

## Il World Wide Web

Il World Wide Web può essere definito come un sistema informativo non lineare che permette agli utenti di cercare e recuperare informazioni in qualsiasi parte del mondo. World Wide Web, infatti, è un'enorme rete di documenti contenenti testi, immagini ma anche suoni ed animazioni, distribuiti tra i vari nodi di Internet e collegati tra loro a formare un'infinita trama di percorsi. Lungo questa trama ogni utente può costruire i suoi percorsi di lettura, guidato dai suoi interessi e dalla sua curiosità.

Per realizzare questi documenti sono utilizzate una serie di convenzioni e artifici, che nel loro insieme costituiscono il linguaggio HTML, o HyperText Markup Language.

Poiché la dimensione e la struttura del Web richiamano l'immagine di un oceano di informazione, la consultazione delle sue pagine viene comunemente detta navigazione o, in modo ancora più fantasioso, surfing, ovvero fare del surf. La navigazione, chiaramente richiede una barca; nel nostro caso la barca è un apposito programma, la cui funzione è quella di richiamare dalla rete le pagine che l'utente desidera consultare e di mostrare sul video il loro contenuto.

Nel gergo telematico questi programmi sono chiamati browser, dall'inglese "To Browse" (Scorrere), poiché essi permettono appunto di scorrere i documenti; il browser visualizza le pagine Web in modalità grafica, nel senso che le varie sezioni del testo possono avere stili, colori e posizioni diverse ed essere combinate con immagini ed animazioni.

Alcune porzioni di testo o parti di immagini in una pagina Web corrispondono a collegamenti (link) ad altre pagine o ad altre sezioni della stessa pagina; se si tratta di testo in genere il link è evidenziato tramite un cambiamento nel colore del carattere, eventualmente associato alla sottolineatura. Per attivare un collegamento è sufficiente posizionare il puntatore su di esso, e premere il tasto sinistro del mouse.

Accanto alla navigazione fatta seguendo i link, è inoltre possibile visualizzare una determinata pagina specificandone l'indirizzo: basta digitarlo nell'apposito campo.

Nel giro di pochi anni World Wide Web ha avuto una crescita esponenziale. Si immagini che oggi si contano più di 50 milioni di pagine, che trattano degli argomenti più disparati, e che sono realizzate da fornitori di ogni tipo: si va dalle università alle industrie private (grandi e piccole),

dai centri di ricerca ai negozi, dalle imprese editoriali ai partiti politici. Vi sono poi le numerosissime home page personali realizzate dagli utenti della rete.

## Il concetto di Iper testo

È opportuno, a questo punto, prima di introdurre le caratteristiche di funzionamento di WWW, soffermarsi sul concetto di Iper testo.

Ogni strumento comunicativo determina la forma dei contenuti che veicola ed i modi con cui si può accedere a tali contenuti. Il libro, ad esempio, è un oggetto composto da fogli di carta limitati e rilegati in modo che una pagina segue l'altra, così come ogni pagina è composta da una serie di linee di scrittura che hanno un orientamento ben definito.

È naturale, dunque, pensare che l'informazione contenuta in un testo debba avere un'organizzazione sequenziale; ciò significa che il lettore di un libro, se desidera ricostruire il suo contenuto, è in un certo senso obbligato a seguire un percorso prefissato, linea dopo linea, pagina dopo pagina.

Un iper testo, invece, è un testo non necessariamente sequenziale, o, come suggerisce il prefisso "iper", un testo a più dimensioni.

È come se, leggendo un brano e spinti dalla curiosità, si potesse, attraverso una parola, passare in un altro brano, collegato logicamente al primo, e di lì ricominciare a leggere o decidere di fare un salto verso un altro brano ancora, e così via; ogni brano, infatti, può contenere uno o più collegamenti che lo mettono in relazione ad altri brani. Un iper testo, in altre parole, è costituito da un insieme di frammenti di testo interconnessi tra loro.

Per sfruttare questa nuova struttura dell'informazione, gli iper testi sono dotati di un sistema di interfaccia specifico. Per informare il lettore della presenza di un collegamento, detto comunemente link, alcune parole o simboli del brano visualizzato sono evidenziati attraverso un colore predefinito o una sottolineatura; agendo sul link, ad esempio tramite il clic del mouse, è possibile spostarsi verso un altro frammento di testo.

