

Storia e organizzazione di Internet

Internet: un primo sguardo

Per il principiante, **Internet**, il *World Wide Web* (**WWW**) e la rete del proprio ufficio sono più o meno la stessa cosa: una serie di calcolatori collegati tra loro che scambiano tra loro e con i loro utenti delle risorse, in particolare risorse di tipo informativo. È importante, invece, imparare a distinguere i vari livelli in cui la rete si struttura, per comprendere la logica interna dello sviluppo di questi livelli, e, dunque, le loro potenzialità, le loro criticità, i loro centri decisionali.

Scopo di questo intero modulo è descrivere la struttura e la storia della rete globale, rimandando invece ai moduli seguenti una discussione approfondita dei servizi forniti dalla rete stessa e delle tecniche per ricercare ed immettere in rete l'informazione. È un modulo più discorsivo e meno tecnico di altri. Alcuni approfondimenti discuteranno in modo tecnico alcune delicate questioni implementative.

Sappiamo già che una rete di calcolatori è una struttura di telecomunicazione in cui più calcolatori (in genere eterogenei, cioè diversi per *hardware* e sistema operativo) sono collegati allo scopo di condividere risorse e scambiarsi informazioni. Anche le reti sono tra loro eterogenee:

- per dimensioni: **LAN** (*local area network*), **MAN** (*metropolitan area network*), e **WAN** (*wide area network*);
- per supporto di telecomunicazione: cavo *ethernet*, cavo coassiale, doppino telefonico, fibra ottica, connessione senza fili (infrarosso, radio, UMTS, ecc.);
- per topologia della connessione: a stella, ad anello, a *bus* (tutti i nodi connessi su un'unica linea di comunicazione), punto a punto;
- per stabilità della connessione: connessioni dedicate (sempre attive), connessioni commutate (attivate su domanda di uno dei due nodi collegati), connessioni mobili (per esempio un cellulare UMTS).

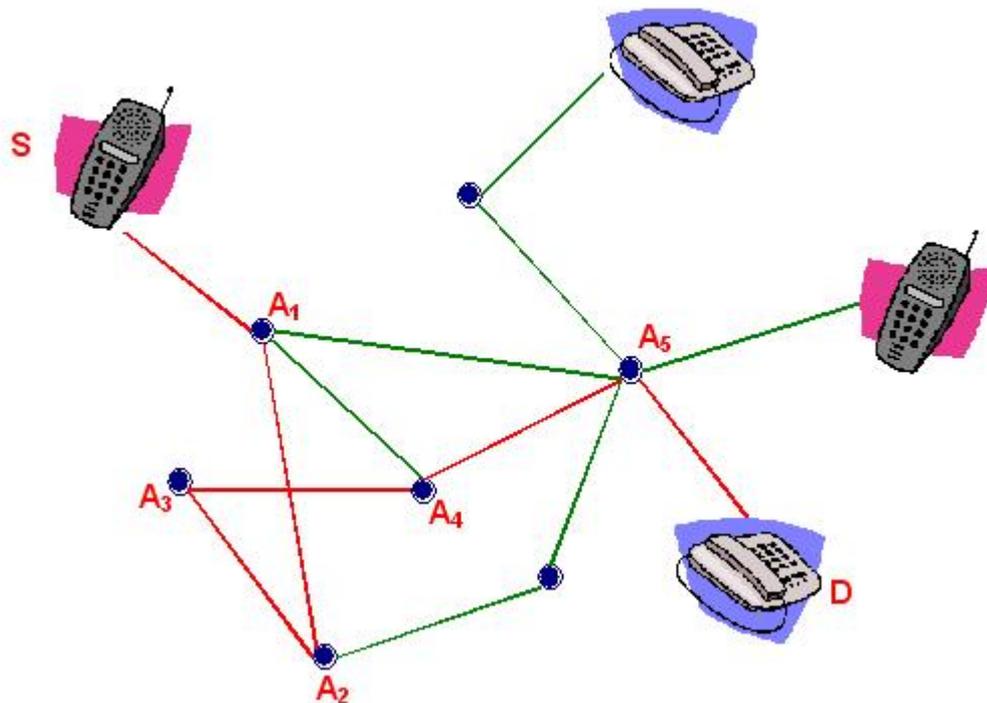
L'eterogeneità delle reti, però, non impedisce la loro comunicazione. Sebbene alcune di queste reti usino al proprio interno protocolli di comunicazione specifici, mediante l'uso di un protocollo comune, le reti possono essere collegate tra loro a formare reti di reti (internets). Chiamiamo **Internet** (con la "I" maiuscola!) la rete planetaria di tutte le reti collegate tra loro e che comunicano con lo stesso protocollo. Nell'ultima sezione di questo modulo commenteremo la definizione ufficiale di **Internet**: per comprenderne a pieno le implicazioni, tuttavia, è necessario analizzare nelle prossime sezioni la struttura di questa rete di reti, come e quando questa si è costituita, come si sviluppa, chi la gestisce, chi la paga.

La struttura pervasiva di **Internet** rende la rete un veicolo formidabile per la trasmissione dell'informazione. Il *World Wide Web* è la principale architettura informativa basata su **Internet**: si tratta di un insieme di ipertesti collegati tra loro e che risiedono su nodi fisicamente diversi e molto distanti tra loro. Anche del **WWW** presenteremo la struttura e la storia del suo sviluppo.

Reti a commutazione di circuito

Internet si basa sulla interconnessione di diverse centinaia di migliaia di calcolatori e reti altrimenti indipendenti ed eterogenee. Come possono dialogare due nodi di questa rete? È evidente che non è ragionevole connettere direttamente ogni nodo ad ogni altro nodo. La connessione può però essere stabilita in modo indiretto:

per andare dal nodo S (sorgente) al nodo D (destinazione), passiamo prima per il nodo A1, poi per A2, ecc. fino a raggiungere il nodo An e da qui raggiungere D.



Si tratta della stessa situazione della normale telefonia vocale: per effettuare una chiamata dal numero (nodo) S al numero (nodo) D, la chiamata passa per la centrale a cui è direttamente collegato S (questa sarà il nodo A1), viene instradata attraverso altre centrali (i nodi A2, A3, ecc.) fino a raggiungere la centrale An a cui D è direttamente collegato e raggiungere da qui, finalmente, D.

Nella telefonia vocale, le varie centrali (il cui ruolo era un tempo svolto da operatori manuali) funzionano come interruttori (deviatori): viene stabilito un collegamento "fisico" tra S e D, che, al momento in cui la connessione è stabilita, consiste in un (complesso) circuito elettrico che collega S a D. Per questo motivo questa tecnica si chiama connessione a commutazione di circuito. Le risorse (canali di comunicazione, interruttori, ripetitori, ecc.) che si trovano sul percorso S, A1, A2, ..., An, D sono assegnate alla connessione tra S e D e non sono disponibili per altri fino a quando S e D non le rilasciano, al termine della telefonata. A questo punto, tutte le risorse ritornano disponibili per altre telefonate.

