



Fondamenti di Informatica
Corso di laurea in Psicologia
Università di Roma "La Sapienza"
a.a. 2000/2001

Dispensa 5 Internet e World Wide Web

Riccardo Rosati

VERSIONE PRELIMINARE

Indice

1 Reti di computer	4
1.1 Classificazione delle reti	4
1.2 Larghezza di banda di una rete	4
2 Internet	5
2.1 Trasmissione dei dati su Internet	5
2.2 Servizi e protocolli di alto livello per Internet	6
2.3 Indirizzi Internet e DNS	7
2.4 Domini	7
2.5 Name server	8
2.6 Posta elettronica	9
3 Connettersi a Internet	10
3.1 Internet provider	10
3.2 Modalità di connessione	11
3.3 Collegamenti multilink	12
4 Il World Wide Web	12
4.1 Web server e web client	13
4.2 Le URL	14
4.3 HTML, ipertesti, multimedia	15
4.4 Pagine web e siti web	15
5 Il linguaggio HTML	15
5.1 Sintassi di HTML	16
5.2 Struttura di un documento HTML	17
5.3 Semantica di un documento HTML	18
5.4 I principali tag di HTML	18
5.5 Limiti di HTML	24
5.6 HTML e il web prossimo venturo	25
6 Web browser: programmi per navigare sul Web	26
6.1 I web browser più diffusi	27
6.2 Internet Explorer	28
6.3 La barra dei pulsanti di Internet Explorer	28
6.4 Navigare negli ipertesti	29
6.5 Specificare la URL	30
6.6 Connessione ad un Web server	30
6.7 Aprire le risorse	31
7 Motori di ricerca	31
8 Creare un documento HTML	33

Sommario

In questa dispensa presenteremo le nozioni di base relative a Internet e alla navigazione su rete. Introduciamo pertanto i concetti di reti di calcolatori e Internet, quindi descriveremo i principali servizi informativi basati su Internet, con particolare riguardo per il World Wide Web, che è la modalità più diffusa di accesso a Internet. In particolare, presenteremo i principali componenti dell'architettura del World Wide Web: web server, web client (browser), URL, dominio, protocollo HTTP, ipertesti e linguaggio HTML.

1 Reti di computer

I computer possono essere collegati tra di loro tramite *reti* di interconnessione, che permettono lo scambio di dati tra computer.

Il vantaggio di avere a disposizione una rete sta nel fatto che tutti i computer collegati in rete hanno la possibilità di condividere tra di loro dati, programmi e dispositivi periferici. Ad esempio, è possibile in una rete locale che i computer collegati in rete “condividano” la stessa stampante, cioè utilizzino tutti la stessa stampante.

Rispetto allo schema classico dell’elaboratore, la rete può quindi essere vista come un ulteriore dispositivo che permette l’input e l’output dei dati: la differenza con gli altri dispositivi periferici sta nel fatto che la rete fa sì che siano collegati al computer tutti i dispositivi che si trovano sulla rete stessa.

1.1 Classificazione delle reti

A seconda delle loro dimensioni, le reti si distinguono in:

- *reti locali* o LAN (Local Area Network), di dimensioni dell’ordine delle decine o centinaia di metri. Ad esempio, i computer di un ufficio situato in un unico edificio possono essere connessi tra loro da una rete locale;
- *reti metropolitane* o MAN (Metropolitan Area Network), di dimensioni dell’ordine dei chilometri, che tipicamente coprono l’area di una città. Ad esempio, le reti dei telefoni cellulari sono reti metropolitane;
- *reti geografiche* o WAN (Wide Area Network), che coprono aree ancora più vaste delle reti metropolitane. Ad esempio, la rete telefonica italiana è una rete geografica.

1.2 Larghezza di banda di una rete

L’unità di misura principale della capacità di una rete è data dalla sua *larghezza di banda*. La larghezza di banda determina la quantità di informazione che la rete è in grado di trasmettere in un certo tempo. La larghezza di banda si misura in *bit al secondo* (bps), e nei suoi multipli:

- kilobit per secondo (Kbps), pari a 1000 bit per secondo
- megabit per secondo (Mbps), pari a 1000 kilobit per secondo
- gigabit per secondo (Gbps), pari a 1000 megabit per secondo

Maggiore è la larghezza di banda, minore è il tempo necessario per trasmettere l’informazione. Ad esempio, un file di 1MB può essere trasmesso

in 8 secondi da una rete che ha larghezza di banda di 1 megabit per secondo ($1\text{MB} = 8 \text{ megabit}$), mentre viene trasmesso in 160 secondi da una rete che ha larghezza di banda di 50 kilobit per secondo.

2 Internet

Internet è una “rete di reti”, cioè una rete che collega tra loro altre reti (locali, metropolitane e geografiche), ed ha dimensioni mondiali. Tale è l'importanza di Internet che ormai viene usualmente chiamata “la” rete di computer.

Milioni di computer sono attualmente connessi a Internet: ciò significa che ognuno di tali computer è potenzialmente in grado di scambiare informazioni con tutti gli altri computer connessi alla rete. Pertanto, Internet può anche essere vista come una gigantesca sorgente informativa, perché mette a disposizione la massa di informazioni “in rete”, cioè memorizzate sui computer di tutto il mondo connessi ad Internet.

Internet è una rete in continua crescita: decine e decine di milioni di utenti di tutto il mondo ogni giorno usano Internet per scambiare informazione. La velocità della diffusione di Internet come mezzo di comunicazione di massa, che non ha precedenti nella storia dell'uomo, sta provocando rapidi e profondi cambiamenti sia nel modo stesso di comunicare che nelle attività più disparate (studio, lavoro, commercio, ecc.).

2.1 Trasmissione dei dati su Internet

L'accesso a Internet è regolato da diversi *protocolli*, cioè convenzioni che permettono lo scambio di informazioni attraverso la rete.

Il protocollo di basso livello, cioè quello che regola la trasmissione dei dati attraverso Internet, si chiama *TCP/IP* (Transmission Control Protocol/Internet Protocol). Tale protocollo stabilisce la modalità di comunicazione tra nodi della rete. Tutti coloro che vogliono comunicare via Internet devono usare il protocollo TCP/IP, così come tutti coloro che vogliono comunicare attraverso il servizio postale devono usare le regole imposte dalle Poste italiane.

In particolare, nel protocollo TCP/IP:

- i dati che devono essere trasmessi da un computer all'altro vengono divisi in “pacchetti” di dati, che viaggiano nella rete in modo indipendente l'uno dall'altro
- ogni pacchetto di dati contiene l’“indirizzo” (cioè il nodo di destinazione della rete) al quale il pacchetto deve giungere. Tale indirizzo è l'indirizzo Internet, che verrà descritto nel seguito

- a causa di interferenze e malfunzionamenti sulla rete, i pacchetti trasmessi possono non giungere a destinazione, o giungere in modo errato: il protocollo tuttavia si preoccupa di garantire (attraverso programmi di gestione delle comunicazioni in rete, che ritrasmettono i pacchetti di dati quando necessario) che tutti i pacchetti di dati giungano a destinazione, e che arrivino nello stesso ordine nel quale sono stati trasmessi

Il motivo per cui i dati sono suddivisi in pacchetti è dovuto principalmente al fatto che, a causa dei problemi fisici di trasmissione dei segnali sulla rete, più il messaggio è lungo, maggiore è la probabilità che il messaggio contenga errori quando viene ricevuto. La scomposizione in pacchetti facilita la gestione degli errori: infatti, se il messaggio contiene errori, vengono ritrasmessi solo i pacchetti contenenti errori, non tutto il messaggio.

Inoltre, così come in una città esistono diversi percorsi per arrivare da un punto all'altro, su Internet esistono diversi collegamenti fisici per andare da un nodo all'altro: la scomposizione del messaggio in pacchetti fa sì che ogni pacchetto possa viaggiare in modo indipendente attraverso la rete, cercando sempre la strada “migliore” (in base al “traffico” sulla rete) per arrivare a destinazione.

2.2 Servizi e protocolli di alto livello per Internet

Ad un livello più alto del protocollo TCP/IP, esistono molti protocolli per la comunicazione su Internet, progettati per i vari tipi di *servizi* che si vuole offrire, come scambio di file, posta elettronica, connessione remota, ecc.

