

Codifica di Testi - Introduzione XML Markup a.a. 2024-2025

Angelo Mario Del Grosso

`angelo.delgrosso@ilc.cnr.it`

CNR-ILC

Istituto di Linguistica Computazionale “A. Zampolli”,
18th March 2025

Contenuto della lezione

- 1 I linguaggi di codifica
- 2 Fondamenti del linguaggio XML
- 3 Validare un documento XML e Definire uno Schema
 - Document Type Definition (DTD)
 - XML Schema Definition (XSD)
 - RELAX NG
- 4 Conclusioni

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

I linguaggi di codifica

introduzione

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Definizione di codifica digitale del testo

Per **codifica** digitale dei testi intendiamo la *rappresentazione formale* di un **testo** ad un qualche livello descrittivo, su di un supporto digitale, in un formato utilizzabile da un elaboratore (*Machine Readable Form*) mediante un opportuno **linguaggio informatico** (F. Ciotti).

I linguaggi di codifica

Riassumendo

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Impostazione teorico-pratica

- un testo è molto di **più della sequenza di caratteri** che lo compongono
- per mezzo della codifica vogliamo **rendere esplicite le caratteristiche** che vogliamo analizzare
- solo quello che è esplicito può essere **interpretato ed elaborato dal computer**
- vogliamo codificare il **testo per quello che è**, non per quello che sembra
- codifica da effettuare mediante **linguaggio di markup**

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Il markup

Il termine **markup** è stato utilizzato in passato per denotare i **segni grafici** che accompagnavano un testo apposti sul documento per **indicare correzioni o modalità grafiche di stampa**.

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

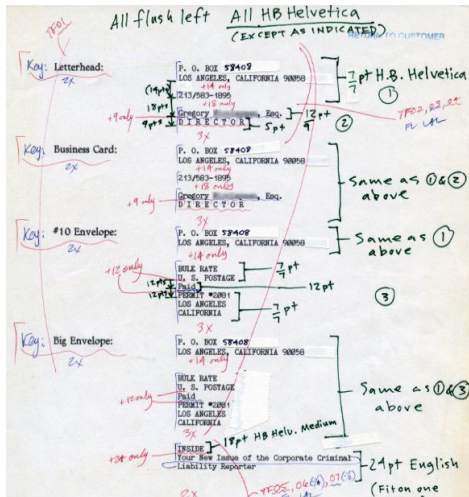
Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni



I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML





































































Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

		Delete			Flush Left		Set in Bold Face Italic
		Insert			Flush Right		Set in Light Face
		Join			Center Horizontally		Wrong Font
		Move closer			Center Vertically		Hyphen
		Space			Move to the next line		En Dash
		Add Space			Move to the preceding line		Em Dash
		Delete Space			Indent 1 em		Superscript
		Transpose Word			Indent 2 ems		Subscript
		Transpose Letters			Paragraph		Comma
		To separate two or more marks			All Caps		Apostrophe
		Let it Stand (ignore correction)			Small Caps		Period
		Move Left			Caps & Small Caps		Semicolon
		Move Right			Capital Letter		Colon
		Move Up			Lower Case		Quotation Marks
		Move Down			Set in Roman		Parentheses
		Align Vertically			Set in Italic		Brackets
		Align Horizontally			Set in Bold Face		

I linguaggi di codifica

Linguaggi di marcatura

Il markup

La codifica con linguaggi di marcatura (markup) è in sostanza **un insieme di convenzioni**, rese attraverso specifiche **sequenze di caratteri, etichette, codici**, (detti *tags*) **intercalati nel testo** per permettere agli elaboratori elettronici di distinguere le varie parti di un documento.

Il markup formale

Un linguaggio di markup è un **sistema formale** per *scambiare* e *pubblicare* informazioni in **formato testo in modo strutturato**.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Il markup formale

Markup formale: costituito da un **sistema non ambiguo** di istruzioni, ognuna delle quali è **dotata di una specifica semantica e sintassi**.

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Diversi tipi di markup

Esistono diversi linguaggi di markup, per rappresentare diversi tipi di documenti.

