



Struttura di una pagina HTML

Le parti fondamentali che compongono una pagina web

Sviluppo di siti web – UD03

prof. Giovanni Borga

Macrostruttura della pagina

Una pagina HTML è sostanzialmente composta di una riga che definisce un **tipo di documento** tra i diversi standard definiti nel corso degli anni, e il **documento** stesso che a sua volta è composto di intestazione e corpo. **Nel corpo viene inserita la gran parte dei contenuti da visualizzare.**

Definizione tipo documento

Documento

Intestazione

Corpo

La pagina HTML tipo

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
    <title>Titolo della pagina</title>
  </head>
  <body>
    contenuto ....
  </body>
</html>
```

La pagina tipo in HTML 5

```
<!doctype html>  
<html lang="en">  
  <head>  
    <meta charset="utf-8">  
    <title>Titolo della pagina</title>  
  </head>  
  <body>  
    contenuto ...  
  </body>  
</html>
```

Tipo del documento: la «doctype declaration»

Detta DTD (Document Doctype Declaration), la prima riga di codice di una pagina HTML definisce quale tipo di definizione si intende utilizzare.

I più comuni doctype sono:

- **HTML 4.01 Strict**
- **HTML 4.01 Transitional**
- **HTML 4.01 Frameset**
- **XHTML 1.0 Strict**
- **XHTML 1.0 Transitional**
- **XHTML 1.0 Frameset**
- **XHTML 1.1**
- **HTML 5**

*Le **versioni** sono sostanzialmente 3: HTML 4 e 5 e XHTML già descritte nella prima lezione.*

*Le **varianti** sono: Strict, Transitional, Frameset*

La variante Strict non contiene gli elementi «deprecati» a differenza della Transitional. Frameset è uguale Transitional ma consente in più l'uso dei frames.

Il doctype HTML 4.1 Transitional è ancora molto usato

Tuttavia si è pienamente affermato il **HTML 5 che è quello che utilizzeremo nel corso.**

Doctype declaration 4.1

La DTD è la prima riga del listato HTML:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"  
"http://www.w3.org/TR/html4/loose.dtd">
```

!DOCTYPE

Apertura del doctype

HTML

Indica che il linguaggio utilizzato per il documento corrente è l'HTML

PUBLIC

Indica che il documento è pubblico

-//W3C//DTD HTML 4.01 Transitional//EN

Indica che il documento si riferisce ad una specifica del W3C, che il DTD è HTML versione 4.01 Transitional e che la lingua è l'inglese

http://www.w3.org/TR/html4/loose.dtd

E' la URL alla quale è raggiungibile la specifica relativa alla DTD del documento; questa indicazione è facoltativa (e nella prassi frequentemente omessa).

L'intestazione (HEAD)

Nell'HEAD della pagina vengono inseriti elementi non immediatamente percepibili, ma che riguardano il modo in cui il documento deve essere letto e interpretato dal browser.

`<head></head>`

In particolare in questa sezione vengono inseriti 4 fondamentali elementi:

- Il **Title** (il titolo della pagina che compare sulla barra della finestra del browser)
- I **meta-tag** (ad esempio quelli a beneficio dei motori di ricerca)
- Gli **script** di carattere generale (come ad esempio le definizioni di funzioni)
- I riferimenti ai **fogli di stile**

Il tag head non ha attributi.

Il tag TITLE

Il TITLE è il titolo della pagina e compare in alto sulla barra del browser.

`<title>Titolo della pagina</title>`

È bene compilarlo da subito, onde evitare di dimenticarlo e di avere pagine senza titolo. Questo elemento è strategico per almeno due motivi:

- 1) Viene visualizzato sulle **barre delle finestre** ma soprattutto delle **schede** di molti browser e consente l'individuazione dei diversi contenuti a colpo d'occhio senza dover visualizzare necessariamente la pagina;
- 2) Il valore del tag è la parola chiave che assume l'importanza maggiore nell'indicizzazione automatica effettuata dai **motori di ricerca**.

Set di caratteri

Si utilizzano i CHARSET per indicare al browser quale set di caratteri utilizzare nel rendering:

<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">

Questo meta-tag indica l'utilizzo del set iso-8859-1, ovvero il set di caratteri latini utilizzati nei paesi occidentali.

<meta charset="utf-8">

Questo meta-tag (espresso nella forma breve prevista dall'HTML 5) indica l'utilizzo del set esteso utf-8, detto anche Unicode, che comprende il set base dell'iso-8859-1 più una serie di altri caratteri usati in quasi tutte le lingue vive e in alcune lingue morte, nonché simboli matematici e chimici, cartografici, l'alfabeto Braille, ideogrammi ecc.

In teoria, qualsiasi codifica dei caratteri può essere utilizzata, ma nessun browser è in grado di interpretarle tutte. L'utilizzo di una codifica molto ampia come l'unicode aumenta la possibilità che la maggioranza dei browser visualizzino correttamente i contenuti.

Il corpo (BODY)

Nel BODY si inserisce la gran parte del contenuto di una pagina.

<body></body>

Il tag BODY può essere pensato come il «foglio» su cui disponiamo i contenuti ovvero la superficie sopra la quale vengono renderizzati tutti gli altri elementi.

Tale superficie dispone dei seguenti attributi:

- BGCOLOR (colore di sfondo)
- BACKGROUND (immagine di sfondo)
- LEFTMARGIN e TOPMARGIN (margini in orizzontale e in verticale)
- **TEXT (colore del testo)**
- LINK (colore dei collegamenti ipertestuali)
- ALINK (colore dei link attivi)
- VLINK (colore dei link visitati)

Oltre ai cosiddetti «attributi globali», in particolare quelli per i fogli di stile: STYLE, CLASS e ID

Va detto però che tutti gli attributi di cui sopra, tranne TEXT, sono considerati DEPRECATI in virtù dell'adozione dei fogli di stile