

Web: linguaggi, e strumenti di produzione

Web: linguaggi e strumenti di produzione

1 HTML

Introduzione storica
Standard di riferimento
Sintassi
Analisi ed esempi di strutture

2 I fogli di stile

3 XHTML e I box model

Il protocollo HTTP

HTTP - HyperText Transfer Protocol

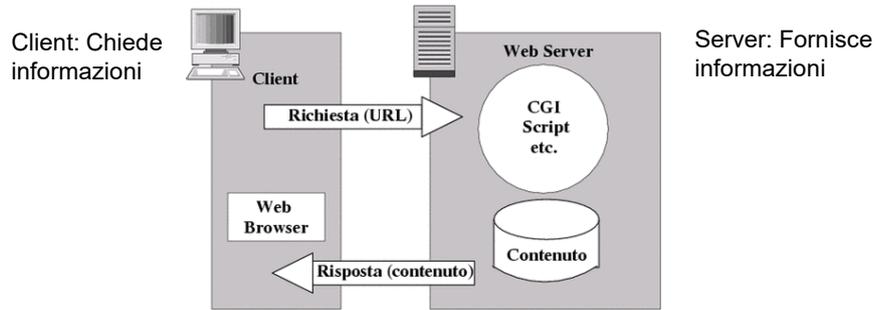
- È il protocollo principale che interconnette quella vastissima collezione di siti Internet nota come **World Wide Web (WWW)**.
- Viene generalmente utilizzato per veicolare i documenti codificati in linguaggi di marcatura, per esempio l'**HTML**, ma permette anche il trasferimento di file e l'esecuzione di procedure via rete.
- Si basa sul protocollo **TCP/IP (Internet Protocol)** permettendo l'apertura di più connessioni contemporanee.
- Attualmente è il protocollo di alto livello di **IP** più usato in assoluto.

Il protocollo HTTP

- Consente l'accesso a documenti ipertestuali e multimediali in cui vengono realizzati dei **link** tra file di vario genere (non solo testuali) fisicamente residenti anche su **host** differenti.
- È gestito da un software, server **HTTP** (**HTTPD**), residente sugli **host** che intendono essere fornitori di informazioni.
- Per accedere alle informazioni fornite dal server HTTP si deve utilizzare un software **client** (**Browser**) in grado di interpretare le informazioni inviate dal server.
- HTTP è un protocollo "**stateless**": ad ogni richiesta si effettua una connessione al server che viene chiusa al termine del trasferimento dell'oggetto richiesto (pagina HTML, immagine, ecc.).

Il protocollo HTTP

Sodalizio Client – Server



Il Server HTTP



È un programma che "gira" su un server in attesa di una richiesta di connessione sul suo socket (la porta assegnatagli, tipicamente la 80)

- Il server HTTP svolge un ruolo di interfaccia tra il client remoto che effettua delle richieste (attraverso le URL) sul socket ed il sistema che lo ospita (la macchina server), su cui sono disponibili le risorse (file, pagine web, applicativi)
- URL (**Uniform Resource Locator**): è il modo formalizzato per la locazione e l'accesso alle risorse di internet (indirizzi WEB)

`http:// sparc2.iac.rm.cnr.it :80/doc/index.htm`
Protocollo Host Dominio Porta (socket) Documento

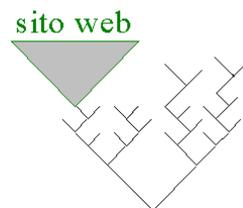
Il Server HTTP

Il server HTTP svolge tre compiti fondamentali

- invia al client le risorse disponibili localmente, richieste mediante l'indicazione di una URL;
- richiama eventuali procedure esterne con cui comunica mediante l'interfaccia CGI (*Common Gateway Interface*) per lo scambio di parametri e per ottenere in risposta informazioni in formato HTML
- effettua, ove esplicitamente richiesto dalla sua configurazione, l'autenticazione degli utenti mediante *username* e *password*.

