

# Presentazione del Corso Codifica di Testi a.a. 2019-2020

Angelo Mario Del Grosso

`angelo.delgrosso@ilc.cnr.it`

*CNR-ILC*

Istituto di Linguistica Computazionale “A. Zampolli”,  
22nd February 2020

# Piano della presentazione

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

1 Presentazione

2 Introduzione

3 Codifica dei Testi

4 Ecosistema XML e Strumenti

5 Conclusioni

# Progress status

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

1 Presentazione

2 Introduzione

3 Codifica dei Testi

4 Ecosistema XML e Strumenti

5 Conclusioni

# Di cosa mi occupo

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Filologia Digitale e Computazionale

Attività di ricerca per lo sviluppo di sistemi di linguistica e filologia digitale e computazionale volti alla produzione, rappresentazione, analisi, fruizione e interrogazione di testi di tradizione medievale, a stampa e di autori moderni e contemporanei.

# Presentazione del Corso

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia



# Progress status

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

**Introduzione**

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

1 Presentazione

**2 Introduzione**

3 Codifica dei Testi

4 Ecosistema XML e Strumenti

5 Conclusioni

# Introduzione al Corso di Codifica di Testi

Obiettivi, competenze e conoscenze

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Obiettivo

Illustrare i principi di modellazione e le prassi di codifica del testo per una adeguata rappresentazione ed elaborazione digitale di risorse testuali.

## Razionale

Fornire gli strumenti e le conoscenze necessarie per progettare e realizzare criticamente una codifica digitale di testi complessi, in particolare testi letterari e di interesse storico-culturale, seguendo le linee guida messe a punto dalla Text Encoding Initiative (TEI).

# Argomenti trattati

Obiettivi, competenze e conoscenze

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

Alla fine del corso saprete:

- valutare il metodo di codifica più appropriato allo scenario d'interesse
- creare uno schema di codifica TEI
- usare gli strumenti più idonei per la codifica di una risorsa testuale
- elaborare e visualizzare il testo codificato tramite un browser web



# Principali Argomenti

Obiettivi, competenze e conoscenze

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

- Rappresentazione digitale di caratteri e di testi
- I linguaggi di markup e introduzione a XML
- Creazione di schemi di codifica
- Le norme TEI (Text Encoding Initiative)
- Alcuni specifici Moduli TEI
- Definizione di schemi di codifica personalizzati
- Introduzione ai fogli di stile XSLT
- Elaborazione documenti XML-TEI (XSLT2.0, DOM)
- Esempi, esercitazioni e seminari

# Bozza piano del Corso: Calendario 2019/2020

Lunedì dalle 14.15 alle 15.45 - Giovedì dalle 16.00 alle 17.30

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

- Settimana 1: Introduzione al corso e strumenti software utilizzati
- Settimana 2: Elementi teorici relativi alla rappresentazione del testo
- Settimana 3: Elementi di XML e introduzione alla definizione degli Schemi XML
- Settimana 4: Elementi di Codifica TEI
- Settimana 5: Elementi di Codifica TEI

# Bozza piano del Corso: Calendario 2019/2020

Lunedì dalle 14.15 alle 15.45 - Giovedì dalle 16.00 alle 17.30

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

- Settimana 6: Moduli Editoriali TEI
- Settimana 7: Modulo di approfondimento TEI
- Settimana 8: Elaborazione documenti XML-TEI (XSLT e DOM)
- Settimana 9: Elaborazione documenti XML-TEI (XSLT e DOM) Esercitazione e discussione sui progetti
- Settimana 10: Esercitazione e discussione sui progetti
- (Settimana 11-15): Eventuali esercitazioni, attività laboratoriali, recupero argomenti

# Perché è importante la codifica dei testi

## Motivazioni pratiche

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Perché codificare i testi

Per rendere disponibile l'immenso patrimonio testuale tramite l'uso di sistemi digitali e computazionali è necessario effettuare una trasposizione/transcodifica\* dei testi dal loro supporto originario verso il nuovo supporto elettronico (*Machine Readable Form and Machine Actionable Form*).

