**XML**

**Che cos’è?**

* E**x**tensible **M**arkup **L**anguage
* È un linguaggio che consente la **rappresentazione** **di dati su supporto digitale**
* È **potente** e **versatile** nella creazione, memorizzazione e distribuzione di documenti digitali
* Ha una **sintassi** **rigorosa** e **flessibile**

**SGML (Standard Generalized Markup Language):**

È il **padre** sia di **HTML** che di **XML.**

**Vantaggi:**

* Potente e flessibile (standard ISO, fortemente strutturato, espandibile, non proprietario, indipendente dalla piattaforma)

**Svantaggi:**

* Struttura pesante
* Sono obbligatori DTD e uno stylesheet
* Obbligatoria la validazione del documento
* Le istanze SGML sono troppo pesanti e non robuste per applicazioni web

**Caratteristiche di XML:**

* Indipendente dal tipo di piattaforma hardware e software utilizzato
* Permette la rappresentazione di qualsiasi tipo di documento e di struttura indipendentemente dalle finalità applicative
* È uno standard di pubblico dominio
* Ci sono numerose applicazioni e librerie open source per la manipolazione di dati in formato XML basate su diversi linguaggi di programmazione
* È indipendente dai dispositivi di archiviazione e visualizzazione
* una applicazione in grado di elaborare dati in formato XML viene definita **elaboratore XML**
* Ci sono regole sintattiche e semantiche (l’ultimo non è obbligatorio):
  + **Regole** **sintattiche**: **Come** dobbiamo scrivere le informazioni all’interno del documento
  + **Regole** **semantiche**: **Cosa** possiamo scrivere in un documento XML
* Un documento XML:
  + Può essere archiviato su qualsiasi tipo di supporto digitale
  + Può essere visualizzato su qualsiasi dispositivo di output
  + Facilmente trasmesso via Internet tramite i protocolli HTTP, FTP, STP

**Quando si dice che un documento XML è valido?**

**Validazione DTD:**

Rispetta le regole della sintassi XML e soddisfa i vincoli specificati della **DTD** **(Document Type Definition).** Le dichiarazioni di una DTD definiscono**:**

* Gli elementi strutturali di un documento mediante un identificatore generico
* Il modello di contenuto di ogni elemento, ovvero gli elementi che contiene ed i loro rapporti (un elemento può essere vuoto)
* La lista degli attributi associati a ciascun elemento e il loro tipo

**Problemi delle DTD:**

* Difficili da comprendere
* Non permettono di definire i tipi di dati
* Non consentono di specificare il numero minimo o massimo di occorrenze di un tag in un documento

**Validazione XML schema (XSD):**

Comprende:

* Namespace di riferimento
* Dichiarazione di elementi e attributi

**Quando si dice che un documento XML è ben formato?**

* Deve contenere una dichiarazione corretta
* Il corpo deve avere un unico elemento radice (esclusi eventuali commenti).
* Ogni elemento deve avere un tag di apertura e uno di chiusura.
* Se l’elemento è vuoto si può utilizzare la forma abbreviata (<nometag/>)
* Gli elementi devono essere opportunamente identificati, cioè i tag di chiusura devono seguire l’ordine inverso dei rispettivi tag di apertura
* I nomi dei tag di apertura e chiusura devono coincidere, anche in termini di maiuscole e minuscole
* I valori degli attributi devono sempre essere racchiusi tra singoli o doppi apici

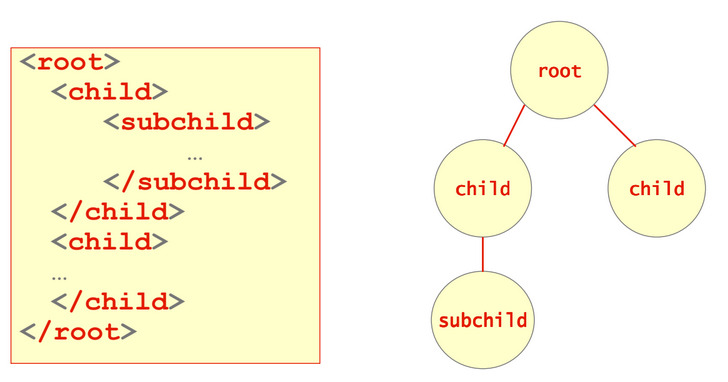
**Struttura LOGICA di un documento XML:**

