



XML per l'editoria elettronica

Sistemi per il recupero delle informazioni

Laurea Magistrale in Editoria e Giornalismo

Gabriele Pozzani

materiale prodotto dalla Dott.ssa Barbara Oliboni



Sommario

- Documenti orientati alla narrazione.
 - XML come formato documentale.
 - Struttura dei documenti XML orientati alla narrazione.
 - TEI (Text Encoding Initiative).
 - DocBook.
 - Trasformazione e presentazione.
- Documenti orientati ai dati.
 - XML come formato per i dati.
 - Applicazioni XML per la programmazione.



eXtensible Markup Language

- XML è un linguaggio di marcatura proposto dal W3C
- XML definisce una sintassi generica per contrassegnare i dati di un documento elettronico con marcatori (tag) semplici e leggibili
- La sintassi XML viene utilizzata in contesti molto diversi:
 - pagine web
 - scambio di dati elettronici
 - grafica vettoriale
 - cataloghi di prodotti
 - sistemi di gestione di messaggi vocali
 - ...
- <http://www.w3c.org/XML>



XML come formato documentale

- Utilizzare XML per:

- Pagine Web
- Libri
- Articoli
- Poemi
- Storie brevi
- Manuali di riferimento
- Corsi
- Libri di testo
- Citazioni legali
- Contratti
- Manuali di istruzioni



Documenti
leggibili
dall'uomo



XML per...

- XML soddisfa le esigenze di:
 - Programmatori (o utenti esperti) addestrati a lavorare con le strutture rigide tipiche delle applicazioni orientate ai dati.
 - Scrittori (o utenti normali) che preferiscono la forma libera di un libro o di un articolo.
- XML soddisfa le esigenze di entrambe le comunità in maniera equa e soddisfacente.



Documenti XML: struttura

- Alberi

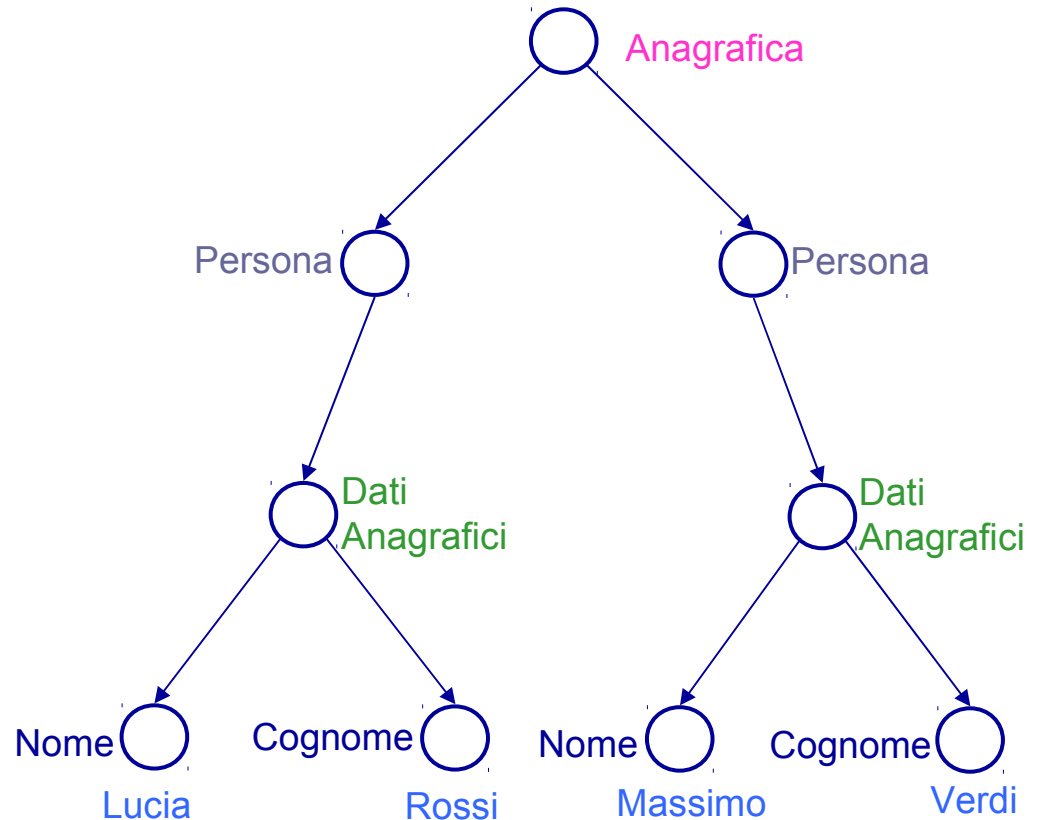
- Senza particolari restrizioni su:
 - modo in cui i nodi sono ordinati
 - come i nodi sono connessi fra loro
 - verso quali nodi può connettersi ogni nodo

