

Concetti avanzati di HTML

Liste

Il formato **HTML** mette a disposizione 5 **Tag** per definire delle liste, supportando liste numerate, non numerate, di descrizione, menu e *directory*.

Liste Numerate e Ordinate

Una lista numerata si crea usando i seguenti **Tag**:

1. usare il **Tag** di *open list* ``
2. usare il **Tag** `` seguito dall'oggetto della lista;
3. iterare questo punto per ogni elemento da listare;
4. chiudere la lista con il **Tag** ``.

Sia le liste ordinate che quelle non ordinate vengono visualizzate indentate di qualche spazio rispetto al testo normale ed un carattere particolare viene messo prima degli elementi della lista (per le liste ordinate é un num. progressivo).

Liste non numerate

Una lista non numerata si crea usando i seguenti **Tag**:

- usare il **Tag** di *open list* ``
- usare il **Tag** `` seguito dall'oggetto della lista;
- iterare questo punto per ogni elemento da listare;
- chiudere la lista con il **Tag** ``.

Liste di menu

Questo tipo di lista dovrebbe essere fatta di piccoli paragrafi (solitamente 1 linea per elemento). Essa viene visualizzata con uno stile piu' compatto rispetto alle liste non ordinate. Una lista di menù si crea usando i seguenti **Tag**:

- usare il **Tag** di *open list* `<MENU>`;
- usare il **Tag** `` seguito dall'oggetto della lista;
- iterare questo punto per ogni elemento da listare;
- chiudere la lista con il **Tag** `</MENU>`.

Questo tipo di lista dovrebbe essere composta di piccoli elementi, tipicamente di 20 caratteri. Questi possono essere incolonnati lungo la pagina. Una lista di *directory* si crea usando i seguenti **Tag**:

- usare il **Tag** di *open list* `<DIR>`;
- usare il **Tag** `` seguito dall'oggetto della lista;
- iterare questo punto per ogni elemento da listare;
- chiudere la lista con il **Tag** `</DIR>`.

Liste di descrizione

Una lista di descrizione é solitamente un alternanza di titoli (`<DT>`) e descrizioni (`<DD>`). Si crea usando i seguenti **Tag**:

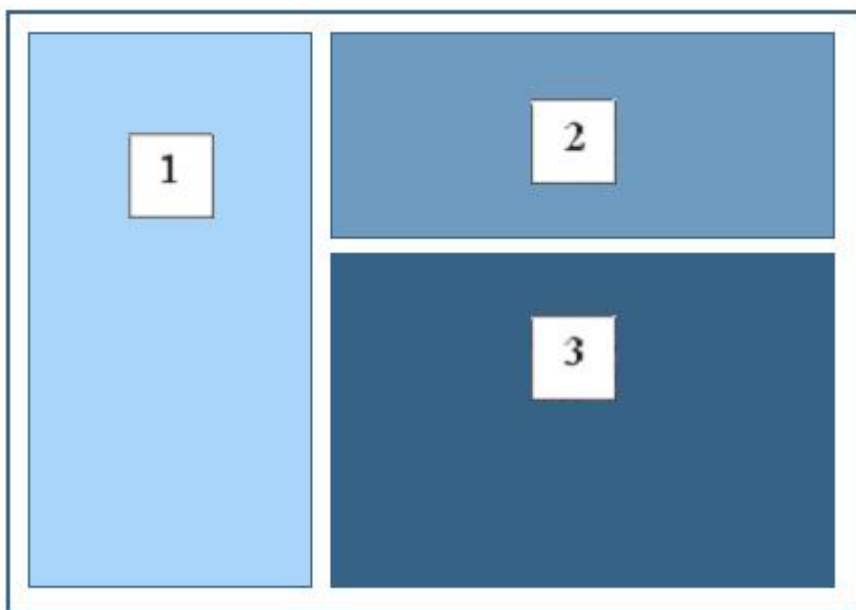
- usare il **Tag** di *open list* `<DL>`

- inserire il titolo del primo oggetto della lista dopo il **Tag** <DT>
- inserire la descrizione del corrispondente oggetto dopo il **Tag** <DD>
- iterare questo punto per ogni elemento da listare;
- chiudere la lista con il **Tag** </DL>.

