

XML

eXtensible Markup Language

eXtensible

- Chiunque può personalizzare il linguaggio aggiungendovi nuovi elementi
- Metalinguaggio: insieme di regole per la definizione di un linguaggio
- Esistono diversi “dialetti” xml, che condividono le regole grammaticali e sintattiche del linguaggio xml
- Bisogno di uno sviluppo “coordinato” dei linguaggi XML

Markup

- Annotazione inserita all'interno del testo di un documento utilizzata per inserire informazioni aggiuntive (grafiche, strutturali, semantiche, ...)
- In generale un *linguaggio di markup* descrive i meccanismi di rappresentazione (strutturali, semantici o presentazionali) del testo che, utilizzando convenzioni standardizzate, sono utilizzabili su più supporti (Wikipedia)

World Wide Web Consortium

- Organizzazione il cui scopo è lo sviluppo delle tecnologie destinate al World Wide Web
- Standardizzazione
- Recommendation
 - HTML
 - CSS
 - SGML

SGML

- **Standard General Markup Language**
- Matalinguaggio che permette di definire diversi linguaggi di marcatura
- Permette di definire sia aspetti strutturali che di formattazione
- Estremamente complesso
- Linguaggi derivati da SGML
 - HTML
 - XML
 - ...

All'inizio era HTML

- Linguaggio “spurio”, che descrive sia elementi strutturali che di formattazione
- Difficoltà di analisi automatica dei documenti
- Necessità di separare i due aspetti
- Riutilizzo delle impostazioni di formattazione (CSS, ...)

... poi, arrivò XML

- Nuovo linguaggio standardizzato, estremamente flessibile e sintatticamente rigoroso
- Molto più semplice dell'SGML, conserva comunque le caratteristiche salienti di tale linguaggio
- Permette di comporre documenti comprensibili sia all'uomo che alla macchina

XML

- In HTML la rappresentazione dei dati si basa sulla forma
- In XML la rappresentazione dei dati è resa attraverso la struttura
- Indipendenza dei dati → netta separazione tra il contenuto e la sua rappresentazione

XML e Web

- Largo impiego dell'XML nel Web
- Spesso utilizzato come standard per la comunicazione client/server
- Di recente utilizzato anche come formato per la conservazione ed il recupero di dati strutturati (Database XML e mapping)

Tag

- elemento sintattico con cui si marcano porzioni di un documento nei linguaggi di markup (Wikipedia)
- In XML forma ed uso dei tag sono ispirati agli analoghi tag dell'HTML
- Parole chiave racchiuse tra parentesi angolari

Tag

- I tag di un documento XML sono sempre in coppia
- Tag di apertura
- Tag di chiusura

`<primotag>`

... (contenuto vario)

`</primotag>`

Tag

- La disposizione dei tag è analoga a quella delle parentesi in un'espressione algebrica
- Ogni elemento di un documento XML è racchiuso tra un tag di apertura ed il corrispondente tag di chiusura

Tag

- Inserimento degli elementi “a scatole cinesi”
- L'ordine di chiusura dei tag è inverso a quello di apertura

Tag

```
<primotag>  
  <secondotag>  
    <terzotag>  
      ...  
    </terzotag>  
  </secondotag>  
</primotag>
```

Tag

- La particolare indentazione degli elementi di un documento XML serve soprattutto a semplificare la lettura del documento
- Linguaggio estendibile perché permette di inventare sempre nuovi tag
- Case sensitive
- L'autore del documento è libero di adoperare qualsiasi parola per un tag
- Unico vincolo: il rispetto delle regole sintattiche dell'XML

Struttura del documento XML

- La struttura di un documento XML è omomorfa a quella di un albero
- Ciascun documento ha un elemento radice
- L'elemento radice è uno ed uno solo
- Ciascun elemento del documento può contenere altri elementi (uno o più)
- Un elemento che non contiene altri elementi è una foglia

Struttura del documento XML

- Esistono diversi tipi di elementi in un documento XML
 - Testo
 - Contenuti particolari
 - Altri nodi
 - Interi sottoalberi

Struttura del documento XML

- Una coppia di tag può anche non racchiudere alcun contenuto

```
<root>
```

```
  <null></null>
```

```
</root>
```

