

HyperText Markup Language

Introduzione al linguaggio per la creazione di ipertesti sul World-Wide Web

Marco Liverani
liverani@mat.uniroma1.it

26 Marzo 1997

Sommario

- Introduzione
- Cronologia dello sviluppo del World-Wide Web
- Formattazione del testo e dei paragrafi
- Liste ed elenchi ; immagini ; tabelle
- Collegamenti ipertestuali
- Form di inserimento dati
- Indicazioni per la creazione di pagine HTML
- Bibliografia

Introduzione

- Internet è una infrastruttura che interconnette un grandissimo numero di reti di calcolatori utilizzando il protocollo TCP/IP.
- Sopra al protocollo TCP/IP viaggiano informazioni codificate secondo altri protocolli, stabiliti dalle applicazioni utilizzate dagli utenti.
- Uno di questi protocolli di alto livello è l'*HyperText Transfer Protocol*, che consente l'accesso a risorse distribuite su diversi *host* sulla Rete con un approccio di tipo *ipertestuale*.
- I documenti ipertestuali sono codificati secondo le specifiche dettate dallo standard HTML : *HyperText Markup Language*.
- L'HTML è una applicazione dell'SGML (*Standard Generalized Markup Language*), uno standard internazionale per la formattazione di documenti (norma ISO 8879).
- HTTP identifica le risorse in rete mediante *URL (Universal Resource Locator)*:

http://magritte.sci.uniroma1.it:80/doc/index.html

The diagram shows the URL **http://magritte.sci.uniroma1.it:80/doc/index.html** with yellow brackets underneath. Below each bracket is a label: 'Protocollo' under 'http://', 'Host' under 'magritte', 'Dominio' under 'sci.uniroma1', 'Porta (socket)' under ':80', and 'Documento' under 'doc/index.html'.

Cronologia del WWW

- 1980 Tim Berners-Lee (CERN) sviluppa il programma "Enquire-Within-Upon-Everything" che consentiva di effettuare link tra diversi nodi.
- 1989 Tim Berners-Lee diffonde due documenti per raccogliere opinioni sul suo lavoro presso il CERN.
- 1990 Il documento di TBL viene riformulato con l'appoggio ufficiale del CERN ; viene coniato il nome World-Wide Web.
- 1991 Sviluppo dei primi client ed apertura del WWW server del CERN.
- 1992 Sviluppo del client con interfaccia X ; viene rilasciata la lista dei primi 26 server WWW.
- 1993 Viene rilasciato il primo browser per Macintosh ; viene rilasciato X-Mosaic di Marc Andreessen (NCSA) ; i server HTTP sono circa 50.
- 1994 Viene fondata la Mosaic Corporation (oggi Netscape Corp.) ; i siti WWW sono 1.500. A dicembre si tiene il primo meeting del "W3 Consortium" presso il MIT ; il CERN, a causa di problemi di budget, decide di sospendere il supporto del progetto WWW.
- 1995 Viene fondata la "Web Society" a Graz (Austria) : partecipano la Technical University di Graz, il CERN, l'Università del Minnesota e l'INRIA (Francia).

Generalità sull'HTML

- Ogni documento in formato HTML è costituito da una intestazione (HEAD) e da un corpo (BODY).
- Un documento HTML è un file di testo ASCII dotato di opportuni *tag* che svolgono diverse funzioni :
 - ☛ formattazione dello stile del testo ;
 - ☛ rimandi (link) ad altre risorse disponibili sul WWW ;
 - ☛ definizione di maschere di inserimento dati ;
 - ☛ inclusione di altri oggetti complessi.
- I *tag* HTML sono racchiusi tra i simboli di maggiore e minore : `<tag>`
- Alcuni *tag* prevedono un identificatore di inizio (`<tag>`) e di fine marcatura (`</tag>`), altri sono dei tag singoli (es. : `<HR>`, `
`).
- Esistono delle codifiche per rappresentare in modo universale ed indipendente dalla piattaforma i caratteri ASCII non standard (es. : lettere accentate).

