouseSite/)
- Nel 1989, *Berners-Lee* del CERN specifica un sistema di gestione di documenti ipertestuali distribuiti.
- Alla fine del 1990 viene realizzato un prototipo.
- Nel febbraio 1993 venne rilasciato *Mosaic*, un *browser* scritto dallo studente *M. Andreessen*.
- Alla fine del 1994 *Andreessen* fonda *Netscape Corporation* (nel 1999 verrà acquisita da AOL, *America On Line*).
- Nel dicembre 1994 nasce il W3C.

Dimensioni del www

Riportiamo alcune cifre che danno l'idea delle dimensioni del *Web* e della sua continua vertiginosa espansione.

All'inizio dell'anno 2000 c'erano *on-line* un miliardo di pagine circa. Alla fine del 2000 c'erano *on-line* 2.5 miliardi di pagine circa. Alla fine del 2001 4 miliardi di pagine circa. Con una crescita *media* di più di 7 milioni di pagine al giorno.

Alla fine del 2001, il motore di ricerca *Google* classifica 3 miliardi di pagine, inclusi *newsgroup* e immagini.

Il mondo produce annualmente da uno a due *exabyte* (10<sup>18</sup> *byte*) di dati, ovvero all'incirca 250 *megabyte* per ogni abitante della terra.

I documenti stampati, di qualsiasi genere, sono solo lo 0.003% del totale. Le informazioni memorizzate su supporto magnetico sono di gran lunga le più numerose, e crescono rapidamente: ogni anno raddoppiano.

Concetti Logici di base: ipertesti e multimedialità

Il *Web* in estrema sintesi è un gigantesco ipertesto multimediale distribuito.

Il concetto di multimedialità si riferisce alla tipologia dell'informazione rappresentata, mentre il concetto di ipertesto si riferisce alla struttura ed alla organizzazione dell'informazione stessa.

Ipertesto

Dal punto di vista logico un ipertesto è un sistema di organizzazione delle informazioni (testuali, ma non solo) in una struttura non sequenziale, bensì reticolare.

Ciò significa che un ipertesto permette di saltare da un punto ad un altro del documento o addirittura saltare da un documento ad un altro seguendo *link* specifici messi a disposizione dall'autore del documento.

Dal punto di vista della implementazione concreta, un ipertesto digitale si presenta come un documento elettronico in cui alcune porzioni di testo o immagini presenti sullo schermo, evidenziate attraverso artifici grafici (icone, colore, tipo e stile del carattere), rappresentano i diversi collegamenti disponibili nella pagina. Questi funzionano come dei pulsanti che attivano il collegamento e consentono di passare, sullo schermo, al documento di destinazione. Il pulsante viene premuto attraverso un dispositivo di *input*, generalmente il *mouse*, una combinazione di tasti, o un tocco su uno schermo *touch-screen*. L'altro aspetto che fa dell'ipertesto elettronico uno strumento comunicativo dalle enormi potenzialità è la interattività che esso consente al fruitore, non più relegato nella posizione di destinatario più o meno passivo del messaggio, ma capace di guidare e indirizzare consapevolmente il suo atto di lettura.

Multimedia

Con multimedialità ci si riferisce alla possibilità di inserire contemporaneamente in uno stesso documento informazione di tipo diverso. Ad esempio, immagini, suoni e filmati. Questa informazione ovviamente viene rappresentata in forma digitale appropriata usando *standard*, linguaggi e strumenti tra loro molto diversi.

Distribuito

Un ipertesto multimediale può essere spezzato in più parti (memorizzate in *file*). Ognuna di queste parti è tipicamente memorizzata su un *host* di rete. Se un documento è memorizzato su più *host* viene detto distribuito.

L'incontro tra ipertesto, multimedialità e interattività rappresenta dunque la nuova frontiera delle tecnologie comunicative. Il problema della comprensione teorica e del pieno sfruttamento delle enormi potenzialità di tali strumenti, specialmente in campo didattico, pedagogico e divulgativo è

naturalmente ancora in gran parte aperto: si tratta di un settore nel quale vi sono state negli ultimi anni, ed è legittimo aspettarsi negli anni a venire, innovazioni di notevole portata.