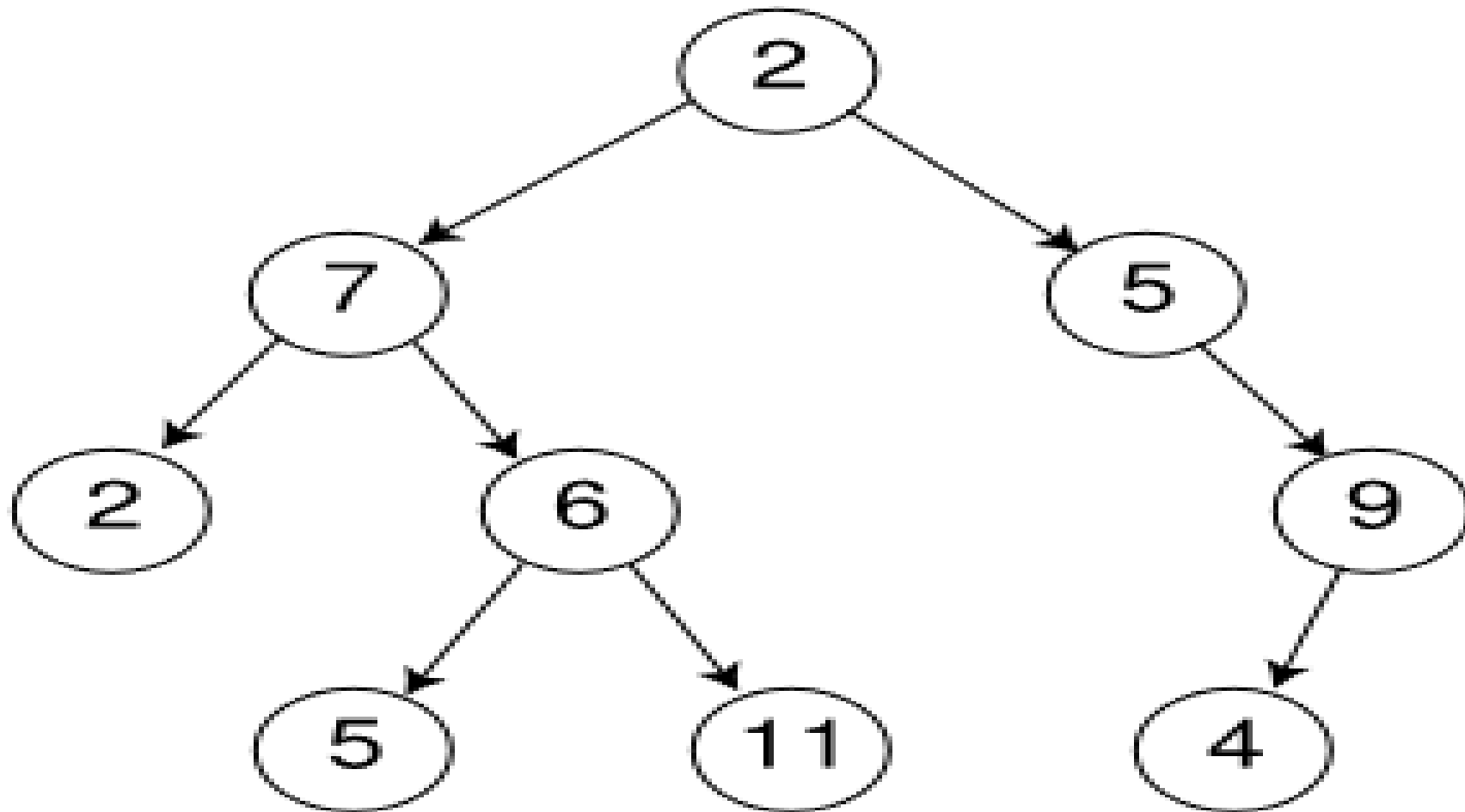
- In tali casi si utilizza spesso una forma contratta per rappresentare la coppia dei tag di apertura e chiusura

```
<root>
```

```
  <null/>
```

```
</root>
```

Struttura del documento XML



Struttura del documento XML

```
<root>  
  <person>  
    <name>John</name>  
    <surname>Smith</surname>  
  </person>  
  <person>  
    <name>Mary</name>  
    <surname>Brown</surname>  
  </person>  
  ...  
</root>
```

Gli attributi

- Ciascun tag può contenere degli attributi
nomeattributo="valoreattributo"
- Il valore di un attributo va sempre messo tra virgolette

```
<person sex="M">
```

```
  <name>John</name>
```

```
</person>
```

Processing Instruction

- I dati contenuti in una istruzione di elaborazione vengono passate all'applicazione che usa il documento XML
- Solo delimitate dai caratteri `<? E ?>`
- Sono formate da un *target* e da un *valore*
`<?xml:stylesheet type="text/xsl" href="usage.xsl" ?>`

Intestazione del documento XML

- Tutti i documenti XML devono contenere una dichiarazione nel primo rigo

```
<?xml version="1.0" standalone="yes" ?>
```

(intestazione)

```
<myMessage>
```

```
    <message>hello</message>
```

```
</myMessage>
```

Intestazione del documento XML

- Anche l'intestazione è un esempio di Processing Instruction
- Essa indica la versione dell'XML in cui è stato scritto il documento
- A volte indica anche se il documento è da ritenersi come una risorsa a se stante o se legata ad altre

Le sezioni CDATA

- Tali sezioni possono contenere testi, caratteri riservati (come < e >) e caratteri vuoti
- Tali caratteri NON sono elaborati dal parser

```
<![CDATA[
```

```
 2<3
```

```
]]>
```

I commenti

- I commenti sono compresi tra i caratteri `<!-- e -->`

`<root>`

...

`<!-- questo è un commento -->`

...

`</root>`

I namespace

- Chiunque può definire i propri tag
- Conflitti di nomi
- Posso definire dei namespace
- `<prefissonamespace:tag>`

```
<radice xmlns:image="urn:deitel:imageInfo">  
  <image:file filename="bunny.jpg" />  
</radice>
```