Questa modalità di connessione è concettualmente semplice, ma mal si adatta alla comunicazione tra due calcolatori, perché:

- ci vuole troppo tempo per realizzare la connessione tra S e D.
La commutazione del circuito è estremamente lenta rispetto alla velocità con la quale S produce i dati e a quella con cui D è capace di riceverli. In altre parole, il tempo speso per l'effettiva trasmissione dei dati tra S e D è quasi trascurabile rispetto alla somma del tempo necessario ad attivare il circuito e di quello che serve per rilasciarlo al termine della comunicazione.
- comporta un grande spreco di risorse
La comunicazione tra calcolatori procede tipicamente per "picchi" di dati. Dopo che un picco è stato trasmesso, passa un certo tempo prima che il prossimo picco sia a sua volta inviato. Nel tempo che passa tra i due picchi il circuito rimane comunque impegnato e le risorse che lo compongono non possono essere rese disponibili per altre connessioni.
- è molto dipendente da eventuali guasti al circuito
Se il circuito si guasta occorre "disfare" tutta la comunicazione e attivare la commutazione di un nuovo circuito. Inoltre non è semplice per D capire se la mancanza di dati che rileva sulla

linea corrisponde al silenzio di S o ad un guasto intermedio.

Reti a commutazione di pacchetto

Per ovviare ai problemi delle commutazione di circuito, viene usata un'altra tecnica, cruciale per il successo di **Internet**: la commutazione di pacchetto.

Siamo nella situazione seguente:

S vuole inviare a D il messaggio M costituito dai caratteri c_1, c_2, \dots, c_k .

Invece di richiedere alla rete la commutazione di un circuito che raggiunga D, S suddivide M in pacchetti, ciascuno composto di una quantità fissata (e piccola) di caratteri. Contraddistingue ogni pacchetto con la propria firma, con il numero d'ordine del pacchetto all'interno di M (così che D possa poi ricostruire il messaggio) e con un indirizzo che identifica in modo unico D sulla rete. A questo punto ogni pacchetto inizia una propria vita autonoma: S invia i pacchetti, uno alla volta, ad uno dei calcolatori che gli sono più vicini sulla rete, diciamo ad A1. Quando A1 riceve un pacchetto, si accorge di non essere il destinatario finale e così lo inoltra ad uno dei suoi vicini (diciamo ad A2), che a sua volta si comporta nello stesso modo, fino a quando il pacchetto riceve D. Via via che D riceve i pacchetti, li rimette in ordine, scarta eventuali duplicati e se necessario richiede a S la ritrasmissione di qualche pacchetto che non fosse giunto a destinazione.

Osserviamo che:

- ogni pacchetto utilizza tutta la banda di comunicazione disponibile in quel momento;
- ogni pacchetto che arriva su un nodo, viene memorizzato prima di essere ritrasmesso (meccanismo *store-and-forward*)
- la presenza di molti pacchetti che devono essere trasmessi su una specifica connessione può causare congestione: i pacchetti si accodano aspettando di usare la connessione;
- non è assolutamente garantito che tutti i pacchetti che compongono uno stesso messaggio compiano lo stesso percorso.

È evidente che questa descrizione lascia molte cose poco chiare:

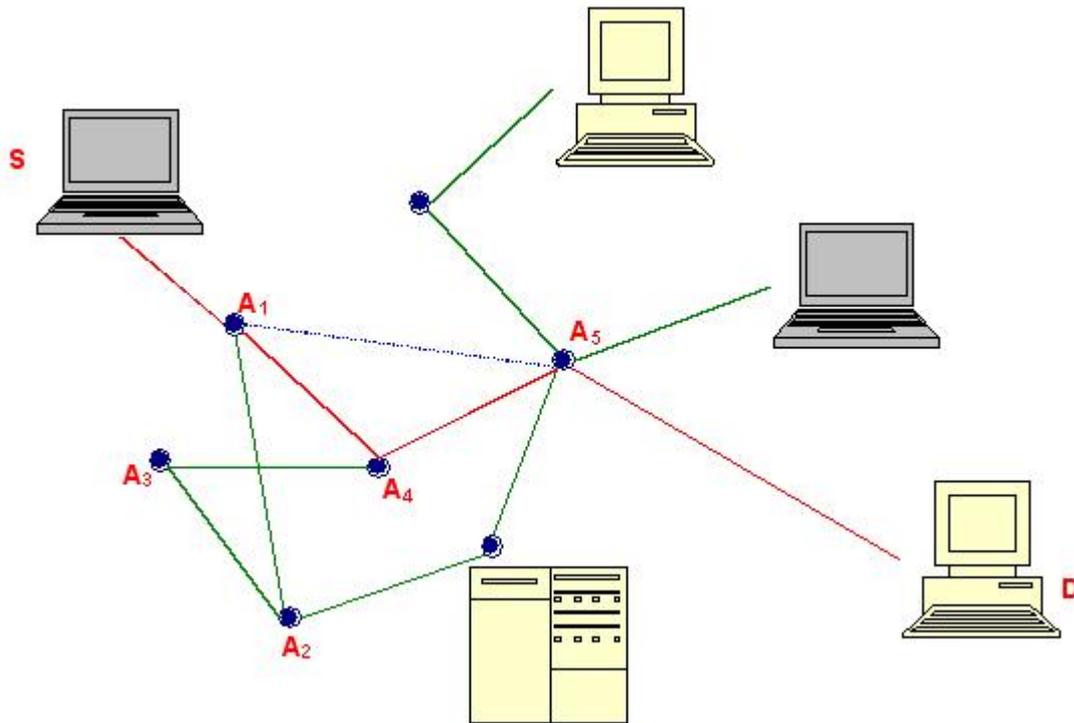
- come decide un nodo a quale vicino trasmettere un pacchetto, in modo che, in qualche modo, il pacchetto si avvicini alla destinazione D?
- come viene indicato il codice che identifica D in modo univoco?
- come conosce S questo codice?
- come viene gestita l'eventuale presenza di errori durante la trasmissione?
- come viene gestita la congestione (per esempio con la scelta di cammini alternativi)?
- chi si occupa di tutti questi dettagli?

Non certo l'utente finale, che non sa nulla di tutto questo. Forse le diverse applicazioni (browser, clienti di posta elettronica, ecc.)? O il sistema operativo?

Ad alcune di queste domande risponderemo nel seguito. Per comprendere la struttura di **Internet**, tuttavia, la descrizione che abbiamo dato è sufficiente. Osserviamo che:

- non c'è spreco di tempo per connettere S a D, perché nessuna connessione viene stabilita. Viene solo impegnata una connessione da un nodo al proprio vicino.
- il circuito di comunicazione tra un nodo ed il vicino è usato per il messaggio M solo e quando si trasmette un pacchetto. Tra un pacchetto e l'altro possono essere trasmessi altri pacchetti di altri messaggi tra altri nodi.
- la comunicazione può avvenire anche se un circuito si guasta: basta aggirare il guasto mediante un altro percorso sulla rete.

Nel disegno che segue i pacchetti tra S e D possono seguire un qualunque cammino che collega sorgente e destinazione. In particolare, la connessione guasta che collega A1 e A5 può essere facilmente aggirata.



La gerarchia dei protocolli

Rispondiamo in questa sezione alla domanda: chi si occupa dei dettagli della comunicazione?

La risposta è più complessa di quanto ci si possa aspettare, perché dipende dal livello al quale la domanda è posta. Nel caso di **Internet**, possiamo distinguere cinque livelli. Ogni livello corrisponde alla gestione di una parte di dettagli (cioè ad un protocollo) e alla predisposizione di una serie di servizi per il livello superiore (il quale, dunque, non si preoccupa dei dettagli del livello inferiore).

Un protocollo definisce il formato e l'ordine delle comunicazioni spedite fra entità di rete, nonché le azioni da compiere al momento della trasmissione e del ricevimento della comunicazione.

