

Sistemi Distribuiti

HTML e XHTML

Questi lucidi sono basati sul materiale didattico preparato dal prof. G. Mecca dell'Università della Basilicata (<http://www.difa.unibas.it/users/gmecca/corsi/sviluppoWeb>)

Sommario

- Storia di HTML
- Principali problemi tecnologici
- Standard rilevanti
 - ⇒ HTML 1.0 e CSS
 - ⇒ Modularizzazione e HTML 1.1
- XHTML Basic
 - ⇒ principali moduli
 - ⇒ convalida del codice
 - ⇒ uso degli standard

La Storia di HTML

- HTML
 - ⇒ un linguaggio per descrivere documenti ipermediali
 - ⇒ documenti fatti di titoli, paragrafi, testo ecc.
 - ⇒ collegamenti (link) ad altri documenti
 - ⇒ contenuti di altri media (es: immagini)
- Linguaggio di marcatura
 - ⇒ obbedisce ad un DTD di SGML

La Storia di HTML

○ SGML

- ⇒ sintassi per descrivere la struttura gerarchica di documenti
- ⇒ il “padre” di XML (anni '70); più complesso
- ⇒ DTD: schemi per i documenti

○ Meno restrittivo di XML

- ⇒ i tag di chiusura possono mancare
- ⇒ maiuscole e minuscole
- ⇒ apici per gli attributi

La Storia di HTML

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
<HTML>
  <HEAD>
    <TITLE> Il portale universale </TITLE>
  </HEAD>
  <BODY>
    <H1>Portale Universale</H1>
    <P>Questa &egrave; la pagina principale del
    <B>Portale Universale</B>. <img src=portale.gif>
    <A href="materiale.html">Materiale Interessante</A>
    ...
    <I>Ultima modifica Aprile 2002</I>
  </BODY>
</HTML>
```

La Storia di HTML

○ HTML 1.0: La prima versione di HTML

- ⇒ Tim Berners Lee, Dave Raggett, giugno '93

○ Il DTD di HTML 1.0

- ⇒ titoli (“headings”)
 - Es. <h1>, <h2>, <h3>, ...
- ⇒ paragrafi e formattazione del testo
 - Es. <p>,
, , <i>, ...
- ⇒ liste e glossari
 - Es. , , , ...
- ⇒ collegamenti ad altre pagine
 - Es. <a>
- ⇒ Immagini
 - Es.

La Storia di HTML

○ Primi browser testuali

- ⇒ Lynx, 1991 – 1992

○ Primi browser grafici

- ⇒ Mosaic 1.0
novembre '93, Marc Andreessen, NCSA-UIUC
- ⇒ Mosaic 2.0 v. alfa
gennaio '94, introduce le maschere (“form”)

○ 1994

- ⇒ Andreessen e Jim Clark fondano la Netscape

La Storia di HTML

- W3C
 - ⇒ ottobre '94, fondato dal MIT
 - ⇒ comincia l'attività di standardizzazione
- HTML 2.0 (RFC 1866)
 - ⇒ tra l'aprile '94 e il settembre 1995
 - ⇒ standardizza le maschere e le mappe
- Netscape Navigator 1.0
 - ⇒ dicembre '94, supporto HTML 2.0 e nuovi tag

La Storia di HTML

- Netscape 1.1
 - ⇒ aprile '95, introduce le tabelle e nuovi tag
- Internet Explorer 1.0
 - ⇒ agosto '95, distribuito gratuitamente
- Nel 1995
 - ⇒ Netscape ha l'**80%** del mercato
 - ⇒ nascono i primi motori di ricerca
 - ⇒ inizia la "guerra dei browser" (*browser war*)

La Storia di HTML

- HTML 3.0 (draft W3C)
 - ⇒ marzo '95 – settembre '95, tabelle
- Internet Explorer 2.0
 - ⇒ novembre '95, tabelle <marquee>
- Netscape Navigator 2.0