Normalmente i sistemi ipertestuali sono dotati di strumenti di navigazione che aiutano il lettore ad esplorare l'iper testo senza correre il rischio di perdersi: pulsanti di ritorno indietro, elenco delle pagine o dei frammenti di testo visitati, indici o mappe dei contenuti.

Si tenga presente, però, che la potenza espressiva di un iper testo può spesso sfociare nei cosiddetti problemi di navigazione, ovvero quei problemi classici di disorientamento dell'utente all'interno dello spazio delle

informazioni:

- sforzo di concentrazione per tentare di collocare l'attuale posizione;
- difficoltà di accesso ad un'informazione che si suppone esista nella rete;
- difficoltà nel rivedere un'informazione già presente nel percorso compiuto;
- difficoltà nel percepire tutti i link che partono dalla pagina visualizzata.

Ad un ipertesto, infine, possono essere associati svariati strumenti comunicativi, quali immagini, suoni o filmati; in tal caso, per indicare questo complesso veicolo informativo, si utilizza la parola Ipermedia, nata dall'associazione dei termini ipertesto e multimedia.

## Architettura del WWW

L'architettura alla base del funzionamento di WWW è di tipo client-server. In generale, si parla di architettura client-server tutte le volte che un'entità (il server) offre un servizio ed un'altra (il client) accede a quel servizio secondo una serie di regole di comunicazione.

Tutte le informazioni su Internet hanno origine all'interno dei server, computer come collegati in permanenza alla rete, sui quali girano particolari programmi (detti anch'essi server) che hanno il ruolo di fornire i dati a chi li richiede; qualsiasi computer, direttamente connesso ad Internet, può fornire informazioni, assumendo così il ruolo di server. Storicamente, i server erano principalmente macchine con il sistema operativo Unix, ma ora sono sempre più usati i sistemi Macintosh e Windows NT ed altri che sfruttano il sistema operativo gratuito Linux, disponibile anche su personal computer economici.

Un programma client WWW, chiamato in genere browser, può essere eseguito su diverse piattaforme, ma per sfruttare a fondo le caratteristiche multimediali di questo sistema, è consigliabile disporre di una macchina che abbia capacità grafiche e audio, oltre ad una connessione ad Internet abbastanza veloce da consentire il trasferimento di immagini, animazioni e suoni in tempi tollerabili.

Come fanno il server ed il client a comunicare tra loro?

Per ottenere un corretto scambio di informazioni, occorre definire il formato dei messaggi, la loro lunghezza, i caratteri di controllo e le procedure da seguire per recuperare gli errori di trasmissione; occorre cioè definire un insieme di regole che permettano di effettuare correttamente lo

scambio di informazioni.

Con il termine di Protocollo, allora, si intende proprio quell'insieme di regole che determinano lo scambio di informazioni attraverso un supporto fisico ben definito, generalmente ininfluente. Con l'introduzione e lo sviluppo dello standard OSI (di cui si parlerà in Appendice), la struttura di un protocollo è stata a sua volta suddivisa in sette livelli, ognuno dei quali non è conscio del significato del messaggio che gli proviene dal livello superiore e di quali siano le elaborazioni che su di esso compie il livello inferiore.

Il protocollo standard di comunicazione che tutti i computer collegati ad Internet devono usare è il protocollo TCP/IP. Oltre al TCP/IP, vi sono i protocolli applicativi, come HTTP, posti "sopra" TCP/IP nel senso che definiscono come dev'essere fatto il contenuto dei pacchetti TCP a seconda della specifica applicazione.

## TCP/IP

Il protocollo TCP/IP (Transmission Control Protocol/Internet Protocol) è in realtà costituito da due protocolli: il protocollo TCP suddivide all'origine i messaggi in diversi pacchetti, per poi ricomporli in unità una volta arrivati a destinazione, il protocollo IP ha il compito di fare da postino, indirizzando le singole porzioni di dati attraverso nodi multipli.

Affinché il protocollo IP possa svolgere il suo compito, ad ogni computer con accesso a tempo pieno in rete è assegnato un Indirizzo IP (IP Number o Quaterna Puntata). Esso è costituito da quattro byte, separati da un punto; per convenzione ogni byte è tradotto nel corrispondente valore decimale, compreso da 0 (00000000) a 255 (11111111). Un tipico indirizzo IP potrebbe essere il seguente: 195.31.190.31.