- **Linguaggi procedurali** (specific markup languages)
- **Linguaggi dichiarativi** (generic markup languages)

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Linguaggi procedurali

- **Orientati al documento**, indicando come deve essere elaborato e *disposto il testo* in **resa grafica**
- Istruzioni da inserire nel testo per connotarne specifiche *caratteristiche di visualizzazione*
- Font, dimensione, spaziatura del carattere, posizionamento nella pagina, colore, etc.

Esempi: TeX e LaTeX, RTF

I linguaggi di codifica

Linguaggi di marcatura procedurale

Esempio RTF

```
{\rtf1\ansi\deff0\adefflang1025
{\fonttbl{\f0\froman\fprq2\fcharset0 Times New Roman;}
{\f1\froman\fprq2\fcharset0 Times New Roman;}
{\f2\fnil\fprq2\fcharset0 Lucida Sans Unicode;}
{\colortbl;\red0\green0\blue0;\red128\green128\blue128;}
{\stylesheet{\s1\cf0{\*\hyphen2\hyphlead2\hyphtrail2\hyphmax0}
\rtlch\af5\afs24\lang255\ltrch\dbch\af2\afs24\langfe255
\loch\f0\fs24\lang1040\next1 Standard;}}
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Esempio LaTeX

```
\documentclass [a4paper,10pt] {article} \usepackage[utf8] {inputenc}
\usepackage[T1] {fontenc}
\usepackage[italian] {babel}
\title {Il mio primo documento}
\author {Angelo Mario Del Grosso}
\begin {document}
\maketitle
\begin {abstract}
Primo tentativo di scrivere in \LaTeX .
\end {abstract}
\section {titolo della sezione}
Questo documento è vuoto.
\footnote {nota a piè di pagina.}

\end {document}
```

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Il markup procedurale

L'unico utilizzo di un testo codificato tramite un linguaggio procedurale è la **creazione di un output orientato alla visualizzazione.**

I linguaggi di codifica

Linguaggi di marcatura procedurale

Scrivere la tesi di laurea con $\text{\LaTeX} 2_{\epsilon}$

Dipartimento di Ingegneria Meccanica,

Nucleare e della Produzione

Università di Pisa

56126 Pisa PI

Sommario

Lo scopo del presente articolo è fornire gli strumenti per scrivere una tesi di laurea utilizzando $\text{\LaTeX} 2_{\epsilon}$. Tale obiettivo è conseguito analizzando i problemi tipici incontrati durante la stesura della tesi e le possibili soluzioni; si pone particolare attenzione ai pacchetti da usare nelle varie circostanze. I singoli argomenti non vengono approfonditi nei dettagli ma si rimanda alla letteratura specifica o ad i manuali dei pacchetti suggeriti, ove necessario.

*Ringrazio in primo luogo Fabiano Busdraghi che ha collaborato alla scrittura delle sezioni riguardanti le figure e gli oggetti flottanti. Ringrazio inoltre tutti coloro che mi hanno consigliato durante la stesura e la revisione di questo documento ed in particolare Claudio Beccari, Gustavo Cevolani, Massimo Guiggiani, Maurizio Himmelmann, Lorenzo Pantieri e Emiliano Vavassori.

I linguaggi di codifica

Linguaggi di marcatura dichiarativi

Linguaggi dichiarativi

Orientati al testo, annotano la *struttura*, la *funzione* ed il *significato* degli elementi costitutivi del testo, **tralasciandone l'aspetto**.

- La posizione che il brano in questione occupa all'interno del documento (**markup strutturale**)
- Peculiarità del testo stesso (**markup semantico**)
- I fogli di stile definiscono la formattazione dell'output
- *Molteplici usi del medesimo testo*

Esempio: famiglia SGML, XML

I linguaggi di codifica

Linguaggi di marcatura dichiarativi

Markup dichiarativi: contenuto e presentazione

La **separazione tra contenuto e presentazione** non solo è intenzionale, ma è la **caratteristica principale** di questi sistemi di marcatura: essa permette di concentrarsi sull'**annotazione logica-semantica** per funzioni di *ricerca e di analisi*, lasciando ad altro (ai fogli di stile) la resa grafica.

Unico testo più usi

In questo modo si ha inoltre la possibilità di utilizzare uno **stesso testo codificato con finalità o formattazioni differenti**, a seconda delle varie esigenze.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

I linguaggi di codifica

Markup dichiarativi: esempio SGML

Standard Generalized Markup Language