\* *procedimento di conversione dei dati codificati secondo un sistema verso un sistema diverso*

# Elementi di Codifica dei Caratteri

## Definizioni

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

### Rappresentare il testo in formato digitale

L'adozione di metodologie e tecnologie informatiche per il trattamento di documenti testuali richiede in primo luogo la disponibilità di un'adeguato sistema di rappresentazione digitale dei dati - presenti nella risorsa originale - e conseguentemente un formalismo adatto a tale rappresentazione.

# Perché è importante la codifica dei testi

## Dove troviamo i testi

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

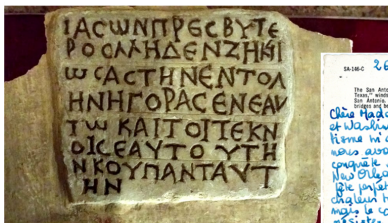
Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia



# Elementi di Codifica dei Caratteri

## Problemi di rappresentazione

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia



lasciato bianca 42r e ha scritto ancora su 43r  
capovolto.

¶ Nous commençons par dire que  $\tau$  était la "forme faible" de  $\tau$ .

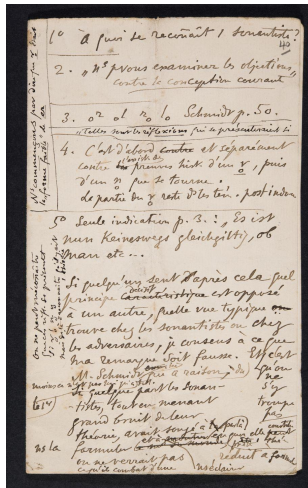
- 1° À quoi se reconnaît un sonantiste?
2. nous pouvons examiner les objections contre la conception courante.
3. or  $\tau$  et  $\tau_0$  Schmidt p. 50.
4. C'est d'abord  $\tau$  et séparément contre l'existence de preuves historiques d'un  $\tau$ , puis d'un  $\tau_0$  qui se tourne.
- 5° Seule indication p. 3: "Es ist nun keineswegs gleichgültig [sic], ob man etc. ..."

Si quelque'un sent après cela quel principe décisif est opposé à un autre, quelle vue typique on trouve chez les sonantistes ou chez les adversaires, je consens à ce que ma remarque soit fautive. Et qu'on ne s'y trompe pas c'est M. Schmidt ici qui a raison, (du moins ce n'est pas lui qui a le premier tort.)

Si quelque part les sonantistes, tout en menant grand bruit de leur théorie, n'avaient songé à nous la formuler et à nous éclairer par là en quoi elle constitue une théorie on ne verrait pas réduit à formuler ce qu'il combat d'une.

¶ On ne peut méconnaître que ces réflexions se présentent si  $\tau$   $\tau_0$   $m$   $n$  [...] mais de là à reconnaître qu'il y ait une théorie.

- ¶ Si l'on ne peut méconnaître que ces réflexions se présentent si  $\tau$   $\tau_0$   $m$   $n$  [...] mais de là à reconnaître qu'il y ait une théorie.
- ¶ Si l'on ne peut méconnaître que ces réflexions se présentent si  $\tau$   $\tau_0$   $m$   $n$  [...] mais de là à reconnaître qu'il y ait une théorie.
- ¶ Si l'on ne peut méconnaître que ces réflexions se présentent si  $\tau$   $\tau_0$   $m$   $n$  [...] mais de là à reconnaître qu'il y ait une théorie.
- ¶ Si l'on ne peut méconnaître que ces réflexions se présentent si  $\tau$   $\tau_0$   $m$   $n$  [...] mais de là à reconnaître qu'il y ait une théorie.
- ¶ Si l'on ne peut méconnaître que ces réflexions se présentent si  $\tau$   $\tau_0$   $m$   $n$  [...] mais de là à reconnaître qu'il y ait une théorie.



