



Università Degli Studi di Salerno

Dipartimento di Scienze del Patrimonio Culturale
Corso di Laurea in Discipline delle Arti Visive, della Musica e dello Spettacolo

ANNO ACCADEMICO 2019/2020

CALABRO' CATERINA
D'ANZILIO SILVIA
SALERNO ASSUNTA

XML (EXTENSIBLE MARKUP LANGUAGE)

- Consente la rappresentazione di documenti su supporto digitale e ne permette la creazione, memorizzazione e distribuzione
- È un **linguaggio a marcatori** → Composto da istruzioni definite tag o marcatori, che descrivono la struttura e la forma di un documento. Ogni marcatore o coppia di marcatori identifica un **elemento**
- Sia il testo, sia i marcatori sono memorizzati in formato ASCII
- XML usa la codifica dei caratteri Unicode

L'XML è:

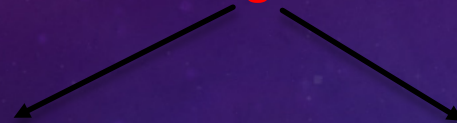
- **Case Sensitive** → nei nomi dei tag distingue fra maiuscole e minuscole
- **Portabile** → indipendente dal tipo di piattaforma hardware e software e può essere utilizzato per qualsiasi finalità

- Nell' XML si può definire, secondo una precisa struttura logica, una **grammatica**



è un insieme di regole che indica quali vocaboli (elementi) possono essere utilizzati e con che struttura è possibile comporre frasi (documenti)

- All'interno di un XML vi sono delle **regole**



Sintattiche: come dobbiamo scrivere le informazioni all'interno dei documenti

Semantiche: cosa possiamo scrivere in un documento XML

- Un documento XML che rispetta le regole sintattiche si dice **ben formato** (well-formed)
- Un documento XML che rispetta le regole sintattiche e le regole semantiche si dice **valido**
- Un documento ben formato può non essere valido rispetto ad una grammatica, mentre un documento valido è necessariamente ben formato

Un documento XML si divide in due parti:

- **PROLOGO** → contiene una **xml declaration**:
 1. Versione: per ora solo 1.0
 2. Set di caratteri (opzionale):
 - ASCII: set di caratteri a 7 bit
 - UTF-8: unicode a 8 bit
 - UTF-16: unicode a 16 bit

Può contenere riferimenti a documenti esterni utili per il trattamento del documento

- **CORPO** → è il documento XML vero e proprio

STRUTTURA LOGICA

- Un documento XML è strutturato in modo gerarchico ed è composto da **elementi**



rappresentano un componente logico del documento e possono contenere un frammento di testo oppure altri elementi (**sotto-elementi**)

- Ad un elemento possono essere associate informazioni descrittive chiamate **attributi** → rappresentati sotto forma di coppie nome-valore all'interno dei tag
- Gli elementi sono organizzati ad albero con **radice (root)**
- Un documento può inoltre contenere spazi bianchi, a capo e commenti XML
- Il mark-up è separato dal contenuto testuale mediante caratteri speciali: **< > &** (parentesi angolari e ampersand)
 - I **commenti** possono apparire ovunque in un documento XML (sia nel prologo che nel corpo) e sono utili per spiegare la struttura del documento XML e commentare parti del documento durante le fasi di sviluppo e di test del nostro software

Esempio XML nell'ambito della catalogazione di specie agricole

RADICE
(MARCATORE DI INIZIO)

PROLOGO

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Catalogo SYSTEM "catalogo.dtd">

<Catalogo>
  <SpecieAgricola tipo="Cereale">
    <Nome>Mais</Nome>
    <NomeScientifico>Zea mays L.</NomeScientifico>
    <PeriodoColtivazione>Aprile</PeriodoColtivazione>
    <PeriodoRaccolto>Agosto</PeriodoRaccolto>
    <PeriodoRaccolto>Settembre</PeriodoRaccolto>
    <MetodiDiCottura>Forno</MetodiDiCottura>
    <MetodiDiCottura>Vapore</MetodiDiCottura>
    <MetodiDiCottura>Griglia</MetodiDiCottura>
    <Conservazione></Conservazione>
  </SpecieAgricola>
  <SpecieAgricola tipo="Barbabietola">
    <Nome>Barbabietola da zucchero</Nome>
    <NomeScientifico>Beta vulgaris</NomeScientifico>
    <PeriodoColtivazione>Autunno</PeriodoColtivazione>
    <PeriodoColtivazione>Primavera</PeriodoColtivazione>
    <PeriodoRaccolto>Primavera</PeriodoRaccolto>
    <PeriodoRaccolto>Estate</PeriodoRaccolto>
    <Conservazione>In luogo fresco e ombroso</Conservazione>
  </SpecieAgricola>
</Catalogo>