Il Server HTTP: locazione delle risorse

- La home directory rappresenta il punto di partenza per i visitatori che esplorano il sito ed è la directory di livello superiore nella struttura di directory pubblicate (rese pubbliche).
- Essa contiene un file indice o una home page introduttiva in cui viene dato il benvenuto ai visitatori e sono inclusi i collegamenti ad altre pagine (o risorse). La home directory è mappata nel nome di dominio del proprio sito.
- Tutti i file contenuti nella home directory e nelle relative sottodirectory sono automaticamente disponibili ai visitatori che si connettono al sito. Però se non si conosce esattamente il nome del file che si vuole raggiungere, e nel sito non sono presenti link ad esso è impossibile recuperarlo.



Il Browser

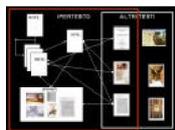
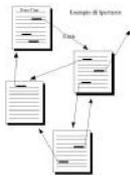


Il browser è l'applicazione *client* di questo sistema ad architettura "*client/server*". Gira sulla macchina dell'utente remoto, legge ed interpreta l'output in formato **HTML**.

- Visualizza o gestisce i file in formati a lui noti o attraverso **plug-in** (es.: immagini **GIF** o **JPEG**, filmati **QuickTime**, **Flash**) o rimanda ad altri programmi esterni presenti sulla macchina client per la gestione di formati non conosciuti (es.: documenti **Word**, documenti **Postscript**, ecc.).
- Il browser consente di impaginare l'output *indipendentemente dalla piattaforma che lo ospita* (**X11**, **Macintosh**, **Windows**, ecc.).
- I programmi in linguaggio **Java** e **Javascript** vengono invece scaricati sul client, compilati in tempo reale ed eseguiti su di esso.
- Le procedure **CGI non** vengono eseguite sulla macchina client.

Il Linguaggio HTML

HyperText Markup Language E' il linguaggio dei browsers http



Il concetto di **ipertesto** è quello di un documento che contenga, oltre al testo, dei rimandi dinamici ad altre parti del testo nonché ad altri documenti (**hyperlinks**). Sono inoltre previste anche parti di diversa natura, come immagini, suoni, applicazioni, aree interattive. È un semplice *linguaggio di marcatura* del testo (come il LaTeX), ereditato da **SGML** (*Standard Generalized Markup Language*).

Breve storia del linguaggio HTML

- Le origini dell'**HTML** (**HyperText Markup Language**) possono essere fatte risalire ad un'idea di **Vannevar Bush** (~1940) che descrisse un sistema per collegare in modo associativo le informazioni.
- Bush definì questo sistema **memex** (**memory extension**), e lo propose come mezzo per aiutare la mente umana a far fronte ad un sempre crescente numero di informazioni.
- Nel 1965, **Ted Nelson** conìò il termine **ipertesto** per descrivere un testo che seguisse le idee di Bush.
- L'**ipertesto**, come descritto da Nelson, avrebbe dovuto collegare dei documenti per creare una rete di relazioni con l'intento di estendere ed accrescere il significato di un testo "piatto" con collegamenti ad altri testi.

Breve storia del linguaggio HTML

- Le idee di Vannevar Bush e di Ted Nelson furono riprese in un progetto del 1989 da **Tim Berners-Lee**, un ricercatore del **CERN** di Ginevra, che propose un sistema basato sull'ipertesto per permettere una più efficiente condivisione delle informazioni tra i membri della comunità scientifica dell'istituto per cui lavorava.
- La proposta di Berners-Lee (il cui titolo era "**HyperText and CERN**") aveva come obiettivi:
 - la creazione di un'interfaccia utente che potesse essere utilizzata su tutte le piattaforme, per permettere all'utente di accedere alle informazioni da diversi computer;
 - la possibilità di accedere ad una grande varietà di tipi di documenti e di protocolli di informazione;
 - la predisposizione per un "accesso universale", che avrebbe permesso ad un qualunque utente della rete di accedere ad un qualunque tipo di informazione.