Interazione con il client

- È il *browser* utilizzato dall'utente ad occuparsi di tradurre correttamente i *tag* HTML visualizzando una pagina leggibile ed includendo eventuali oggetti complessi (quali immagini, animazioni, suoni).
- Esempio:

```
<HTML>
  <HEAD>
    <TITLE>La mia pagina</TITLE>
  </HEAD>
  <BODY BGCOLOR=#ffffff>
    <H1>Benvenuti!</H1>
    Questa è la mia pagina.
    Visita il sito di <A HREF="http://www.mat.uniroma1.it">
    Matematica</A>.
  </BODY>
</HTML>
```



Formattazione del testo

- Titoli : `<H1> . . . </H1>`, `<H2> . . . </H2>`, ..., `<H6> . . . </H6>`
- Paragrafo: `<P ALIGN=RIGHT|LEFT|CENTER> . . . </P>`
- Fine riga : `
`
- Linea orizzontale : `<HR SIZE=n WIDTH=m ALIGN=x NOSHADE>`
- Dimensione, colore e tipo dei caratteri :
` . . . `
- Stile del testo : **Bold** (`...`), *Italic* (`<I>...</I>`), Typewriter (`<TT>...</TT>`), *Enfaticizzato* (`...`), Sottolineato (`<U>...</U>`)
- Testo a spaziatura fissa preformattato: `<PRE> . . . </PRE>`
- Ingrandimento (`<BIG> . . . </BIG>`) o riduzione (`<SMALL> . . . </SMALL>`) della dimensione del carattere
- Apici (`^{. . .}`) e pedici (`_{. . .}`)
- Lettere accentate : à=à è=è é=é
ì=ì ò=ò ù=ù
- Simboli : > = > ; < = < ; & = &

Liste ed elenchi

- Elenchi puntati :

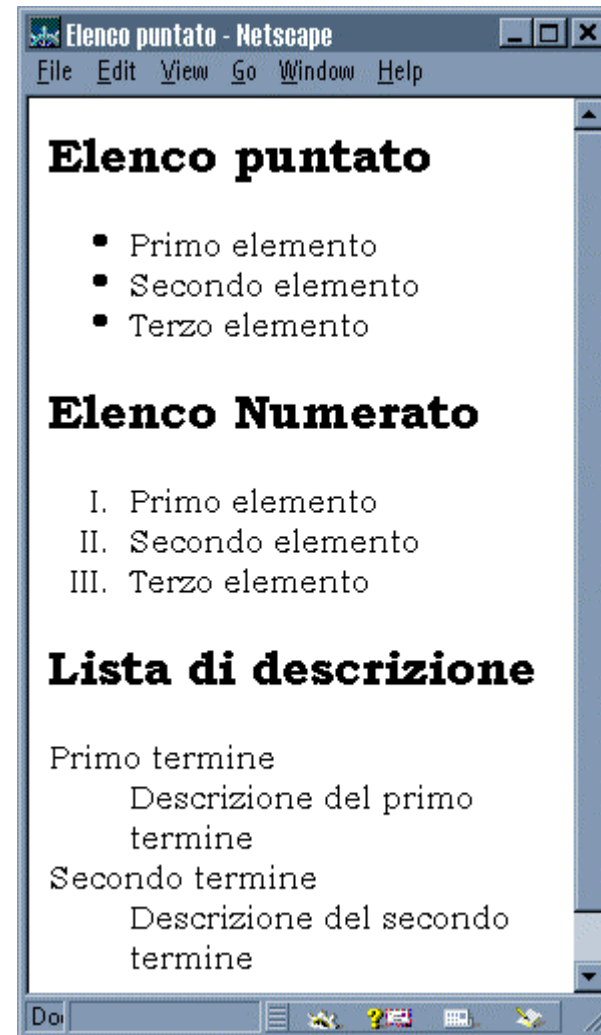
```
<UL TYPE=DISC | CIRCLE | SQUARE>  
  <LI> testo  
  <LI> testo  
</UL>
```

- Elenchi numerati :

```
<OL TYPE=A | a | I | i | 1>  
  <LI> testo  
  <LI> testo  
</OL>
```

- Liste di descrizione :

```
<DL>  
  <DT> Termine  
  <DD> Descrizione  
</DL>
```



