

Seminario XML/TEI - FeDHLab

Angelo Mario Del Grosso

`angelo.delgrosso@ilc.cnr.it`

CNR-ILC

Istituto di Linguistica Computazionale “A. Zampolli”,
9th October 2024

Outline della lezione

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

1 Introduzione e Strumenti di lavoro

2 Rappresent. Digitale dei Testi

3 Ecosistema XML

4 Fondamenti XML

5 Conclusioni XML

Progress status

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

1 Introduzione e Strumenti di lavoro

2 Rappresent. Digitale dei Testi

3 Ecosistema XML

4 Fondamenti XML

5 Conclusioni XML

Argomenti del Seminario

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML



Introduzione al Seminario

Obiettivi, competenze e conoscenze

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

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 con senso critico una **codifica digitale di testi** complessi, in particolare testi letterari e di interesse storico-culturale, adottando **le linee guida introdotte dalla Text Encoding Initiative (TEI)**.

Principali Argomenti

Obiettivi, competenze e conoscenze

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

- Introduzione XML come linguaggio di markup
- Validazione di un documento XML
- Le norme TEI (Text Encoding Initiative)
- Alcuni Moduli specifici delle linee TEI
- Introduzione ai fogli di stile XSLT
- Elaborazione documenti XML-TEI (XSLT2.0, XQuery)
- Esempi ed esercitazioni

Materiale Didattico e strumenti di lavoro

Studio e approfondimento

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Studio

Sono disponibili **numerose risorse**, anche gratuite e online, per studiare ed approfondire i principali temi trattati durante il seminario

Pratica e strumenti di lavoro

Utilizzeremo **strumenti open source**, non vincolati da licenze a pagamento, per sviluppare gli argomenti oggetto di studio

Materiale di studio

Open Edition Book - Lou Burnard

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso


Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi


Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

**OpenEdition Press**

HOME
ENCYCLOPÉDIE NUMÉRIQUE
READ/WRITE BOOK
SCIENCES PO | MÉDIALAB
BRÉSIL / FRANCE | BRASIL / FRANÇA



CERCA NEL LIBRO **OK**

INDICE DEI CONTENUTI

CITA CONDIVIDI

CITATO DA

WHAT IS THE TEXT ENCODING INITIATIVE?

How to add intelligent markup to digital resources

Lou Burnard
Encyclopédie numérique


...


The Text Encoding Initiative (TEI) Guidelines have long been regarded as the de facto standard for the preparation of digital textual resources in the scholarly research community. For the beginner, they offer a daunting range of possibilities, reflecting the huge range of potential applications for text encoding, from traditional scholarly editions, to language corpora, historical lexicons, digital archives and beyond.


Drawing on many examples of TEI-encoded text from a variety of res...


[→ Leggi il seguito](#)

LEGGI

**ACCESSO APERTO**

**LETTURA CONFORTEVOLE**

**EPUB**

**PDF DEL LIBRO**

Editore: OpenEdition Press
Luogo di pubblicazione: Marseille
Pubblicato su OpenEdition Books: 11 avril 2014
EAN digitale: 9782821834606
Numero di pagine: 114 p.

Collana: Encyclopédie numérique | 3
Anno di pubblicazione: 2014
EAN (edizione cartacea): 9782821834590
DOI: 10.4000/books.oep.426

<https://books.openedition.org/oep/426>

Materiale di studio

Linee Guida della TEI - Online

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML



TEI: Guidelines for Electronic Text Encoding and Interchange

P5 Version 4.8.0. Last updated on 2nd September 2024, revision 4e6e41b0b

[English] [Deutsch] [Español] [Italiano] [Français] [日本語] [한국어] [中文]



Front Matter

Title

- I. Releases of the TEI Guidelines
- II. Dedication
- III. Preface and Acknowledgments
- IV. About These Guidelines
- V. A Gentle Introduction to XML
- VI. Languages and Character Sets

Back Matter

- Appendix A Model Classes
- Appendix B Attribute Classes
- Appendix C Elements
- Appendix D Attributes
- Appendix E Datatypes and Other Macros
- Appendix F Bibliography
- Appendix G Deprecations
- Appendix H Prefatory Notes
- Appendix I Colophon