Breve storia del linguaggio HTML

- Da questo momento in poi ha inizio la storia dell'HTML, che diventa di pubblico dominio grazie all'introduzione nel febbraio del 1993 del browser **Mosaic** per il sistema **X-Window** (**UNIX**).
- **Mosaic** fu sviluppato dall'**NCSA** (**National Center for Supercomputing Applications**) sotto la guida di **Marc Andreessen** che in seguito sarebbe diventato uno dei fondatori di **Netscape**.
- Dal 1994 tutte le modifiche ad **HTML** sono avvenute sotto l'egida del World Wide Web Consortium (**W3C**) consorzio nato grazie agli sforzi del **MIT** e del **CERN**.
- Il consorzio comprende i più importanti centri di ricerca e le principali aziende del settore informatico che stanno guidando l'**HTML** nella sua evoluzione verso la definizione di uno standard per lo sviluppo della rete

Il linguaggio HTML

HyperText Markup Language: Markup

L'HTML è un linguaggio in cui ogni zona da formattare, ogni opzione di impaginazione, ogni caratteristica del testo vengono impostati attraverso dei **marcatori** (tags) di inizio e fine. Per esempio, il grassetto (bold) si ottiene inserendo il testo da evidenziare tra i tag **** e ****

****Questa parte è in grassetto**** → Questa parte è in grassetto

Quindi il tag **** identifica la **marcatura di inizio** del testo in grassetto, mentre il tag **** ne **identifica la fine**.

Document Type Definition

La 'grammatica' di un linguaggio di markup si chiama **Document Type Definition**

Il **DTD** definisce gli elementi del linguaggio i loro attributi e i possibili valori le combinazioni possibili degli elementi.