Immagini

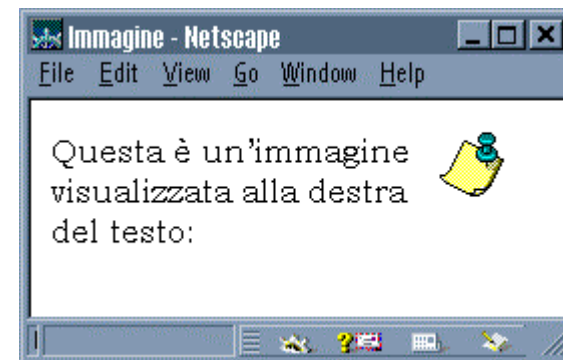
- È possibile includere nella pagina delle immagini ; è necessario indicare il nome del file in formato GIF o JPEG specificandone la collocazione mediante una URL :

```
<IMG SRC="http://host.domain/file.gif" ISMAP HEIGHT=h WIDTH=w  
BORDER=n ALIGN=RIGHT | LEFT | TOP | MIDDLE | BOTTOM ALT="testo  
alternativo">
```

- Esempio :

```
<IMG SRC="/gif/nota.gif"  
ALT="Nota!">
```

Questa è visualizzata
alla destra del testo:



- È possibile associare una immagine anche allo sfondo dell'intero documento :

```
<BODY BACKGROUND="http://host.domain/file.gif" BGCOLOR=#rrggbb  
TEXT=#rrggbb LINK=#rrggbb VLINK=#rrggbb ALINK=#rrggbb>
```

Tabelle

- Le tabelle consentono di articolare il *lay-out* della pagina, stabilendo dove collocare i diversi componenti (testi, titoli, immagini) :

```
<TABLE>
  <CAPTION> . . . </CAPTION>
  <TR>
    <TH BGCOLOR=#RRGGBB ALIGN=LEFT|CENTER|RIGHT
      VALIGN=TOP|MIDDLE|BOTTOM COLSPAN=n ROWSPAN=n
      NOWRAP> . . . </TH>
    <TH BGCOLOR=x ALIGN=y VALIGN=z > . . . </TH>
  </TR>
  <TR>
    <TD BGCOLOR=x ALIGN=y VALIGN=z > . . . </TD>
    <TD BGCOLOR=x ALIGN=y VALIGN=z > . . . </TD>
  </TR>
  . . .
</TABLE>
```

Collegamenti ipertestuali

- Definizione di una *hotword* con collegamento ad una URL :
`hotword`
- Definizione di un'*ancora* da utilizzare per un *link* ad un punto specifico all'interno del documento corrente :
`testo`
- Collegamento ipertestuale ad un'*ancora* definita all'interno di un documento :
`hotword`
- Definizione di un *link* ad un indirizzo di posta elettronica per l'invio di una E-mail :
`Nominativo`

Form di inserimento dati

- È possibile costruire delle maschere (*form*) di input di dati da passare come parametri a procedure CGI appositamente predisposte sul server web.
- Si deve indicare la URL che identifica la procedura CGI sul server che verrà eseguita come azione conseguente l'invio dei dati della *form* da parte dell'utente ; si deve anche indicare il metodo secondo cui vengono inviati i dati attraverso l'interfaccia CGI : attraverso una variabile di ambiente (metodo GET) ovvero attraverso lo *standard input* (metodo POST, utile per grosse quantità di dati).
- Per ogni campo di input presente nella *form* deve essere indicato il tipo, il nome della variabile ad esso associata ed un eventuale valore di default.

Form di inserimento dati - 2

- Campo di input di testo :

`<INPUT NAME=nome SIZE=n TYPE=HIDDEN MAXLENGTH=m>`

- Campo di input di testo su più righe : `<TEXTAREA NAME=nome COLS=n ROWS=m WRAP=PHYSICAL | VIRTUAL> . . . </TEXTAREA>`

- Liste di opzioni :

`<SELECT NAME=nome SIZE=n>`

`<OPTION VALUE="valore"> Etichetta`

`. . .`

`</SELECT>`

- Checkbox : `<INPUT TYPE=CHECKBOX NAME=nome VALUE="valore" CHECKED>`
- Radio button : `<INPUT TYPE=RADIO NAME=nome VALUE="valore" CHECKED>`
- Bottone di invio : `<INPUT TYPE=SUBMIT VALUE="Etichetta">`
- Bottone di ripristino : `<INPUT TYPE=RESET VALUE="Etichetta">`