Attenzione: sono ammesse liste nidificate di qualsiasi livello ma l'*output* risultante varia a seconda del *browser* usato, e alcuni non le supportano.

I frame

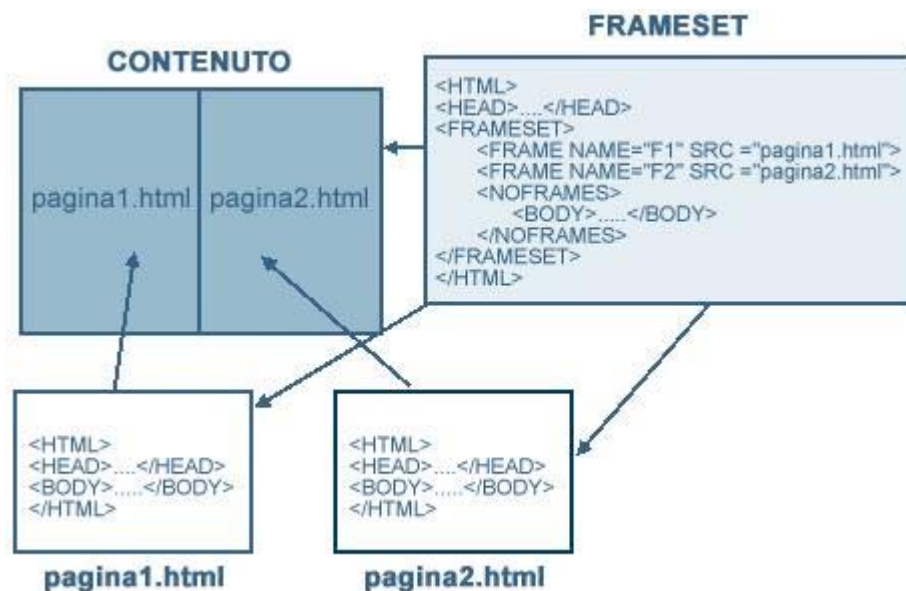
E' utile impostare la pagina **HTML** in modo da suddividerla in sezioni, ciascuna delle quali è a sua volta un documento **HTML**. In tal modo, è possibile agire sulla struttura in fase di visualizzazione, per esempio, modificando il contenuto di un riquadro della pagina principale.



Una situazione di questo tipo è piuttosto frequente nel caso in cui si voglia mantenere sempre disponibile un menu (per esempio nel riquadro 1) ed altre informazioni (per esempio una intestazione nel riquadro 2), mentre si vuol modificare dei contenuti visualizzati (per esempio nel riquadro 3). A ogni riquadro viene dunque assegnato inizialmente una pagina da visualizzare, pagina che può essere diversa nel tempo.

I marcatori per i Frame

Quando si vuole definire la struttura a riquadri di una pagina, i marcatori sostituiscono il corpo del contenitore. Ovvero non comparirà il marcatore <BODY> della pagina principale. Il marcatore che definisce le righe o colonne di una struttura è <FRAMESET>, in cui saranno riportati dei parametri che indicano la dimensione delle righe o colonne che compongono la griglia. La dimensione può essere definita in *pixel*, in percentuale, o inserendo il carattere * che indica come quella riga (o colonna) comprende lo spazio rimasto dopo aver impostato le righe o colonne precedentemente definite. Ogni riquadro dovrà definire almeno la pagina da visualizzare attraverso il parametro SRC (*source*).



L' esempio seguente definisce due colonne (la prima a dimensioni fisse da 100 *pixel*, la seconda per lo spazio rimanente). La prima colonna contiene inizialmente la pagina `index.html` e il riquadro è stato chiamato *menu* che rappresenta il nome per intervenire sul riquadro, per esempio per modificare la pagina visualizzata. La seconda colonna è a sua volta suddivisa in due righe (di dimensione pari a 110 la prima, a completamento della pagina la seconda).