In **HTML** il **DTD** viene specificato attraverso **!DOCTYPE**:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
```

significa che il documento **HTML** che segue è riferito alla specifica **HTML 4.01**, emessa dal **W3C** e denominata **Transitional**

HTML 4.01 Transitional è una specifica di transizione tra **HTML** e **XHTML**

Alcuni tag ancora molto diffusi sono **deprecated** (deplorate) in questa specifica, cioè destinati ad essere abbandonati nelle successive

Versioni di HTML

❖ 0.9 (?): il linguaggio di HTML disponibile sul browser WWW aveva strutture di base per intestazioni, paragrafi e stili base, più ovviamente il **tag A** che ne costituiva la novità.

❖ 1.0 (1992): I primi browser shareware e freeware al mondo (il più importante di questo periodo fu Arena, ma esistevano anche MacWeb ed altri) implementavano alcune versioni di HTML leggermente diverse. Tra queste ebbe un certo successo la proposta HTML+. Viene introdotto il **tag IMG** e il supporto per il **GIF**.

❖ 2.0 (1994): La prima versione veramente nota di HTML. Questa è quella implementata su Mosaic, da cui deriverà Netscape. E' la prima versione ad essere formalizzata su un DTD SGML, invece che ispirarsi vagamente ad SGML. Introduce i **form**.

Versioni di HTML

- ❖ 3.0 (1995): Questa versione non è mai stata ufficialmente approvata. Durante la sua discussione vennero proposte molte aggiunte. Alcune di queste vennero implementate prima di raggiungere un consenso (tabelle), altre (ad es., supporto per la matematica) mai prese in considerazione.
- ❖ 3.2 (1997): Quando divenne chiaro che i browser non avrebbero supportato tutto il 3.0, si lavorò per generarne un sottoinsieme su cui ci fosse consenso, e che tenesse conto delle aggiunte proprietarie dei vari produttori. Questa versione include **tabelle**, **applet**, **script** e altre migliorie, ma non i **frame**, sebbene Netscape e Microsoft le avessero già implementate fin dal 1995.
- ❖ 4.0 (1997): supporto per l'internazionalizzazione, per gli **style sheet**, per i **frame**, tabelle molto più ricche, il **tag object**, ecc.

Versioni di HTML

- ❖ La versione 4.01, pubblicata il 24 dicembre [1999](#) e penultima versione ufficiale, costituisce il frutto dei diversi ampliamenti e miglioramenti del decennio precedente, in particolare nella separazione del livello presentazionale della formattazione, cioè quella che descrive gli aspetti grafici del documento, in un'entità separata dall'HTML, i fogli di stile a cascata ([CSS](#)), definiti nello stesso anno nella loro prima versione, il livello 1^[3]. Tale distinzione, tra contenuto e aspetto finale del documento, permette a browser e dispositivi differenti di rappresentare gli stessi contenuti in maniera consona alle diverse capacità grafiche e dimensioni disponibili. Se da una parte questo impone agli [sviluppatori web](#) la creazione di personalizzazioni dei fogli di stile, dall'altra ha garantito la massima diffusione del web ed evitato che esso diventasse un [medium](#) di [élite](#).

Versioni di HTML

- ❖ Nel gennaio 2000, viene pubblicato come standard [XHTML](#) 1.0, una variante di HTML 4 che usa [XML](#) 1.0 anziché in [SGML](#) come metalinguaggio di markup, per migliorarne l'interoperabilità con altri linguaggi. Il W3C decide di abbandonare HTML e di continuare lo sviluppo solo su XHTML e nel maggio 2011 XHTML 1.1 diventa uno standard ufficiale.
Nel 2004 nasce un gruppo di lavoro alternativo al consorzio: [Apple](#), [Mozilla Foundation](#), [Opera Software](#) e, in un secondo momento, anche [Google](#), si riuniscono nel [Web Hypertext Application Technology Working Group](#) (WHATWG) e cominciano lo sviluppo di una nuova versione dell'HTML, preoccupati dal disinteresse per l'HTML dimostrato dal consorzio.

Versioni di HTML

- ❖ Nel 2006 il W3C si rassegna a partecipare allo sviluppo di [HTML5](#) e i due gruppi collaborano fino al 2011, quando si rendono conto di avere obiettivi inconciliabili: il consorzio voleva tracciare una linea e pubblicare una nuova versione delle specifiche standard, mentre il WHATWG voleva uno standard in continua evoluzione.
- ❖ Il W3C ha pubblicato la quinta revisione delle specifiche il 28 ottobre 2014.

Caratteri ed entità

Per garantire la massima compatibilità e trasportabilità del documento HTML utilizza la codifica ASCII standard a 7 bit (caratteri da 1 a 127)

Per rappresentare simboli speciali o caratteri nazionali utilizza alcune sequenze speciali di caratteri

Esempi:

` = à	´ = é	î = î
ü = ü	ñ = ñ	å = å
æ = æ	± = ±	© = ©

Struttura della pagina HTML

Esistono tanti tipi di tag che riguardano principalmente:

Struttura, Impaginazione, Caratteristiche del Testo, Immagini, Link, Form.

Per convenzione, l'intero documento va racchiuso tra i tag `<HTML>` e `</HTML>`.

La pagina è strutturata in due parti principali:

- l'intestazione del documento (`<HEAD>` e `</HEAD>`)

- il corpo del documento (`<BODY>` e `</BODY>`)

```
<HTML>
<HEAD>
...
</HEAD>
<BODY>
...
</BODY>
</HTML>
```

Titolo della pagina

Ogni pagina HTML deve avere un titolo. Questo va inserito nel blocco `<HEAD> ... </HEAD>` per mezzo del tag `<TITLE> ... </TITLE>`