1. Livello fisico: i nodi sono fisicamente connessi da cavi e dispositivi di comunicazione (p.e modem); sono gestiti i parametri che dipendono dal mezzo di comunicazione (per esempio: qual è il voltaggio di 0 e di 1, quanti *bit* sono trasmessi al secondo, ecc.).
2. Livello *data-link*: i nodi condividono un certo codice di trasmissione che consente loro di scambiarsi delle informazioni digitali e di correggere elementari errori di trasmissione che possono avvenire a livello fisico; fornisce al livello 3 la gestione dell'accesso al mezzo di comunicazione.
3. Livello rete: i pacchetti sono inviati sulla rete e viaggiano attraverso più nodi; gestisce i servizi che aprono e chiudono le connessioni e, in particolare, gestisce l'instradamento (*routing*) dei messaggi: a quale nodo devo inoltrare un pacchetto? In **Internet**, questo è il livello del protocollo **IP** (*Internet Protocol*).
4. Livello trasporto: gestisce i dati che riceve dal livello 3, li organizza in pacchetti, li invia sulla rete sfruttando i servizi del livello 3; gestisce la ricostituzione del messaggio. In **Internet**, questo è il livello del protocollo **TCP** (*Transmission Control Protocol*).
5. Livello applicazione: le applicazioni dialogano tra loro per scambiarsi dati e documenti. Sfruttando i servizi del livello 4, richiedono, scambiano e ricevono dati. Questo è il livello dei

protocolli **HTTP** (per lo scambio di servizi sul **WWW**), **SMTP** (per l'invio della posta elettronica), ecc.

Un approfondimento: i livelli ISO/OSI

L'Organismo Internazionale degli *Standard* ha definito una gerarchia *standard* di protocolli per reti, il cosiddetto modello ISO/OSI. Rispetto ai livelli che abbiamo elencato, ne aggiunge due, intermedi tra il livello trasporto e il livello applicazione. Sono il livello sessione (sopra a quello trasporto) e il livello presentazione (tra quello sessione e quello applicazione). **Internet** adotta il modello semplificato che abbiamo discusso; inoltre **TCP** e **IP** non rispettano del tutto lo *standard* ISO/OSI per i rispettivi livelli.

Il cuore di **Internet** (e la ragione del suo successo) è costituito dalla coppia di protocolli **TCP/IP**.

Una rete di reti eterogenee

La strutturazione della comunicazione secondo una gerarchia di protocolli spiega come in **Internet** possano essere collegate tra loro reti indipendenti ed eterogenee.

Affinché la rete R1 possa comunicare con R2 è sufficiente che un nodo "di frontiera" di R1 (un *gateway*, o, come si dice più modernamente, un *router*) e un nodo di frontiera di R2 comunichino tra loro sfruttando la coppia di protocolli **TCP/IP**. I *router*, ciascuno sfruttando i livelli inferiori della propria gerarchia, gestiscono le differenze tra le due reti (velocità di trasmissione, dimensione dei pacchetti, condizioni d'errore, ecc.). Nessuna delle due reti deve essere modificata per essere collegata: basta che i *router* siano in grado di interagire mediante **TCP/IP**. Le due reti possono avere diversi comportamenti quanto a velocità e solidità. La commutazione di pacchetto permette alla rete rete più veloce di non sincronizzarsi con la più lenta, e mette in capo al destinatario (e non alla gestione della rete nel suo complesso) la responsabilità di rimettere insieme il messaggio e di scartare i pacchetti duplicati. Aggiungere una nuova rete a quelle già collegate è estremamente semplice: non occorrono modifiche alle reti esistenti e ai loro collegamenti; basta che la nuova rete abbia un *router* che supporta **TCP/IP** e che lo si possa collegare alla interconnessione di reti (*inter-network*, da cui **Internet**) già esistente. L'unica cosa necessaria è la possibilità di indirizzare un pacchetto verso un nodo della nuova rete.

Indirizzare un nodo su Internet

Ogni nodo della rete deve poter essere individuato in modo univoco. La soluzione più semplice sembrerebbe quella di avere un'autorità centrale che gestisce tutta la rete e assegna indirizzi univoci a chi li richiede, per esempio numeri assegnati in successione.

Questa soluzione centralizzata è semplice solo in apparenza. In particolare entra in conflitto con la scelta progettuale (evidente dalle sezioni precedenti) di mantenere **Internet** una rete aperta, cioè facilmente estensibile. La scelta di **Internet** è quella di avere un indirizzamento gerarchico:

- l'universo **Internet** è suddiviso in reti fisiche;
- ad ogni rete fisica è assegnato in modo centralizzato un certo numero (indirizzo);
- a ciascun nodo della rete fisica è assegnato un indirizzo composto dall'indirizzo della rete fisica concatenato con un altro numero, che individua in modo univoco il nodo all'interno della rete;
- se la rete fisica è suddivisa in sottoreti, l'assegnamento di indirizzi ai suoi nodi può avvenire, a sua volta, in modo gerarchico.

In questo modo esiste un'autorità centrale che si preoccupa di assegnare numeri alle reti; mentre i numeri ai nodi delle reti fisiche possono essere assegnati dai gestori delle reti stesse.

Indirizzi IP

Vediamo in dettaglio come è composto l'indirizzo di un nodo.

Gli indirizzi dei nodi sono definiti e gestiti a livello rete e dunque dal protocollo **IP**, che specifica che ogni nodo sia univocamente identificato da un numero di 32 *bit*, il suo indirizzo **IP**.

Un indirizzo **IP** viene in genere indicato come sequenza di 4 numeri decimali, ciascuno compreso tra 0 e 255, separati da un punto. Ad esempio, la macchina su cui sono state composte queste note ha indirizzo 130.136.2.37.

Un *byte* sono 8 *bit*. Un indirizzo **IP** è dunque composto da 4 *byte*. Con 8 *bit* si possono indicare i numeri decimali tra 0 e $2^8=255$. Ciascuno dei quattro *byte* dell'indirizzo **IP** viene convenzionalmente letto come un numero decimale e separato con un punto dal successivo.

Quali dei quattro numeri costituiscono l'indirizzo della rete e quali l'indirizzo del nodo all'interno della rete?

Dipende dall'importanza e dalla dimensione della rete. Le reti sono classificate in tre categorie:

- Classe A: la rete è indicata dal solo primo *byte*; i restanti 3 *byte* indicano i nodi al suo interno. Sono reti di grandi dimensioni: con 3 *byte* a disposizione per i nodi, ciascuna rete di classe A può avere 224 nodi; per contro, possono esistere solo 256 reti di classe A.
- Classe B: la rete è indicata da due *byte*, i nodi sono indicati dai restanti 2 *byte*; ciascuna rete di classe B può avere 216 nodi.
- Classe C: la rete è indicata da tre *bit*, i nodi sono indicati dall'unico *byte* restante; ciascuna rete di classe C può avere solo $2^8=256$ nodi.

Si comprenderà come l'assegnamento della classe ad una rete sia un'operazione di grande delicatezza. Una rete di classe C non ha grandi prospettive di crescita. Una rete di classe A è un risorsa estremamente rara, visto che ne possono esistere solo 256. La gestione dell'assegnamento di un numero (classe) - non ancora utilizzato - ad una rete è affidata ad un organismo centrale, la **Internet Assigned Number Authority**, IANA, www.iana.org. IANA delega poi analoghi organismi regionali all'assegnamento dei numeri **IP** all'interno delle relative *zone* geografiche. Per l'Europa, l'organismo di riferimento è la **RIPE NCC** (**R**éseaux **IP** **E**uropéens), www.ripe.net. I gestori delle singole reti fisiche sono responsabili dell'assegnamento dei numeri ai loro nodi.

