

## Da HTML a XHTML

A. Ferrari

Alberto Ferrari

## HTML

- L'**H**yper**T**ext **M**arkup **L**anguage (HTML) (traduzione letterale: linguaggio di marcatura per ipertesti) è un linguaggio usato per descrivere la struttura dei documenti ipertestuali disponibili nel World Wide Web.
- Tutti i siti web sono scritti in HTML, codice che viene letto ed elaborato dal browser, il quale genera la pagina che viene visualizzata sullo schermo del computer.
- L'HTML non è un linguaggio di programmazione, ma un linguaggio di **markup**, ossia descrive il contenuto, testuale e non, di una pagina web.

Wikipedia

Alberto Ferrari

## XHTML

- L'XHTML (acronimo di eXtensible HyperText Markup Language, Linguaggio di marcatura di ipertesti estensibile) è un linguaggio di marcatura che associa alcune proprietà dell'**XML** con le caratteristiche dell'**HTML**: un file XHTML è un pagina HTML scritta in conformità con lo standard XML.
- Il linguaggio prevede un **uso più restrittivo dei tag HTML** sia in termini di validità che in termini di sintassi, in modo da descrivere solo la **struttura logica** della pagina, mentre il **layout** e la resa grafica sono imposti dai fogli di stile a cascata (Cascading Style Sheets, CSS).
- L'XHTML è nato ufficialmente il 26 gennaio 2000 come standard del World Wide Web Consortium (W3C).

Wikipedia

Alberto Ferrari

## Passare a XHTML. Perché?

- Compatibilità con i diversi browser:
  - aderire allo standard XHTML consente di ottenere una maggiore **compatibilità** con i diversi browser e con i diversi user agent che navigano sul Web.
  - User agent (definizione): dispositivi che accedono al Web: telefonini, PDA, lettori braille, sintetizzatori vocali.
- Motori di ricerca
  - gli spider dei motori di ricerca operano sicuramente meglio con codice che segue una sintassi basata su XML.
- Separazione contenuto da forma (layout)
  - Tag HTML per definire il contenuto e CSS per il layout. Primo passo per Web semantico.
- Validazione
  - Possibilità di controllare la struttura sintattica di un documento

Alberto Ferrari

## XHTML - Sintassi

- Scrivere Xhtml richiede una chiara sintassi html
- Alcune regole di sintassi xhtml
  - Chiusura degli elementi
  - Nomi di elementi e attributi in minuscolo
  - Valori degli attributi tra doppi apici
  - Minimizzazione degli attributi vietata
  - Attributo *id* sostituisce *name*
  - Elementi obbligatori

Alberto Ferrari

## XHTML - Chiudere gli elementi

- Un tag di chiusura per ogni tag di apertura
  - Tag `<p>` e `</p>` per aprire e chiudere paragrafi
  - Lo stesso vale per `<i>` e tutti gli altri tag
  - Per compatibilità con i vecchi browser, meglio aggiungere uno spazio prima del simbolo "/" nei tag semplici (senza corpo)
- No
  - Paragrafo 1. `<p>Paragrafo 2.`
  - `<b><i>This is wrong</b></i>`
  - `<br>`
- Sì
  - `<p>Paragrafo 1. </p><p>Paragrafo 2. </p>`
  - `<b><i>This is right</i></b>`
  - `<br />`

Alberto Ferrari

## XHTML - Lettere minuscole

- No
  - `<HTML>`
- Si
  - `<html>`
- No
  - `<a HREF="http://w3.org">`
- Si
  - `<a href="http://w3.org">`

Alberto Ferrari

## XHTML - Valori degli attributi

- No
  - `<a href=http://w3.org>`
- Si
  - `<a href="http://w3.org">`

Alberto Ferrari

## XHTML Minimizzazione degli attributi

- No
  - `<input checked>`
  - `<input readonly>`
  - `<input disabled>`
  - `<option selected>`
  - `<frame noresize>`
- Si
  - `<input checked="checked" />`
  - `<input readonly="readonly" />`
  - `<input disabled="disabled" />`
  - `<option selected="selected" />`
  - `<frame noresize="noresize" />`