# Complessità della Codifica dei Caratteri

## Un Esempio

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

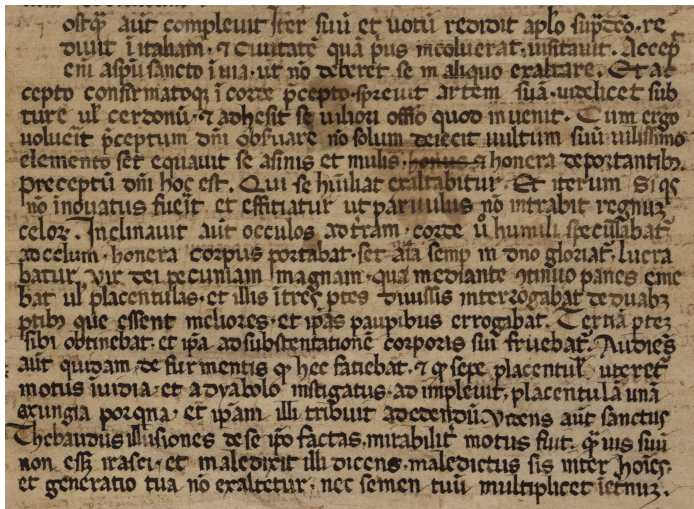
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia





# Complessità della Codifica dei Caratteri

## Un Esempio

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

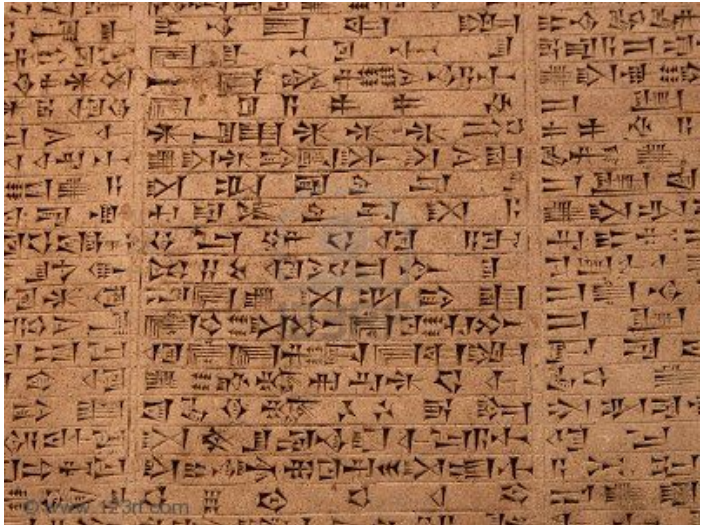
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia



# Perché è importante la codifica dei testi

## In sintesi

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Rappresentare il testo

Il focus del corso sarà incentrato sulla rappresentazione digitale del testo.

## Esistono dibattiti e controversie

Per ottenere una rappresentazione digitale del testo ci sono diversi formati, formalismi e prassi:  
la nostra scelta ricade sulle norme suggerite dal consorzio TEI.

Molte questioni non sono risolte altre sono controverse, sia dal punto di vista teorico-metodologico, sia pratico-tecnologico.

# Perché è importante la codifica dei testi

Ma in definitiva

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Diversi formati e formalismi per il testo

*Le differenze di formato sono più che altro estetiche e non sostanziali*

## Human Readable e Machine Readable

**Ma anche l'occhio umano vuole la sua parte**

# Progress status

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

1 Presentazione

2 Introduzione

**3 Codifica dei Testi**

4 Ecosistema XML e Strumenti

5 Conclusioni

# Rappresentazione digitale dei testi

basso e alto livello di codifica

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Codificare un testo

La codifica dei caratteri evidentemente non esaurisce i problemi per una opportuna rappresentazione delle caratteristiche interne ed esterne di un testo.