Indirizzi simbolici di dominio

Gli indirizzi **IP** sono pensati per essere usati da... **Internet** Protocol!

Per l'uso umano essi sono difficili da ricordare e soggetti ad errori di trascrizione.

Sono stati pertanto introdotti indirizzi simbolici di dominio (o nomi logici), che indicano un nodo della rete con una sequenza di stringhe di caratteri (etichette, *label*) separate da punti. Per esempio, la macchina su cui sono state composte queste note ha nome logico papageno.cs.unibo.it. L'insieme e la struttura di questi nomi costituiscono il *Domain Name System*, o **DNS**, di **Internet**.

Anche **DNS** è strutturato in modo gerarchico. La struttura a livelli si legge a partire da destra. Nel nostro esempio, la sequenza (etichetta) [it](http://papageno.cs.unibo.it) contraddistingue il livello (il dominio) più alto nella gerarchia (è un TLD, *Top Level Domain*, o anche dominio di primo livello) e corrisponde al dominio italiano; [unibo](http://papageno.cs.unibo.it) indica il livello successivo (il dominio di secondo livello), quello dell'università di Bologna; [cs](http://papageno.cs.unibo.it) è il Dipartimento di Scienze dell'Informazione; [papageno](http://papageno.cs.unibo.it) è la singola

macchina. Non vi è alcun limite al numero di livelli di un nome logico.

Ad un nome logico corrisponde un unico indirizzo **IP**, ma può accadere che ad uno stesso indirizzo **IP** siano assegnati più nomi logici. Per esempio i nomi pop.cs.unibo.it, www.cs.unibo.it, leporello.cs.unibo.it corrispondono tutti al medesimo nodo, con indirizzo **IP** 130.136.1.110.

La traduzione (in gergo: resolving) del nome logico di dominio nel corrispondente indirizzo **IP** è compito di un particolare servizio svolto da alcuni nodi della rete. Un **Domain Name Resolver** è un servizio fornito da un nodo che, mantenendo specifiche tabelle, è in grado di rispondere alla domanda: qual è l'indirizzo **IP** corrispondente ad un certo nome logico?

Un sinonimo di **Domain Name Resolver** è **Domain Name Server**, **DNS**. Si osservi che la sigla **DNS** non è usata in modo coerente. Talvolta indica l'insieme dei nomi simbolici (la S sta per *System*); altre volte indica un *resolver* (la S sta per *Server*). In particolare **DNS** come un particolare *resolver* è usata nelle finestre di configurazione di alcuni sistemi operativi (o meglio del loro modulo di rete che gestisce i protocolli **TCP/IP**). In quel contesto nella finestrella indicata con **DNS** deve essere indicato l'indirizzo **IP** del *resolver* che verrà utilizzato da quella macchina.

La vastità della rete non permette ad un singolo *resolver* di mantenere tabelle complete. Ma se non conosce la risposta per uno specifico dominio, un *resolver* sa quale altro *resolver* potrebbe avere l'informazione, e gira la domanda a quest'ultimo. In caso di bisogno può sempre risalire ad uno dei *resolver* radice (ce ne sono 13 sparsi per il mondo), ciascuno dei quali conosce l'indirizzo dei *resolver* di tutti i top *level domain*.

La presenza dei *resolver* permette di svincolare gli utenti dagli indirizzi **IP**. Il gestore di una rete può cambiare la macchina fisica che corrisponde ad un nome logico senza che gli utenti dei servizi di quella macchina se ne accorgano. Basta comunicare alla gestione del **DNS** che un certo nome logico corrisponde ora ad un nuovo indirizzo **IP**.

Non vi è legame tra la gerarchia dei nomi logici e la gerarchia degli indirizzi **IP**; per esempio, i nodi del dominio it non fanno parte della medesima rete fisica.

Chi assegna i nomi logici di dominio

Come nel caso degli indirizzi **IP**, la responsabilità dell'assegnamento dei nomi di dominio è suddivisa in modo gerarchico:

- la **Internet Corporation for Internet Names and Numbers** (www.icann.org) è un organismo internazionale indipendente che sovrintende alla ripartizione e alla definizione dei domini di primo livello. Agli albori di **Internet** esistevano solo pochi i domini:
 - edu (*educational*: università e scuole);
 - com (commerciale);
 - mil (militare);
 - gov (governativo);
 - int (internazionale);
 - net (fornitori di connettività).

Successivamente sono stati introdotti domini nazionali (come it, fr, jp, ecc.) e, più recentemente, altri domini quali:

- biz (*business*);
- coop (*cooperative*);
- museum;
- pro (professioni).

- ogni dominio di primo livello ha un organismo di gestione:
- sovrintende alla registrazione dei nomi simbolici di quel dominio.
- una volta registrato quel nome può venire usato solo da chi l'ha registrato; è a questo livello che viene risolto, quando possibile, il contenzioso sulla legittimità della registrazione di un particolare nome (per esempio, chi è autorizzato a registrare il dominio martini.it, visto che Martini non è solo un cognome ma anche un marchio di fabbrica tutelato dalla legge?).
- i domini di terzo livello sono assegnati dal titolare del dominio di secondo livello;
- e così via.

Il dominio .it

Il dominio it gestito da <http://www.nic.it/>, dove hanno sede due organismi:

- *Registration Authority* Italiana: responsabile dell'assegnazione dei nomi e della gestione del **DNS** primario per il dominio .it;
- *Naming Authority* Italiana: organismo che stabilisce le procedure operative ed il regolamento in base al quale opera la *Registration Authority*.

Fisicamente, la *Registration Authority* italiana ha sede a Pisa, presso un istituto del Consiglio Nazionale delle Ricerche. La *Registration Authority* non interagisce direttamente con chi intende registrare un dominio. Il servizio di registrazione dei domini, per privati e aziende, svolto da intermediari, detti *maintainer*. Alcuni *maintainer* per il dominio it sono:

- <http://www.register.it/>
- <http://www.dominio.it/>

Chi coordina Internet

Si potrebbe pensare che la gestione di una rete così complessa e articolata richieda uno stretto controllo centralizzato. Nulla sarebbe più lontano dalla realtà.

Se si eccettuano le questioni tecniche che abbiamo discusso nelle prime sezioni, la potenza di impatto di **Internet** si deve in buona misura a due aspetti:

- la pubblicità delle sue scelte tecniche;
I protocolli usati e tutte le scelte tecniche che permettono ai vari nodi e alle varie applicazioni di cooperare sono completamente pubblici e accessibili a chiunque abbia voglia di dar loro un'occhiata; lo erano sin dai primi momenti sfruttando il *File Transfer Protocol* (FTP); lo sono oggi in quanto pubblicati sul **WWW**. Nessuno deve pagare diritti per usarli. Torneremo su questo punto quando parleremo di **RFC**, che sono la modalità con la quale questa pubblicità si attua.
- la presenza di organismi di coordinamento che hanno gestito le sole scelte dalla quali dipendeva la interoperabilità della rete (per esempio l'assegnamento degli indirizzi e dei domini).
Internet, nella sua globalità, non è proprietà di nessuno. Tuttavia vi sono delle scelte tecniche che devono essere necessariamente condivise, pena la scomparsa stessa della rete. Queste scelte sono prese da alcuni organismi che, mediante procedure consultive delle parti coinvolte (tecnici, ricerca, aziende, commercio, enti pubblici) definiscono una sorta di *standard* per molte questioni.
- Le organizzazioni che coordinano **Internet** sono state spesso riorganizzate, per rispondere meglio alle mutate condizioni della rete (il coordinamento di **ARPANET** era cosa ben diversa dalla gestione dell'**Internet** contemporaneo).

