

# Codifica di Testi - Introduzione XML Markup a.a. 2018-2019

Angelo Mario Del Grosso

`angelo.delgrosso@ilc.cnr.it`

*CNR-ILC-LicoLab*

Istituto di Linguistica Computazionale “A. Zampolli”,  
4th October 2018

# Contenuto della lezione

- 1 I linguaggi di codifica
- 2 Fondamenti del linguaggio XML
- 3 Validare XML
- 4 Document Typed Definition - DTD
- 5 XML Schema Definition - XSD
- 6 Conclusioni

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

# Progress status

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare XML

4 Document Typed Definition - DTD

5 XML Schema Definition - XSD

6 Conclusioni

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

# I linguaggi di codifica

## introduzione

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

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

### 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

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

# I linguaggi di codifica

## Linguaggi di marcatura

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

## Il markup

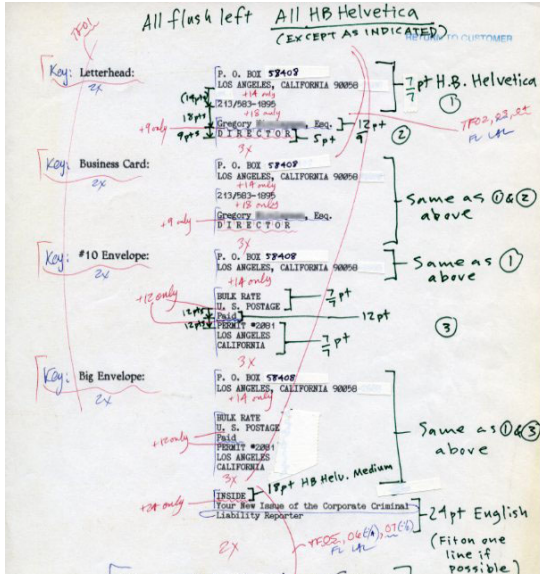
Il termine markup è stato utilizzato in passato per denotare i segni grafici che accompagnavano un testo apposti sul documento per indicare correzioni o modalità grafiche di stampa.

## Linguaggi di marcatura

2018-2019

A.M. Del  
Grosso

## I linguaggi di codifica



# I linguaggi di codifica

## Linguaggi di marcatura

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

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 definitiva 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.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

# I linguaggi di codifica

## Linguaggi di marcatura

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

## Il markup formale

Markup formale: costituito da un sistema ben preciso 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.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

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.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

## Linguaggi procedurali

- Orientati al documento, indicano come deve essere elaborato e disposto il testo
- Istruzioni da inserire nel testo per specificarne specifiche caratteristiche
- Font, dimensione, spaziatura del carattere, posizionamento nella pagina, colore, etc.

*Esempi: TeX e LaTeX, RTF*

# I linguaggi di codifica

## Linguaggi di marcatura procedurale

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

## Esempio RTF

```
{\rtf1\ansi\deff0\adeflang1025
{\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\snext1 Standard;}
```

# I linguaggi di codifica

## Linguaggi di marcatura procedurale

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

## Esempio LaTeX

Ciao

# I linguaggi di codifica

## Linguaggi di marcatura

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

## Il markup procedurale

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

## Scrivere la tesi di laurea con $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X} 2_{\epsilon}$

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  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X} 2_{\epsilon}$ . 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



# 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

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

### 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.

# I linguaggi di codifica

## Markup dichiarativi: esempio SGML

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

```
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
```

# I linguaggi di codifica

## Markup dichiarativi vs Markup procedurali

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

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.

# 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. Uno degli esempi più famosi è HTML che tra le sue etichette mescola istruzioni di tipo procedurale per indicare come devono essere rese determinate porzioni di testo, a istruzioni di tipo dichiarativo che hanno una base semantica e vengono rese in modo differente a seconda del browser utilizzato.

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

# Progress status

- 1 I linguaggi di codifica
- 2 Fondamenti del linguaggio XML
- 3 Validare XML
- 4 Document Typed Definition - DTD
- 5 XML Schema Definition - XSD
- 6 Conclusioni

# Progress status

- 1 I linguaggi di codifica
- 2 Fondamenti del linguaggio XML
- 3 Validare XML
- 4 Document Typed Definition - DTD
- 5 XML Schema Definition - XSD
- 6 Conclusioni

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni

# Progress status

- 1 I linguaggi di codifica
- 2 Fondamenti del linguaggio XML
- 3 Validare XML
- 4 Document Typed Definition - DTD**
- 5 XML Schema Definition - XSD
- 6 Conclusioni



# Progress status

- 1 I linguaggi di codifica
- 2 Fondamenti del linguaggio XML
- 3 Validare XML
- 4 Document Typed Definition - DTD
- 5 XML Schema Definition - XSD**
- 6 Conclusioni

# Progress status

- 1 I linguaggi di codifica
- 2 Fondamenti del linguaggio XML
- 3 Validare XML
- 4 Document Typed Definition - DTD
- 5 XML Schema Definition - XSD
- 6 Conclusioni**

Codifica di  
Testi -  
Introduzione  
XML Markup  
a.a.  
2018-2019

A.M. Del  
Grosso

I linguaggi di  
codifica

Fondamenti  
del linguaggio  
XML

Validare XML

Document  
Typed  
Definition -  
DTD

XML Schema  
Definition -  
XSD

Conclusioni