Si supponga che una persona, dal suo studio privato di Milano, decida di connettersi alla pagina Web dell'Università di Napoli. Nel trasferimento della pagina, i pacchetti di dati non viaggiano necessariamente lungo lo stesso percorso; infatti, poiché tale percorso potrebbe diventare intasato o non funzionante, l'IP deve assicurare che i pacchetti trovino in ogni caso percorsi alternativi verso la destinazione finale; affinché questo avvenga, il protocollo IP deve essere privo di connessione, cioè non deve richiedere un collegamento diretto tra mittente e destinatario in quanto i dati possono attraversare decine di reti diverse prima di giungere a destinazione; per tale motivo i pacchetti possono perdere l'ordine in cui sono stati inviati o possono addirittura andare persi. Bisogna fare in modo che questo network di comunicazione, piuttosto inaffidabile, diventi affidabile, perché i computer possano comunicare tra loro.

Occorre allora l'altro strato di protocollo chiamato TCP che ha il compito di garantire il successo della trasmissione. Quando il server (l'università di Napoli) è pronto ad inviare i dati al client, il TCP suddivide il file in pacchetti, ognuno con un indirizzo e con un numero per consentirne l'ordinamento finale; sul lato destinatario, il TCP riassembla i pacchetti nell'ordine corretto e chiede la ritrasmissione dei pacchetti eventualmente mancanti.

In sintesi, il protocollo TCP gestisce i pacchetti, mentre il protocollo IP li invia in rete.

Per comprendere meglio il funzionamento del TCP/IP, si supponga di voler mandare un libro ad un amico, ed il solo modo in cui ciò possa avvenire è attraverso l'invio di cartoline. Si dovrebbero tagliare le pagine del libro ed attaccare le stesse alla cartolina, tenendo conto che non tutte le cartoline avrebbero il numero della pagina di provenienza; l'amico potrebbe quindi ricevere le pagine senza alcun ordine. Sarebbe una buona idea numerare le cartoline in modo che l'amico possa metterle insieme in ordine. Si tenga conto, inoltre, che alcune di queste cartoline potrebbero andar perse, per cui si dovrebbero conservare le copie da inviare all'amico nel caso che non ne riceva qualcuna.

Come si fa a sapere quali cartoline l'amico ha ricevuto? Sarebbe una buona idea se l'amico scrivesse una cartolina ogni tanto e dicesse: "Ho ricevuto tutte le cartoline eccetto la numero 102"; ma bisogna tenere presente che la cartolina che dice quante cartoline l'amico ha ricevuto potrebbe a sua volta andare smarrita. Dunque, se non si riceve alcuna cartolina che indica quante altre ne sono state ricevute, si potrebbe cominciare a mandare copie delle cartoline delle quali non sono state registrate le ricevute. Quando, alla fine, si registrerà la ricevuta di tutte le cartoline, allora si potranno buttar via le copie tenute da parte nel caso fossero servite. Questo è tutto ciò che succede nel protocollo TCP.

L'indirizzo IP è formato dall'host number (che individua la macchina a cui va mandato il pacchetto) e un numero aggiuntivo, il port number, che identifica il programma ricevente tra quelli in esecuzione su quella macchina.

## HTTP

Il protocollo HTTP (HyperText Transfer Protocol) è utilizzato per il trasferimento di formati di dati tra server e client (solitamente un browser); i formati di dati comprendono il testo semplice, l'ipertesto, le immagini, i suoni e qualsiasi altra forma specificata nei tipi MIME.

Esso, operando su Internet, si basa sul protocollo TCP/IP ed è detto, per questo motivo, privo di stato.

Se un utente richiede un file selezionando un collegamento, HTTP opera secondo questi passaggi:

1. il browser del computer richiedente imposta una connessione con il computer trasmittente attraverso il protocollo TCP/IP, secondo una modalità client/server,
2. il browser (client) richiede il file al computer trasmittente (server),
3. il computer trasmittente (server) trasmette il file,
4. il computer trasmittente (server) conclude la connessione.

Il protocollo HTTP viene definito privo di stato in quanto non esiste traccia della transizione avvenuta in nessuno dei due computer; se l'utente vuole, infatti, rivedere la stessa pagina, il computer emittente la rinvierà completamente, a meno che essa non sia stata memorizzata nella memoria cache del computer ricevente.