Tra i protocolli più usati su Internet citiamo:

- *FTP* (File Transfer Protocol): è il protocollo Internet più usato per lo scambio di file attraverso Internet. In altre parole, è il protocollo usato per trasferire un file da un computer ad un altro, se entrambi i computer sono connessi a Internet
- *SMTP* (Simple Mail Transfer Protocol): è il protocollo Internet più usato per lo scambio di *email*, cioè messaggi di posta elettronica. Precisamente, SMTP stabilisce l'interazione tra due sistemi di posta elettronica e il formato dei messaggi che tali sistemi si scambiano per trasmettere la posta elettronica
- *TELNET*: è il protocollo Internet più usato per il servizio di connessione di “terminale remoto”. In pratica, attraverso il protocollo TELNET è possibile, per un utente che lavora su un computer, connettersi in modalità “terminale”, cioè attraverso una interfaccia a caratteri, ad un computer remoto connesso a Internet, ed interagire in questo modo con il computer remoto esattamente come se si trovasse davanti a tale computer

- *HTTP* (HyperText Transfer Protocol): è il protocollo Internet usato per lo scambio di documenti ipertestuali, del quale parleremo estesamente nel seguito

2.3 Indirizzi Internet e DNS

Ogni computer connesso a Internet ha un proprio *indirizzo Internet*, che è l'identificatore che viene usato sulla rete per accedere a tale computer.

Ogni indirizzo Internet è composto da una sequenza di 4 numeri compresi tra 0 e 255 e separati tra loro da un punto. Tale sequenza di numeri è l'analogo del numero di telefono sulla rete telefonica: così come per poter comunicare per telefono con un amico è indispensabile conoscere il suo numero di telefono, per poter comunicare con un certo computer è necessario conoscere il suo indirizzo Internet.

Per facilitare l'utilizzo della rete da parte degli utenti, esiste un sistema, detto *Domain Name System* (DNS) che permette di associare nomi simbolici ai computer su Internet. Attraverso il DNS è quindi possibile accedere ad un computer non tramite il suo indirizzo Internet, ma tramite un nome. Così ad esempio, per poter accedere al computer che contiene la versione elettronica di questa dispensa, è più facile ricordare il nome

`www.dis.uniroma1.it`

piuttosto che la sequenza di numeri

`151.100.16.20`

corrispondente all'indirizzo Internet di tale computer.

Pertanto, il DNS permette di disaccoppiare il nome mnemonico dall'indirizzo "fisico" del computer (cioè il suo indirizzo Internet): in tal modo, il nome mnemonico resta lo stesso anche quando, per ragioni tecniche, viene cambiato l'indirizzo Internet corrispondente.

2.4 Domini

Il DNS è basato sui concetti di *dominio* e di *nome di dominio*. Esiste infatti un insieme predefinito di nomi di domini di *primo livello*, che vengono usati in base alla nazione di provenienza e/o dal tipo di organizzazione che possiede il computer. Tali domini sono suddivisi in due gruppi:

- domini specializzati in base al tipo di organizzazione:
 - com per le società private di tipo commerciale
 - edu per le scuole e università statunitensi
 - gov per le istituzioni governative degli Stati Uniti
 - net per le risorse di rete

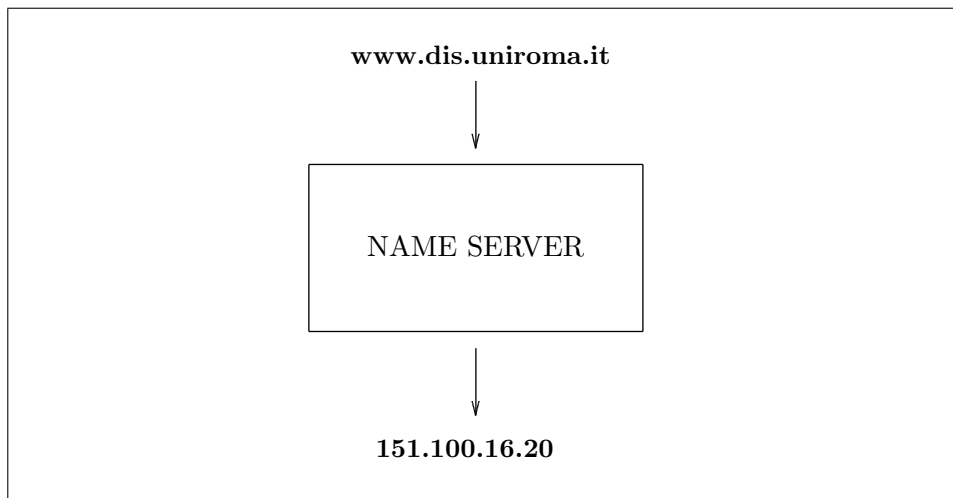


Figura 1: Il name server

- **mil** per le organizzazioni dell'esercito degli Stati Uniti
- **org** per le organizzazioni non-profit
- domini relativi alle singole nazioni del mondo. Tali domini hanno un nome corrispondente ad un codice a due lettere: per l'Italia tale codice è **it**

Ad ognuno di tali domini di primo livello può essere associato un insieme di nomi di dominio di secondo livello, ad ognuno dei quali è associato un insieme di nomi di terzo livello, e così via. Il nome del computer è ottenuto componendo i nomi dei domini dei vari livelli a cui appartiene. Tali nomi sono separati da un punto, e compaiono in ordine decrescente di livello.

Ad esempio, il nome

www.uniroma1.it

ha **it** come dominio di primo livello e **uniroma1** come dominio di secondo livello.

2.5 Name server

Per poter utilizzare i nomi simbolici del DNS, è necessario essere in grado di tradurre (o *risolvere*) ogni nome nell'indirizzo Internet corrispondente. Per far ciò si ricorre ai cosiddetti *name server*. Un name server non è altro che un computer che memorizza le corrispondenze tra nomi e indirizzi Internet per un certo insieme di nomi. Ogni zona dello spazio dei nomi dei domini è “coperta” da uno o più name server, che risolvono gli accessi relativi a tale insieme di nomi. Ad esempio, esiste un name server che risolve tutti i nomi che hanno come suffisso **dis.uniroma1.it**.

Quando si utilizza un nome per l'accesso a Internet, come prima operazione il nome viene trasformato in un indirizzo Internet. Per fare questo, si accede, dal computer che effettua l'accesso ad internet, ad un "proprio" name server, cioè ad un name server che gestisce, tra gli altri, anche il nome del computer che effettua l'accesso: se il nome cercato è "locale", cioè relativo allo stesso dominio o insieme di nomi risolti dal proprio name server, il name server comunica al computer l'indirizzo Internet trovato, altrimenti il name server inoltra la richiesta di indirizzo verso il name server che è in grado di risolvere il nome richiesto. Tale richiesta in genere "naviga" attraverso una catena di name server intermedi, che passano di volta in volta la richiesta ricevuta al name server successivo. Una volta raggiunto un name server che è in grado di risolvere il nome, l'indirizzo Internet corrispondente viene rispedito al name server iniziale che lo comunica al computer che effettua l'accesso.

Ad esempio, se dal computer `assi.dis.uniroma1.it` voglio accedere al computer il cui nome è `www.psi.uniroma1.it`, per prima cosa il computer accede al proprio name server. In questo caso, il name server risolve tutti i nomi del dominio `dis.uniroma1.it`, quindi non è direttamente in grado di risolvere il nome richiesto. Pertanto, inoltra la richiesta ad un name server successivo, che eventualmente interrogherà altri name server per ottenere l'indirizzo richiesto. In genere, una volta ottenuto l'indirizzo, questo viene memorizzato localmente dal name server per un certo periodo di tempo, durante il quale il name server è quindi in grado di risolvere direttamente nuove richieste relative al nome `www.psi.uniroma1.it`.

2.6 Posta elettronica

Un servizio su Internet molto usato è la posta elettronica o *email*. Il servizio di posta elettronica consente di spedire messaggi ad altri utenti raggiungibili via Internet.