```
1 <!DOCTYPE testo [  
2 <!ELEMENT testo (titolo?, paragrafo+)>  
3 <!ELEMENT titolo (#PCDATA)>  
4 <!ELEMENT paragrafo (#PCDATA)>  
5 ]>  
6 <testo>  
7   <titolo> Questo è il titolo del documento</titolo>  
8   <paragrafo> Questo è un paragrafo </paragrafo>  
9 </testo>  
10
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

I linguaggi di codifica

Markup dichiarativi vs Markup procedurali

resa a video della frase

Le *Guidelines for Electronic Text Encoding and Interchange* sono *molto* complete e descrivono uno standard di *markup* del testo basato su XML.

Le `\textit{Guidelines for Electronic Text Encoding and Interchange}` sono `\textit{molto}` complete e descrivono uno standard di `\textit{markup}` del testo basato su XML.

`<titolo>`Le Guidelines for Electronic Text Encoding and Interchange`</titolo>` sono `<enfasi>`molto`</enfasi>` complete e descrivono uno standard di `<linguastraniera>` markup`</linguastraniera>` del testo basato su XML.

LaTeX vs SGML

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

I linguaggi di codifica

Linguaggi di marcatura

linguaggi semi-dichiarativi e/o semi-procedurali

Esistono anche linguaggi che possono essere definiti **semi-procedurali**, o **semi-dichiarativi**, che come si intuisce utilizzano le istruzioni sia per una codifica di tipo procedurale, sia per una codifica di tipo descrittivo o dichiarativo.

HTML

HTML ha tra le sue etichette istruzioni di tipo procedurale per indicare come devono essere rese determinate porzioni di testo, e istruzioni di tipo dichiarativo che hanno una base semantica.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

- Document Type Definition (DTD)
- XML Schema Definition (XSD)
- RELAX NG

4 Conclusioni

Elementi per la definizione degli schemi xml

principi

Validare il contenuto di un documento XML

Se vogliamo condividere efficacemente informazioni bisogna avere dei meccanismi per controllare che i dati trasmessi rispettino una ben precisa struttura e abbiano un ben preciso e coerente modello dei contenuti (*rispettino una grammatica*).

Validare il contenuto di un documento XML

Per essere sicuri che un documento XML sia corretto da un punto di vista della struttura e del contenuto, cioè sia **valido**, bisogna riferirsi ad uno *schema*.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

principi

Schemi XML

Per condividere quindi efficacemente un vocabolario XML bisogna definire delle regole che controllino come utilizzare correttamente gli elementi e gli attributi del vocabolario.

Schemi XML

Gli strumenti per descrivere le regole relative ad una corretta compilazione di un documento XML sono principalmente: *Document Type Definition* (DTD) oppure *XML Schema Definition* (XSD).

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

principi

XML schema come contratto

Gli schemi XML possono essere visti come un contratto (formale) condiviso tra chi codifica i dati e chi deve consumarli. In questo modo può avvenire in modo rigoroso la comunicazione per lo scambio delle informazioni codificate attraverso il formato XML.

XML schema come contratto

Validare un documento XML vuol dire verificare che il documento sia aderente al formato definito nel contratto (schema).

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

principi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Documento XML valido

Un documento XML si dice **valido** se esso è ben formato (well formed) e se soddisfa anche le regole specificate all'interno di uno schema XML associato.

Elementi per la definizione degli schemi xml

principi

schemi XML come contratti

- definiscono la struttura di un documento XML
- definiscono le regole per validare il contenuto degli elementi e degli attributi
- permette ad un programma (*validator*, *checker*) di verificare la validità di un documento rispetto allo schema prescelto.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Tipi di formalismi per definire schemi XML

Molti tipi di schemi XML

Nel corso degli anni sono stati proposti molti formalismi per codificare gli schemi XML

- DTD, XSD, RELAX NG
- Schematron
- XDR, SOX, DSD, DCD, DDML

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Document Type Definition (DTD)

Una document type definition (DTD) descrive le regole relative alla struttura e al contenuto di un documento XML.

