

Micro-dispensa

11/11/2013 (prima lezione)

Sito del corso

<http://www.unipa.it/paolo.monella/labinformatica>

Modellizzazione e markup testuale

- Meta!
 - Meta-testo
 - Meta-linguaggio
 - Meta-informazione
 - Markup, o marcatura
- Insieme di caratteri (character set - **charset**) / tabella di caratteri
 - [ASCII](#)
 - Unicode
- Stringa di caratteri (incluso il *carriage return*, ASCII 13)
- linguaggio di marcatura (mark-up language), con una sua sintassi (parentesi uncina o graffe etc.)
- alcuni linguaggi di mark-up
 - Standard Generalized Markup Language - SGML
 - SGML → XML (extensible markup language)
 - entrambi sono meta-linguaggi
 - LaTeX (un markup language, o linguaggio di marcatura, orientato alla tipografia):
 - `\text{Mi illumino d'immenso.}`

XML

XML è un un meta-linguaggio: serve per inventare linguaggi di markup

```
<title>Mattino</titolo>
<body>Mi illumino (spesso)
d'immenso.</body>
<author>G. Ungaretti</author>
```

PML (Paolo Markup Language)

```
<title>
<body>
<author>
```

PML2 (cambia la *semantica* dei tag, cioè il loro significato, ma la *sintassi* XML, fatta di parentesi uncinate e barre resta la stessa)

```
<titolo></titolo>
<corpo>
<autore>
```

PML3

```
<poesia>
<corpo>
<verso>
<titolo>
<autore>
<strofa>
```

```
<poesia>
  <titolo>Mattino</titolo>
  <corpo>
    <strofa>
      <verso tipo="libero" num=1>Mi illumino</verso>
      <verso tipo="libero" num=1>d'immenso.</verso>
    </strofa>
  </corpo>
  <autore>G. Ungaretti</autore>
</poesia>
```

```
<verso tipo="libero" num=1>
  tipo: attributo
  libero: valore dell'attributo
```

Sintassi XML:

- Parentesi uncinata per i tag: <autore>
- Barre (slash) per i tag di chiusura: </autore>
- I figli (gli elementi child) devono essere figli del genitore (elemento parent)
- Tutto il documento dev'essere contenuto in un elemento, "padre di tutti"

Alcuni linguaggi importanti della famiglia XML

HTML

HyperText Markup Language (linguaggio di markup per ipertesti)

- Orientato all'apparenza grafica del testo (far apparire il testo su schermo, sul browser)
- Serve per dare al browser una *procedura* (istruzioni) per visualizzare il testo

- Linguaggio *procedurale*

<h1>Titolo</h1>

<p>

Ciao a tutti, questa parola è in

<i>corsivo</i>

</p>

TEI/XML

- TEI: Text Encoding Initiative
 - TEI Consortium, dalla fine degli anni '80
- Linguaggio dichiarativo (dice cosa sono le cose)
 - dice che un titolo è un titolo, che un verso è un verso etc.

<div>

Una divisione del testo

<div type="poem">

Una suddivisione del testo di tipo "poesia" → una poesia

<l>

Line = verso

<l n=1>

Verso il cui numero di verso è 1

<lg>

Line group = strofa

<author>

Autore

<head>

Titolo

<div type="poem">

<head>Mattino</head>

<lg>

<l type="libero" n=1>Mi illumino</l>

<type="libero" n=2>d'immenso.</l>

</lg>

<author>G. Ungaretti</author>

<div>

Esempio: il Perseus Project