Un documento XML:

* È strutturato in modo gerarchico
* È composto da elementi, ciascun elemento:
  + Rappresenta un componente logico del documento
  + Può avere dei sottoelementi
  + Può avere degli attributi
* Gli elementi sono organizzati in un albero con una radice singola (root)
* Ogni documento XML può essere rappresentato come un albero (document tree)

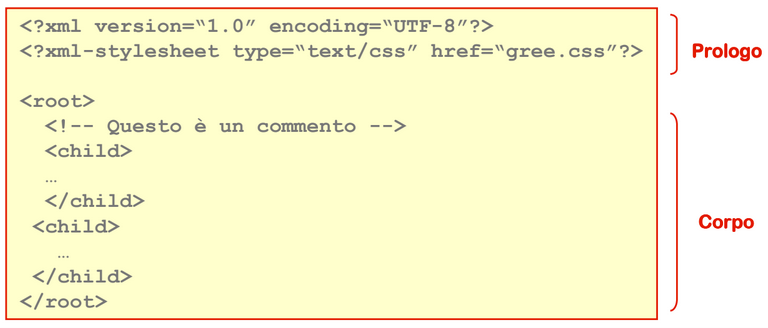
**Struttura FISICA di un documento XML:**

* Un documento XML è un semplice file di testo un semplice file di testo
* La struttura del documento viene rappresentata mediante marcatori
* Gli elementi sono rappresentati mediante tag: coppie di marcatori che racchiudono il contenuto dell’elemento
* I sottoelementi sono tag contenuti all’interno di un altro tag
* Gli attributi vengono rappresentati sotto forma di coppie nome-valore all’interno dei tag
* La radice è un tag che racchiude tutto il resto del documento
* Un documento può inoltre contenere spazi bianchi, a capo e commenti



**Struttura formale di un documento XML:**

1. **Prologo**: contiene una dichiarazione XML ed il riferimento (opzionale) ad altri documenti che ne definiscono la struttura o direttive di elaborazione
2. **Corpo**: è il documento XML vero e proprio



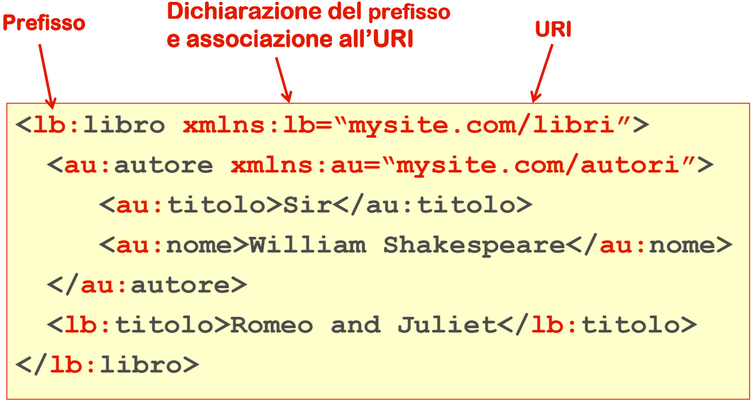
**Elementi e tag:**

* **Elementi**: sono frammenti di testo racchiusi fra star tag e end tag
* **Attributi**: specifiche caratteristiche degli elementi

**Namespace:**

È il concetto con il quale si risolve il conflitto dei nomi:

* Si usano prefissi che identificano il vocabolario di appartenenza di elementi ed attributi
* Ogni prefisso è associato ad un URI (Uniform Resource Identifier) ed è un alias per l’URI stesso
* L’URI in questione è normalmente un URL: si ha quindi certezza di univocità



Sintassi di namespace: **xmlns:NamespacePrefix=”NamespaceURI”**