Document Type Definition (DTD)

Una DTD dichiara gli elementi, gli attributi, le entità e le notazioni ammesse in un documento XML.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

well-formed document \neq valid document

Se un documento XML manca di riferirsi ad una DTD oppure non rispetta le regole di una DTD, esso può essere tutt'al più ben formato, ma sicuramente non può essere valido.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Document Type Definition (DTD)

La validazione dei documenti XML è alla base della condivisione e scambio dati in quanto è possibile confidare sulla natura dei dati trasmessi.

Attenzione: non possiamo validare la correttezza della semantica dei dati!

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Document Type Definition (DTD)

L'associazione tra documento XML e DTD viene realizzata tramite una dichiarazione inclusa nel prologo all'inizio del documento.

Root element and content

La DTD dichiara l'elemento radice del vocabolario e il suo *content model* (children elements).

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Element Declaration (con figli)

```
<!ELEMENT element-name (child-element1,  
                           child-element-2 ...)>
```

Element Declaration (solo testo)

```
<!ELEMENT element-name (#PCDATA)>
```

Il Parsed Character Content designa contenuto testuale piano senza figli

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Child Element Declaration

La dichiarazione di un elemento figlio, è analoga in tutto e per tutto alla dichiarazione dell'elemento radice. Cioè utilizzando l'etichetta `<!ELEMENT >`

Element Declaration (root)

La dichiarazione dell'elemento radice deve sempre essere la prima

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Modificatori

Nella dichiarazione di un elemento possono essere inclusi opzionalmente dei modificatori, per stabilire il numero di occorrenze degli elementi figli.

Modificatori

- + Una o più occorrenze
- ? Zero o una occorrenza
- * Zero o più occorrenze

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Modificatori

```
<!ELEMENT element-name (B, C)+ >  
<!ELEMENT element-name (B+, C) >  
<!ELEMENT element-name (B, C+) >  
<!ELEMENT element-name (B+, C+) >
```

Se un elemento figlio deve presentarsi solo una volta, allora non c'è bisogno di modificatori.

Attenzione l'ordine dei figli nella dichiarazione è significativa

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Esercizio

Definire i seguenti elementi:

- elemento root: **TEI**
- elementi figli:
 - header (obbligatorio una occorrenza)
 - facsimile (opzionale una occorrenza)
 - text (obbligatorio almeno una occorrenza)

Gli elementi header, facsimile e text hanno tutti un content model testuale

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

choice declaration

```
<!ELEMENT element-name (child-a | child-b) >
```

Dichiarazione di Choice

La sintassi della DTD consente di dichiarare una scelta (*choice*) tra due o più elementi come content model di un elemento.

Il choice indica una scelta tra una lista di possibilità

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Attributi degli elementi XML

Un attributo è dichiarato sfruttando l'elemento `<!ATTLIST >`

Cos'è un attributo di un elemento XML

Un attributo è una proprietà, una caratteristica di un elemento e descrive il contenuto dell'elemento stesso.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Attributi: sintassi

```
<!ATTLIST Element-name Attr-name Attr-type  
Attr-state? default-value?>
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Attributi: sintassi

- “Element-name” è il nome dell’elemento a cui l’attributo si riferisce
- “Attr-name” è il nome dell’attributo dichiarato
- “Attr-type” è il tipo di dato atteso dell’attributo
- “Attr-state” indica uno tra i tre stati possibili di un attributo
- “default-value” indica il valore di default per quell’attributo, se non fornito.

Elementi per la definizione degli schemi xml

TABELLA dei tipi

Attribute Value	Description
CDATA	Any character data except characters reserved by XML
<i>enumerated list</i>	A list of possible attribute values
ID	A unique text string
IDREF	A reference to an ID value
IDREFS	A list of ID values separated by white space
ENTITY	A reference to an external unparsed entity
ENTITIES	A list of entities separated by white space
NMTOKEN	An accepted XML name
NMTOKENS	A list of XML names separated by white space
NOTATION	The name of a notation defined in the DTD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Stato di attributi

Lo stato di un attributo può essere uno tra:

- #IMPLIED (attributo opzionale)
- #REQUIRED (attributo obbligatorio)
- #FIXED (valore fisso dell'attributo)

Il valore di un attributo fisso viene fornito come valore di default

Elementi per la definizione degli schemi xml

Choice attributi