Come funziona il Web - Architettura client-server.

La navigazione in **Internet** si basa su interazione di tipo *client-server*. Le funzioni elaborative necessarie per rendere possibile la navigazione in **Internet** sono distribuite tra *client* e *server* in modo da distribuire il carico di lavoro e ottimizzare l'efficienza complessiva del sistema.

## Client

Per navigare in **Internet** è necessario possedere un *software* (applicativo) chiamato *browser* (esempi: *Netscape Communicator*, *MS Internet Explorer*). Il *browser* costituisce lo strumento di interfaccia tra l'utente e il sistema *Web* e si occupa sostanzialmente di:

- ricevere ed eseguire i comandi dell'utente;
- richiedere ad un *server* i documenti;
- interpretare i formati di codifica degli oggetti che costituiscono ogni singolo documento e presentarli all'utente sul suo monitor (fase di *rendering*);
- invocare applicazioni esterne (*helper*);
- riprogrammarsi mediante *plug-in*;
- eseguire codice *JavaScript* o *Java*.

## Server

Un *server Web*, o più precisamente *server http* (vedremo poi il perché di questo nome) si occupa di memorizzare, reperire e inviare i documenti (e tutti gli oggetti digitali che li costituiscono) richiesti dai *client*. Nel momento in cui il *client* (utente) invia una richiesta *http* al *server* opportuno, con l'indicazione del documento che vuole ricevere, il *server* interpreta la richiesta e invia gli oggetti che compongono il documento richiesto al *client*.

## Protocolli del Web

Il *client* e il *server* comunicano e rendono possibile la navigazione. Per comunicare utilizzano protocolli specificatamente progettati a tal fine.

Nel seguito daremo una descrizione sintetica dei meccanismi e dei protocolli principali che permettono di navigare in **Internet**.

Il *Web* è un servizio **Internet** che si basa su tre protocolli fondamentali:

### Uniform Resource Locators (url)

Schema di denominazione di risorse in **Internet** (tipicamente *file* in formato *html*). Quando parliamo di un *url* ci riferiamo al nome completo ed univoco che una certa risorsa ha in **Internet**. Ad esempio diciamo che *www.cs.unibo.it* è l'*url* del sito **Internet** del dipartimento di Scienze dell'Informazione dell'università di Bologna.

### HyperText Transfer Protocol (http)

*Http* è il protocollo di livello applicazione che regola il trasferimento dei *file html* in **Internet**.

- I *server http* distribuiscono documenti ipertestuali.

- I clienti http (*browser*) li visualizzano.

*Server* e *client* per comunicare fanno riferimento al protocollo http.

## HyperText Markup Language (html) per ipertesti

HTML è un linguaggio di *markup*, ossia un linguaggio che permette di specificare come un certo testo deve essere formattato (colori, collegamenti ipertestuali, sfondi e molte altre caratteristiche grafiche) al momento della sua visualizzazione sul nostro *display*.

Siti Web

Un sito *Web* è definito da:

- un indirizzo, o url (esempio <http://www.cs.unibo.it/>) al quale corrisponde un *server* http;
- un insieme di documenti (scritti di solito in html) sotto il controllo del *server* http;
- un insieme di servizi, che è variabile; nel caso la parte di servizi sia preponderante su quella dei documenti il sito si definisce portale.

Come funziona http

Ogni sito *Web* (*server* http corrispondente) ha un processo servente in ascolto sulla porta 80 in attesa di connessioni http; dopo che è stata stabilita una connessione, il cliente invia una richiesta e il servente invia una risposta; quindi la connessione viene abbandonata (http è un servizio non orientato alla sessione).

Esempio: supponiamo di voler seguire il link:

[www.w3.org/hypertext/WWW/TheProject.html](http://www.w3.org/hypertext/WWW/TheProject.html).