- Grafi

- Orientati
 - Etichettati su nodi
 - Etichettati su nodi e archi

Albero XML: esempio 1

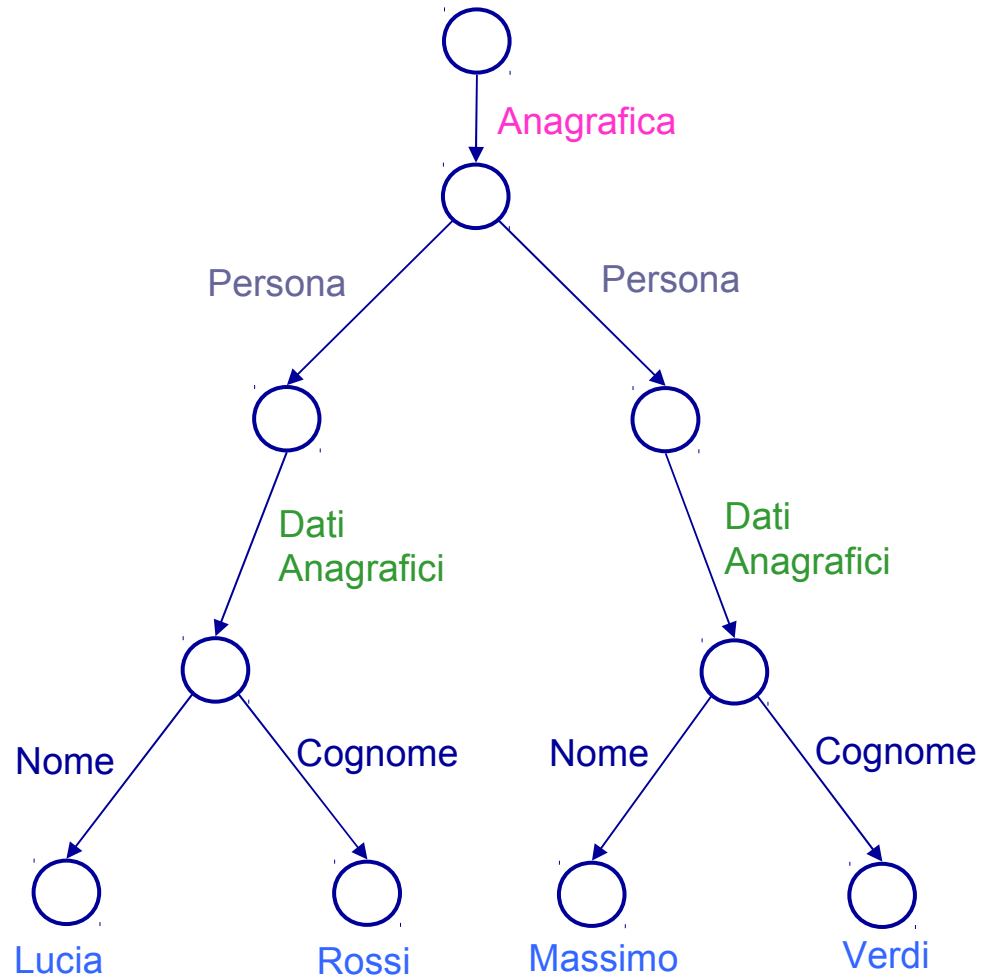
```
<Anagrafica>
  <Persona>
    <DatiAnagrafici>
      <Nome> Lucia </Nome>
      <Cognome> Rossi </Cognome>
    </DatiAnagrafici>
  </Persona>
  <Persona>
    <DatiAnagrafici>
      <Nome> Massimo </Nome>
      <Cognome> Verdi </Cognome>
    </DatiAnagrafici>
  </Persona>
</Anagrafica>
```