## Codificare un testo

Difatti la codifica del testo è una questione molto più complessa di una semplice riproduzione meccanica di un dato.

# Rappresentazione digitale dei testi

basso e alto livello di codifica

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Rappresentare un testo

La rappresentazione digitale di un testo è una operazione intrinsecamente assai difficile perché coinvolge una plethora di aspetti, a varie dimensioni, a varie granularità e a vari livelli di astrazione sia teorici, sia metodologici, sia tecnologici e sia pratici.

# Rappresentazione digitale dei testi

basso e alto livello di codifica

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Rappresentare un testo

**Prima di poter fare qualsiasi ipotesi su come compiere una codifica di un testo e su come rappresentarlo digitalmente, bisogna stabilire cosa si intende per testo.**

# Rappresentazione digitale dei testi

## Modello dati di un testo

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

Un testo non ha una struttura rigida, predefinita:

- Non è rappresentabile solo come un insieme di record di un archivio elettronico.
- Non è rappresentabile solo come un insieme di tabelle di una banca dati.
- Non è rappresentabile solo come un albero o un insieme di sotto-alberi
- Non è rappresentabile solo come un grafo o come un insieme di sotto grafi



# Molteplici modelli per diverse esigenze

Strutture dato e testo

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## La rappresentazione di un testo

- modello lineare: sequenza di dati non strutturati
- modello a record: enumerazione delle proprietà
- modello tabulare: insieme di dati omogenei
- modello ad albero: gerarchie di dati e insiemi di dati
- modello grafo: rete di strutture informative interconnesse tra loro

# Elementi di Codifica del testo

## Tabella Formalismi

### Formalismi

	Data	Text	Hierarchies	Presentation	Validation	References	Annotations	Overlapping
CSV								
JSON								
RDF								
Markdown								
HTML								
HTML+RDFa								
XML								
Overlapping formats								

courtesy of *Fabio Vitali*

# Elementi di Codifica del testo

## Formalismi

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

### Formati di rappresentazione

Un formato è un insieme di regole e convenzioni formali per rappresentare un insieme di dati, nel nostro caso un testo.

### Importanza dei formati

Seppur isomorfi la scelta dei formati condiziona molto l'efficienza delle operazioni e l'efficacia delle dichiarazioni.

# Elementi di Codifica del testo

## Esempio di codifica del testo utilizzando CSV

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

Why it 's worth giving\_up .  
The human\_body is remarkably resilient and will start to repair the damage caused by smoking almost as soon as you start giving\_up .  
The benefits to your health can be seen in as little as 20 minutes from the time you put\_down your last cigarette .  
The whole process remains only marginally\_safe on average for women\_and\_children .  
Both men\_and women want the process to be safer .  
Both men\_and women are involved in developing and exploiting the sciences and technologies to make\_this\_happen .  
The gender\_dynamics of public\_perceptions on DNA\_paternity\_testing are likely to become more important as the tests become more widespread .  
Some perceptions of paternity\_testing will obviously be dependent on the inherently\_different reproductive\_roles of men\_and\_women .

Assume	VB	assume	
that	IN	that	
a	DT	a	
natural	JJ	natural	
cyclic	JJ	cyclic	
phenomenon		phenomenon	phenomenon
has	VBZ	have	
been	VDN	be	
measured	VEN	measure	
but	CC	but	
the	DT	the	
data	NN	data	
is	VBZ	be	
corrupted	VEN	corrupt	
by	IN	by	
errors	NNS	error	
.	SENT	.	

```
<$ s_id="1">
  <w surface="Cathy" lemma="Cathy" pos="N" syn="su:2" />
  <w surface="zag" lemma="zie" pos="V" syn="ROOT:0" />
  <w surface="hen" lemma="hen" pos="Pron" syn="obj1:2" />
  <w surface="wild" lemma="wild" pos="Adj" syn="mod:5" />
  <w surface="zwaalen" lemma="zwaal" pos="N" syn="vc:2" />
  <w surface="." lemma="." pos="Punc" syn="punct:5" />
</$>
```

```
{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "http://574heritago.com/annotations/2/",
  "type": "Annotation",
  "created": "2017-03-12T12:03:45Z",
  "body": {
    "type": "TextualBody",
    "purpose": "tagging",
    "format": "text/plain",
    "language": "en",
    "value": "Where is the origin of this statue?"
  },
  "creator": {
    "id": "http://heritago.com/users/1/",
    "type": "Person",
    "email_sha1": "58bad08927902ff9307b621c54716dcc5083e339"
  },
  "selector": {
    "type": "TextPositionSelector",
    "start": 15,