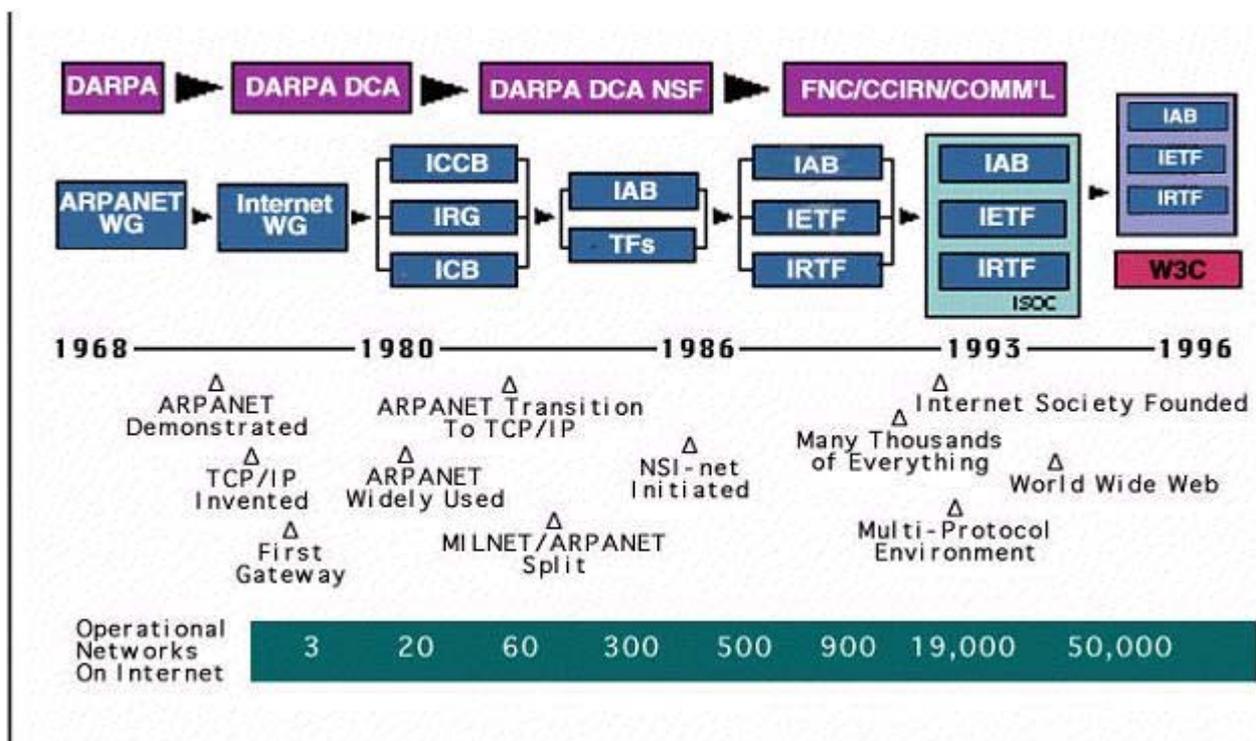
- Le sigle principali che coordinano **Internet** sono oggi:
 - **Internet Society** (www.isoc.org): un'associazione di privati e enti pubblici che ha come scopo lo sviluppo di **Internet** e il supporto dei due organismi tecnici IETF e IRTF;
 - **Internet Architecture Board** (www.iab.org): un comitato parte della **Internet Society** che nomina i membri dei due organismi tecnici;
 - **Internet Engineering Task Force** (www.ietf.org): un comitato che si occupa dei problemi tecnici diretti, in particolare della definizione degli *standard*, mediante la procedura degli **RFC**;
 - **Internet Research Task Force** (www.irtf.org): un comitato che coordina la ricerca di medio periodo;
 - **Internet Corporation for Internet Names and Numbers** (www.icann.org): un organismo internazionale indipendente che gestisce lo spazio dei nomi e degli indirizzi nelle sue trasformazioni di medio periodo.

La storia di Internet su un grafico

La figura che segue riassume in modo sintetico lo sviluppo di **Internet** dai suoi esordi alla fine degli anni 90. A partire dal basso troviamo:

- il numero di reti interconnesse;
- gli eventi più rilevanti;
- la linea del tempo;
- le organizzazioni che coordinano **Internet** da un punto di vista tecnico;
- le organizzazioni che sovrintendono a **Internet** da un punto di vista più vasto, incluso il suo finanziamento.

Fonte: *Internet Society*, <http://www.isoc.org/>



Un approfondimento discute in modo più ampio, tra le altre cose, di alcuni aspetti storici.

Il procedimento di standardizzazione: le RFC

Internet non ha un centro direzionale capace di imporre delle scelte. Le organizzazioni di coordinamento suggeriscono e definiscono degli *standard* comuni, lasciando poi ai produttori, ai fornitori di connettività, agli utenti, il compito di uniformarsi allo *standard*. L'autorevolezza del coordinamento, e la necessità di non esser buttati fuori dal mercato, fa sì che le indicazioni degli *standard* abbiano nei fatti un amplissimo seguito.

Per divenire uno *standard* di **Internet**, una proposta tecnica deve passare al vaglio di un lungo e approfondito procedimento di discussione, distribuito sulla rete e al quale tutti gli utenti, in linea di principio, possono partecipare.

Alla fine degli anni 60 la gestione di **ARPANET** decise che le proposte tecniche dovessero essere pubbliche. Tali rapporti tecnici furono denominati **RFC** (*Request for Comments*, richieste di commenti):

- sono disponibili sulla rete stessa (dapprima mediante ftp, poi mediante il www);
- sono numerati in ordine cronologico; **RFC** 1 è del 7 aprile del 1969; oggi (febbraio 2003) sono quasi 3500;
- tutti gli **RFC** sono disponibili a www.ietf.org/rfc.html;
- la consultazione e l'utilizzo del materiale degli **RFC** sono liberi e gratuiti.

Gli **RFC** sono redatti dai gruppi di lavoro tecnici dello IAB (**Internet** Architecture Board) e, come dice il loro nome, il loro scopo iniziale è quello di descrivere una soluzione tecnica, al fine di suscitare commenti (supporto, critiche, modifiche, ecc.).

Se la proposta contenuta in un **RFC** genera abbastanza interesse e consenso nella comunità, il contenuto passa nello stato di *standard* proposto. Dopo che l'interesse preliminare si è concretizzato in due implementazioni funzionanti ed indipendenti, il documento passa nello stato di *standard* preliminare (*draft standard*) e sottoposto agli organi di governo di IAB. Se le ulteriori sperimentazioni sono *positive* e IAB è convinta che la proposta è sensata e corretta, e che le implementazioni sono funzionanti, lo **RFC** viene approvato e dichiarato *Standard* di **Internet**.

La procedura di standardizzazione è definita essa stessa nel **RFC** 2026.

Non tutti gli **RFC** corrispondono a *standard*. Alcuni sono di tipo informativo o storico, altri descrivono esperimenti. Per esempio **RFC** 2055 descrive la storia degli **RFC**.