Albero con etichette sui nodi

Albero XML: esempio 2

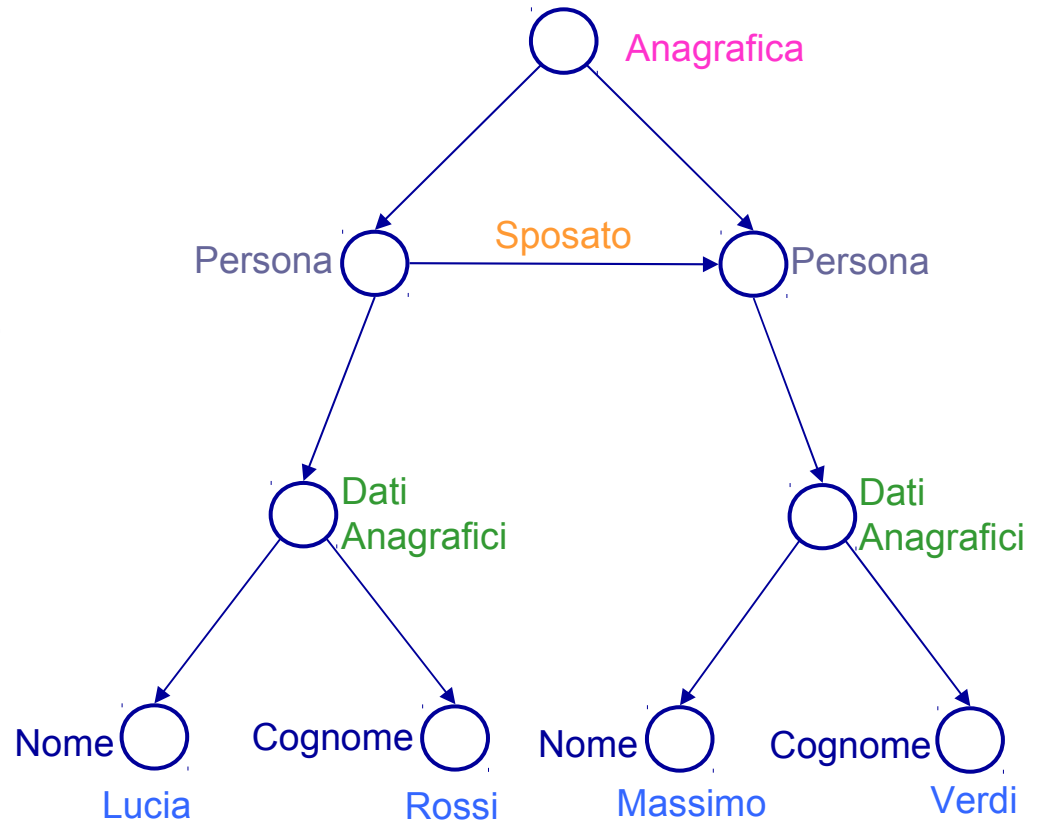
```
<Anagrafica>
  <Persona>
    <DatiAnagrafici>
      <Nome> Lucia </Nome>
      <Cognome> Rossi </Cognome>
    </DatiAnagrafici>
  </Persona>
  <Persona>
    <DatiAnagrafici>
      <Nome> Massimo </Nome>
      <Cognome> Verdi </Cognome>
    </DatiAnagrafici>
  </Persona>
</Anagrafica>
```