```

<<http://example1.com/citingwork>>

**cito:cites** <<http://example2.com/citedwork>> ;

**cito:inTextCitationFrequency** [

a **cito:inTextCitationCount** ;

**cito:inTextCountValue** "10"^^xsd:integer ;

# Elementi di Codifica del testo

## Formalismi

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

### Formati come formalismi

Data l'importanza metodologica il formato del dato diviene un vero e proprio formalismo, si parla cioè di linguaggi di codifica in quanto questi sistemi si basano su un insieme di istruzioni rigorose di codifica.

# Elementi di Codifica del testo

## Formalismi

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

### Formati e formalismi di codifica

Quindi ogni pezzo di informazione aggiunta ad un testo grezzo attraverso l'inserimento di dati metatestuali (markup, annotazione, codifica), costituisce il risultato di una analisi e di una interpretazione che è stata condotta (da un umano o da una macchina) al fine di esplicitare e rappresentare nel modo più accurato e completo possibile le informazioni da veicolare attraverso il formato digitale prescelto (anche in modo incrementale).

# Progress status

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

1 Presentazione

2 Introduzione

3 Codifica dei Testi

**4 Ecosistema XML e Strumenti**

5 Conclusioni

## Metodi e tecniche per la codifica di testi

La riflessione sui metodi e le pratiche migliori per la codifica elettronica dei testi è stato uno dei temi fondamentali della ricerca e della sperimentazione nel dominio dell'Informatica Umanistica per molti anni.



# Markup language e XML

soluzione corrente per la codifica dei testi

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## XML per la descrizione e la codifica

Ad oggi la soluzione considerata ottimale per una corretta rappresentazione del testo è l'adozione dei markup language descrittivi basati su XML.

## TEI-XML

Standard de facto per la codifica dei testi è considerato lo schema XML messo a punto dalla Text Encoding Initiative (TEI-XML).

# Lo schema TEI-XML

## Estratto di documento TEI-XML

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

```
1 <?xml version="1.0" encoding="utf-8" standalone="yes"?>
2 <!DOCTYPE div SYSTEM "divTEIsnippet.dtd">
3 <divTEIsnippet>
4   <div type="destination">
5     <ab>
6       <stamp type="postmark">
7         <placeName>El Paso</placeName>
8         - TX 799 -
9         <date notBefore="1980-07-26">
10          <unclear>PM JUL</unclear>
11        </date>
12      </stamp>
13      <stamp type="postage">
14        Profil masculin, avec un avion et un radar au second
15        plan:
16        <mentioned>US Airmail 21 c.</mentioned>
17      </stamp>
18    </ab>
19    <ab>
20      <address>
21        <addrLine>
22          Madame
23          <name>Lefrère</name>
24        </addrLine>
25        <addrLine>4, allée George Rouault</addrLine>
26        <addrLine>75020 Paris</addrLine>
27        <addrLine>France</addrLine>
28      </address>
29    </ab>
30  </div>
31 </divTEIsnippet>
```

### Perché TEI

La Text Encoding Initiative (TEI) è un autorevole progetto internazionale, a cui afferiscono varie organizzazioni e università, il cui scopo è fornire agli studiosi di informatica umanistica uno strumento il più espressivo e flessibile possibile per rappresentare qualsiasi aspetto di interesse relativo alla risorsa testuale da rappresentare digitalmente.