```
<HTML>
  <HEAD>
    <TITLE> Prima Pagina </TITLE>
  </HEAD>
</HTML>
```

Tag di Separazione

Si riferiscono al blocco `<BODY> ... </BODY>`, e sono

`<P>` Questo è l'indicatore di inizio paragrafo.

La chiusura di questo tag, `</P>` non è obbligatoria.

`
` indica l'andata a capo. La chiusura di questo tag non è prevista.

`<HR>` inserisce nel testo una linea orizzontale.

Proprietà del testo

In HTML è possibile modificare le seguenti proprietà del testo:

- `<H=n> ... </H=n>` Stabilisce le dimensioni dei caratteri (n = 1..6)
- ` ... ` Formatta il testo in grassetto
- `<I> ... </I>` Formatta il testo in italico
- `<U> ... </U>` Sottolinea il testo
- `<CENTER> ... </CENTER>` Centra nella pagina sia il testo che le immagini
- ` ... ` Per modificare il tipo di carattere, la sua grandezza e il colore.

Liste

In HTML abbiamo due tipi di liste, quelle con punto (tag ``) e quelle numerate (tag ``). Ogni elemento è identificato dal tag ``.

```
<UL>
  <LI> Primo elemento </LI>
  <LI> Secondo elemento </LI>
</UL>
```



- Primo elemento
- Secondo elemento

```
<OL>
  <LI> Primo elemento </LI>
  <LI> Secondo elemento </LI>
</OL>
```



- 1 Primo elemento
- 2 Secondo elemento

Un primo esempio

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0//EN">
<HTML>
  <HEAD>
    <TITLE> Titolo del documento </TITLE>
  </HEAD>
  <BODY>
    <H1> Intestazione maggiore </H1>
    <P> Questo &acute; un paragrafo completo di un
      documento. Scrivo e scrivo finch&grave; non
      riempio varie righe, perch&grave; voglio
      vedere come va a capo automaticamente. Certo
      non &acute; un documento molto eccitante. </P>
    <P> Forse che vi aspettavate <B> poesia </B>? </P>
    <HR>
    <P> Una lista di cose importanti da ricordare: </P>
    <UL>
      <LI> Spazi, tabulatori e ritorni a capo </LI>
      <LI> Dichiarazione di tipo del documento </LI>
      <LI> Struttura del documento </LI>
      <LI> Annidamento e chiusura dei tag </LI>
    </UL>
  </BODY>
</HTML>
```

Inserimento Immagini

Per inserire le immagini in un testo si usa il tag ****.

Questo tag ha varie opzioni:

- **SRC** -> immagine
- **ALIGN** -> formattazione
- **ALT** -> testo

```
<IMG SRC="fiore.jpg" ALIGN=middle  
ALT="caricamento fiore">
```



Link

Il tag per i link è **<A>** utilizzato con l'attributo **HREF**. Se invece si usa l'attributo **NAME** questo tag diventa una label all'interno della pagina web (àncora).

```
<A HREF="http://www.html.it/sicurezza.htm"> Sic </A>  
<A HREF="http://www.iac.cnr.it/#pagina"> Pag </A>  
<A NAME="#pagina">
```

Tabelle

Per creare una tabella si usa il tag **<TABLE>** inserendo una per una le righe (tag **<TR>**) formate da celle (tag **<TD>**). *Le celle corrispondenti, sovrapposte, creano le colonne.* Occorre prestare molta attenzione affinché le celle si sovrappongano nella maniera desiderata!

```
<TABLE>
```

```
<TR> riga 1
```

```
<TD> ... </TD> <TD> ... </TD>  
colonna 1 e 2
```

```
</TR>
```

```
<TR> riga 2
```

```
<TD> ... </TD> <TD> ... </TD>  
colonna 1 e 2
```

```
</TR>
```

```
</TABLE>
```

NB: TH definisce una cella di titolo

Per estendere la cella a coprire più colonne o più righe della tabella si usano le opzioni **COLSPAN** e **ROWSPAN**