Da quanto detto finora, si evince che le transazioni fondamentali supportate dal protocollo HTTP sono quattro:

1. Connessione,
2. Richiesta,
3. Risposta,
4. Chiusura della connessione.

## I nomi di computer e di dominio

Anche se i computer connessi ad Internet sono identificati dagli host number, normalmente essi sono individuati dagli utenti tramite nomi. Il nome di computer non è equivalente all'host number, poiché non basta a identificare univocamente una macchina; per questo il sistema assegna a ogni calcolatore anche un dominio.

Il nome di dominio serve per identificare un intero gruppo di computer collegati ad Internet. I domini più grandi sono quelli corrispondenti a intere nazioni; nella tabella seguente sono mostrati alcuni esempi:

it	Dominio italiano
jp	Dominio giapponese

---

uk	Dominio inglese
fr	Dominio francese
edu	Comprende il gruppo nazionale delle istituzioni educative. La maggior parte delle Università e delle scuole americane fanno parte di questo dominio
com	Comprende tutte le organizzazioni commerciali
gov	Comprende tutte le organizzazioni governative

Ovviamente, un dominio grande può comprenderne altri più piccoli; l'insieme dei computer dell'Università di Napoli, per esempio, sta nel sottodominio unina di it, ma questo sottodominio ha a sua volta dei sottodomini, uno dei quali è chiamato idis perché comprende le macchine del Polo della Città della Scienza di Bagnoli. Il nome completo di quest'ultimo sottodominio, in realtà, è idis.unina.it perché occorre elencare, oltre a quello corrente, anche tutti i livelli superiori.

E' a questo nome completo che corrisponde univocamente un host number, cioè 159.149.70.70. Ogni volta che si digita un nome, viene interpellato un apposito programma, il Domain Name System, per eseguirne la traduzione in numero.

## Introduzione al linguaggio HTML

Con il diffondersi del World Wide Web, una delle più diffuse applicazioni disponibili sulla rete Internet, si è sviluppato anche uno speciale linguaggio denominato Hypertext Markup Language (HTML) utilizzato per costruire i documenti presenti nel Web.

HTML, che tradotto in italiano suona come Linguaggio di Contrassegno Ipertestuale, rientra nella categoria dei linguaggi a marcatori; esso è cioè costituito da un insieme di istruzioni (dette marcatori, o tag), che servono a descrivere la struttura del documento.

In altre parole, un documento HTML è un file di testo composto da una serie di tag (e relativi attributi), attraverso i quali il documento presenta testo, immagini statiche, animazioni, suoni, così da creare un documento multimediale.

Ciò che si richiede quindi per scrivere un documento in HTML è la conoscenza della sintassi con cui i marcatori vengono introdotti all'interno del documento.

Un documento HTML non necessita di un particolare software o hardware per essere sviluppato. Ciò significa che tale documento potrà essere reso disponibile su diverse piattaforme e senza il vincolo del possesso di un particolare programma; è sufficiente un semplice elaboratore di testi, come il Blocco Note di Windows o il SimpleText sul Mac, e il salvataggio del relativo file con un'estensione *htm* o *html*.

Esistono alcuni programmi, i cosiddetti editor HTML, che consentono di realizzare documenti HTML in modo molto meno faticoso e con ottimi risultati. È comunque consigliabile acquisire una conoscenza dettagliata del linguaggio HTML prima di affidarsi completamente ad un editor HTML, che in ogni caso presenterà dei limiti intrinseci.

Fra gli editor HTML troviamo ad esempio Netscape Composer, incluso nel pacchetto di Netscape Communicator, Microsoft FrontPage e Macromedia Dreamweaver; anche con le ultime versioni di Microsoft Word è possibile creare una nuova pagina Web a partire da un documento di Word ed usare numerosi modelli che rappresentano i tipi più comuni di pagine Web.

Due sono le caratteristiche fondamentali del linguaggio HTML, riflesse anche dallo stesso nome:

- HTML è un linguaggio ipertestuale che permette di stabilire collegamenti (link) all'interno di uno stesso documento oppure fra più documenti nello stesso sito Web o in siti diversi.
- HTML è un linguaggio a marcatori che incorpora, oltre all'informazione che si vuole comunicare attraverso la pagina Web, particolari sequenze di caratteri (marcatori, tag) che descrivono la struttura e il comportamento del documento.