Le caratteristiche principali della posta elettronica sono le seguenti:

- ogni utente della posta elettronica ha assegnato un proprio *indirizzo email*, che rappresenta a tutti gli effetti l'indirizzo che identifica l'utente per il sistema di posta elettronica
- sebbene il protocollo di posta elettronica sia stato originariamente progettato per il solo scambio di messaggi testuali, è in realtà possibile attraverso la posta elettronica spedire e ricevere non solo messaggi, ma anche qualunque tipo di file, attraverso il meccanismo dell'*allegato* (in inglese *attachment*)
- un messaggio di posta elettronica arriva al destinatario in genere entro pochi minuti (o addirittura secondi) dalla spedizione del messaggio

Per poter ricevere e spedire messaggi di posta elettronica un utente deve, quindi, avere accesso a Internet e avere un proprio indirizzo email; inoltre, deve avere a disposizione un programma per la gestione della posta elettronica: tali programmi permettono sia di leggere le email che l'utente riceve, sia di scrivere e spedire email ad altri utenti.

Il servizio di posta elettronica è gestito da un insieme di *mail server*, cioè programmi (in esecuzione permanente su computer connessi alla rete) che si scambiano tra loro i messaggi di posta. Ogni utente della posta elettronica è “servito” da un mail server, che si occupa, attraverso l'interazione con gli altri mail server presenti sulla rete, di far giungere a destinazione i messaggi spediti dall'utente, e di memorizzare i messaggi destinati all'utente. L'utente, attraverso i programmi di gestione della posta elettronica (detti *mail client*), si connette al proprio mail server e può sia scaricare i messaggi contenuti nella propria “casella postale” (mailbox), cioè i messaggi destinati all'utente e memorizzati dal mail server, che spedire messaggi di posta verso altri utenti.

I principali protocolli utilizzati per il servizio di posta elettronica sono di due tipi:

- protocolli per il trasferimento di messaggi tra mail server: ad esempio, il protocollo SMTP (Simple Mail Transfer Protocol);
- protocolli per il trasferimento di messaggi tra mail server e mail client: ad esempio, i protocolli POP3 (Post Office Protocol 3) e IMAP (Internet Message Access Protocol).

3 Connettersi a Internet

Abbiamo detto che, per poter accedere alle informazioni su Internet, è necessario che il computer sia connesso a tale rete. Internet è in realtà accessibile anche dai computer che non sono direttamente connessi a Internet (come ad esempio i personal computer domestici), tramite un *Internet service provider*, chiamato anche solo *Internet provider*, cioè un “fornitore di servizi Internet”.

3.1 Internet provider

Un Internet provider è una organizzazione che ha un computer (o più computer) connesso a Internet e permette ad un computer “cliente”, che non ha accesso a Internet, di connettersi alla rete tramite un collegamento (tipicamente via modem e linea telefonica) al proprio computer. In tal modo il computer del cliente viene temporaneamente connesso a Internet attraverso questa “appendice di rete” costituita dal collegamento con il computer del provider.

La connessione tra computer cliente e provider è stabilita di solito attraverso la linea telefonica nel seguente modo:

1. il computer del cliente telefona al computer del provider
2. su questa linea telefonica viene instaurata una connessione PPP (Point-to-Point Protocol)
3. tale protocollo assegna al computer cliente un indirizzo Internet per tutto il tempo della connessione: in tal modo il computer cliente è a tutti gli effetti un nodo della rete, finché la connessione PPP è attiva
4. quando il computer cliente decide di chiudere la connessione a Internet, viene chiusa la connessione PPP e quindi il collegamento telefonico col computer del provider

Di solito questo servizio ha un costo per il provider, che quindi fa pagare il cliente. In Italia attualmente la maggior parte dei provider non chiede canoni al cliente: i provider tuttavia incassano una percentuale del costo della connessione telefonica (pagato dal cliente).

Per permettere a tutti i clienti di pagare solo il costo di una telefonata urbana, i provider di solito sono presenti in tutte le maggiori città, il che significa che hanno computer connessi a Internet distribuiti in tutto il territorio. Tali computer vengono anche chiamati POP (Point Of Presence).

3.2 Modalità di connessione

Come detto in precedenza, per gli utenti privati il modo di connettersi a Internet è tramite un Internet provider. Come avviene in pratica tale connessione?

- innanzitutto bisogna avere un modem installato sul proprio computer e connesso alla linea telefonica
- bisogna quindi “registrarsi” (o abbonarsi) al servizio offerto dal provider. Per fare questo, solitamente si deve:
 1. procurarsi ed installare sul proprio computer un software distribuito gratuitamente dal provider (di solito su CD-ROM). Tale software effettua la configurazione del modem
 2. comporre quindi il numero telefonico del provider: a connessione avvenuta, si apre automaticamente un Web browser (ad esempio, Internet Explorer) che visualizza una pagina Web
 3. seguire le istruzioni riportate su tale pagina per effettuare l’abbonamento, o registrazione, al provider: attualmente l’abbonamento è gratuito per tutti i maggiori provider italiani

4. a registrazione avvenuta, chiudere il browser e chiudere la connessione

- una volta effettuata la registrazione, è possibile connettersi a Internet tramite un programma di connessione che, quando eseguito, telefona al POP prescelto del provider (quello corrispondente alla città da cui si chiama) ed instaura la connessione PPP con il computer del provider

Una volta ottenuta la connessione, è possibile usufruire di tutti i servizi informativi basati su Internet (World Wide Web, FTP, posta elettronica, telnet...): basta avere a disposizione sul proprio computer i programmi che permettono l'utilizzo di tali servizi. Solitamente, nel software distribuito dal provider ed installato sul proprio computer è già presente almeno il programma per la navigazione su World Wide Web (web browser).

3.3 Collegamenti multilink

Per velocizzare il collegamento ad Internet via modem, è possibile utilizzare la modalità chiamata *multilink*.

Un collegamento multilink non è altro che un collegamento tra due punti attraverso due o più linee di trasmissione fisiche, ma viste logicamente come un'unica linea di trasmissione. In pratica, i dati che viaggiano su tale linea vengono ripartiti lungo le linee fisiche in modo trasparente all'utente, che "vede" un unico canale di comunicazione: il vantaggio di tale tipo di collegamento sta nella maggiore velocità di comunicazione.

Sebbene sia possibile arrivare fino a 4 collegamenti paralleli, la tipica configurazione del collegamento multilink è di due linee. Per poter effettuare tale tipo di collegamento multilink sono necessari due modem, due linee telefoniche e un Internet provider che fornisca il servizio multilink.

Una volta attivato, tale tipo di collegamento permette una larghezza di banda (e quindi una velocità di trasmissione e ricezione dei dati) pari all'incirca al doppio della larghezza di banda del collegamento tramite un solo modem.

Sebbene il collegamento multilink sia stato progettato per i modem ISDN, può essere utilizzato anche da due o più modem analogici, anche se con tali tipi di modem analogici si possono causare problemi di sovraccarico sulla porta seriale, compromettendo le prestazioni del collegamento multilink.

4 Il World Wide Web

Il modo di accedere a Internet più diffuso attualmente è sicuramente il World Wide Web (o WWW). Con il termine *World Wide Web* si indica un sistema

di accesso a Internet basato sul protocollo HTTP (HyperText Transfer Protocol). Mediante tale protocollo è in pratica possibile accedere a documenti ipertestuali memorizzati sui computer raggiungibili via Internet.

Il World Wide Web nasce nel 1989 al CERN di Ginevra, dove un informatico di nome Tim Berners-Lee inventa il protocollo HTTP e il linguaggio HTML, allo scopo di permettere ai fisici dei vari centri di ricerca (senza competenze informatiche specifiche) di scambiarsi in modo semplice dati sperimentali e documenti tramite Internet.

Come vedremo in dettaglio nel seguito, il vantaggio del World Wide Web sugli altri protocolli per Internet sta nel fatto che questo sistema riesce a coniugare semplicità e potenza di utilizzo: infatti, attraverso i *web browser*, cioè i programmi che permettono l'accesso al World Wide Web, e il concetto di *collegamento ipertestuale* messo a disposizione da HTML, è estremamente semplice accedere ed usare la rete.

I vantaggi del World Wide Web ne hanno sancito un successo travolgente: nei pochi anni passati dalla sua apparizione (il 1990), il World Wide Web è diventato il più popolare sistema di accesso a Internet, e ha contribuito in modo decisivo all'“esplosione” del fenomeno Internet in tutto il mondo.

4.1 Web server e web client

Il World Wide Web è basato sul concetto di *web server*. Un computer connesso ad Internet è infatti accessibile tramite il World Wide Web (o, in altre parole, “sta” sul Web) solo se su tale computer è in esecuzione un programma chiamato appunto web server. Altrimenti, tale computer non è raggiungibile tramite World Wide Web, anche se può essere comunque acceduto via Internet da altri protocolli e servizi. Pertanto, Internet e World Wide Web *non* sono la stessa cosa: Internet è la rete, World Wide Web è un sistema di accesso alla rete.