Le fasi che si succedono tra il click dell'utente e la visualizzazione della pagina sono le seguenti:

- Il *browser* determina l'url (controlla ciò che è stato selezionato).
- Il *browser* chiede al dns l'indirizzo di <http://www.w3.org/>.
- Il dns risponde con 18.23.0.23 (indirizzo fisico della macchina <http://www.w3.org/>).
- Il *browser* apre una connessione *tcp* sulla porta 80 della macchina con indirizzo 18.23.0.23.
- Quindi invia un comando *GET* /hypertext/WWW/TheProject.html.
- Il *server* <http://www.w3.org/> invia il *file* TheProject.html.
- Viene rilasciata la connessione *TCP*.
- Il *browser* ripete il ciclo per ciascuna immagine (o componente di altro tipo) in TheProject.html.

Ogni interazione http è una richiesta ASCII, seguita da una risposta *mime*.

Http è un protocollo *client-server*, progettato per la distribuzione rapida di documenti ipertestuali su reti *TCP/IP*.

È un protocollo *connection-less*, che significa che ogni richiesta di un cliente viene elaborata dal *server* indipendentemente: il *server* non ha memoria della sessione.

Esempio: se un *server* http deve trasmettere un documento html contenente 10 immagini, occorrono 11 connessioni.

Ogni risorsa è specificata da una url.

**Nota.** Sebbene http sia stato inventato per il *Web*, è più generale del necessario, con funzioni pensate per future applicazioni.

Pubblicare pagine sul Web

In estrema sintesi, per pubblicare un documento su *Web* occorre:

- poter usare una url pubblica di un *server Web*;
- poter scrivere in un disco sotto il controllo del *server Web* (ovvero fare *upload* via ftp);
- scrivere il documento direttamente in html oppure usare un *editor* html (esempio *DreamWeaver*);
- preparare, linkare e fare l'*upload* sul *server* delle figure (se presenti);
- fare l'*upload* del documento sul disco del *server*.

Helper e plug-in

Il *browser* può invocare l'aiuto di altre applicazioni per gestire documenti in formati particolari. Quando un *browser* viene progettato non è possibile prevedere quali nuovi formati di *file* e quali tecnologie per la loro gestione saranno introdotti in futuro. Inoltre i *browser* sono applicazioni molto complesse e di grandi dimensioni. Come conseguenza di queste considerazioni i *browser* si appoggiano e usano altre applicazioni per gestire formati di *file* specifici. Queste applicazioni sono sostanzialmente di due tipi: *helper* e *plug-in*.

### Helper

Un *helper* è un programma indipendente dal *browser*, ma da esso invocabile direttamente per gestire la visualizzazione di particolari tipi di documenti. Esempi di *helper*:

- *GhostView* per documenti in formato *PostScript*.
- *RealPlayer* per documenti *streaming* in formato *Realaudio*.

### Plug-in

Un *plug-in* è un componente *software* del *browser* in grado di arricchirne la capacità di gestire documenti in formati speciali. Esempi di *plug-in*:

- *Acrobat plug-in* per documenti in formato pdf.
- *Quicktime plug-in* per filmati in vari formati.

I *plug-in* di solito sono programmi disponibili in rete, caricati automaticamente dal *browser* quando si trova di fronte ad un documento in formato speciale sconosciuto.

La sostanziale differenza tra *helper* e *plug-in* è che l'*helper* è un programma esterno al *browser* che viene attivato al momento di gestire *file* di un certo formato, mentre il *plug-in* è un modulo *software* che si può aggiungere al *browser* e che viene usato al suo interno senza dover attivare nessun altro programma oltre al *browser* stesso.

*Netscape 2.0* è stato il primo *browser* che ha supportato l'idea tecnologica secondo la quale case produttrici diverse da quella del *browser* offrono componenti *software* che lo integrano.

Un buon sito in cui si trovano molti *plug-in* è:

[browserwatch.internet.com/plug-in/plug-in-win.html](http://browserwatch.internet.com/plug-in/plug-in-win.html)

## Html

Un documento html è un documento scritto nel linguaggio di *markup* html, che contiene i *tag* (delimitatori sintattici) necessari ad informare il *browser* sulle modalità di formattazione.

**Nota:** Un documento contenente testo senza alcun *tag* html viene comunque visualizzato dal *browser*, ma senza alcuna formattazione.