Di seguito viene riportato il marcatore `<NOFRAMES>` utile per impostare dei contenuti che saranno visualizzati dai vecchi *browser* che non riconoscono i *frame*.

```
<FRAMESET COLS = "100,*">
  <FRAME SRC="index.html" NAME="menu">
    <FRAMESET ROWS = "110,*" >
      <FRAME SRC="Logobar.html" NAME="Logobar">
      <FRAME SRC="hp.html" NAME="Main">
    </FRAMESET>
  </FRAMESET>
<NOFRAMES>
<BODY>
.....
```

Indirizzamento di un riquadro

La struttura iniziale di una pagina prevede la visualizzazione di determinate pagine all'interno di ogni riquadro. Tuttavia, l'utente può avere necessità di modificare il contenuto di un settore. Sarà necessario utilizzare un marcatore di collegamento (`<A>`) indicando la pagina da visualizzare (attraverso il parametro `HREF`) e lo spazio di destinazione (il parametro è `TARGET=nome_del_frame_indirizzato`), indicandone il nome definito con il parametro `NAME` del marcatore `FRAME`. Per esempio se è stata definita la struttura:

```
<FRAMESET COLS = "85,*">
  <FRAME SRC="menu.html" NAME="menu">
  <FRAMESET ROWS = "110,*" >
    <FRAME SRC="Logobar.html" NAME="Logobar">
    <FRAME SRC="hp.html" NAME="Main">
  </FRAMESET>
</FRAMESET>
```

Attraverso il parametro *TARGET* inserito nei *link* è possibile definire il riquadro di destinazione

della pagina da visualizzare.

```
TARGET = nome | _blank | _self | _parent | _top
<A HREF="nuova.html" TARGET="Main">Nuovo caricamento</A>
```

- **blank**: apre un'altra sessione del *browser*, visualizzando la pagina a schermo intero;
- **self**: la pagina da visualizzare appare nello stesso riquadro da cui è partita la richiesta;
- **parent**: serve a indirizzare il documento nel *frameset* da cui ha avuto origine quel documento. Se il documento non ha un *frameset* di origine, viene caricato nella stessa area da cui parte la richiesta;
- **top**: il documento viene ricaricato a pieno schermo, che diventa un'unica area.

Da notare che l'uso dei *frame* può comportare anche dei svantaggi:

- alcuni vecchi *browser* non consentono la visualizzazione dei *frame*; risulta necessario prevedere la compilazione di siti diversi (o pagine diverse) in funzione della compatibilità dei *frame* ai *browser*.
- Il caricamento di una intera pagina strutturata in *frame* è più lenta, dovendo caricare pagine distinte per ogni riquadro.
- Lo spazio disponibile per il contenuto informativo è ridotto, non essendo presentato a pieno schermo. Le barre di scorrimento possono ulteriormente ridurre tale spazio.
- E' poco agevole la gestione dei riquadri a livello utente (visualizzazione del codice sorgente, aggiunta dei segnalibri).

Tabelle

Le tabelle (*tables*) vengono adoperate nei siti per due ragioni principali:

- la prima, più ovvia, è quella di sistemare le informazioni in una tabella;
- la seconda - meno ovvia - è quella di creare il *layout* della pagina servendosi di tabelle nascoste.

L'uso delle tabelle per dividere la pagina in diverse sezioni è uno strumento estremamente efficace. Quasi tutti i siti più grandi sulla rete si servono di tabelle nascoste per strutturare le pagine. Gli aspetti più importanti della progettazione di una pagina che si possono trattare con l'uso di tabelle sono:

- **la divisione della pagina in sezioni separate.** Una tabella invisibile è una risorsa eccellente per questo scopo.
- **La creazione di menù.** In genere con un colore per la voce principale e un altro per i *link* che seguono nelle righe successive.
- **L'inserimento di form *field* interattivi.** Generalmente un'area grigia contenente un'opzione di ricerca.
- **La creazione di titoli di pagina rapidamente scaricabili.** Una tabella colorata con un testo viene scaricata in un batter d'occhio rispetto ad un *banner* anche semplicissimo.
- **L'allineamento di immagini che sono state tagliate in parti più piccole.**
- **La disposizione del testo in due o più colonne sistemate una accanto all'altra.**

L'importanza delle tabelle nella progettazione non deve essere tuttavia sopravvalutata. Ci sono alcune cose da tenere presenti quando si decide di farne uso. La più importante è che il contenuto della tabella viene mostrato solo quando l'intera tabella è scaricata. Se hai pagine molto lunghe, ti consigliamo di dividerle in più tabelle - in modo che l'utente possa leggerne l'inizio, mentre il resto della pagina viene scaricato.