Internet: la definizione ufficiale

La definizione ufficiale di cosa sia **Internet** è contenuta in una risoluzione approvata all'unanimità il 24 ottobre 1995 dalla FNC:

"Internet" *refers to the global information system that -- (i) is logically linked together by a globally unique address space based on the **Internet** Protocol (**IP**) or its subsequent extensions/follow-ons; (ii) is able to support communications using the Transmission Control Protocol/**Internet** Protocol (**TCP/IP**) suite or its subsequent extensions/follow-ons, and/or other **IP**-compatible protocols; and (iii) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein.*

Con **"Internet"** si indica il sistema informativo globale che: (i) è logicamente connesso mediante un unico spazio globale di indirizzi basato sul protocollo **IP** o sulle sue estensioni; (ii) permette di supportare le comunicazione utilizzando la coppia di protocolli **TCP/IP** o le sue estensioni e/o altri protocolli compatibili con **IP**; (iii) fornisce, utilizza o rende accessibili, in modo pubblico o privato,

servizi ad alto livello sfruttando i livelli di comunicazione e le infrastrutture che sono stati descritti ai punti precedenti.

Comprendiamo adesso come (i) e (ii) si riferiscano alla complessa, ma elegante organizzazione tecnica responsabile della comunicazione e della connessione. Ma essenziale è anche il punto (iii): **Internet**, al livello delle applicazioni, appare come un grande, universale sistema informativo. La descrizione approfondita dei servizi di **Internet** a questo livello è compito dei moduli successivi.

Tra tutti i servizi, tuttavia, il *World Wide Web* ha assunto negli ultimi 10 anni un ruolo centrale e cruciale. Discuteremo la sua storia e le sue caratteristiche principali nelle prossime sezioni.

Intranet

La tecnologia con la quale **Internet** è realizzata (cioè la coppia di protocolli **TCP/IP** ed le applicazioni che su di essi sono basate) può essere usata anche per realizzare reti di dimensioni più ridotte.

Una rete aziendale, per esempio, può essere realizzata collegando calcolatori e *server* e facendoli dialogare usando **TCP/IP**. Anzi: questo è oggi il metodo più diffuso, e relativamente meno costoso, per realizzare una rete.

Una rete di questo tipo

- cioè una rete basata sulla tecnologia di interconnessione, sul *software* e le applicazioni che sono usati per **Internet**;
- realizzata per le esigenze di un privato;

è chiamata intranet.

Una intranet può essere del tutto scollegata dalla rete pubblica, oppure vi può essere connessa per il tramite di appositi *server* (*firewall*) che limitano sia l'accesso dall'esterno che l'uscita da parte degli utenti interni.

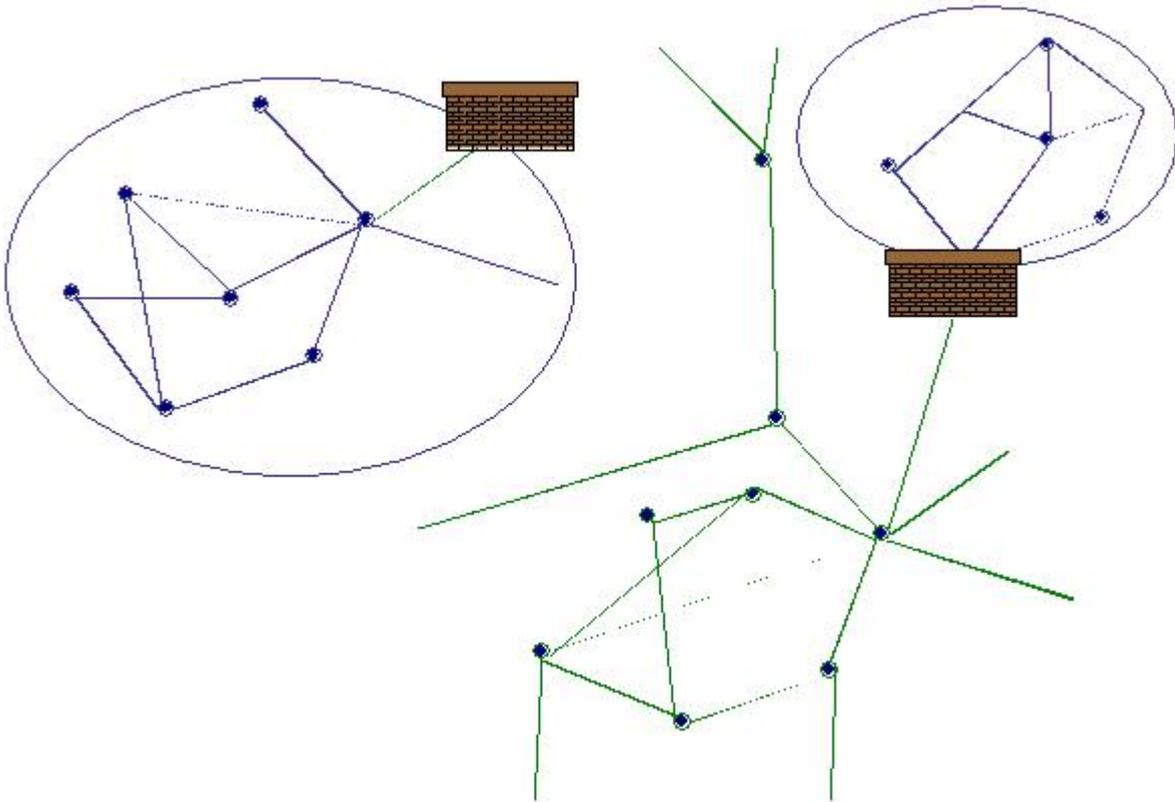
Un'azienda o un'istituzione può così basare i propri strumenti interni di comunicazione e cooperazione sulle applicazioni *standard* (posta elettronica, *web* browsers, ecc.), mantenendo rigidamente riservato l'accesso alla propria intranet.

Talvolta due sottoreti della stessa intranet sono collegate tra loro per mezzo di connessioni facenti parte della rete pubblica di **Internet**. Le questioni di sicurezza delle comunicazioni lungo i canali pubblici sono risolte mediante tecniche crittografiche.

I vantaggi di questa organizzazione:

- non è richiesto un investimento ulteriore (di *software*, *hardware* e di formazione del personale) per l'uso degli strumenti di cooperazione;
- lo sfruttamento delle connessioni pubbliche riduce i costi di connessione e di investimento nelle infrastrutture.

Nella figura che segue i due ovali azzurri costituiscono due sottoreti della stessa intranet. Le due sottoreti comunicano attraverso la parte pubblica di **Internet** (connessioni in verde). Sono indicati i *firewall* di separazione.



Extranet

Un'istituzione o un'azienda hanno spesso la necessità di condividere alcune informazioni (anche molto complesse e strutturate) con alcuni partner. Tali informazioni sono disponibili sulla intranet, ma non possono essere rese disponibili sulla parte pubblica di **Internet**.

Chiamiamo extranet l'utilizzo della tecnologia **Internet**/intranet per creare reti utilizzate da un insieme esteso di clienti o partner. Una extranet è sempre connessa alla parte pubblica di **Internet** e, come una intranet, si trova in genere al di là di uno o più *firewall*. È chiusa al pubblico generale, ma è aperta ad alcuni partner che non hanno accesso alla intranet.

I partner esterni alla intranet usano la parte pubblica pubblica di **Internet** per accedere al *firewall*; mediante una procedura di autenticazione (nel caso più semplice: nome e *password*) superano il *firewall* e accedono alla extranet. La loro autenticazione tuttavia non permette loro di accedere a tutte le informazioni riservate che fanno parte della intranet.