```
<!ATTLIST element attribute (value1 | value2 |  
value3 | ...) type default >  
<!ATTLIST name title (Mr. | Mrs. | Ms.)  
#IMPLIED 'Mr.'>
```

Attenzione non è possibile avere un valore di default se un attributo è marcato #REQUIRED

Elementi per la definizione degli schemi xml

Attributi ID IDREF IDREFS

```
<!ATTLIST element attribute ID >  
<!ATTLIST order orderID ID #REQUIRED>  
<!ATTLIST customer orders IDREFS>  
<!ATTLIST package orderRef IDREF>
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Mixed content - DTD

```
<!ELEMENT element-name (#PCDATA|child-element)* >
```

Mixed content XML

```
<p>Ieri pomeriggio sono andato a  
<placeName>Pisa</placeName>, per un giro</p>
```

Elementi per la definizione degli schemi xml

Esercizio

root: TEI

Figli:

- header(obbligatorio una volta sola)
- facsimile(opzionale una volta sola)
- testo(obbligatorio una o più volte)
- * testo è un mixed content con possibile elemento <seg>

Attributi:

- header: type:(fixed, CDATA "intestazione"); lang(opzionale, NMTOKEN)
- facsimile: source:(obbligatorio); ref(opzionale, IDREFS)
- testo: id(obbligatorio, ID) type(opzionale contenuto testuale)

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Empty elements

Un elemento XML può essere vuoto (empty). La dichiarazione di un elemento vuoto si realizza con la parola chiave **EMPTY**.

Empty content

```
<!ELEMENT element-name EMPTY>
```

Empty content

```
<!ELEMENT lb EMPTY>  
<lb />
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Any elements

E' possibile dichiarare anche elementi che hanno qualsiasi tipo di content model.

A tal proposito viene impiegata la parola chiave "ANY".

Any content

```
<!ELEMENT element-name ANY >
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Dichiarare il tipo del documento XML

La dichiarazione della DTD viene inserita attraverso una URL nel prologo del documento XML, tra la dichiarazione del documento XML e l'elemento radice.

Dichiarare il tipo del documento XML

Grazie al sistema di dichiarazione della DTD è possibile massimizzare il riuso e collegare lo schema a tutti i documenti che si vuole validare.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

DOCTYPE

```
<!DOCTYPE root-element SYSTEM ‘‘External DTD’s  
URL’’ [Internal DTD ]>  
    <!DOCTYPE root-element [Internal DTD] >  
<!DOCTYPE root-element SYSTEM ‘‘Ext-DTD URL’’ >
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

DOCTYPE PUBLIC

```
<!DOCTYPE root PUBLIC \id" \uri">  
standard//owner//description//language  
-//W3C//DTD XHTML 1.0 Strict//EN
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Esercizi

- includere all'interno di un documento XML la dichiarazione del tipo, definire internamente gli elementi e gli attributi e validare.
- inserire nel prologo di un documento XML la dichiarazione del tipo di documento e validare.

Creare un file esterno con estensione .dtd prima di includerlo nel prologo XML.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Entity

Per includere dati da diverse fonti, DTD prevede l'uso di entità. Due tipologie di entità sono state definite: general entities e parameter entities.

Entity: generiche e parametriche

- le general entities vengono espase nel documento XML
- le parameter entities vengono espase nel documento DTD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Entity

- Le general entities si possono classificare in interne ed esterne, a loro volta possono essere parsed oppure unparsed.
- Le parameter entities si possono classificare in interne ed esterne; che possono essere solo parsed.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Entity

Le entità generiche interne aiutano ad includere nel documento XML quei caratteri speciali che altrimenti causerebbero errori al passaggio del parser.