Il programma web server risponde alle richieste che gli giungono via Internet da altri programmi detti *web client*: tipicamente tali programmi sono i web browser (quali Microsoft Internet Explorer e Netscape Navigator) in esecuzione su qualunque macchina connessa a Internet.

Il web server e il web client si “parlano” tramite il protocollo HTTP (HyperText Transfer Protocol). Precisamente, il web client richiede un documento HTML (o, più in generale, una risorsa, come spiegato nel seguito) presente sul computer dove è in esecuzione il web server. Il web server ricerca tale documento e lo trasmette (tramite Internet) al web client, cioè al browser, che a questo punto è in grado di presentarla all'utente che sta utilizzando il browser stesso e che ha richiesto la visualizzazione del documento.

4.2 Le URL

Il documento (o risorsa) richiesto dal web client deve essere specificato attraverso un nome chiamato *URL* (Uniform Resource Locator). La URL è infatti un nome che identifica univocamente ogni documento HTML e risorsa accessibile via World Wide Web.

Ogni URL specifica le seguenti informazioni:

- il protocollo, che è di solito HTTP ma che può essere anche diverso (ad esempio, FTP o email)
- il nome del dominio, cioè del web server che si vuole interrogare
- il nome della risorsa che si vuole accedere: nel caso del protocollo HTTP (ma anche FTP), tale nome corrisponde al nome del file (completo di percorso) nell'albero delle directory accessibile dal web server, che indica al web server quale file deve essere trasferito al web client

Ad esempio, la URL

```
http://www.dis.uniroma1.it/index.html
```

indica che si richiede al web server corrispondente all'indirizzo

```
www.dis.uniroma1.it
```

di trasferire, mediante il protocollo HTTP, il file `index.html`.

Se nella specifica della URL manca il protocollo, viene automaticamente completata assumendo che il protocollo sia HTTP. Ad esempio, la URL precedente può anche essere specificata nella forma

```
www.dis.uniroma1.it/index.html
```

Infine, anche la parte finale della URL (relativa al nome della risorsa) può essere incompleta e venire completata automaticamente. Ad esempio, se il nome del file che costituisce la risorsa è uno dei seguenti: `index.htm`, `index.html`, `default.htm`, o `default.html`, tale nome può essere omesso nella specifica della URL. Così ad esempio la URL precedente può essere scritta come

```
http://www.dis.uniroma1.it
```

o anche, omettendo il protocollo, soltanto come `www.dis.uniroma1.it`.

La risorsa descritta da una URL può essere non solo un file contenente un documento HTML ma anche, più in generale, un file contenente immagini, suoni o filmati, oppure un file contenente un programma.

Nel caso di protocolli diversi da HTTP e FTP, la URL non è in genere associata ad un file, ma ad un diverso tipo di risorsa. Ad esempio, la URL

```
mailto:rosati@dis.uniroma1.it
```

è associata all'azione di spedizione di un messaggio di posta elettronica all'indirizzo `rosati@dis.uniroma1.it`.

4.3 HTML, ipertesti, multimedia

I documenti su World Wide Web sono scritti in un linguaggio standard, chiamato *HTML* (Hypertext Markup Language). HTML è un linguaggio di formattazione di *ipertesti*, cioè di documenti testuali in cui è possibile inserire *collegamenti* o *link*, cioè riferimenti ad altri documenti o risorse sul Web.

Precisamente, un riferimento ad una risorsa è specificato mediante la URL corrispondente; inoltre, come vedremo più avanti, un collegamento è direttamente utilizzabile per accedere alla risorsa a cui ci si riferisce.

Attraverso l'uso dei collegamenti, un documento HTML può “contenere” immagini, suoni, filmati: in altre parole, un documento HTML è un oggetto *multimediale*, che cioè utilizza varie modalità di comunicazione (testo, audio, video, ecc.).

Infine, un documento HTML può interagire *dinamicamente* con l'utente, può cioè contenere al suo interno riferimenti a programmi che interagiscono con le azioni che l'utente compie, tramite mouse e tastiera, sul documento.

Per dare un'idea del successo e della diffusione del linguaggio HTML, basti pensare che attualmente (anno 2001) si stima che ci siano più di un miliardo di documenti HTML accessibili via World Wide Web, e che ogni giorno almeno un milione di nuovi documenti HTML si aggiunga all'insieme dei documenti *on-line*, cioè disponibili sul web.

4.4 Pagine web e siti web

Ogni singolo documento HTML accessibile dal World Wide Web è chiamato *pagina web*. Data la natura ipertestuale di tali documenti, le pagine web non sono quasi mai documenti “isolati”, ma sono collegate ad altre pagine disponibili in rete.

Un *sito web* è un insieme di pagine web relative ad uno stesso argomento, scritte da uno stesso autore o gruppo di autori, e organizzate logicamente in modo da essere accedute, attraverso i collegamenti ipertestuali, secondo le modalità imposte da chi ha scritto tali documenti.

Ogni sito web ha una sua *pagina iniziale* (o *home page*) che è la pagina da cui partire per “visitare” il sito, cioè per navigare (ricercando informazioni) all'interno delle pagine contenute nel sito.

5 Il linguaggio HTML

In questa sezione esaminiamo più da vicino il linguaggio HTML, che è il linguaggio standard per la specifica di documenti ipertestuali sul Web.

5.1 Sintassi di HTML

Un documento scritto in HTML è un documento testuale, scritto utilizzando il set di caratteri ASCII, cioè l'insieme di caratteri di testo "standard", che comprende i caratteri corrispondenti a lettere minuscole e maiuscole, le 10 cifre decimali, i simboli di interpunzione e alcuni simboli particolari. Pertanto, il documento HTML è in realtà un documento testuale ASCII, e quindi visualizzabile da qualsiasi editor di testi: tuttavia, tale documento testuale viene "interpretato" in termini di un ipertesto da appositi programmi (tipicamente i web browser), che visualizzano in modo particolare alcune parti del documento, sulla base delle marcature in esso presenti, come spiegato nel seguito.

Un documento HTML contiene due tipi di elementi:

- blocchi di testo
- *tag* (coè marcature o, impropriamente, "comandi")

Un tag è una stringa di testo ASCII delimitata dai simboli "<" e ">". Ad esempio, <head> è un tag HTML. Il termine tag significa "marcatura", o "etichetta": un tag viene infatti usato per marcare una parte di testo. HTML fa quindi parte della famiglia dei cosiddetti "linguaggi a marcatura".

Ci sono 2 tipi di tag:

- tag di apertura (marcatore iniziale), la cui forma è del tipo <nome-tag>
- tag di chiusura (marcatore finale), la cui forma è del tipo </nome-tag>

Ad esempio,

```
<h1>Topolino e Pippo</h1>
```

è un frammento di documento HTML in cui la parte di testo Topolino e Pippo è stata marcata con i tag <h1> e </h1>.

Ogni occorrenza di tag (di apertura) può contenere *assegnazioni di attributi*. Ad ogni tag è infatti associato un diverso insieme di possibili *attributi* che possono assumere diversi valori. Una assegnazione di attributo ha la forma

nome-attributo = valore

e la struttura di un tag con attributi è la seguente:

```
<nome-tag attributo1 = valore1 attributo2 = valore2 ...>
```

Ad esempio,

```
<font face="arial" size="+1">
```


è un tag HTML con assegnazione dei due attributi **face** e **size** rispettivamente ai valori "arial" e "+1". Come mostrato in questo esempio, i valori a cui gli attributi sono assegnati vanno scritti in genere tra virgolette (doppi apici).

Alcuni attributi del tag sono obbligatori (vanno sempre assegnati). Ad esempio, per il tag font è obbligatorio assegnare l'attributo **face**: pertanto, ad esempio, il tag

```
<font> e <font color="white">
```

non è sintatticamente corretto, mentre il tag

```
<font face="courier">
```

lo è.

Per HTML i tag (e i nomi degli attributi) sono “case-insensitive”, cioè possono indifferentemente essere scritti sia in minuscolo che in maiuscolo (ad esempio , e corrispondono allo stesso tag e sono tutti accettati dal linguaggio).