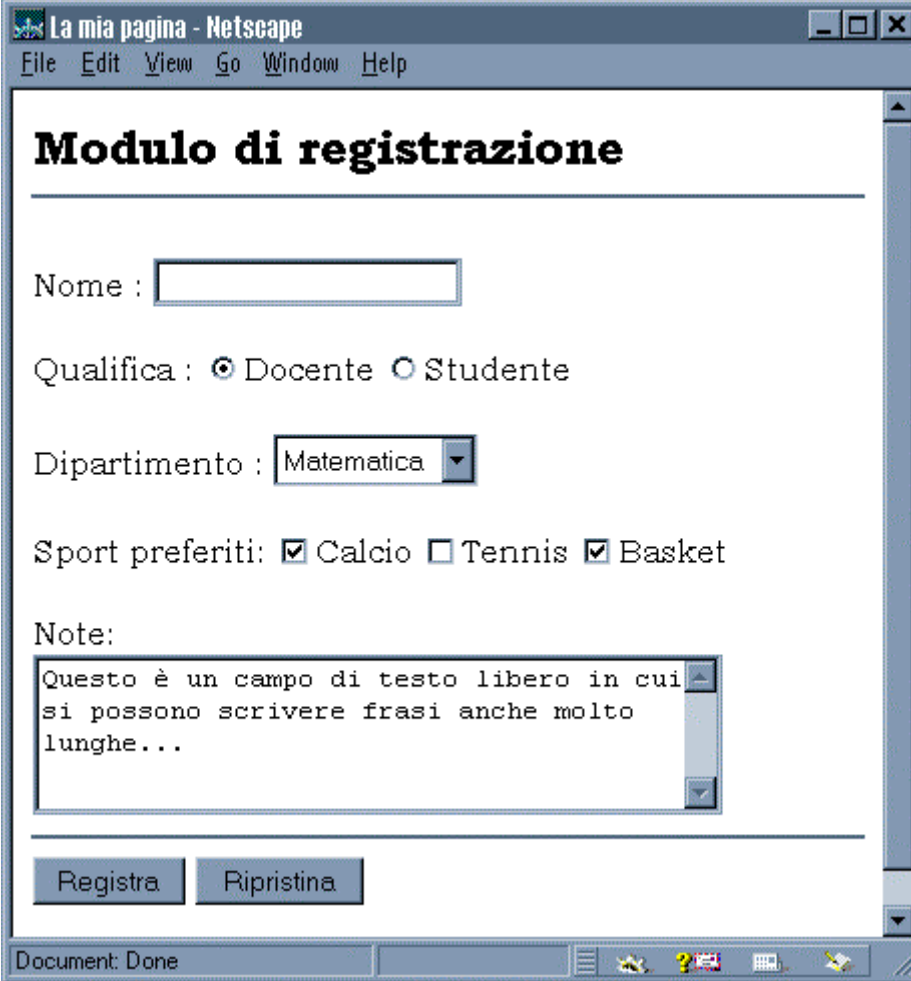
Form di inserimento dati - 3

- Esempio :

```
<FORM ACTION="/cgi-bin/registra.cgi" METHOD=GET>
  Nome : <INPUT NAME=nome SIZE=20> <BR>
  Qualifica :
  <INPUT TYPE=RADIO NAME=qualifica VALUE="prof"
  CHECKED>Docente
  <INPUT TYPE=RADIO NAME=qualifica
  VALUE="stud">Studente <BR>
  Dipartimento :
  <SELECT name=dip>
    <OPTION VALUE="chimica">Chimica
    <OPTION VALUE="fisica">Fisica
    <OPTION VALUE="matematica" SELECTED>Matematica
  </SELECT><BR>
  <INPUT TYPE=SUBMIT VALUE="Registra">
  <INPUT TYPE=RESET VALUE="Ripristina">
</FORM>
```

Form di inserimento dati - 4

- Esempio :



The screenshot shows a Netscape browser window titled "La mia pagina - Netscape". The browser's menu bar includes "File", "Edit", "View", "Go", "Window", and "Help". The main content area displays a registration form with the following elements:

- Modulo di registrazione**: The title of the form, underlined.
- Nome**: A text input field.
- Qualifica**: Radio buttons for "Docente" (selected) and "Studente".
- Dipartimento**: A dropdown menu currently showing "Matematica".
- Sport preferiti**: Checkboxes for "Calcio" (checked), "Tennis" (unchecked), and "Basket" (checked).
- Note**: A text area containing the text: "Questo è un campo di testo libero in cui si possono scrivere frasi anche molto lunghe...".
- Buttons**: "Registra" and "Ripristina" buttons at the bottom of the form.