Internal General Entity: Sintassi

```
<!ENTITY entity-name ‘‘replacement-string’’ or  
          ‘‘hexadecimal-code’’ >
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Internal General Entity: Sintassi

Per usare le entità all'interno del documento XML basta prefissare al nome dell'entità una "e commerciale" (&) e aggiungere alla fine come suffisso un "punto e virgola" (;):
`&entity-name;`

Una entità può contenere un frammento XML ben formato

Internal General Entity: Sintassi

```
<!ENTITY firma ‘‘<i>Angelo Mario Del Grosso</i>’’>
    <p><salutation>&firma;<salutation></p>
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Internal General Entity: Esempio con UNICODE

Spesso le entità vengono utilizzate per dare un nome ai riferimenti a carattere.

Internal General Entity: Sintassi

```
<!ENTITY amaiuscola ‘‘&#65;’’>
<!ENTITY amaiuscola ‘‘&#x0041;’’>
<p><salutation>&amaiuscola;ddio<salutation></p>
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

External General Entity

Un documento XML può essere composto da altri pezzi di XML distribuiti in diversi luoghi.

Grazie alle entità generiche esterne è stata implementata questa caratteristica.

External General Entity: Sintassi

```
<!ENTITY entity-name SYSTEM ‘‘URL’’ >  
<!ENTITY salutation SYSTEM ‘‘salut.xml.ent’’ >
```

L'URL punta al luogo dove risiede la external entity

Elementi per la definizione degli schemi xml

Principi Document Type Definition

External General Entity

Le entità esterne non possono contenere una DTD per ovvi motivi di gestione dei potenziali conflitti.

E' possibile utilizzare altre entità all'interno delle entità esterne.

General Entity

Le entità generiche vengono dichiarate nella DTD, ma possono essere utilizzate solo all'interno di un documento XML e non nella DTD stessa.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Parameter Entity

Le parameter entity sono entità sfruttabili all'interno del documento DTD. Ma non possono essere utilizzate all'interno del documento XML.

Eseistono due tipi di entità parametriche:

1) **internal** parameter entities 2) the **external** parameter entities.

Parameter Entity: Sintassi

```
<!ENTITY % entity-name 'replacement-string'>  
<!ENTITY % parameter-name SYSTEM 'URL' >
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Parameter Entity: Impiego

```
<!ENTITY % biblinfo ‘‘(title,author?,cost?)’’>  
<!ELEMENT biblInfo %biblinfo;>
```