Le tabelle sono definite con il tag `<table>`.

Per inserire una tabella sulla tua pagina si devono semplicemente aggiungere questi tag nel punto in cui vuoi metterla.

```
<table>
</table>
```

La tabella qui sopra non può funzionare, perché non ha né righe né colonne.

Formattazione delle tabelle

I tag per inserire righe in una tabella sono `<tr>` e `</tr>`, e per le colonne `<td>` e `</td>`:

```
<table>
  <tr>
    <td>Prima riga, sinistra.</td>
    <td>Prima riga, destra.</td>
  </tr>
  <tr>
    <td>Seconda riga, sinistra.</td>
    <td>Seconda riga, destra.</td>
  </tr>
</table>
```

Prima riga, sinistra. Prima riga, destra.

Seconda riga, sinistra. Seconda riga, destra.

Esistono diverse opzioni per personalizzare le tabelle. In particolare è possibile formattare le tre zone in cui è divisa una tabella, oltre alle celle:

- spazio tra le celle;
- spazio (spessore) del bordo;
- spazio interno tra celle e testo;
- l'allineamento, dimensioni e colori per ognuna delle *zone* citate.

Tali attributi vanno utilizzati all'interno delle righe (colonne) a cui si vuole attribuire l'aspetto desiderato. La tabella sottostante riporta all'interno di se stessa gli attributi che la descrivono. Ci sono attributi per il bordo (`border="0"`), per la larghezza (`width="72%"`), per lo spazio tra le celle (`cellspacing="1"`), per lo spazio tra celle e testo (`cellpadding="3"`) e per le celle (`width="50%" bgcolor="#C7E1F7" height="17"`)

attributo	nome e valore
bordo	<code>border="0"</code>
larghezza	<code>width="72%"</code>
spazio tra le celle	<code>cellspacing="1"</code>
spazio tra testo e cella	<code>cellpadding="3"</code>
in ogni cella	<code>width="50%" bgcolor="#C7E1F7" height="17"</code>

Alcuni attributi (il colore per esempio) vanno definiti all'interno dello spazio di ciascuna cella. Altri invece (bordo) sono globali per la tabella.

Formato dei caratteri

Singole parole o intere frasi possono essere messe in risalto tramite l'uso di speciali Stili. L'**HTML** supporta due tipi di Stili: quello logico e quello fisico. Il formato logico dà uno stile al testo a cui si riferisce in accordo con il suo significato (quello del tag) e può variare da *browser* a *browser*. I due tipi di stile sortiscono lo stesso effetto, ma quello logico sarebbe da preferire.

Stile fisico

stile	tag
Italico	<I>testo</I>
Grassetto (Bold)	testo
Sottolineato	<U>testo</U>

Stile logico

stile	tag
Associato all'italico	<DEN>
Per dare enfasi	
Per titoli di libri, film, ecc	<CITE>
Per scrivere codice	<CODE>
Per gli <i>input</i> da tastiera	<KBD>
Per comunicazioni dal <i>computer</i> all'utente	<SAMP>
associato al grassetto	
Solitamente in Italico.	<VAR>

Sequenze di *Escape* e altri Tag

Qui di seguito vi sono alcune **escape sequence** per inserire caratteri particolari (quali < &, eccetera) che altrimenti verrebbero considerati come facenti parte di **Tag**: &seguito, senza spazi, da:

- `lt;` per <.
- `gt;` per >.
- `amp;` per &.
- `quot;` per ".
- `egrave;` per è.
- `Egrave;` per È.

Attenzione: le sequenze di *escape* sono case sensitive cioè vi è differenza nell'usare un carattere in minuscolo dall'usarlo in maiuscolo.