Alberto Ferrari

## XHTML - Attributi id e name

- L'attributo *id* sostituisce l'attributo *name*
  - HTML 4.01 definisce un attributo *name* per molti elementi:
    - *a, applet, frame, iframe, img, map*
  - In XHTML l'attributo *name* è deprecato, sostituito da *id*
- No
  - ``
- Si
  - ``
- Transitoriamente, per interoperare con i vecchi browser, si può assegnare lo stesso valore a *name* e *id*
  - ``

Alberto Ferrari

## XHTML - Attributo lang

- L'attributo *lang* si applica a quasi tutti gli elementi `xhtml`
  - Specifica la lingua del contenuto di un elemento
  - L'attributo *xml:lang* è lo standard per xml, ma per compatibilità con i browser meglio definire entrambi
  - `<div lang="no" xml:lang="no">Heia Norge!</div>`

Alberto Ferrari

## XHTML - Elementi obbligatori

- La dichiarazione *DOCTYPE* deve essere presente
  - Non è considerata parte del documento xhtml stesso
  - Non è un elemento xhtml e non deve avere un tag di chiusura
- Gli elementi *html*, *head* e *body* devono essere presenti
- L'elemento *title* deve essere presente dentro *head*
- L'attributo *xmlns* nel tag *html* è necessario
- ```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Title goes here</title>
  </head>
  <body>
  </body>
</html>
```

Alberto Ferrari

## XHTML

- **DOCTYPE**
  - perché un documento possa essere validato, deve contenere un elemento chiamato DTD (Document Type Definition o Document Type Declaration)
  - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">`
- **xmlns**
  - Namespace XML e lingua del documento nel tag html
  - `<html xmlns="http://www.w3.org/1999/xhtml" lang="it">`

Alberto Ferrari

## Versioni di XHTML

- **XHTML 1.0 Transitional**: nato per favorire la migrazione dalla vecchia versione HTML 3.2 o per uso insieme a link e frame in-line. Accetta come validi anche i tag HTML che sono stati definiti come deprecati in XHTML e è tollerante rispetto ad alcune non conformità sintattiche.
- **XHTML 1.0 Strict**: rispetto alla versione Transitional non accetta i tag HTML definiti deprecati, non è tollerante a non conformità sintattiche e prevede controlli più rigorosi anche rispetto al valore di alcuni attributi dei tag (per esempio, l'attributo id deve avere valori univoci all'interno dello stesso documento).
- **XHTML 1.0 Frameset**: nato per motivi di compatibilità per suddividere la finestra visualizzata dal browser in diversi frame (sottofinestre), pratica un tempo diffusa ma ora deprecata dal World Wide Web Consortium.
- L'ultima versione di XHTML è la **XHTML 1.1**, che è una riformulazione dell'XHTML Strict, quindi ha una sola DTD. Non accetta i tag e gli attributi più criticati della precedente versione (come la struttura a frameset). Questa versione invece supporta i tag del cosiddetto markup «ruby», che consente di visualizzare i set di caratteri delle lingue orientali.

Alberto Ferrari

## Validazione

[\(http://validator.w3.org/\)](http://validator.w3.org/)

- Un documento per essere validato deve contenere un elemento chiamato DTD (Document Type Definition o Document Type Declaration), da posizionare preferibilmente all'inizio del documento.
- I tipi più comuni di DTD, e le relative sintassi dichiarative, sono:
- XHTML 1.0 Strict
  - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">`
- XHTML 1.0 Transitional
  - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">`
- XHTML 1.0 Frameset
  - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">`
- XHTML 1.1
  - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">`

Alberto Ferrari

## Link

- Lista tag (ordinati per funzione)
  - [http://www.w3schools.com/tags/ref\\_byfunc.asp](http://www.w3schools.com/tags/ref_byfunc.asp)
- Validazione codice
  - <http://validator.w3.org/>
- Lista tag (in italiano)
  - [http://lau.csi.it/risorse/xhtml1\\_tagtag/index.shtml](http://lau.csi.it/risorse/xhtml1_tagtag/index.shtml)

Alberto Ferrari