<!--
  <SpecieAgricola tipo="">
    <Nome></Nome>
    <PeriodoColtivazione></PeriodoColtivazione>
    <PeriodoRaccolto></PeriodoRaccolto>
    <MetodiDiCottura></MetodiDiCottura>
    <Conservazione></Conservazione>
  </SpecieAgricola>
-->
```

ELEMENTO

SOTTOELEMENTI

ATTRIBUTO

MARCATORE DI FINE

CORPO

COMMENTO

XML realizzato mediante il programma Notepad ++

DTD: Document Type Definition

- È un linguaggio per definire la grammatica che descrive la composizione degli elementi costituenti una certa classe di documenti XML
- Fornisce uno strumento per la validazione dei documenti XML
- Non è un linguaggio XML (non rispetta la sintassi XML)
- Per applicare un DTD ad un documento XML nel suo prologo dobbiamo inserire una dichiarazione con questa sintassi: `<!DOCTYPE root-element SYSTEM "filename">`

nome dell'elemento radice

definisce documenti di
utilizzo locale

è il file che contiene
il DTD

- La dichiarazione va posta sotto l'XML Declaration

Le dichiarazioni di un DTD definiscono:

1. Gli **elementi** di un documento XML



Per dichiarare un elemento si usa la sintassi:

```
<!ELEMENT element-name  
content-model>
```

2. Il modello di contenuto di ogni elemento (**content model**), ovvero gli elementi che contiene e le loro relazioni (un elemento può essere vuoto)



Il contenuto può essere di 4 tipi:

EMPTY

ANY

Content-model Children

Content model Mixed

Tutte le dichiarazioni sono globali

Un elemento può essere indicato una sola volta

3. La lista degli **attributi** associati a ciascun elemento



Definiti attraverso **tipi e valori**

CONTENT MODEL

EMPTY → parola chiave che indica l'elemento vuoto

ANY → indica che si può inserire testo o elementi qualsiasi

Content-model Children → Può specificare: 1. una **sequenza** es .Titolo, Autore, Anno, Genere
2. una **scelta** es. Titolo | Autore | Anno | Genere

In entrambi i casi si può stabilire quante volte è visibile un determinato elemento grazie agli **OPERATORI**

- * Zero o più volte
- ? Zero o 1 volta
- + 1 o più volte

Content-model Mixed → Elenco di elementi senza un ordine preciso, perciò si scrive una lista scelta che deve comparire zero o più volte (*)

(es. #PCDATA Anno | Genere | Autore | Titolo) *

Deve essere sempre il primo elemento della lista scelta

ATTRIBUTI

1. TIPI:

- CDATA (testo)
- Valore scelto (e1|e2|...|en)
- ID (identificatore univoco a livello di documento)

2. VALORI:

- #REQUIRED (obbligatorio)
- #IMPLIED (opzionale)
- #FIXED (attributo deve avere un valore prefissato)

Esempio DTD nell'ambito della catalogazione di specie agricole

ATTRIBUTI

OPERATORI

```
<!ELEMENT Catalogo (SpecieAgricola)*>  
<!ELEMENT SpecieAgricola (Nome, NomeScientifico, PeriodoColtivazione+ PeriodoRaccolto+, MetodiDiCottura*, ValoriNutrizionali*, Conservazione+)>  
<ATTLIS! SpecieAgricola  
  tipo (Barbabietola|PiantaForaggera|PiantaDaFibra|Cereale|Patata) #REQUIRED>  
<!ELEMENT Nome (#PCDATA)>  
<!ELEMENT NomeScientifico (#PCDATA)>  
<!ELEMENT PeriodoColtivazione (#PCDATA)>  
<!ELEMENT PeriodoRaccolto (#PCDATA)>  
<!ELEMENT MetodiDiCottura (#PCDATA)>  
<!ELEMENT Conservazione (#PCDATA)>
```

DTD realizzato mediante il programma Notepad ++