Albero con etichette sugli archi

Albero XML: esempio 3



```
<Anagrafica>
  <Persona id="obj1">
    <Sposato idref="obj2">
      <DatiAnagrafici>
        <Nome> Lucia </Nome>
        <Cognome> Rossi </Cognome>
      </DatiAnagrafici>
    </Persona>
  <Persona id="obj2">
    <DatiAnagrafici>
      <Nome> Massimo </Nome>
      <Cognome> Verdi </Cognome>
    </DatiAnagrafici>
  </Persona>
</Anagrafica>
```



Grafo con etichette su nodi e archi



Documenti XML orientati alla narrazione: struttura (1)

- La struttura ad albero risulta poco identificabile.
- Il documento ha una radice che potrebbe essere assimilata al documento stesso.
 - Se il documento fosse un **libro**  Radice = **book**
 - Se il documento fosse un **articolo**  Radice = **articolo**



Documenti XML orientati alla narrazione: struttura (2)

- Generalmente l'elemento radice contiene anche gli elementi che forniscono meta-informazioni.
 - Titolo opera.
 - Nome autore.
 - Data di stesura del documento.
 - Data ultima modifica.
 - ...

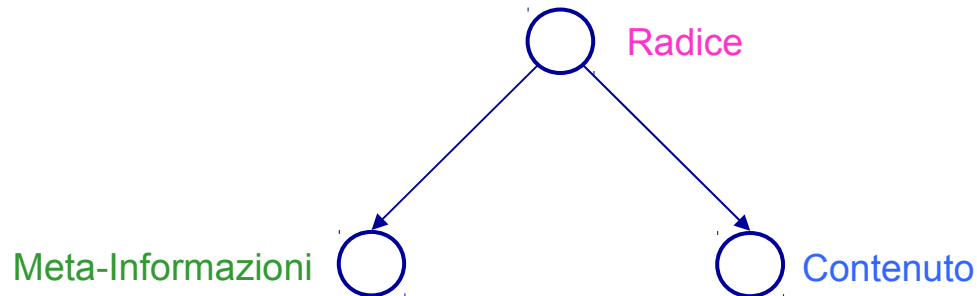


Documenti XML orientati alla narrazione: struttura (3)

- I documenti di grosse dimensioni sono solitamente suddivisi in sezioni.
 - Capitoli di un libro.
 - Sezioni di un articolo.
 - Citazioni di un documento legale.
- Tipi differenti di sezioni.
 - Una per l'indice capitoli.
 - Una per l'indice termini.
 - Una per ognuno dei capitoli di un libro.

Documenti XML orientati alla narrazione: struttura (4)

- **Meta-informazioni** in un elemento figlio della **radice**.
- **Contenuto** del documento in un altro elemento figlio della **radice**.
- Stessa struttura di HTML.





Documenti XML orientati alla narrazione: struttura (5)

- Le sezioni di un documento possono essere suddivise in sottosezioni.
- Le sottosezioni possono essere ulteriormente suddivise.
- Ogni sezione o sottosezione generalmente ha:
 - Un titolo.
 - Elementi e attributi che forniscono meta-informazioni sulla (sotto)sezione stessa.



Documenti XML orientati alla narrazione: struttura (6)

- I documenti orientati alla narrazione non sono composti da una sequenza di caratteri qualsiasi, ma da **parole**.
- Contengono un testo concepito per essere letto da persone.
- Sono caratterizzati da un flusso narrativo.



Documenti XML orientati alla narrazione: struttura (7)

- Il **testo vero e proprio** potrebbe essere suddiviso in:
 - Paragrafi.
 - Altri elementi di tipo blocco:
 - Titoli.
 - Figure.
 - Note laterali.
 - Note a piè di pagina.
- I DTD di documenti generici (DocBook) non sono in grado di specificare precisamente cosa conterranno tali elementi.



Documenti XML orientati alla narrazione: struttura (8)

- I paragrafi e gli altri elementi di tipo blocco contengono principalmente una sequenza di parole, ovvero del **testo**.
- Alcuni di essi possono contenere **marcatori** all'interno del testo.
 - Marcatore per indicare che una certa stringa rappresenta una data o un indirizzo.
- La maggior parte del testo non viene caratterizzata da marcatori.