 - ⇒ marzo '96, frame, plug-in, applet, Javascript
- CSS1, Cascading Style Sheets Livello 1
 - ⇒ novembre '95 – dicembre '96

La Storia di HTML

- HTML 3.2
 - ⇒ standard W3C, maggio '96 – gennaio '97
 - ⇒ introduce script e tag di presentazione
- Netscape Navigator 3.0
 - ⇒ agosto '96, tag <multicol>, <spacer>
- Internet Explorer 3.0
 - ⇒ agosto '96, JScript, VBScript

La Storia di HTML

- Netscape Navigator 4.0
 - ⇒ giugno '97, layer, CSS1
- Internet Explorer 4.0
 - ⇒ ottobre '97, versione migliorata, CSS1
- Gennaio 1998
 - ⇒ la quota di Netscape è scesa al **40%** circa
 - ⇒ il browser diventa gratuito
 - ⇒ strategia “open source” (nasce Mozilla)

La Storia di HTML

- Si affacciano altre tecnologie
 - ⇒ DHTML “Dynamic HTML”
 - ⇒ combinazione di DOM, CSS e JavaScript
 - ⇒ possibilità di animare elementi della pagina
- Macromedia Flash
 - ⇒ grafica vettoriale
 - ⇒ animazioni
- Alcuni fallimenti autorevoli

La Storia di HTML

- HTML 4.0
 - ⇒ giugno '97 – aprile '98
 - ⇒ introduce i fogli di stile, standardizza i frame
 - ⇒ DTD “strict”: solo elementi strutturali
 - ⇒ DTD “transitional”
 - ⇒ DTD “frameset”
- CSS2
 - ⇒ maggio '98, posizionamento assoluto

La Storia di HTML

- Internet Explorer 5.0
 - ⇒ marzo '99, supporto per XML, CSS1 e CSS2
- Netscape 6
 - ⇒ novembre 2000, abbastanza instabile
- Ottobre – Novembre 2001
 - ⇒ Netscape 6.2
 - ⇒ Internet Explorer 6
 - ⇒ Opera 6

Principali Problemi del Linguaggio

- Problema 1: Evoluzione molto rapida
 - ⇒ mancanza di metodologie consolidate
 - ⇒ struttura e presentazione sono indipendenti
- Indicazione del Consorzio
 - ⇒ separare struttura e presentazione
 - ⇒ CSS livello 1 e CSS livello 2

Principali Problemi del Linguaggio

- Problema 2: Codice scorretto
 - ⇒ i browser sono “onnivori”
 - ⇒ i DTD dello standard non sono rispettati
 - ⇒ problemi per le applicazioni
- Indicazione del Consorzio
 - ⇒ adottare la sintassi XML
 - ⇒ XHTML 1.0: una riformulazione di HTML 4.0 secondo lo standard di XML

Principali Problemi del Linguaggio

- Problema 3: Diverse capacità dei browser
 - ⇒ diverse versioni di HTML, diversi DTD
 - ⇒ diverse piattaforme (palmari, WebTV, ...)
- Indicazione del Consorzio
 - ⇒ modularizzazione di XHTML e negoziazione
 - ⇒ principio di “accessibilità”
 - ⇒ estensibilità
 - ⇒ “Modularization of XHTML”, XHTML1.1

Standard Rilevanti

- XHTML 1.0
 - ⇒ gennaio 2000
- CSS livello 1, CSS livello 2
 - ⇒ dicembre '96, maggio '98
- Modularization of XHTML, XHTML 1.1
 - ⇒ aprile 2001, maggio 2001
- XHTML Basic
 - ⇒ dicembre 2000

XHTML 1.0

- Riformulazione di HTML 4.0
 - ⇒ formato XML
- DTD di riferimento
 - ⇒ riformulazione dei DTD di HTML 4.0
 - ⇒ DTD strict XHTML1-strict.dtd
 - ⇒ DTD transitional XHTML1-transitional.dtd
 - ⇒ DTD frameset XHTML1-frameset.dtd

XHTML 1.0

- Vincoli sulla sintassi
 - ⇒ validità rispetto ad uno dei DTD