Parameter Entity: Impiego

```
<!ENTITY % biblInfo SYSTEM ‘‘biblInfo.dtd’’ >  
<!ELEMENT listBibl (bib+) >  
    %biblInfo;
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Parameter Entity: utilità

Quando una entità parametrica viene inserita in una DTD, essa viene rimpiazzata dal suo contenuto a tempo di esecuzione.

Parameter Entity: utilità

Ciò permette di facilitare lo sviluppo della DTD e di ottimizzarne la manutenibilità.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Parameter Entity: utilità

Le entità parametriche esterne facilitano la modularità di grandi DTD e permettono un linking dinamico ai vari documenti di definizione.

Parameter Entity: utilità

Grazie a questi tipi di entità è possibile includere pezzi di DTD residenti in posizioni remote e formare un completo e unico documento DTD a tempo di esecuzione.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

DTD: Pros

- sono compatte e facilmente comprensibili
- sono definibili inline all'interno del documento XML
- possono definire entità
- sono utilizzate da quasi tutti i vocabolari esistenti
- sono supportate da quasi tutti i parser esistenti

Elementi per la definizione degli schemi xml

Principi Document Type Definition

DTD: Cons

- non sono scritte con una sintassi XML
- richiedono parser specifici
- non supportano i namespaces
- non hanno un vero meccanismo per i tipi di dati
- la validazione del contenuto di un elemento è molto limitato e limitante
- non ci sono meccanismi per indicare esattamente il numero di figli che può contenere un elemento.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

principi XML Schema Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Cos'è uno schema XML

Uno schema XML è un documento XML standard che descrive come deve essere realizzato un altro documento XML. Ci riferiamo a questa tecnologia con l'acronimo XSD.

A cosa serve uno Schema XML

I documenti XSD sono usati per validare documenti XML. Tuttavia un documento XSD viene realizzato tramite l'uso di un vocabolario predefinito riferibile attraverso un namespace ad un URI standard.

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XSD Schema

Il termine XSD o XML Schema denota un documento XML che descrive e valida la struttura e il contenuto di un altro documento XML.

XSD Schema

Dichiarazione del documento (declaration) e istanza del documento (instance).

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XSD elemento root

L'elemento radice di uno schema XSD è sempre l'elemento
“<schema>”.

Tutte le definizioni devono seguire quindi l'elemento
“<schema>”.

XSD Schema

Tutti gli elementi e gli attributi dello schema sono dichiarati
all'interno del namespace

“http://www.w3.org/2001/XMLSchema.”.

Tutti i documenti XSD contengono la dichiarazione a questo
namespace con prefisso convenzionale **xsd** oppure **xs**.

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XSD componenti di base

I componenti di base di uno Schema XSD sono le dichiarazioni degli elementi e le dichiarazioni degli attributi.

XSD Schema

Le dichiarazioni più complesse si poggiano su queste unità: elementi e attributi.

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XSD dichiarazioni

Scrivere un pezzo di codice XSD per descrivere e validare un elemento per un documento XML è detto *element declaration*.

XSD dichiarazioni di base

XSD permette di dichiarare elementi, attributi e di specificare il numero di figli, le occorrenze, l'ordine di apparizione, e i tipi di dati del *content model*.

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Element Types: simple and complex

La dichiarazione di un elemento può avere un tipo semplice (*simple type*) oppure un tipo complesso (*complex type*) a seconda della sua struttura e del suo contenuto.

Simple Type e Complex Type

La dichiarazione di un elemento ha un tipo semplice se non possiede **né figli né attributi**.

La dichiarazione di un elemento ha un tipo complesso in tutti gli altri casi.

Elementi per la definizione degli schemi xml

principi XSD

XSD esempio

```
<xsd:schema
xmlns:xsd='http://www.w3.org/2001/XMLSchema''>
<xsd:element name='text' />
</xsd:schema>
```

XSD esempio elemento di tipo semplice

```
<text>Il primo documento XML Validato</text>
```

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XML XSD esempio

Il documento XML istanza dello schema XSD per essere valido deve contenere un elemento radice. Validare il documento XML con il relativo XSD con XMLlint.

XMLlint

```
xmlLint xmlfirst.xml --schema  
../schema/xsd/xsdfirst.xsd
```

Elementi per la definizione degli schemi xml

principi XSD

Element Complex Types: esempio

```
<xsd:schema
xmlns:xsd='http://www.w3.org/2001/XMLSchema'>
<xsd:element name='Employee'> <xsd:complexType>
<xsd:attribute name='FirstName' />
</xsd:complexType> </xsd:element> </xsd:schema>
```

Element Complex Types: esempio

Il documento XML istanza dello schema:

```
<Employee FirstName="Jacob"/>
```

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

XML per rappresentare il testo

- I markup language per supportare la rappresentazione, memorizzazione, pubblicazione di un testo.
- XML è un markup language flessibile e potente.
- le istruzioni dei markup language sono per lo più dichiarazioni indicando particolari funzioni del dato.
- le istruzioni sono etichette visibili.

XML per rappresentare il testo

- Una sintassi e una grammatica regolano l'applicabilità del linguaggio di marcatura
- Sintassi: documento well formed (ben formato)
- Grammatica: documento valido

XML per rappresentare il testo

- XML deriva dal linguaggio SGML.
- XML è una specifica del consorzio W3C.
- XML è un meta-linguaggio.
- XML è plain text.
- XML è portabile.

Elementi XML

Conclusioni

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2024-2025

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XML per rappresentare il testo

- XML definisce markup dichiarativi e descrittivi.
- XML ha un modello dati ad albero ordinato.
- XML può avere associato un tipo di documento (DTD) o uno schema di codifica (XSD).