5.2 Struttura di un documento HTML

La struttura ad alto livello di un documento HTML è la seguente:

```
<html>
<head>
...
</head>
<body>
...
</body>
</html>
```

I tag <html> e </html> denotano rispettivamente l'inizio e la fine del documento. I tag <head> e </head> delimitano invece la cosiddetta *intestazione* (o *parte intestazione*) del documento, mentre i tag <body> e </body> delimitano il *corpo* del documento.

La parte intestazione del documento contiene *meta-informazione*, cioè informazione riguardante il documento stesso (ad esempio, autore del documento, parole chiave del documento, titolo associato al documento). Tale informazione ha quindi scopi di documentazione, ed è particolarmente importante per i motori di ricerca e più in generale per tutti i programmi che visitano il Web allo scopo di classificare i documenti in rete, per semplificare la ricerca di informazioni da parte degli utenti.

Il corpo del documento contiene invece il documento vero e proprio, cioè per intenderci quello che viene visualizzato dal browser.

5.3 Semantica di un documento HTML

Dal punto di vista semantico, in HTML esistono fondamentalmente due tipi di tag:

- tag di formattazione (per cambiare l'aspetto ad una parte di testo), come ad esempio il tag ``. Infatti, il frammento di codice HTML seguente:

```
<font face="courier" color="red">ciaio mamma</font>
```

viene interpretato nel seguente modo: la parte di testo “ciaio mamma” deve essere visualizzata all'interno del documento utilizzando il colore rosso e usando un particolare tipo di carattere (Courier).

- tag “semantici” (per dare “struttura” o “significato” ad una parte di testo), ad esempio il tag `<address>`. Il seguente frammento di codice:

```
<address>via dei Marsi, Roma</address>
```

ha il seguente significato: la parte di testo “via dei Marsi, Roma” deve essere considerata come un “indirizzo”. Questo tag non ha pertanto un significato direttamente associabile ad un attributo di presentazione su video, anche se in realtà i browser evidenziano (rispetto al resto del documento) le parti di testo delimitate dal tag `<address>`.

5.4 I principali tag di HTML

Vediamo ora i principali tag di HTML (oltre ai tag `<html>`, `<head>`, `<body>`, usati per definire la struttura ad alto livello del documento). Tra i tag definiti nel seguito, i tag di intestazione vengono usati nella parte intestazione del documento, mentre tutti gli altri tag vengono utilizzati nel corpo del documento.

Tag di intestazione. Sono i tag che vengono usati nella parte intestazione del documento, cioè all'interno dei tag `<head>` e `</head>`. I principali tag di intestazione sono:

- `<meta>`, che permette di definire “meta-informazione”, cioè informazione riguardante il documento stesso, all'interno della parte intestazione. I due attributi maggiormente utilizzati all'interno del tag `<meta>` sono **name**, che dichiara il tipo di meta-informazione contenuta nel tag, e **content**, che specifica la meta-informazione vera e propria. Ad esempio:

```
<meta name="author" content="Riccardo Rosati">
```

- `<title>`, che permette di assegnare un “titolo” al documento HTML. Tale titolo viene visualizzato sulla barra del titolo della finestra del browser (e non all’interno del documento stesso).

Tag di formattazione fisica. Sono i tag che permettono di formattare il documento per un browser visuale. Tra i principali tag di questo tipo citiamo:

- ``, che permette di scegliere il tipo di carattere (font) utilizzato, il colore e la grandezza del font di una parte di testo;
- `` (“bold”), che permette di visualizzare in neretto una parte di testo;
- `<i>` (“italic”), che permette di visualizzare in corsivo una parte di testo;
- `<u>` (“underlined”), che permette di sottolineare una parte di testo;
- `<hr>` (“horizontal row”), che permette di inserire una riga orizzontale nel testo;
- `
` (“break”), che permette di andare a capo nel testo.

Tag di strutturazione logica. Sono i tag che permettono di strutturare in varie forme il documento. Tra i principali tag di tale tipo citiamo:

- `<h1>`, `<h2>`, `<h3>`, `<h4>`, `<h5>`, `<h6>`, che permettono di dichiarare sei diversi livelli di “titoli” all’interno del documento;
- `` e ``, che permettono rispettivamente di “enfaticizzare” e “rafforzare” una parte di testo. Questi tag vengono chiamati tag di formattazione *logica*, a differenza dei tag fisici di formattazione visti in precedenza, in quanto non viene direttamente dichiarato il tipo di visualizzazione che si vuole ottenere, ma viene soltanto dichiarato in modo più astratto che si intende dare un attributo di formattazione “astratto” alla parte di testo stessa. In tal senso, la visualizzazione di una parte di testo racchiusa tra tali tipi di tag è dipendente dal contesto in cui tale parte di testo occorre: ad esempio, all’interno di un testo scritto in corsivo, una parte di testo marcata con `` viene visualizzata come testo non corsivo.
- `<address>`, già visto in precedenza, che permette di marcare come “indirizzo” una parte di testo;
- `<blockquote>`, che permette di dichiarare che una parte di testo è tratta da un’altra fonte;

- `<cite>`, che permette di citare una fonte di informazione;
- `<p>`, che permette di dichiarare che una parte di testo corrisponde ad un singolo paragrafo all'interno del testo;
- `<table>`, che permette di definire una tabella nel documento, per inserire informazione strutturata in forma tabellare (griglia di righe e colonne). In particolare, all'interno di una zona di testo delimitata dai tag `<table>` e `</table>`, è possibile utilizzare i tag `<th>` (“table header”) e `<td>` (“table data”) per inserire informazione in un elemento (o cella) della tabella, e il tag `<tr>` (“table row”) per suddividere in righe la tabella. Ad esempio, il seguente frammento di codice HTML

```
<table border=1>
<tr>
<th>nome</th> <th>cognome</th> <th>città</th>
</tr>
<tr><td>Mario</td> <td>Rossi</td> <td>Roma</td></tr>
<tr><td>Paola</td> <td>Bianchi</td> <td>Milano</td></tr>
</table>
```

viene visualizzato dal browser in forma tabellare, nel seguente modo:

nome	cognome	città
Mario	Rossi	Roma
Paola	Bianchi	Milano

- `` (che sta per “ordered list”) e `` (“unordered list”), che permettono rispettivamente di creare liste numerate e non numerate di elementi. Ad esempio, una lista numerata è racchiusa tra `` e ``, e ogni elemento della lista è a sua volta racchiuso tra `` e `` (`` sta per “list item”). Così il frammento di codice HTML

```
<ol>
<li>primo elemento della lista</li>
<li>secondo elemento della lista</li>
<li>terzo elemento della lista</li>
</ol>
```

viene rappresentato dal browser come una lista numerata, nel modo seguente:

1. primo elemento della lista
2. secondo elemento della lista
3. terzo elemento della lista

Tag per collegamenti ipertestuali. Il tag utilizzato per inserire collegamenti ipertestuali nel documento è `<a>`, chiamato tag “anchor” o “ancora”. Ad esempio, nel seguente frammento di codice HTML:

```
<a href="http://www.dis.uniroma1.it/~rosati/">ciao a tutti</a>
```

la URL

```
http://www.dis.uniroma1.it/~rosati/
```

viene associata alla parte di testo `ciao a tutti`. In tal modo è possibile, attraverso un browser, accedere alla URL collegata a tale parte di testo, direttamente cliccando sulla parte di testo stessa, e realizzando così, mediante l’uso del browser, la metafora dell’ipertesto. La parte di testo delimitata dal tag anchor viene evidenziata dal browser (di solito viene sottolineata ed è di colore diverso dal resto del testo).

Tag di inclusione di immagini e programmi. Esistono infine tag per associare immagini e programmi al documento HTML. Tali tag sono:

- ``, che permette di includere una immagine al documento HTML. Ad esempio, il tag

```

```

permette di includere nel documento (precisamente nel punto del documento in cui tale tag occorre) l’immagine contenuta nella URL

```
http://www.dis.uniroma1.it/~rosati/pippo.jpg
```

Ovviamente, la URL che viene assegnata all’attributo `src` del tag `` deve corrispondere ad un file contenente una immagine in uno dei formati immagine riconosciuti dal browser;

- `<applet>` e `<script>` sono invece tag che permettono di includere programmi nel documento HTML. Precisamente, il tag `<applet>` permette di includere nel documento un *applet*, cioè un particolare tipo di programma scritto nel linguaggio JAVA: come per le immagini, tale programma non è contenuto nel documento HTML, ma risiede in una URL a parte, che viene collegata al documento stesso tramite il tag `<applet>`. Il tag `<script>` permette invece di dichiarare, *all’interno del documento stesso*, programmi scritti in un particolare linguaggio di programmazione chiamato JavaScript. Tali programmi sono usati per aggiungere aspetti “dinamici” al documento (ad esempio, effetti di animazione grafica). Questi aspetti verranno approfonditi in una successiva dispensa.