The browser's status bar at the bottom shows "Document: Done" and several system icons.

Indicazioni per la creazione di pagine HTML

- Riportare sempre un riferimento all'autore o al responsabile delle informazioni contenute nella pagina (*webmaster*), possibilmente anche con un link ad un indirizzo di posta elettronica.
- Indicare la data di emissione del documento e lo stato delle informazioni in esso contenute (*definitive, bozza, ecc.*)
- Meglio creare un link a risorse presenti altrove in Rete, anziché copiarle sul nostro sito.
- Rendere omogenee e coerenti le pagine di uno stesso sito, anche mediante *icone di navigazione*, per evitare che l'utente si perda dopo poche pagine.
- Prima di iniziare lo sviluppo di un sito è bene progettare uno schema della struttura che assumerà il sito stesso (tipicamente un albero con dei link trasversali).
- Non creare documenti troppo lunghi (lenti da trasferire) ; meglio suddividere il documento in capitoli su file separati e collegati tra loro mediante indici ipertestuali, icone di navigazione ed *hotword*.

Indicazioni per la creazione di pagine HTML - 2

- Non segmentare eccessivamente il documento in pagine piccole : la navigazione ipertestuale è faticosa se il passaggio da una pagina ad un'altra è ingiustificato.
- Esplicitare sempre il contesto in cui si colloca ogni pagina del sito : il lettore potrebbe essere giunto direttamente su una certa pagina senza aver seguito una sequenza da noi immaginata.
- Il titolo del documento deve essere breve, ma deve anche richiamare il contesto in cui è inserita la pagina.
- Il lay-out della pagina non deve essere progettato basandosi esclusivamente sull'output prodotto da un particolare browser : meglio effettuare un test del documento su più browser di produttori diversi.
- Non utilizzare caratteristiche tipiche di un particolare client in modo esclusivo, senza offrire una seconda possibilità agli utenti che non potranno utilizzare o vedere determinati oggetti (font specifici, *applet Java*, immagini prive del *tag ALT*, *frame*, *Java-script*, controlli *Active-X*, suoni in *background* o altri oggetti che richiedano particolari *plug-in*).

Bibliografia

Guide e manuali di riferimento sull'HTML

- HTML 3.2 Reference Specification,
<http://www.w3.org/pub/www/TR/REC-html32.html>, a cura del World Wide Web Consortium
- HTML 2.0 (RFC1866)
<ftp://ds.internic.net/rfc/rfc1866.txt>, di Tim Berners-Lee e Dan Connolly, Novembre 1995.
- The Hypertext Transfer Protocol (HTTP)
<http://www.w3.org/pub/www/Protocols>, documenti a cura del World Wide Web Consortium
- A Beginner's Guide to HTML
<http://www.ncsa.uiuc.edu/General/Internet/www/HTMLPrimer.html>, di Marc Andreessen.
- Introduction To HTML and URLs
<http://www.utoronto.ca/webdocs/HTMLdocs/NewHTML/intro.html>, di Ian Graham
- The Bare Bones Guide to HTML
<http://werbach.com/barebones/>, di K. Werbach

Bibliografia - 2

- Style Guide for online hypertext
<http://www.w3.org/pub/WWW/Provider/Style/>, di Tim Berners-Lee
- Cascading Style Sheets, level 1
<http://www.w3.org/pub/WWW/TR/REC-CSS1>, draft del W3 Consortium che riporta le specifiche dello standard CSS1 (W3C Recommendation 17 Dec 1996)
- WWW History
<http://www.w3.org/pub/WWW/History.html>

Organizzazioni ed associazioni

- The World Wide Web Consortium (<http://www.w3.org>)
- Web Design Group (<http://www.htmlhelp.com/>)
- Internet Society (<http://www.isoc.org>)

Programmazione

- JavaScript Authoring Guide (<http://home.netscape.com/eng/mozilla/3.0/handbook/javascript/>)
- HTTP, interfaccia CGI e linguaggio Perl (<http://www.isinet.it/~marco/HTTP/>, di Marco Liverani)