Come si capisce da ciò che abbiamo appena detto, il termine extranet non costituisce un concetto tecnico preciso. Spesso lo si utilizza per indicare la compresenza di reti pubbliche e *private*.

Si tratta in ogni caso di un modello organizzativo di largo uso.

Intranet e extranet: un esempio

Una università ha una propria intranet, costituita dalla rete su cui cooperano le varie sezioni amministrative (stipendi, personale, gestione delle carriere degli studenti, ecc.)

Le intranet corrispondenti alle sedi decentrate della stessa università sono collegate mediante connessioni pubbliche. I dati sensibili (per esempio tutti quelli riguardanti la carriera degli studenti) viaggiano in modo sicuro mediante tecniche crittografiche e sono protetti da ogni accesso dall'esterno della intranet.

Da **Internet** non è possibile accedere direttamente a tale intranet, per ovvi motivi di sicurezza e riservatezza.

Agli studenti, attraverso **Internet**, è però permesso accedere alla propria carriera, per la modifica di alcuni dati anagrafici e la visione della carriera stessa. La rete dedicata a questi aspetti costituisce (parte del-) la extranet della rete che stiamo considerando.

I costi di Internet

I costi di **Internet** sono di due tipi:

- costo della struttura di interconnessione e dell'effettivo trasferimento dei dati;
- costo dei servizi, cioè delle informazioni messe a disposizione sulla rete stessa.

I costi della seconda categoria non ci interessano qui: ogni fornitore di servizi li vende (o li regala) ai propri clienti al costo e con le modalità di pagamento (a forfait, a blocchi, a tempo, a dimensione, ecc.) che ritiene più adeguati.

I costi della prima categoria sono, invece, i costi intrinseci di **Internet**: senza queste spese **Internet** semplicemente non esisterebbe.

Una delle scelte progettuali originali di **Internet** è che i suoi costi intrinseci siano ripartiti tra tutti gli utenti, secondo il tipo di connessione di cui dispongono.

Ciascun utente paga una porzione di connessione: quella che lo collega al proprio (od ai propri) **ISP**. Un utente domestico pagherà la propria compagnia telefonica per la connessione commutata, a tempo o a forfait. Un utente più sofisticato pagherà il canone della fornitura ADSL. Un'azienda potrebbe pagare una connessione dedicata in fibra ottica.

A sua volta un **ISP** compra connettività:

- da altri **ISP** (comprando grandi quantità di banda otterrà prezzi più bassi del piccolo utente e questo gli permetterà di rivenderla con qualche margine);
- dai proprietari della connessione fisica.

Parte di quello che l'utente domestico paga alla propria compagnia telefonica per una piccola connessione commutata viene così girato al proprietario della connessione internazionale satellitare, il cui costo è effettivamente ripartito tra tutti gli utenti.

Non tutti gli utenti pagano però la stessa somma per lo stesso servizio.

La ripartizione terminale dei costi è fatta in ultima analisi dalle compagnie telefoniche, che spesso sono anche le proprietarie delle connessioni fisiche. In paesi dove la connessione telefonica è costosa, anche **Internet** viene pagato di più, a parità di servizio (cioè di banda disponibile e tempo di connessione).

La connessione telefonica è tanto più costosa quanta meno concorrenza vige in un certo mercato. OCSE stimava nel 1996 che il costo della connessione di base fosse tre volte più costoso nei paesi con servizio telefonico in regime di monopolio rispetto a quelli in regime di libera concorrenza. Sempre nel 1996, il costo di un'utenza **Internet** domestica italiana era di circa 10 volte il costo della stessa utenza in USA o in Canada.

I servizi basati su Internet

I servizi principali che **Internet** fornisce sono l'oggetto principale dei prossimi moduli. Qui ci basta ricordare che, fino ai primi anni 90, i servizi più usati erano:

- la posta elettronica;
- lo scambio di dati (sotto forma di *file: file transfer*);
- l'accesso remoto (*remote login*).

Quest'utilizzo era consistente con la diffusione di **Internet** principalmente tra centri di ricerca; la maggiore applicazione commerciale era la posta elettronica, un servizio che MCI vendeva ai propri clienti americani.

Nei primi anni 90 fa la sua comparsa un nuovo servizio, che sarà destinato a diventare in brevissimo tempo il più diffuso e pervasivo: il *World Wide Web*, o **WWW**.

Si tratta di un sistema di (iper-)testi multimediali memorizzati in modo distribuito sulla rete, e che la tecnologia di **Internet** permette di visualizzare e visitare - seguendone i collegamenti - in modo semplice e intuitivo.

Nelle sezioni che seguono discuteremo brevemente di questi concetti. Alcuni moduli seguenti sono dedicati alle tecniche di progetto e di realizzazione di questi ipertesti.

Osserviamo intanto che il successo e la pervasività del **WWW** sono tali che, per molti, **Internet** è sinonimo di **WWW**. Si tratta di un errore tanto grave quanto diffuso.

Ipertesti

Un ipertesto è un documento che contiene al suo interno collegamenti ad altri documenti o a sezioni dello stesso documento. Le informazioni sono organizzate non in modo sequenziale, ma reticolare; l'utente può saltare da un punto all'altro del documento (o da un documento all'altro) seguendo i collegamenti.

Alcuni commenti critici di opere letterarie hanno la struttura di ipertesto:

- contengono collegamenti a passi paralleli della stessa opera;
- contengono riferimenti ad altre opere di critica, sotto forma di citazioni;
- contengono rimandi ad altre sezioni del commento.

La possibilità di creare ipertesti era ben nota all'ambiente letterario. Ma la struttura tradizionale del libro mal si presta a "realizzare" un ipertesto e a seguire i collegamenti interni ad esso.

Il **WWW** è un gigantesco ipertesto multimediale distribuito dotato di un'interfaccia di facile uso:

- ipertesto: la tecnologia informatica è usata per realizzare un ipertesto per il quale seguire un collegamento sia semplice quanto la lettura sequenziale di un testo;
- multimediale: ogni documento può essere composto non solo di testo, ma anche di informazioni espresse con *media* diversi: immagini, *video*, suono, ecc.
- distribuito: le diverse parti di questo gigantesco ipertesto risiedono su calcolatori diversi e distanti tra loro; l'utente non necessariamente è a conoscenza di dove e come le informazioni sono memorizzate;
- dotato di interfaccia di facile uso: la fruizione di questo ipertesto è mediata da un'interfaccia che si prende carico di visualizzare in modo uniforme i dati multimediali e di risolvere il problema del raggiungimento dei dati remoti. La disponibilità di questa interfaccia è la chiave

di volta del successo del **WWW**. Chiamiamo *browser* le applicazioni che realizzano questa interfaccia per l'utente.

Il Web: un primo assaggio

Un ipertesto è realizzato sul *Web* come un documento in cui alcune parti (porzioni di testo o immagini) svolgono contemporaneamente due ruoli:

- descrivono l'informazione che denotano;
- rappresentano un collegamento.

Un'opportuno modo di evidenziare tali porzioni di testo permette all'utente di individuare i collegamenti e, se vuole, di seguirli.

I collegamenti sono attivabili: svolgono il ruolo di pulsanti, che, se premuti, permettono di seguire il collegamento stesso.

La struttura del documento - che comprende la posizione ed il valore dei collegamenti - è descritta in un apposito linguaggio, lo *Hyper Text Markup Language (HTML)*.