```
<!DOCTYPE html
PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"xhtml1-strict.dtd">
```
 - ⇒ l'elemento radice è <html>
 - ⇒ il namespace del documento è


```
http://www.w3.org/1999/xhtml
```
 - ⇒ utilizzo delle minuscole per i nomi

XHTML 1.0

- Vincoli sulla sintassi
 - ⇒ il documento deve essere ben formato
 - ⇒ elementi vuoti;


```
es: <br />
```
 - ⇒ valori degli attributi tra apici;


```
es: xmlns="http://www.w3.org/1999/xhtml"
```
 - ⇒ codifica delle entità;


```
es: &amp; ;
```

Latin-1 characters, Special characters, Symbols

XHTML 1.0 e CSS

- Separazione tra struttura e presentazione
 - ⇒ i tag descrivono solo la struttura
 - ⇒ fogli di stile CSS: associano caratteristiche di visualizzazione agli elementi (>>)
 - ⇒ foglio di stile standard del browser;


```
es: testo in carattere Times 10 punti tondo
```

Modularizzazione

- Modulo (“astratto”)
 - ⇒ collezione di elementi, attributi e definizioni di contenuto per gli elementi
- Implementazione di un modulo
 - ⇒ DTD che descrive elementi e attributi
- Modularizzazione di XHTML
 - ⇒ frammentazione in moduli
 - ⇒ modalità per definire nuovi moduli

Elenco dei Moduli

- | | |
|---|--|
| <ul style="list-style-type: none"> ○ Nucleo (“Core”) <ul style="list-style-type: none"> ⇒ Structure, Text, Hypertext, List ○ Estensioni del testo <ul style="list-style-type: none"> ⇒ Presentation, Edit, Bidirectional Text ○ Maschere <ul style="list-style-type: none"> ⇒ Basic Forms, Forms ○ Tabelle <ul style="list-style-type: none"> ⇒ Basic Table, Table ○ Immagini <ul style="list-style-type: none"> ⇒ Image | <ul style="list-style-type: none"> ○ Mappe <ul style="list-style-type: none"> ⇒ Client Side Map, Server Side Map ○ Frame <ul style="list-style-type: none"> ⇒ Frame, Target, Iframe ○ Fogli di stile <ul style="list-style-type: none"> ⇒ Stylesheet, Link ○ Altri <ul style="list-style-type: none"> ⇒ Intrinsic Events, Metainformation, Scripting, Object, Base, Name, Legacy |
|---|--|

Modularizzazione e XHTML 1.1

- Standard per definire altri moduli
- XHTML come “linguaggio ospite”
 - ⇒ linguaggi derivati combinando moduli (DTD)
 - ⇒ devono contenere almeno i moduli “core” (Structure, Text, Hypertext, List)
- XHTML 1.1: “Module Based XHTML”
 - ⇒ selezione di moduli di XHTML 1.0
 - ⇒ nuovi moduli (Ruby)

XHTML Basic

- Insieme minimale di moduli
 - ⇒ per tutti i browser (cHTML e WML)
 - ⇒ ragionevole capacità di formattazione
- Organizzazione
 - ⇒ moduli del nucleo
 - ⇒ fogli di stile esterni
 - ⇒ immagini
 - ⇒ semplici maschere
 - ⇒ semplici tabelle

XHTML Basic

- Elenco completo
 - ⇒ Structure, Text, Hypertext, List
 - ⇒ Image
 - ⇒ Basic Forms
 - ⇒ Basic Tables
 - ⇒ Link
 - ⇒ Object
 - ⇒ Meta, Base
- Mancano
 - ⇒ script (es: JavaScript)
 - ⇒ presentazione (es: b)
 - ⇒ mappe
 - ⇒ tabelle nidificate
 - ⇒ frame
 - ⇒ "legacy" tag (es: font)

XHTML Basic

- Nella lezione
 - ⇒ Structure, Text, Hypertext, List
 - ⇒ Image
 - ⇒ Basic Tables
 - ⇒ Meta
- Nelle lezioni successive
 - ⇒ Basic Forms
 - ⇒ Link
 - ⇒ Object
 - ⇒ Base

XHTML Basic

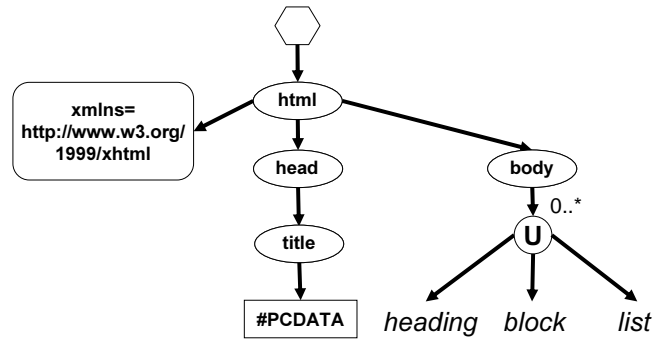
○ Vincoli sulla sintassi di XHTML 1.0

```
<?xml version="1.0" ?>
<!DOCTYPE html
  PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
  "DTD/xhtml11-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
...
</html>
```

XHTML Basic

- Attributi di uso comune
 - ⇒ id
 - ⇒ class
 - ⇒ title
 - ⇒ style
 - ⇒ xml:lang
- Entità
 - ⇒ Latin, Special, Symbol

Modulo di Struttura



Modulo di Struttura

elemento	attributo	modello di contenuto
html	xmlns ...	(head, body)
head	...	(title)
title	...	#PCDATA
body	...	(<i>heading</i> <i>block</i> <i>list</i>)*

Modulo di Struttura

