

IL LINGUAGGIO XML

PER LA CATALOGAZIONE DI CONTENUTI
MULTIMEDIALI

CHE COS'È?

XML (eXtensible Markup Language) è un metalinguaggio creato *dal World Wide Web Consortium (W3C)* nel 1998 come una semplificazione dello Standard Generalized Markup Language (SGML), esso è orientato all'utilizzo sul WWW per la rappresentazione di dati e documenti strutturati su un supporto digitale.

Ogni documento deve presentare un prologo per la dichiarazione della specifica XML che si usa, le eventuali informazioni di codifica e il richiamo esplicito alla DTD in uso.



La struttura vera e propria è composta da un elemento radice che raggruppa altri elementi tramite dei tag che descrivono la struttura e la forma del documento; ogni marcatore (o coppia) identifica un elemento (al quale possono essere associati uno o più attributi). I tag devono essere scritti seguendo tre indicazioni importanti:

Si aprono e si chiudono in ordine inverso (LAST IN FIRST OUT), ovvero, l'ultimo tag che si apre sarà il primo a chiudersi.

Devono essere comprensibili in funzione dello scopo.
(e per facilitare la comprensione agli altri utenti)

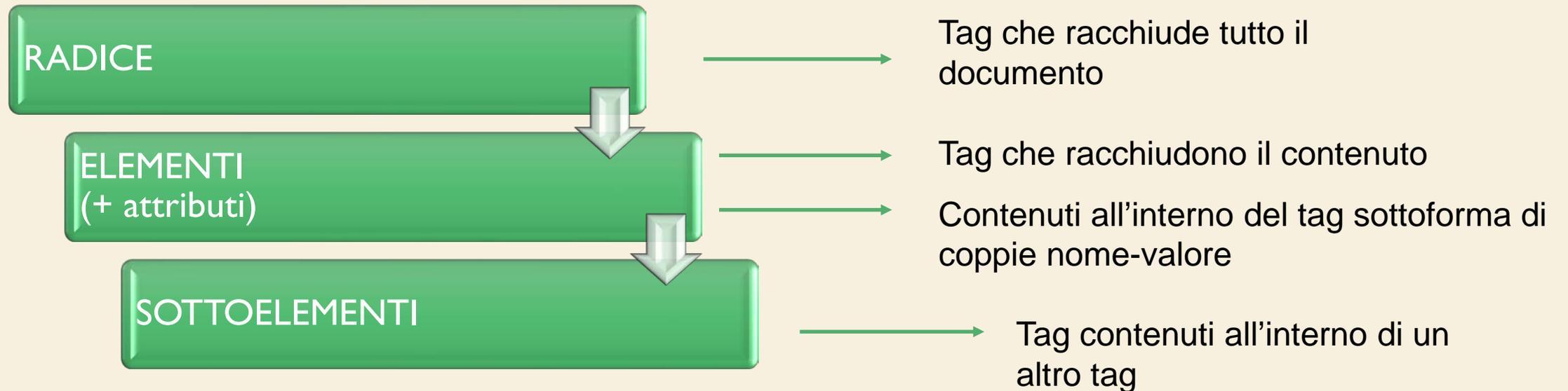
devono rispettare delle regole, come la differenza tra maiuscolo e minuscolo, non possono iniziare con numeri o caratteri speciali e non possono contenere spazi.

Inoltre, in un documento XML ci sono delle:

- **REGOLE SINTATTICHE** che ci indicano **COME** dobbiamo scrivere le informazioni all'interno dei documenti;
- **REGOLE SEMANTICHE** (opzionali): che ci indicano **COSA** possiamo scrivere all'interno dei documenti.

Un documento XML che rispetta le regole sintattiche si dice *ben formato* (ma può non essere valido rispetto ad una grammatica), mentre uno che rispetta sia le regole sintattiche che quelle semantiche si dice *valido* (ed è necessariamente valido rispetto ad una grammatica).

RACICE, ELEMENTI, SOTTO-ELEMENTI E ATTRIBUTI



Attenzione alla differenza tra ELEMENTI e ATTRIBUTI!

- Un elemento è estendibile in termini di contenuto e attributo
- Un attributo non è estendibile, può solo modellare una proprietà di un elemento in termini di valore.

DOCUMENT TYPE DEFINITION (DTD)

Un DTD è un documento di validazione costituito da un elenco di dichiarazioni (markup declarations) che descrivono la struttura del documento XML.

Le dichiarazioni di un DTD definiscono, quindi:

ELEMENTI

- **EMPTY** (elemento vuoto);
- **ANY** (indica che si può inserire testo o elementi).

CONTENT MODEL (Il modello di contenuto di ogni elemento)

- **CONTENT-MODEL CHILDREN** (elenco di elementi figli specifici con ordine);
 - **CONTENT-MODEL MIXED** (elenco di elementi figli senza ordine specifico).
- } ? 0 o più elementi
+ 1 o più elementi
* 0 o più elementi
- } (#PCDATA | es1 | es2 | es3 |)*
oltre al testo può contenere qualsiasi numero di elementi.

ATTRIBUTI Associati a ciascun elemento

- **CDATA** corrisponde al testo;
- (**es1 | es2 | es3**) lista di enumerazione;
- **ID** identificatore univoco a livello documento;
- **VALUE** valore di default;
- **#REQUIRED** deve essere presente;
- **#IMPLIED** è opzionale
- **#FIXED** valore fisso

LINGUAGGIO XML PER LA CATALOGAZIONE DI FILE MULTIMEDIALI:

Il potere del linguaggio XML risiede nella sua semplicità, nonostante riesca a contenere grandi quantità di informazioni in un solo documento, esso ci permette di strutturarle e quindi organizzarle in modo tale che sia semplice effettuare ricerche ed elaborazioni.