Installare e configurare un motore di ricerca per un sito Web

Quando un sito *Web* contiene diverse pagine diventa necessaria la presenza di un motore di ricerca. Individuare il giusto motore di ricerca non è cosa banale. Nel passato le ricerche erano fatte mediante indici non strutturati che erano generati automaticamente dalla strutturazione del *file system*. Negli ultimi anni i motori di ricerca hanno generato indici basandosi sui collegamenti ipertestuali nelle pagine *Web*. Questo approccio ha introdotto il problema di controllare le pagine a

cui accedono i motori di ricerca.

Le principali caratteristiche che deve possedere un motore di ricerca sono:

- indicare velocemente un grande numero di documenti in formato differente, tra cui testo, **HTML** e **XML**; per altri tipi di *file*, come PDF, *gzip*, o *postscript*, usare filtri;
- utilizzare *Web spider* per indicare documenti remoti su HTTP;
- consentire di memorizzare nel *file* di indice anche le proprietà dei documenti (di solito definiti come META o elementi **XML**) e ritornare queste proprietà come risultato della ricerca;
- possibilità di ritornare sunti dei documenti trovati;
- usare espressioni regolari per selezionare i documenti;
- possibilità di limitare in maniera semplice le ricerche a parti dei siti *Web*;
- possibilità di memorizzare i risultati per rilevanza o secondo un numero di proprietà in ordine crescente o decrescente;
- limitare le ricerche a parti di documenti, come certi tag **HTML** (META, TITLE, commenti, etc.) o a elementi **XML**.

Tra i motori di ricerca più importanti si ricordano, [google](#), [Yahoo](#), [eXcite](#), [Swish-e](#), [altavista](#), [MSN serach](#).

Google

In questa sezione analizziamo come installare e configurare *google*. Le corrispondenti operazioni per gli altri motori sono simili.

Google aggiunge nuovi siti all'indice ed aggiorna quelli esistenti dopo ogni scansione della rete. In alternativa, è possibile inviare direttamente l'URL attraverso la pagina (http://www.google.it/intl/it/add_url.html). Per segnalare un sito, è necessario inserire l'URL completo, compreso il prefisso `http://`; ad esempio : `http://www.google.com/`. È possibile anche aggiungere commenti o parole chiave che descrivano il contenuto della pagina. Per evitare rimozioni ingiustificate di pagine dall'indice, *Google* può escludere le pagine di un sito dai propri indici **solo su richiesta del Webmaster responsabile del sito**.

Rimozione di un sito

È possibile escludere un sito *Web* o una parte di *server (directory)* dall'indicizzazione di *Google*. A tal fine è necessario installare nella *directory* principale del *server* un *file* denominato **robots.txt**. Ad esempio, le seguenti istruzioni in `robots.txt`

```
User-Agent: *
Disallow: /
```

evitano che *Google* e altri motori di ricerca possano effettuare operazioni di scansione sul sito. Per ulteriori informazioni su `robots.txt`, consultare il sito *Web* <http://www.robotstxt.org/wc/norobots.html>.

Rimozione di singole pagine

Per impedire a tutti i *file robots* di indicizzare singole pagine del sito, inserire il seguente metatag nel codice della pagina **HTML**:

```
<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">
```

Per escludere le pagine solo dall'indicizzazione di *Google* e permetterne l'indicizzazione da parte di altri motori di ricerca, utilizzare il seguente tag:

```
<META NAME="GOOGLEBOT" CONTENT="NOINDEX, NOFOLLOW">
```

Rimozione di una sintesi

Per sintesi si intende il testo relativo a un risultato della ricerca in cui tutti i termini ricercati sono evidenziati in grassetto. Le sintesi permettono all'utente di esaminare il contesto in cui sono utilizzati i termini prima di selezionare la pagina. Se per una pagina è disponibile una sintesi, l'utente sarà più propenso a selezionarla.

Per evitare che *Google* visualizzi una sintesi per ciascuna delle pagine, utilizzare il seguente tag:

```
<META NAME="GOOGLEBOT" CONTENT="NOSNIPPET">
```