## Perché XML

- separazione dei dati dall'applicativo di authoring/editing
- separazione della rappresentazione dei dati dalla presentazione dei dati
- possibilità di trasformare i dati in qualsiasi altro formato compatibile
- leggibilità dei documenti XML da parte di esseri umani

## Perché XML

- standard w3c testuale, aperto, personalizzabile e liberamente utilizzabile
- semplicità di condivisione e scambio dati (interoperabilità e portabilità)
- adatto per codificare dati semistrutturati oltre che dati strutturati
- validazione del documento attraverso grammatiche formali

# Ecosistema XML

Le tecnologie XML per la definizione ed elaborazione di documenti XML

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

- XSD: XML Schema Definition Language
- XPath: XML Path Language
- XSL: eXtensible Stylesheet Language
- XSL-T: XSL – Transformations
- XSL-FO: XSL – Formatting Objects
- XQuery: XML Query Language for XML Databases
- XInclude: XML inclusion Language
- DTD: Document Type Definition Language
- RelaxNG: Regular Expression Language for XML (New Generation)

# Linguaggio di marcatura XML

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Perché XML

Adottando la tecnologia e il linguaggio XML abbiamo la possibilità di creare linguaggi di marcatura personalizzati e specifici per ogni esigenza e dominio.

# XML come linguaggio per la codifica di testi

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

## Vantaggi

Attraverso XML è possibile strutturare i dati, gestire in modo nativo strutture gerarchiche, elaborare e presentare i dati con strumenti XML nativi, validare i tipi di strutture e i tipi di dati, gestire riferimenti incrociati tramite opportuni meccanismi di dereferenziazione, aggiungere e gestire annotazioni a vari livelli di granularità.



# Progress status

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

1 Presentazione

2 Introduzione

3 Codifica dei Testi

4 Ecosistema XML e Strumenti

5 Conclusioni

# Prerequisiti

Cosa è bisogna sapere

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

- Buona conoscenza dell'inglese
- Nozioni di HTML / CSS / JS
- Computer personale a disposizione per svolgere gli esercizi assegnati
- Uso di editor di testo
- Uso di editor per programmatori o editor XML
- Eseguire comandi da shell

# Strumenti

Cosa è consigliato usare

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

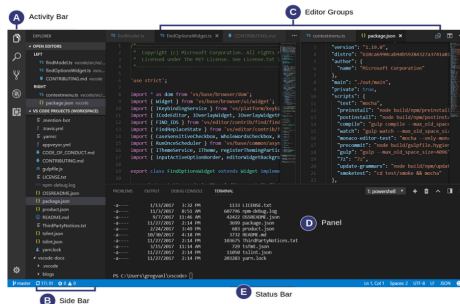
- Editor di testo professionali (syntax highlighting per XML, workspace)
- Editor XML + processore XSLT (normalmente è integrato nell'editor XML)
- Navigatore web
- Manuali di codifica (Guidelines TEI P5)
- Materiale didattico (slide, esempi, esercizi)

# Editor di testo

## Visual Studio Code

### User Interface

At its heart, Visual Studio Code is a code editor. Like many other code editors, VS Code adopts a common user interface and layout of an explorer on the left, showing all of the files and folders you have access to, and an editor on the right, showing the content of the files you have opened.



Edit

#### IN THIS ARTICLE

- Basic Layout
- Side by Side Editing
- Minimap - outline view
- Explorer
- Open Editors
- Views
- Command Palette
- Configuring the Editor
- Tabs
- Preview mode
- Editor Groups
- Grid editor layout
- Working without Tabs
- Window Management
- Next Steps
- Common Questions
- Twit
- Subscribe
- Ask questions
- Follow @code
- Request features
- Report issues
- Watch videos

### Basic Layout

Home page del tool: <https://code.visualstudio.com/>

# Validatore XML

## XMLlint

Presentazione  
del Corso  
Codifica di  
Testi

a.a.