Infatti, il linguaggio XML viene definito portabile (proprietà di programmi che possono essere impiegati in hardware diversi senza subire modifiche) e flessibile, quindi, può essere impiegato nelle applicazioni Web e di contenuti multimediali come: catalogazione di spartiti musicali, opere teatrali e cinematografiche. Ciò permette di offrire numerosi e consistenti vantaggi agli sviluppatori e agli utenti.

LIBRERIA

NETFLIX

Netflix è una piattaforma per la fruizione di contenuti multimediali d'intrattenimento che offre legalmente la visione di prodotti come film, serie TV, show, documentari in streaming su Internet, in modalità on demand.

Sulla base di ciò che è stato detto precedentemente, quindi, si prenderà in esame la costruzione di un documento XML di catalogazione e relativo DTD di validazione di una libreria Netflix, in cui:

- *La libreria può contenere uno o più film (+)*
- *Un film è descritto da: TITOLO, REGISTA, ATTORI(+) E IL GENERE(*)*
- *Un film è dotato di proprietà quali: CODICE DI CATALOGAZIONE (REQUIRED), CODICE DI CLASSIFICAZIONE (REQUIRED), ANNO (IMPLIED), FORMATO (REQUIRED), ORIGINALE NETFLIX (SI | NO).*

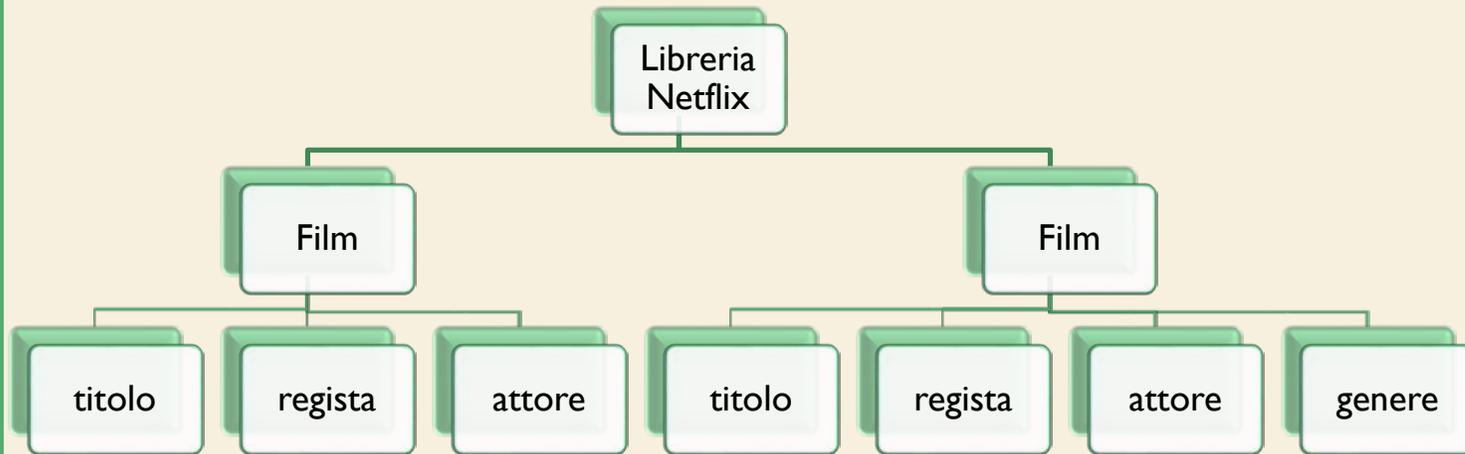
```
<!ELEMENT Liberia Netflix (Film+)>
<!ELEMENT Film (Titolo, Regista, Attore+, Genere*)
<!ELEMENT Titolo (#PCDATA)>
<!ELEMENT Regista (#PCDATA)>
<!ELEMENT Attore (#PCDATA)>
<!ELEMENT Genere (#PCDATA)>
<!ATTLIST Film cod classificazione ID #REQUIRED
  cod catalogazione #REQUIRED
  anno CDATA #IMPLIED
  formato (HD|STANDARD) #REQUIRED
  originale netflix (SI|NO)>
```

DTD

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE Liberia Netflix SYSTEM "liberianetflix.dtd"
<Libreria Netflix>
  <Film cod classificazione="TI123" cod catalogazione="VM14" anno="2019" formato="HD" originale netflix="si"
    <Titolo>"The Irishman"</Titolo>
    <Regista>"Martin Scorzese"</Regista>
    <Attore>"Robert DeNiro, Al Pacino"</Attore>
  </Film>
  <Film cod classificazione="H456" cod catalogazione="T" formato="HD" originale netflix="no"
    <Titolo>"Her"</Titolo>
    <Regista>"Spike Jonze"</Regista>
    <Genere>"drammatico"</Genere>
  </Film>
</Libreria Netflix>
```

XML

DOCUMENT TREE



Un documento XML può essere scritto, oltre che nella sua struttura fisica, anche in una di tipo logico, comunemente chiamata «document tree».

GLI STUDENTI

ARIANNA DE VITO
DANILO FUSCO