

XHTML: Extensible HTML

Mario Bravetti

Introduzione

- ◆ Extensible HyperText Markup Language (XHTML) è l'implementazione W3C di HTML 4.01 in XML.
- ◆ XHTML è in pratica il linguaggio HTML 4.01 al quale sono applicate tutte le caratteristiche strutturali e sintattiche di XML
 - Possibilità di usare i tool e i parser per XML anche con HTML (che invece deve usare un parser apposito)
- ◆ Documenti XHTML sono quindi veri e propri documenti XML (con estensione ".html") che:
 - sono conformi a tutte regole sintattiche e strutturali di XML: sono documenti well-formed
 - sono associati a specifiche DTD: sono documenti validi

Introduzione

- ◆ Il termine "extensible" presente nel nome del linguaggio fa riferimento alla possibilità di estensione fornita dalle caratteristiche di XML.
- ◆ Questo significa che è possibile estendere il linguaggio HTML standard con elementi definiti dall'utente, purché basati su XML.
- ◆ Esempio: uso di specifiche DTD che vengono utilizzate in combinazione con quelle pubbliche sulle quali sono basati tutti i documenti XHTML.
 - es Scalable Vector Graphics (SVG), MathML,...

Regole sintattiche per XHTML

- ◆ In aggiunta well-formedness XML anche **vincoli sintattici specifici per tag HTML (in DTD)**, ad es.:
 - elementi scritti solo in **minuscolo**
 - gli elementi `<a>`, `<label>` e `<form>` **non possono essere innestati con sé stessi**
 - l'elemento `<button>` **non può contenere** `<form>`, `<input>`, `<select>`, `<textarea>`, `<button>`, `<fieldset>`
 - l'elemento `<pre>` **non può contenere** ``, `<object>`, `<big>`, `<small>`, `<sup>`, `<sub>`
 - in `<a>`, `<applet>`, `<form>`, `<frame>`, `<iframe>`, `` e `<map>` si deve **usar attributo "id" al posto di "name"**

Versioni di XHTML

- ◆ Il W3C ha standardizzato **due versioni del linguaggio XHTML**, entrambe le quali sono attualmente utilizzate:
 - **XHTML versione 1.0** (www.w3.org/TR/xhtml1): corrisponde alla riscrittura in chiave XML del linguaggio HTML 4.01.
 - **XHTML versione 1.1** (www.w3.org/TR/xhtml11): è una versione di XHTML che
 - impone **regole più restrittive** sull'uso dei tag HTML e consente di ottenere documenti **maggiormente conformi a una sintassi rigida**;
 - introduce inoltre il **concetto di modularità nelle DTD**.
- ◆ Le versioni **si differenziano per le DTD che utilizzano** per definire la validità dei documenti.

XHTML 1.0

- ◆ XHTML 1.0 si basa su **tre diverse DTD pubbliche** tramite le quali è possibile creare documenti XHTML validi.
- ◆ Le DTD che si possono usare sono:
 - DTD XHTML 1.0 Transitional: **supporta l'intero linguaggio HTML 4.01, compresi gli elementi deprecati** (spesso legati alla formattazione di presentazione) per cui non introduce particolari limitazioni; è anche detta **loose** ("lasca").
 - dichiarazione DTD:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

XHTML 1.0

- DTD XHTML 1.0 Strict: è basata su HTML 4.01 puro, e non prevede il supporto agli elementi **deprecati** (, <align>, <center>, ecc.). Le caratteristiche di presentazione possono essere impostate solo tramite i **fogli di stile CSS**.
 - consente di scrivere **documenti HTML** che soddisfano i **requisiti di accessibilità**, uso di dispositivi non standard (ad esempio screen reader e voice browser).
 - dichiarazione DTD:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

XHTML 1.0

- DTD XHTML 1.0 Frameset: è **utilizzata per realizzare pagine che utilizzano i frame**; è in pratica una DTD Transitional nella quale si può utilizzare il tag <frameset> al posto del tag <body>.

- dichiarazione DTD:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">
```


XHTML 1.1

- ◆ XHTML 1.1 nasce con lo scopo di rendere maggiormente conformi a una sintassi e a una struttura univoca i documenti HTML.
- ◆ Perciò utilizza una sola DTD molto restrittiva (quindi simile alla DTD Strict di XHTML 1.0)
 - esclude l'uso di numerosi elementi del linguaggio HTML 4.01 (in particolare quelli deprecati).
 - dichiarazione DTD:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"  
"http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
```

XHTML 1.1

- ◆ In aggiunta, la versione 1.1 **introduce il concetto di modularità**: caratteristiche del linguaggio sono specificate mediante DTD realizzate sotto forma di **moduli che possono essere composti tra loro**.
- ◆ L'autore può quindi specificare per le singole pagine i **moduli** (cioè i frammenti di DTD) **che devono essere utilizzati**, ottenendo in tal modo funzionalità particolari
 - esempio: **supporto a dispositivi non standard** come palmari e voice browser

Struttura di un documento XHTML

```
<!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"  
    xml:lang="it" lang="it">
```

```
<head>  
    <title> ... </title>  
</head>  
  
<body>  
  
    ...  
  
</body>  
  
</html>
```

Dichiarazione del
tipo di documento

Impostazione del namespace
e della lingua del documento

Corpo del documento

Convalida

- ◆ W3C ha predisposto strumento **XHTML validator** (<http://validator.w3.org>), che **analizza il documento e genera un rapporto di controllo**.
- ◆ Se il documento supera la verifica di validità è possibile **aggiungere ad esso il logo ufficiale W3C di XHTML 1.0 o di XHTML 1.1**:



- ◆ Il validator del W3C tuttavia **non esegue un lavoro completo** in quanto non controlla tutte le caratteristiche richieste, ma solo alcune.