2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

```
XMLLINT(1)                                xmlint Manual                                XMLLINT(1)
NAME
    xmlint - command line XML tool
SYNOPSIS
    xmlint [--version | --debug | --shell | --xpath "XPath_expression" | --debugent | --copy | --recover | --noent | --noout | --nomet |
    --path "PATH(S)" | --load-trace | --htmlout | --nowrap | --valid | --postvalld | --dtdvalid URL | --dtdvalidfpi FPI | --timing |
    --output FILE | --repeat | --insert | --compress | --html | --xsltout | --push | --memory | --maxmem MBBYTES | --nowarning | --noblanks |
    --nocdata | --format | --encode ENCODING | --droptdtd | --nsclean | --testfio | --catalogs | --nocatalogs | --auto | --xinclude |
    --noinclude | --loadtd | --dtdattr | --stream | --walker | --pattern PATTERNVALUE | --chkregister | --relaxng SCHEMA |
    --schema SCHEMA | --c14n] [XML-FILE(S)... ] -)
    xmlint --help
DESCRIPTION
    The xmlint program parses one or more XML files, specified on the command line as XML-FILE (or the standard input if the filename provided is -
    ). It prints various types of output, depending upon the options selected. It is useful for detecting errors both in XML code and in the XML
    parser itself.
    xmlint is included in libxml(3).
OPTIONS
    xmlint accepts the following options (in alphabetical order):
    --auto
        Generate a small document for testing purposes.
    --catalogs
        Use the SGML catalog(s) from SGML_CATALOG_FILES. Otherwise XML catalogs starting from /etc/xml/catalog are used by default.
    --chkregister
        Turn on node registration. Useful for developers testing libxml(3) node tracking code.
    --compress
        Turn on gzip(1) compression of output.
    --copy
        Test the internal copy implementation.
    --c14n
        Use the W3C XML Canonicalisation (C14N) to serialize the result of parsing to stdout. It keeps comments in the result.
    --dtdvalid URL
        Manual page xmlint(1) line 1 (press h for help or q to quit)
```

Home page del tool: <http://www.xmlsoft.org/>

# Editor XML

Cosa è consigliato usare

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

- un buon editor open source: XML Copy Editor (<http://xmlcopy-editor.sourceforge.net/>)
- per Mac: XMLSpear, Textmate, Eclipse, IntelliJIDEA
- un ottimo editor: Oxygen (<http://www.oxygenxml.com/>)  
- multi piattaforma, ma non gratuito (in prova gratuita per un mese)
- altri editor: funzioni fondamentali sono la validazione, l'autocompletamento, l'esecuzione di fogli di stile

# Editor XML

## Editix

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

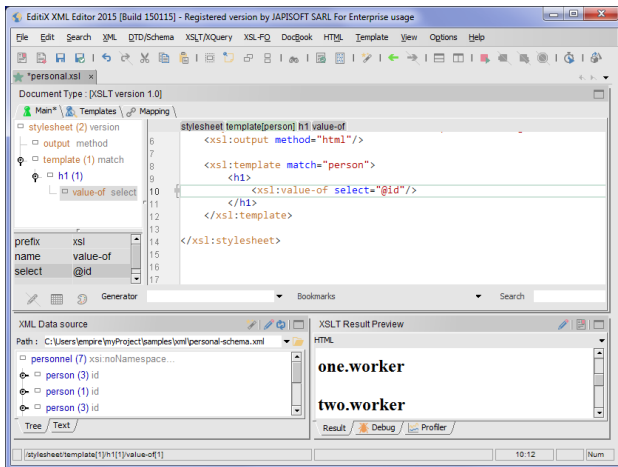
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia



Home page del tool: <http://www.editix.com>

# Programma d'esame

Cosa bisogna fare per superare l'esame

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

- Studiare i lucidi delle lezioni
- Padroneggiare gli esercizi svolti durante il corso
- Studiare i testi indicati dal docente
- Studiare i capitoli delle Guidelines TEI che riguardano il corso e il progetto
- Realizzare il progetto di codifica concordato con il docente