Text Body

- 1 The TEI Infrastructure
- 2 The TEI Header
- 3 Elements Available in All TEI Documents
- 4 Default Text Structure
- 5 Characters, Glyphs, and Writing Modes
- 6 Verse
- 7 Performance Texts
- 8 Transcriptions of Speech
- 9 Computer-mediated Communication
- 10 Dictionaries
- 11 Manuscript Description
- 12 Representation of Primary Sources
- 13 Critical Apparatus
- 14 Names, Dates, People, and Places
- 15 Tables, Formulae, Graphics, and Notated Music
- 16 Language Corpora
- 17 Linking, Segmentation, and Alignment
- 18 Simple Analytic Mechanisms
- 19 Feature Structures

TEI sourcecode

- Getting and Using the TEI Sources.
- TEI GitHub Repository
- Bug Reports, Feature Requests, etc.

<https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML
3RD EDITION



COMPREHENSIVE

Carey :: Vodnik

New Perspectives on XML, Comprehensive, 3rd Edition.
Carey/Vodnik. Cengage Learning.

Strumenti di lavoro

esercitazioni, esempi, progetto

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Overview

SETUP

GET STARTED

Intro Videos

Tips and Tricks

User Interface

Themes

Settings

Key Bindings

Display Language

USER GUIDE

LANGUAGES

NODE.JS /

JAVASCRIPT

PYTHON

JAVA

AZURE

EXTENSION

AUTHORING

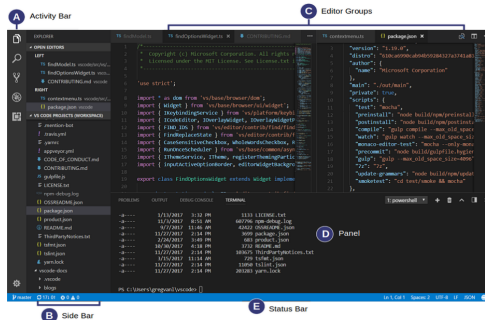
EXTENSIBILITY

REFERENCE

OTHER

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.



Basic Layout

Edit

IN THIS ARTICLE

Basic Layout

Side by Side Editing

Minimap - outline view

Explorer

Open Editors

Views

Command Palette

Configuring the Editor

Themes

Preview mode

Editor Groups

Grid editor layout

Working without Tabs

Window Management

Next Steps

Common Questions

Twitter

Subscribe

Ask questions

Follow @code

Request features

Report issues

Watch videos

<https://code.visualstudio.com/>

Strumenti di sviluppo

esercitazioni, esempi, progetto

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

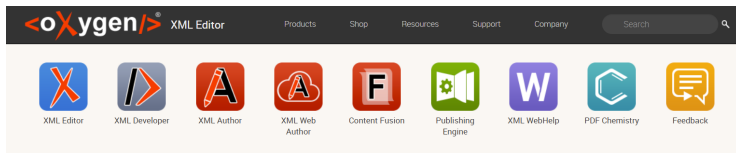
Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML



Oxygen XML Editor 23

The Premier All-In-One XML Editing Suite

Oxygen XML Editor is **the best XML editor available** and provides a comprehensive suite of XML authoring and development tools. It is designed to accommodate a large number of users, ranging from beginners to XML experts. It is available on multiple platforms, all major operating systems, and as a standalone application or an Eclipse plug-in. You can use Oxygen XML Editor in conjunction with all XML-based technologies and it includes a large variety of powerful tools for creating, editing, and publishing XML documents.

[SEE WHAT'S NEW](#)

[Read more](#)

[Download](#)

[Buy \(from \\$99\)](#)



<https://www.oxygenxml.com/>

Strumenti di sviluppo

esercitazioni, esempi, progetto

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

The screenshot shows the eXist-db website homepage. At the top, there is a navigation bar with the eXist-db logo, links for 'About', 'Support', and 'Community', and social media icons for GitHub, LinkedIn, Twitter, Facebook, and YouTube. The main heading is 'Vitamins for your Applications', followed by the tagline 'Try the all-in-one solution for application building.' Below this is the Open Source Initiative logo and the text 'eXist-db is Open Source Software licensed under the LGPL'. A prominent blue button labeled 'Download eXist-db' is centered. To the right of the text is an image of a red apple with a green leaf, wrapped in a spiral of orange and yellow fruit slices. At the bottom left, a 'Latest News' section features a red header and a news item dated 'February 05 2023' for 'eXist-db 6.2.0'. On the bottom right, there is a timeline graphic showing the project's history, with a date 'July 29 2014' and a label 'eXist-db 1.0'.