A titolo di esempio, nella figura 2 è riportato un semplice documento HTML, e nella figura 3 è riportata la sua visualizzazione sul browser Internet Explorer.

```

<html>
<head>
<meta name="author" content="Riccardo Rosati">
<meta name="keywords" content="informatica, Internet, HTML">
<title>Dispensa 5 - Internet e World Wide Web</title>
</head>
<body>
<h1>Reti di computer</h1>
<p>I computer possono essere collegati tra di loro tramite
<em>reti</em> di interconnessione, che permettono lo scambio di
dati tra computer. </p>
<p>Il vantaggio di avere a disposizione una rete sta nel fatto
che... </p>
<h1>Internet</h1>
<p><em>Internet</em> &egrave; una "rete di reti", cio&egrave; una
rete che... </p>
In particolare, nel protocollo TCP/IP:
<ul>
<li> i dati che devono essere trasmessi...</li>
<li> ogni pacchetto di dati ... </li>
<li>a causa di interferenze e malfunzionamenti sulla rete,...</li>
</ul>

<hr width="100%">



<font face="arial" size="+1">Per ulteriori informazioni</font>
consultare la <a href="http://www.dis.uniroma1.it/~rosati/psico">home page
del corso</a>.
</body>
</html>

```

Figura 2: Esempio di documento HTML



Figura 3: Visualizzazione del documento HTML di figura 2

5.5 Limiti di HTML

Come abbiamo visto, il significato di un documento HTML è la combinazione di due diverse componenti:

1. “aspetto” del documento (cioè la sua visualizzazione mediante un browser)
2. struttura logica (o astratta) del documento, imposta dai tag utilizzati

Poiché i documenti HTML vengono acceduti dagli utenti attraverso i browser “visuali” (o “grafici”) come Netscape e Internet Explorer, la semantica “immediata” di un documento HTML per un utente umano è la sua visualizzazione sul browser. Per tale motivo, la stragrande maggioranza dei documenti su web è stata realizzata sulla base di tale significato “visuale”. Tuttavia, in tal modo il significato di un documento HTML diventa dipendente dal browser utilizzato, in quanto differenti browser interpretano HTML in modo diverso, cioè visualizzano lo stesso documento in modo diverso. Inoltre, si ha una perdita del significato legato al “contenuto” del documento. In particolare, utilizzare HTML essenzialmente come un linguaggio di formattazione di un testo ha come effetto quello di rendere poco o per nulla comprensibile il contenuto del documento ad “utenti” del Web diversi dagli utenti umani in grado di vedere il documento attraverso un browser grafico. Esempi di tali utenti sono:

- i *web crawlers*, coè i programmi che visitano il Web classificando tutti i documenti in base al loro contenuto
- gli utenti umani non vedenti, che accedono ai documenti sul Web tramite browser a sintesi vocale
- i browser testuali o quasi-testuali, utilizzati su dispositivi a ridotte capacità di visualizzazione (ad esempio i browser per i telefonini WAP)

Per comprendere l'importanza di questo problema, basti pensare che, data la mole di documenti presenti sul Web, la stragrande maggioranza degli accessi a documenti HTML da parte degli utenti umani viene fatta attraverso la consultazione di un motore di ricerca o di un catalogo, che sono costruiti sulla base di una classificazione automatica dei documenti effettuata da programmi (i *web crawlers*). Tali programmi non sono in grado di dare una interpretazione al documento sulla base della sua presentazione visuale su un browser, ma effettuano l'analisi delle sequenze di caratteri che costituiscono il codice HTML sorgente. Se quindi il contenuto del documento non viene correttamente interpretato dal programma che ha costruito l'indice del motore di ricerca o del catalogo, sarà molto difficile che tale documento possa essere acceduto dagli utenti realmente interessati a consultarlo,

in quanto il motore o il catalogo (utilizzati dagli utenti per la ricerca) lo avranno classificato in modo non corrispondente al suo reale contenuto.

In pratica, a causa dell'uso di HTML come linguaggio di formattazione, l'unica forma di classificazione dei documenti che è possibile effettuare in modo automatico su documenti HTML è in base alle occorrenze delle parole nel testo del documento, il che è molto lontano da una comprensione ad alto livello del contenuto (e quindi del significato) del documento.

5.6 HTML e il web prossimo venturo

Da quanto sopra esposto è evidente che, per chi scrive un documento HTML, nasce il problema di distinguere tra *contenuto* del documento e *presentazione* (o aspetto) del documento, in quanto, come abbiamo appena visto, è molto importante poter individuare il contenuto di un documento *indipendentemente* dalla formattazione del documento.

Nelle intenzioni, HTML doveva mantenere separati questi due aspetti. Nella realtà, non è così, in quanto i tag sono usati sia al fine di dare una struttura (e quindi un qualche “significato” astratto) al testo, che al fine di dare attributi di formattazione al testo. Questo in pratica ha significato che i documenti ipertestuali scritti in HTML mescolano, al loro interno, formattazione e struttura (o contenuto astratto). La maggiore conseguenza di questo fatto è che tali documenti non sono adatti per i dispositivi di accesso al web diversi dai browser visuali attuali. In prospettiva, ciò significa che la stragrande maggioranza dei documenti ipertestuali attualmente sul web è destinata a diventare obsoleta nel giro di pochissimi anni.

Per superare questi problemi, l'organismo preposto alla definizione degli standard relativi al World Wide Web, il *W3C* (World Wide Web Consortium), ha da tempo adottato nuovi linguaggi relativi alla definizione di ipertesti, che integrano il linguaggio HTML al fine di garantire la separazione tra forma e contenuto, e quindi permettere la fruizione dei documenti anche ai nuovi dispositivi di accesso al web.

I due più importanti linguaggi in tal senso sono:

1. il linguaggio di specifica di *fogli di stile* chiamato *CSS* (Cascading Style Sheets). Un foglio di stile non è altro che un insieme di specifiche di formattazione di un documento HTML. Mediante l'uso dei fogli di stile, è possibile distinguere tra forma e contenuto del documento HTML: in particolare, si divide la specifica del documento nel documento HTML vero e proprio (che non contiene alcun attributo di formattazione) e nel foglio di stile ad esso associato, che invece specifica le modalità di presentazione del documento stesso. Tale separazione ha il vantaggio che lo stesso documento HTML può essere acceduto da più dispositivi di accesso (per ognuno di essi va definito uno specifico foglio di stile da associare al documento), e in tal modo diventa una sorgente di

informazione indipendente dal mezzo che accede a tale sorgente, e quindi non obsoleta al variare dei mezzi di accesso al web.

2. il linguaggio *XML* (eXtensible Markup Language), destinato nel prossimo futuro a rimpiazzare HTML come linguaggio per la definizione di ipertesti sul web. Tale linguaggio, che può essere considerato come una “evoluzione” di HTML, fa propria ed estende l’idea, alla base dei fogli di stile, della separazione tra forma e contenuto del documento: inoltre, le caratteristiche del linguaggio XML permettono, rispetto ad HTML, di dare maggiore “significato astratto” al documento ipertestuale. In tal modo, XML dovrebbe segnare l’avvento del cosiddetto “web semantico”, cioè permettere la comprensione del vero “significato” dei documenti ipertestuali sul web non solo da parte degli esseri umani, ma anche da parte dei programmi che navigano sulla rete e classificano i documenti. Come abbiamo visto, da tale classificazione dipende in pratica l’esito delle ricerche di informazioni sul web, che vengono sempre effettuate mediante uno strumento di classificazione automatica come un motore di ricerca o un catalogo.

6 Web browser: programmi per navigare sul Web

Come detto in precedenza, il più diffuso tipo di “cliente” sul World Wide Web è il cosiddetto *web browser*, o programma per la navigazione in rete. Un web browser permette infatti di accedere ai documenti HTML (e, più in generale, a qualunque URL) accessibili dal Web in modo molto semplice. Attualmente i più diffusi web browser sono Microsoft Internet Explorer, sui sistemi operativi della famiglia Windows, e Netscape Navigator, sugli altri sistemi operativi.