Documenti XML orientati alla narrazione: struttura (9)

- Possibili modelli di contenuto.
 - **Contenuto di tipo carattere.**
 - **Contenuto costituito da altri elementi (figli).**

```
<nome>Mario Rossi</nome>  
<indirizzo>  
  <via> Via Mazzini </via>  
  <civico> 10 </civico>  
  <città> Verona </città>  
</indirizzo>
```

- **Contenuto misto.**

```
<dati_anagrafici>Il sottoscritto signor  
  <persona>  
    <nome>Mario Rossi</nome> residente in <indirizzo>  
      <via> Via Mazzini </via>  
      <civico> 10 </civico>  
      <città> Verona </città>  
    </indirizzo>  
  </persona>  
</dati_anagrafici>
```

Usato per
il testo vero
e proprio.





TEI (Text Encoding Initiative)

<http://www.tei-c.org/>

- Applicazione per la marcatura della letteratura classica.
 - Primo esempio di DTD orientato a documenti di tipo narrativo.
 - Include elementi per:
 - Le strutture letterarie più comuni (capitoli, scena, stanza, ...).
 - La tipografia.
 - Le strutture grammaticali.
- } Utili per l'analisi
scolare dei testi.



TEI example: Plato, Parmenides (I)

From <http://www.perseus.tufts.edu/>

...

```
<revisionDesc>
```

```
  <change>
```

```
    <date>July, 1992</date>
```

```
    <respStmt>
```

```
      <name>WPM</name>
```

```
      <resp>(n/a)</resp>
```

```
    </respStmt>
```

```
    <item>
```

```
      Tagged in conformance with Prose.e dtd.
```

```
    </item>
```

```
  </change>
```

```
</revisionDesc>
```

...



TEI example: Plato, Parmenides (II)

...

```
<text><group>
<text n="Parm.">
<body>
  <head>Parmenides</head>
  <castList>
    <castItem type="role">
      <role>Cephalus</role>
    </castItem>
    <castItem type="role">
      <role>Antiphon</role>
    </castItem>
    <castItem type="role">
      <role>Aristoteles</role>
    </castItem>
  </castList>
```

...



DocBook

<http://www.oasis-open.org/docbook/>

- Applicazione per documenti nuovi (non vecchi).
- Formato per la creazione del testo (non di un prodotto finito e pronto per essere presentato al pubblico).
- Utilizzato nella documentazione relativa al campo informatico.
- Sintassi molto semplice.
- Modulare e quindi utilizzabile solo in parte (porzioni e strutture che servono).



DocBook example (I)

```
<!DOCTYPE article PUBLIC "-//OASIS//DTD DocBook
V4.1//EN">
<article>
  <articleinfo>
    <title>An Example Article</title>
    <author>
      <firstname>Your first name</firstname>
      <surname>Your surname</surname>
      <affiliation>
        <address>
          <email>foo@example.com</email>
        </address>
      </affiliation>
    </author>
    <copyright>
      <year>2000</year>
      <holder>Copyright string here</holder>
    </copyright>
```



DocBook example (II)

```
<abstract>
  <para>If your article has an abstract
    then it should go here.</para>
</abstract>
</articleinfo>
<sect1>
  <title>My First Section</title>
  <para>This is the first section
    in my article.</para>
  <sect2>
    <title>My First Sub-Section</title>
    <para>This is the first sub-section
      in my article.</para>
  </sect2>
</sect1>
</article>
```