# Modalità d'esame

In cosa consiste l'esame

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia

- Invio del progetto (obbligatoriamente tramite github)
- Colloquio
  - Discussione del progetto
  - Verifica delle conoscenze di base XML, XSD, XSL
  - Verifica delle basi teoriche
  - Conoscenza di TEI P5 (moduli principali, parti spiegate a lezione, moduli particolari utilizzati nel progetto)

### Slide

- Slide delle lezioni corso a.a. 2019-2020
- Materiale integrativo fornito dal docente
- Repo github del corso  
<https://github.com/angelodel80/corsoCodifica>

### Libri

- Burnard, L. (2014). What is the Text Encoding Initiative? How to add intelligent markup to digital resources. Marseille: OpenEdition Press.
- Ciotti, F. (2007). Il testo e l'automa: saggi di teoria e critica computazionale dei testi letterari. Aracne.
- Goldberg, K. H. (2010). XML: Visual QuickStart Guide. Pearson Education.
- Carey, P., and Vodnik, S. (2014). New Perspectives on XML, Comprehensive. Cengage Learning.

### Siti Web

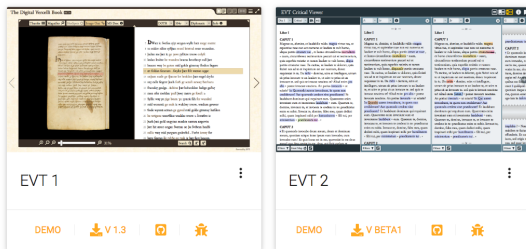
- <http://www.tei-c.org/>
- <http://teibyexample.org/>
- <https://www.w3.org/standards/xml/>

# Tirocini, Tesi, EVT

## Pronti a far parte della comunità?

### Edition Visualization Technology

A light-weight, open source tool specifically designed to create digital editions from XML-encoded texts, freeing the scholar from the burden of web programming and enabling the final user to browse, explore and study digital editions by means of a user-friendly interface.



DH Awards: <http://dhawards.org/dhawards2019/voting/>

# References

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione  
Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia



Lenci, A., Montemagni S., and Pirrelli V. (2016). Testo e Computer. Elementi Di Linguistica Computazionale. Aulamagna. Carocci.



Pierazzo, E. (2015). Digital Scholarly Editing : Theories, Models and Methods. Farnham Surrey: Ashgate.



Orlandi, T. (2010). Informatica testuale: teoria e prassi. Laterza.



Driscoll, M. J., and Pierazzo, E. (Eds.). (2016). Digital Scholarly Editing: Theories and Practices (Vol. 4). Open Book Publishers.



Ciotti F., e Crupi G, a c. di. (2012). Dall'Informatica umanistica alle culture digitali. ROMA : Casa Editrice Università La Sapienza. open access: <http://www.editricesapienza.it/node/7688>



Williams, I. (2009). Beginning XSLT and XPath: Transforming XML Documents and Data. Wiley.



Kay, M. (2011). XSLT 2.0 and XPath 2.0 Programmer's Reference. Wiley.

# References

Presentazione  
del Corso  
Codifica di  
Testi  
a.a.  
2019-2020

A.M. Del  
Grosso

Presentazione

Introduzione

Codifica dei  
Testi

Ecosistema  
XML e  
Strumenti

Conclusioni

Bibliografia



*XML Standards Reference, MSDN.*

[https://msdn.microsoft.com/en-us/library/ms256177\(v=vs.110\).aspx](https://msdn.microsoft.com/en-us/library/ms256177(v=vs.110).aspx)



*IBM XML Tutorial,*

<https://www.ibm.com/developerworks/xml/tutorials/xmlintro/xmlintro.html>



*W3School Tutorial* <https://www.w3schools.com/xml/default.asp>