Precisamente, un web browser svolge due funzioni:

1. permette di richiedere, in modo estremamente semplice, documenti HTML (e più in generale URL di qualunque tipo) a qualsiasi web server su Internet;
2. permette di visualizzare il documento HTML richiesto (e più in generale di “utilizzare” la risorsa corrispondente alla URL richiesta).

Inoltre, sfruttando i link ipertestuali presenti nei documenti HTML, il web browser permette l’immediata “navigazione” attraverso i documenti in rete collegati tra loro dai link ipertestuali. Più precisamente:

- se il documento 1 contiene un collegamento al documento 2, è possibile visualizzare il documento 2 semplicemente cliccando con il mouse in corrispondenza del collegamento sul documento 1: in corrispondenza a tale azione, il browser richiede il documento 2 al corrispondente web server indicato nella URL specificata nel collegamento

- attraverso opportuni comandi è possibile muoversi con facilità tra i documenti acceduti, ad esempio è possibile “tornare indietro” e visualizzare nuovamente i documenti precedentemente visitati nella navigazione

6.1 I web browser più diffusi

Esistono moltissimi web browser (programmi per la navigazione su web). I due programmi di gran lunga più diffusi sono *Microsoft Internet Explorer* e *Netscape Navigator*.

Questi due programmi presentano molte caratteristiche simili:

- sono entrambi programmi distribuiti gratuitamente via Internet (o direttamente distribuiti insieme al sistema operativo): non è quindi necessario acquistarli
- le operazioni che permettono di svolgere sono in pratica le stesse, come anche le modalità per svolgere tali operazioni e l'interfaccia grafica che mettono a disposizione

Accanto a tali somiglianze, Internet Explorer e Netscape Navigator presentano però alcune importanti differenze:

- il modo in cui vengono visualizzati i documenti HTML (in particolare, alcuni tag del linguaggio HTML) è differente. In pratica, questo significa che alcune pagine web hanno un aspetto diverso a seconda se sono visitate con Internet Explorer oppure con Netscape Navigator;
- a causa di tali differenze nella visualizzazione, alcuni siti web vengono progettati “ottimizzando” la visualizzazione per uno dei due browser, a scapito dell'altro
- essendo un software sviluppato dalla Microsoft, Internet Explorer presenta una maggiore “integrazione” con i sistemi operativi della famiglia Windows e con i programmi per l'automazione di ufficio della Microsoft, cioè i programmi della famiglia “Office”, tra cui l'editor di testi Word e il foglio elettronico Excel. In altre parole, usando Internet Explorer è più semplice interagire con questi programmi Microsoft piuttosto che usando Netscape.

Per quanto detto sopra, e in virtù del fatto che entrambi i programmi sono gratuiti, la scelta migliore per un utente che naviga sul web è probabilmente quella di avere installati entrambi i browser sul proprio computer, e di utilizzare l'uno o l'altro a seconda delle necessità.

6.2 Internet Explorer

Nel seguito vedremo in dettaglio le principali modalità di utilizzo di Internet Explorer, con l'avvertenza che, date le profonde analogie con l'interfaccia di Netscape, una volta che si è in grado di utilizzare Internet Explorer, si è in pratica in grado di usare anche Netscape.

La finestra del programma Internet Explorer è visualizzata in figura 3. In tale finestra, sono presenti un insieme di barre orizzontali. In particolare, è possibile visualizzare e quindi utilizzare le seguenti barre:

- nella parte superiore della finestra:
 - la barra dei menu
 - le seguenti barre degli strumenti:
 - * barra dei pulsanti, che permettono di svolgere le principali operazioni con il browser
 - * barra degli indirizzi, che riporta la URL correntemente acceduta dal browser
 - * barra dei collegamenti, mediante la quale è possibile accedere ad un insieme di collegamenti a URL (predefiniti o definiti da utente)
 - * barra radio, che permette di gestire l'accesso a siti web “radiofonici”, che cioè trasmettono programmi radio via Internet
- nella parte inferiore della finestra, la barra di stato, che visualizza informazioni relative alle operazioni che il browser sta eseguendo

Tali barre possono essere visualizzate o meno, utilizzando il menu **Visualizza** di Internet Explorer. Ad esempio, nella finestra riportata in figura 3 sono presenti la barra dei menu, la barra dei pulsanti, la barra degli indirizzi e, in fondo alla finestra, la barra di stato.

6.3 La barra dei pulsanti di Internet Explorer

La barra dei pulsanti di Internet Explorer contiene, da sinistra verso destra, i seguenti pulsanti:

- i pulsanti **Indietro** e **Avanti**, per muoversi all'interno delle pagine web accedute, come spiegato nel seguito;
- il pulsante **Termina**, per terminare il caricamento della pagina web corrente. Questa operazione può essere necessaria quando il caricamento della pagina è molto lento, oppure quando ci si rende conto che il contenuto della pagina web che è in corso di caricamento non è interessante, e non si vuole perdere ulteriormente tempo nel caricamento di tale pagina;

- il pulsante **Aggiorna**, per ricaricare la pagina web corrente. Questa operazione può essere necessaria quando il precedente caricamento della pagina non è stato completato correttamente;
- il pulsante **Pagina iniziale**, per caricare la pagina iniziale definita, come spiegato nel seguito;
- i pulsanti **Cerca**, **Preferiti** e **Cronologia**, che visualizzano nella parte sinistra della finestra di Internet Explorer delle sottofinestre che possono essere usate per effettuare rispettivamente operazioni di ricerca, accesso ai siti web preferiti e visualizzazione delle pagine web accedute in ordine cronologico;
- il pulsante **Posta**, che permette di eseguire operazioni per la gestione della posta elettronica (realizzata in genere attraverso il programma Outlook Express);
- il pulsante **Stampa**, per stampare la pagina web corrente;
- il pulsante **Modifica**, che permette di editare (cioè di modificare) la pagina web corrente attraverso un editor HTML (o un altro editor), come spiegato nel seguito.

6.4 Navigare negli ipertesti

Internet Explorer (così come gli altri web browser) permette di navigare facilmente sul World Wide Web, attraverso i collegamenti ipertestuali presenti nelle pagine web.

Innanzitutto, i collegamenti presenti nella pagina visualizzata vengono evidenziati rispetto al testo normale. In genere, i collegamenti vengono rappresentati nella pagina da parti di testo (singole parole o frasi) che vengono visualizzate dal browser in modo speciale (sottolineate e/o di colore diverso dal resto del testo); inoltre, se posizioniamo il mouse sopra ad una parte di testo corrispondente ad un collegamento, il puntatore del mouse cambia aspetto.

Per poter accedere ad un collegamento, cioè ad una pagina collegata alla pagina correntemente visualizzata, basta cliccare una volta con il tasto sinistro del mouse sul collegamento: la nuova pagina web verrà così visualizzata sulla finestra.

Per poter “tornare indietro” nella visualizzazione, cioè tornare a visualizzare la pagina precedente, è sufficiente cliccare sul pulsante **Indietro** della barra degli strumenti.

Dopo uno o più ritorni indietro, è possibile riandare “avanti” nella navigazione, cioè visualizzare le pagine successive a quella corrente e già visitate, attraverso il pulsante **Avanti** della barra degli strumenti.

È infine possibile visualizzare in Internet Explorer il codice HTML del documento, selezionando l'opzione **HTML** del menu **Visualizza**. Attivando tale opzione, viene visualizzato (in una finestra a parte) il codice HTML del documento correntemente visualizzato nella finestra del browser.

6.5 Specificare la URL

Oltre che tramite i collegamenti presenti nella pagina correntemente visualizzata, è possibile accedere ad una pagina web specificando direttamente la URL di tale pagina (che deve essere conosciuta a priori dall'utente).

In Internet Explorer, per inserire direttamente la URL, si può procedere in due modi equivalenti:

- scrivendo la URL direttamente sulla casella della barra degli indirizzi. Dopo aver scritto la URL, spingendo il tasto **Enter** della tastiera o il tasto **Vai** accanto alla casella della barra, il browser accede alla URL e visualizza la pagina web corrispondente;
- selezionando l'opzione **Apri...** del menu **File**. In tal modo si apre una finestra di dialogo in cui è possibile inserire (mediante la tastiera) la URL che si vuole accedere.