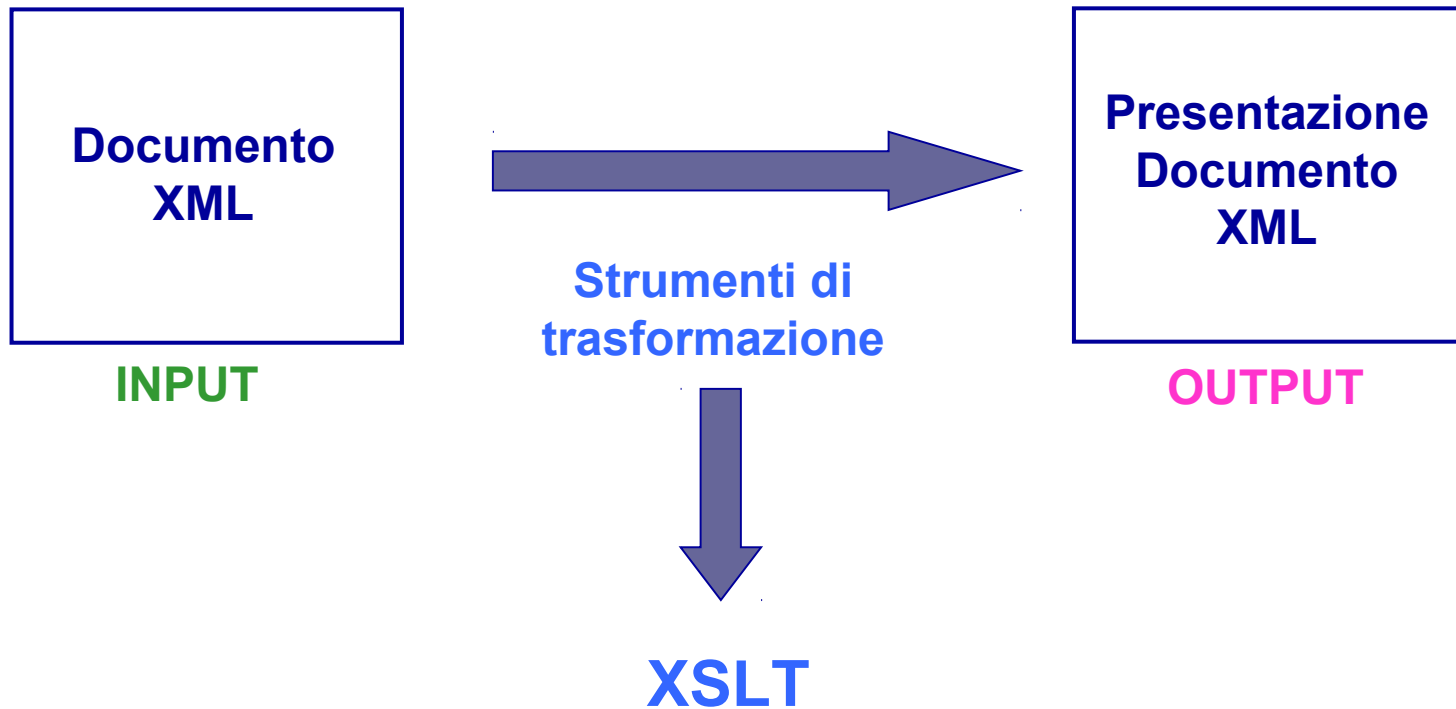



Trasformazione e presentazione (1)

- La marcatura di un documento XML descrive la struttura del documento e non la sua presentazione.
 - Specifica l'organizzazione.
 - Non specifica come deve apparire.
- Un documento XML può essere letto nel suo formato nativo (marcatori + testo).
- Normalmente viene tradotto in un formato differente adatto alla presentazione.
- Per XML il formato di **INPUT** non deve necessariamente corrispondere al formato di **OUTPUT**.
 - Il formato di **INPUT** serve per agevolare chi scrive.
 - Il formato di **OUTPUT** serve per agevolare chi legge.

- ● ●

Trasformazione e presentazione (2)





XSLT (eXtensible Stylesheet Language Transformation)