HTML descrive la struttura del documento, non il modo in cui il documento deve essere stampato o visualizzato. **HTML** "marca" nel documento quello che costituisce un titolo, quello che costituisce un capitolo, un elenco, una citazione, ecc. Marca anche quello che è un collegamento (*link*), insieme al valore di tale collegamento, cioè del documento a cui trasferire l'utente se questi decide di seguire il collegamento.

Sono i *browser* a decidere come visualizzare la struttura descritta da **HTML**: *browser* diversi possono visualizzare in modo diverso lo stesso documento, a patto che rispettino la sua struttura (definita da un'opportuna "raccomandazione del **W3C**").

I collegamenti possono riferirsi a documenti remoti. Per realizzare ciò il *Web* sfrutta il sistema di indirizzamento di **Internet**. Il valore di un collegamento è un Uniform Resource Identifier (**URI**), costituito in sostanza da tre parti:

- un protocollo da utilizzare per recuperare il documento (p.e. http);
- un nome simbolico di dominio (p.e. www.cs.unibo.it);
- un cammino d'accesso, relativo a quel dominio, e che specifica un documento (p.e. /~martini/didattica/index.html).

I documenti che costituiscono l'ipertesto sono residenti su calcolatori su cui girano opportune applicazioni dette *web server*.

Un *web server* risponde alla richiesta di un *browser*, formulata secondo l'opportuno protocollo, e fornisce al *browser* il documento richiesto. Il più comune dei protocolli è lo *Hyper Text Transfer Protocol*, **HTTP**, usato proprio per la richiesta e la trasmissione di ipertesti in modalità non sicura.

Il web è costruito su Internet

Il *Web* non esisterebbe senza la struttura di interconnessione e i servizi di livello inferiore che **Internet** mette a disposizione.

Sentiamo quello che ha da dire a questo proposito Tim Berners-Lee, l'inventore del *Web* (da: www.w3.org/People/Berners-Lee/FAQ.html):

Il **Web** è uno spazio astratto (immaginario) di informazioni. Sul *Net* [**Internet**, n.d.t.] troviamo calcolatori - sul **Web** troviamo documenti, suoni, *video*, ... informazioni. Sul *Net*, le connessioni sono costituite da cavi tra calcolatori; sul **Web** le connessioni sono collegamenti (*link*) ipertestuali. Il **Web** esiste perché i programmi possono comunicare tra calcolatori sfruttando il *Net*. Il **Web** non potrebbe esistere senza *Net*. Il **Web** ha reso *Net* utile perché la gente è interessata all'informazione (senza contare la conoscenza e la saggezza!) e non vuole sapere granché di calcolatori e cavi.

Chi coordina il WWW

Il **WWW**, come **Internet**, è nato in un ambiente scientifico, per promuovere la collaborazione tra ricercatori. Come **Internet**, il **Web** nasce aperto, senza formati proprietari (cioè per l'uso dei quali si debbano pagare diritti) e senza una struttura di controllo centrale che potesse decidere cosa (e come) potesse stare sul **Web** e cosa no.

La rilevanza commerciale del **Web** era peraltro evidente sin dai suoi albori.

Era chiaro che qualche azienda avrebbe potuto introdurre dei formati proprietari. Sarebbe stato un grave handicap per la diffusione globale del **WWW**: formati e protocolli diversi sarebbero entrati in competizione.

Si pensi, per esempio, a quello che sta accadendo per i *video on-line*. Vi sono alcuni formati proprietari (per esempio Real Audio, © RealNetworks, o *Windows Media*, © Microsoft) che possono essere riprodotti solo mediante una specifica applicazione o *plug-in*. I vari formati, proprietari e non, competono per la propria fetta di mercato e di conseguenza l'utente deve disporre di più riproduttori, gratuiti o a pagamento a seconda delle scelte di *marketing* delle aziende proprietarie del marchio. Se la stessa situazione si fosse verificata col formato "standard" per i documenti testuali, la globalità e interoperabilità del **Web** sarebbero state compromesse: alcuni siti sarebbero stati visibili solo con determinati *browser*.

Come nel caso di **Internet**, c'era bisogno di:

- un organismo di supervisione che garantisse l'esistenza unitaria del **Web**;
- contemporaneamente tale organismo non doveva avere il potere di alterare in modo cruciale il carattere aperto e libero del **Web** stesso.

Sullo stile del governo di **Internet**, nel 1994 fu creato il *World Wide Web Consortium*, o **W3C** (www.w3.org), un'organizzazione senza fine di lucro che sovrintendesse allo sviluppo del **Web**, così come ISOC sovrintende a **Internet**.

Il **W3C** ha sede presso il Massachusetts *Institute of Technology*; la sua sede europea è presso lo *European Research Consortium in Informatics and Mathematics* (<http://www.ercim.org/>).

Lo scopo primario del **W3C** è quello di sviluppare protocolli, specifiche, *software* e strumenti che garantiscano l'interoperabilità del **Web**.

La partecipazione al consorzio è aperta a organizzazioni di qualsiasi tipo, commerciali, educative, governative, ecc.

Le decisioni del **W3C** sono prese, in analogia con le **RFC** di ISOC, mediante un processo di proposte, commenti e raccolta di consenso. Quando una proposta ha raggiunto un sufficiente consenso, viene pubblicata come "raccomandazione" (*recommendation*). **W3C** non impone (né potrebbe in alcun modo imporre) le proprie raccomandazioni. Ma la sua autorevolezza le rende per molti aspetti simili a dei veri e propri *standard*.

I tentativi di essere più furbi di altri e di introdurre surrettiziamente caratteristiche proprietarie continuano anche in presenza delle raccomandazioni del **W3C**. Si sono visti (e purtroppo si vedono tuttora) "siti XYZ enhanced" dove XYZ è il nome commerciale di uno specifico *browser*. Si tratta di siti progettati in deroga ai criteri generali di interoperabilità definiti dal **W3C** nelle proprie raccomandazioni e che sfruttano invece caratteristiche "non standard" gestite solo dal *browser* XYZ (e non da ogni *browser* che rispetti le raccomandazioni **W3C**). Per fortuna questa linea di pensiero sembra essere stata sconfitta dallo scarso gradimento degli utenti...

Le raccomandazioni del **W3C** sono liberamente consultabili a www.w3.org/TR/

Conclusioni

Al termine di questa parte introduttiva, dovrebbero essere chiari:

- la struttura di **Internet** come rete aperta;
- le principali organizzazioni che sovrintendono al suo funzionamento e sviluppo, in assenza di un controllo centralizzato;
- il fatto che **Internet** è principalmente un'infrastruttura di comunicazione su cui sono fornite le effettive applicazioni utilizzate dagli utenti finali;
- la particolare rilevanza assunta negli ultimi anni da quell'applicazione che è il **WWW**.

Gli approfondimenti forniti sono di due tipi.

- Un primo approfondimento, fruibile in tutti i percorsi, fornisce complementi sulla storia e l'organizzazione di **Internet**. In particolare, è trattata più in dettaglio la storia di **Internet**, mettendo in luce come le scelte progettuali sono venute gradualmente, ma chiaramente, emergendo.
- Gli altri tre approfondimenti, pensati per il percorso avanzato, hanno invece carattere tecnico e riguardano aspetti relativi ai protocolli di livello rete e di livello trasporto:
 - Il livello rete e il protocollo **IP**;
 - il livello trasporto e il protocollo **TCP**;
 - L'instradamento dei pacchetti.