6.6 Connessione ad un Web server

Vediamo ora più in dettaglio cosa succede quando, cliccando su un collegamento o scrivendo una URL, chiediamo al browser di accedere ad una certa pagina web. Ricordiamo che il browser è un “web client”, e al momento di scaricare una pagina web, chiede ad un “web server” (il cui indirizzo è specificato nella prima parte della URL) di trasmettergli la pagina.

1. Per prima cosa, il browser deve stabilire l'indirizzo “reale”, cioè l'indirizzo Internet, del web server, a partire dalla URL. Per fare ciò, il browser interroga un name server, come spiegato in precedenza. Perciò il browser spedisce la URL al name server. Durante questa fase, sulla barra di stato del browser (cioè nella parte inferiore della finestra) compare il messaggio **Ricerca del sito:....**
2. Il name server restituisce l'indirizzo Internet corrispondente.
3. A questo punto il browser “contatta” il web server a cui fa riferimento la URL, chiedendo la risorsa corrispondente alla URL richiesta dall'utente. Durante questa fase, sulla barra di stato del browser compare dapprima il messaggio **Connessione al sito:....**, e successivamente il messaggio **Sito web rilevato**. In attesa di risposta....

4. Il web server trasmette al browser il contenuto della URL richiesta. Durante tale fase, sulla barra di stato del browser compare il messaggio **Apertura pagina...** mentre al termine del caricamento viene visualizzato il messaggio **Operazione completata**.

6.7 Aprire le risorse

Come già accennato in precedenza, oltre ai documenti HTML corrispondenti alle pagine web, le URL possono fare riferimento ad altri tipi di risorse presenti sul web, come ad esempio documenti non di tipo HTML, immagini, filmati, suoni, o programmi.

Tuttavia, qualunque web browser di per sé è in grado di visualizzare solamente documenti scritti nel linguaggio HTML ed immagini. Come fare quindi per “aprire” una generica URL che non può essere visualizzata direttamente dal browser? la soluzione sta nello specificare un programma associato ad ogni tipo di risorsa non direttamente gestibile dal browser. Così come viene associato un tipo, e quindi un programma, ad un file, con lo stesso meccanismo un web browser può associare un programma ad un certo tipo di URL.

Così ad esempio, è possibile associare un programma per la riproduzione di suoni alle URL corrispondenti alla rappresentazione digitale di suoni: in tal modo, quando viene acceduta una URL di questo tipo, il browser attiva automaticamente tale programma, che riproduce i suoni contenuti nella URL.

Quando viene acceduta una URL che non può aprire direttamente, il browser di solito effettua due diversi tipi di azioni:

- se il browser “riconosce” il tipo del documento, allora chiama automaticamente il programma che è in grado di “aprire” il documento stesso. Se ad esempio con Internet Explorer accediamo ad una URL che corrisponde ad un documento Word, il browser apre automaticamente il programma WordPad (o il programma Word se questo è installato) che è a sua volta in grado di visualizzare il documento
- se il browser non riconosce il tipo della risorsa, in quanto non è stato associato alcun programma a tale tipo di risorsa, allora in genere viene proposto all’utente di salvare il file sulla memoria di massa del computer

7 Motori di ricerca

Una importante classe di siti web è quella dei cosiddetti *motori di ricerca*. I motori di ricerca permettono di cercare informazione su World Wide Web (e su Internet in generale).

Tutti i motori di ricerca attualmente più diffusi effettuano le loro ricerche in base al testo: in parole povere, sono in grado di reperire le URL corrispondenti a documenti che contengono una combinazione di una o più parole. Pertanto, l'utente effettua una richiesta (in inglese *query*) al motore di ricerca inserendo una o più parole (in un'apposita zona, o campo, della pagina visualizzata dal browser), e il motore risponde alla richiesta visualizzando l'elenco delle URL corrispondenti ai documenti che contengono tali parole.

Come fa il motore di ricerca a rispondere ad una tale richiesta? ovviamente, non può accedere in tempo reale a tutti i documenti su Internet, che superano il miliardo secondo le stime attuali (anno 2001). Il motore invece accede a degli *indici*, che vengono costruiti e continuamente aggiornati da appositi programmi, detti *web crawlers*, che “navigano” in continuazione su Internet, visitando via via tutti i documenti accessibili e memorizzando sugli indici la URL del documento e le parole che occorrono in tale documento.

Pertanto, gli indici di un motore di ricerca contengono una quantità enorme di informazione, considerando il numero di documenti presenti su Internet. Ciò nonostante, tali indici sono progettati e costruiti per essere acceduti dal motore di ricerca, in base ad una parola o ad una combinazione di parole, in tempi brevissimi (dell'ordine dei decimi di secondo): il motore è così in grado, in tempo reale, di rispondere ad una richiesta dell'utente.

Esistono moltissimi motori di ricerca accessibili sul web. Alcuni sono specializzati su particolari temi, mentre altri sono generali, ed hanno l'ambizione di consentire la ricerca e l'accesso a tutti o quasi i documenti disponibili via Internet.

Tra i motori di ricerca più noti ed utilizzati citiamo:

- Altavista, accessibile all'indirizzo

`www.altavista.com`

Esiste anche una versione “italiana” di Altavista, all'indirizzo

`it.altavista.com`

- Lycos, accessibile all'indirizzo

`www.lycos.com`

Esiste anche una versione “italiana” di Lycos, all'indirizzo

`www.lycos.it`

- Google, accessibile all'indirizzo

`www.google.com`

- HotBot, accessibile all'indirizzo

`hotbot.lycos.com`

Esiste anche una versione “italiana” di HotBot, all’indirizzo

`www.hotbot.lycos.it`

- un motore di ricerca tutto italiano è Virgilio, accessibile all’indirizzo

`www.virgilio.it`

È interessante notare che la stessa ricerca produce in genere risultati diversi per i vari motori di ricerca. Questo è dovuto principalmente al fatto che gli indici dei vari motori sono aggiornati con tempi diversi, e quindi, ad esempio, documenti creati o modificati in tempi relativamente recenti possono già essere stati indicizzati da un motore e non da un altro. È quindi buona norma eseguire le ricerche usando più di un motore, specie se l’informazione cercata non viene trovata.

8 Creare un documento HTML

Oltre che consultare le pagine web, è possibile *creare* nuovi documenti HTML. Esistono diversi modi per creare un documento HTML:

1. tramite un normale editor di testi: in particolare, i programmi per la creazione di testi più potenti, come ad esempio Word, permettono di “convertire” un normale documento testuale in formato HTML, in modo che tale documento possa essere visualizzato da un web browser. Tuttavia, in tale tipo di documenti è piuttosto difficile includere molte caratteristiche del linguaggio HTML, per cui i documenti generati in questo modo non risultano particolarmente eleganti rispetto alle pagine web create con strumenti appositamente dedicati allo scopo;
2. tramite un *editor HTML*, cioè un programma per la elaborazione di documenti HTML. Questi programmi permettono di creare facilmente un documento HTML, mediante una interfaccia grafica intuitiva, che estende quella di un normale editor di testi con ulteriori opzioni per aggiungere caratteristiche “ipertestuali” al documento (collegamenti, inclusione di immagini e audio, ecc.);
3. scrivendo direttamente il documento nel linguaggio HTML. Tale possibilità è riservata a chi conosce il linguaggio ed è quindi in grado di digitare direttamente i tag HTML, senza passare per il tramite di una interfaccia grafica che traduce l’aspetto grafico del documento in un insieme di tag HTML. Dato che il codice HTML è normale testo in formato ASCII standard, è possibile scrivere un documento HTML utilizzando un semplice editor testuale ASCII (ad esempio, nei sistemi operativi della famiglia Windows, il Blocco Note o Notepad).

Oltre agli editor HTML di livello professionale e che vanno acquistati separatamente, esistono editor HTML disponibili gratuitamente. In particolare, all'interno dei web browser più comuni, è possibile attivare un editor HTML: nel caso di Internet Explorer, tale programma è Microsoft Front-Page Express, nel caso di Netscape Navigator, tale programma è Netscape Composer.

Questi editor HTML, pur essendo piuttosto semplici rispetto agli editor professionali, sono molto utili perché permettono di creare facilmente un documento HTML, mediante una interfaccia grafica molto intuitiva e simile a quella di un normale editor di testi.