```

<?xml version="1.0" ?>
  <!DOCTYPE html
    PUBLIC "-//W3C//DTD XHTML Basic 1.0//EN"
      "DTD-xhtmlbasic/xhtml-basic10.dtd">
  <html xmlns="http://www.w3.org/1999/xhtml">
    <head>
      <title>Esempio</title>
    </head>
    <body>
      ...
    </body>
  </html>
  
```

Modulo di Testo

- Contiene elementi detti di flusso (*"flow"*)
 - ⇒ Heading
 - ⇒ titoli delle sezioni del documento
 - ⇒ Block
 - ⇒ blocchi identificabili del documento
 - ⇒ Possono contenere altri blocchi, inline e testo
 - ⇒ Inline
 - ⇒ porzioni di testo, contengono solo inline o testo

Modulo di Testo

- **Heading**
h1, h2, h3, h4, h5, h6
- **Block**
 - ⇒ p: paragrafi
 - ⇒ pre: testo preformatt.
 - ⇒ address: indirizzi
 - ⇒ blockquote: citazioni (attributo cite)
 - ⇒ div: raggruppamento di elem. in blocco
- **Inline**
 - ⇒ em: enfattizzato
 - ⇒ strong: rafforzato
 - ⇒ br: interruzione linea
 - ⇒ q: citazioni brevi (attributo cite)
 - ⇒ span: raggruppamenti
 - ⇒ ...

Modulo di Testo

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Esempio</title>
  </head>
  <body>
    <h1>Titolo principale</h1>

    <p>Questo &grave; un documento di esempio
    relativo ad <em>XHTML Basic</em> <br />
    Sono andato a capo</p>
  </body>
</html>
```

Modulo Iperestuale

- **Elemento a (ancora)**
 - ⇒ collegamenti ad altri URI
 - ⇒ attributo href con riferimento ad URI
 - ⇒ si aggiunge agli elementi *inline*

elemento	attributo	modello di contenuto
a	href ...	(#PCDATA <i>inline</i>)

Modulo Iperestuale

```
<body>
  <h1>Titolo principale</h1>
  <p>Questo &grave; un documento di esempio
    relativo ad <em>XHTML Basic</em> <br />
    Sono andato a capo</p>
  <a href="http://www.difa.unibas.it">
    <strong>Sito del DIFA</strong>
  </a>
  <a href="documenti/prova.html">
    Una pagina di prova
  </a>
</body>
```

Modulo delle Liste

- Liste di elementi
- Numerate (ol) e non numerate (ul)

elemento	attributo	modello di contenuto
ol	...	(li)+
ul	...	(li)+
li	...	(#PCDATA flow)*

Modulo delle Liste

```
<body> ...
  <a href="http://www.difa.unibas.it">
    <strong>Sito del DIFA</strong>
  </a>
  <a href="documenti/prova.html">
    Una pagina di prova
  </a>
  <ul>
    <li><p>primo elemento di una lista</p></li>
    <li><p>secondo elemento di una lista</p></li>
  </ul>
</body>
```

Modulo delle Immagini

- Elemento `img`
 - ⇒ esprime il collegamento ad un'immagine
 - ⇒ formati ordinari: gif, jpg
 - ⇒ si aggiunge agli elementi *inline*

elemento	attributo	modello di contenuto
img	src alt width, height, ...	EMPTY

Modulo delle Immagini

```
<body>
  ...
  <ul>
    <li><p>primo elemento di una lista</p></li>
    <li><p>secondo elemento di una lista</p></li>
  </ul>
  
</body>
```

Modulo delle Tabelle

○ Tabelle

- ⇒ elemento radice: `table`
- ⇒ didascalia dopo la radice: `caption`
- ⇒ una o più righe: `tr`
- ⇒ fatte di una o più celle
 - ⇒ celle di intestazione: `th`
 - ⇒ celle dei valori: `td`

Modulo delle Tabelle

elemento	attributo	modello di contenuto
table	summary ...	(caption?, tr+)
caption	...	(#PCDATA <i>inline</i>)*
tr	...	(th td)+
th	colspan, rowspan, scope (row column) ...	(#PCDATA <i>flow</i>)* meno altri elementi table
td	colspan, rowspan ...	(#PCDATA <i>flow</i>)* meno altri elementi table

Modulo delle Tabelle

```
<body>
...
<table>
  <caption>Lezioni</caption>
  <tr>
    <th>Data</th><th>Argomento</th><th>Tipo</th>
  </tr>
  <tr>
    <td>Lun 17 feb 03</td><td>HTTP</td><td>Lezione</td>
  </tr>
  <tr>
    <td>Mar 18 feb 03</td><td>XML</td><td>Lezione</td>
  </tr>
  <tr>
    <td>Mer 19 feb 03</td><td>XML</td><td>Esercizi</td>
  </tr>
  ...
</table>
...
</body>
```

Modulo di Metainformazioni

○ Si aggiunge al contenuto di head

- ⇒ head (title, meta+)
- ⇒ Insieme di coppie name, content

es: `<meta name="author" content="Pinco" />`

- ⇒ http-equiv: intestazioni HTTP

es: `<meta http-equiv="Expires" content="<data>" />`

elemento	attributo	modello di contenuto
meta	name, content, http-equiv ...	EMPTY