I Tag si dividono essenzialmente in due categorie:

- strutturali. Il loro utilizzo consente di determinare la forma del documento; essi, in altre parole, definiscono il documento stesso, i paragrafi di intestazione e di corpo, i titoli delle pagine Web ed altri identificatori del documento.
- di formattazione. Il loro utilizzo consente l'organizzazione e la formattazione dei documenti; essi consentono la scrittura di un testo colorato o in corsivo, di disegnare linee orizzontali o aggiungere tabelle.

I tag, per essere riconosciuti da un browser, sono delimitati dai separatori angolari "<" e ">"; alcuni tag, inoltre, hanno bisogno di un tag gemello, detto tag di fine, il cui compito è quello di delimitare la parte del

documento a cui applicare la proprietà del tag stesso.

Per distinguere il tag di inizio da quello di fine, a quest'ultimo viene aggiunto il simbolo "/". Nell'esempio che segue:

```
<B>Questo testo è scritto in <BR>grassetto, </B> il seguente no....
```

la parte di testo delimitata dai tag `<B>` e `</B>` sarà visualizzata in grassetto; la parola grassetto, invece, sarà visualizzata alla riga successiva, poiché preceduta dal tag `<BR>`; se fosse stato omissso il tag `</B>`, anche il resto del documento sarebbe stato visualizzato in grassetto. Si noti, infine, come il tag `<BR>` non sia accompagnato dal relativo tag di chiusura.

La maggior parte dei tag, per essere applicati in maniera ottimale, sono accompagnati da *Attributi* che hanno il compito di fornire informazioni aggiuntive alle proprietà dei tag stessi. Un attributo è definito in genere da un nome e da un valore, separati da un segno di uguale (=), anche se, in alcuni casi, è richiesto il solo nome dell'attributo.

Ad esempio, il tag `<BODY>`, comprensivo dell'attributo per la specificazione del colore giallo dello sfondo, assume la seguente forma:

```
<BODY BGCOLOR=#FFFF40>
```

La struttura fondamentale di un documento HTML è la seguente:

```
<HTML>
<HEAD>
----
----
</HEAD>
<BODY>
----
----
</BODY>
</HTML>
```

Il tag `<HTML>` indica l'inizio del documento e il corrispondente tag di chiusura `</HTML>` ne indica la fine; questo tag comunica al browser che il file prelevato è stato redatto secondo quel linguaggio di contrassegno che è l'HTML.

Se si omettessero i due tag `<HTML>`, alcuni browser potrebbero non caricare il documento o rappresentarlo in maniera non corretta; la causa di

tale anomalia è che HTML non è l'unico linguaggio di contrassegno esistente, per cui esso potrebbe essere confuso, ad esempio, con il linguaggio XML (Extensible Markup Language).

Una volta definiti i limiti del documento si passa alla suddivisione dello stesso in testata e corpo.

tag `<HEAD>` e `</HEAD>` racchiudono l'intestazione del documento dove si possono inserire informazioni che non vengono visualizzate nella pagina, ma che ne descrivono caratteristiche come titolo, autore e parole-chiave. Quando il browser incontra questi simboli capisce che si tratta d'istruzione da eseguire e non di testo da visualizzare sullo schermo.

I tag `<BODY>` e `</BODY>` delimitano l'inizio e la fine del contenuto della pagina Web; fra questi tag è compreso tutto il testo che compare nella pagina Web, inclusi i vari collegamenti.

L'intestazione di un file HTML è definita attraverso i tag di apertura e chiusura `<HEAD>` e `</HEAD>`. In essa possono essere contenute alcune informazioni fondamentali sul file stesso come il titolo, il formato del file, la data dell'ultima modifica o le parole chiave; tali informazioni possono essere utilizzate dai motori di ricerca senza dover caricare tutto il documento.

La sezione `<BODY>` è il corpo principale del documento HTML dove vanno inseriti tutti i contenuti che devono apparire nella pagina web. Alcune caratteristiche generali come lo sfondo della pagina, il colore del testo ed il colore dei collegamenti ipertestuali, possono essere specificate mediante i relativi attributi.