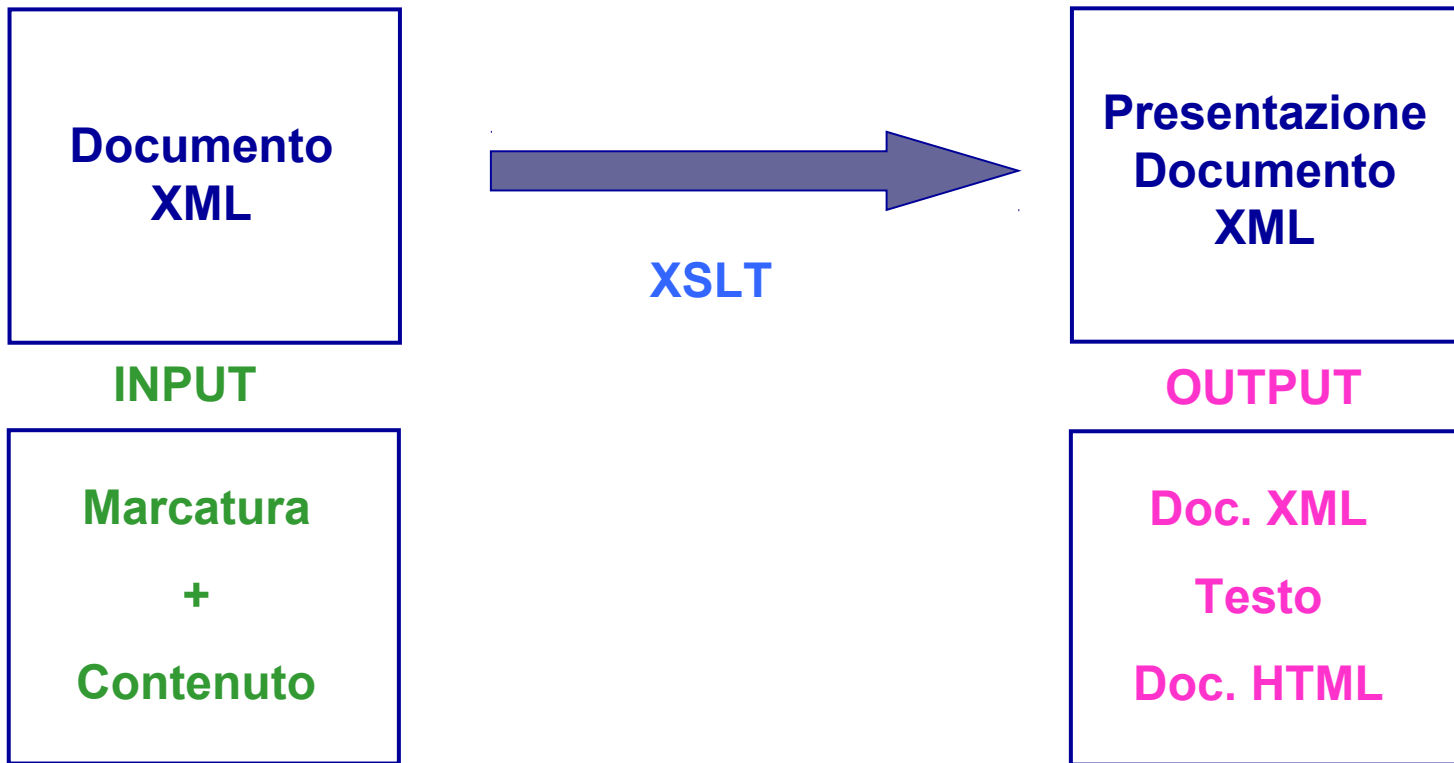
- Un documento XSLT contiene un elenco di modelli.
- Ogni modello contiene un pattern che specifica quali elementi e quali nodi vengono messi in corrispondenza.
- XSLT definisce dei fogli di stile.



XSLT: funzionamento (1)

- Legge documento in INPUT.
- Quando incontra qualcosa (del documento in INPUT) che corrisponde a un modello del foglio di stile.
- Produce in OUTPUT il modello con il relativo contenuto.
- Il modello rappresenta il modo in cui il contenuto deve venir presentato.

XSLT: funzionamento (2)





XML come formato per i dati

- XML si sta imponendo come standard per lo trasmissione e la memorizzazione di informazioni su Internet.



Applicazioni XML per la programmazione

- Comunicazione tra sistemi differenti.
- Memorizzazione dello stato di oggetti.
- Formato per la memorizzazione di informazioni e per il recupero dei dati.
- Descrizione contenuto documenti.