<http://exist-db.org/exist/apps/homepage/index.html>

Progress status

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

1 Introduzione e Strumenti di lavoro

2 Rappresent. Digitale dei Testi

3 Ecosistema XML

4 Fondamenti XML

5 Conclusioni XML

Perché è importante la codifica dei testi

In sintesi

Rappresentare il testo

Il focus del seminario 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.

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

I linguaggi di codifica

introduzione

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

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).

Elementi di Codifica del testo

Formalismi

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Formati di rappresentazione

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

Importanza dei formati

La scelta dei formati condiziona molto l'efficienza delle operazioni e l'efficacia delle dichiarazioni.

Markup language e XML

soluzione corrente per la codifica dei testi

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

TEI-XML

Le specifiche messe a punto dalla Text Encoding Initiative (TEI-XML) sono considerate ad oggi lo **standard de facto** per una corretta **rappresentazione digitale dei testi** con prospettiva scientifica.

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 trattare digitalmente**.

I principi fondamentali della TEI

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML



- Le linee guida della TEI privilegiano il “significato” (**meaning**) del testo piuttosto che l’“aspetto” (layout); privilegia il modello del testo, piuttosto che il formato
- La TEI è stata progettata per essere **indipendente** dagli strumenti software che la usano per la creazione oppure per l'elaborazione dei documenti elettronici
- La TEI cresce, matura, si evolve sulla base delle indicazioni e delle ricerche dalla propria comunità di riferimento (**community-driven**)

I principi fondamentali della TEI

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML



- Le linee guida della TEI privilegiano il “significato” (**meaning**) del testo piuttosto che l’“aspetto” (layout); privilegia il modello del testo, piuttosto che il formato
- La TEI è stata progettata per essere **indipendente** dagli strumenti software che la usano per la creazione oppure per l’elaborazione dei documenti elettronici
- La TEI cresce, matura, si evolve sulla base delle indicazioni e delle ricerche dalla propria comunità di riferimento (**community-driven**)

I principi fondamentali della TEI

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML



- Le linee guida della TEI privilegiano il “significato” (**meaning**) del testo piuttosto che l’“aspetto” (layout); privilegia il modello del testo, piuttosto che il formato
- La TEI è stata progettata per essere **indipendente** dagli strumenti software che la usano per la creazione oppure per l’elaborazione dei documenti elettronici
- La TEI cresce, matura, si evolve sulla base delle indicazioni e delle ricerche dalla propria comunità di riferimento (**community-driven**)

Intro Text Encoding Initiative

TEI

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

TEI Guidelines: versione P5

- basata su **XML**, schema RelaxNG (e DTD tradizionale)
- pubblicata alla **fine del 2007**, aggiornata due volte l'anno
- molte novità interessanti (in particolare: maggior **modularità**)
- <http://www.tei-c.org/Guidelines/P5/>

Intro Text Encoding Initiative

TEI

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

TEI Guidelines: Obiettivi

- better **interchange** and **integration** of scholarly data
- support for **all texts**, in **all languages**, from **all periods**
- guidance for the *perplexed*: **what to encode** - hence, a user-driven codification of existing best practice
- assistance for the *specialist*: **how to encode** - hence, a loose framework into which unpredictable extensions can be fitted

Intro Text Encoding Initiative

TEI

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

TEI: Struttura modulare

- si scelgono soltanto i **moduli che corrispondono alle proprie esigenze**, in modo da realizzare rapidamente uno schema di codifica appropriato
- ogni modulo contiene un certo numero di **elementi** (*tagset*)
- gli **elementi** sono organizzati in **classi** (strutturali, semantiche)
- gli **attributi** sono organizzati in **classi** (globali e specifici)

Progress status

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

1 Introduzione e Strumenti di lavoro

2 Rappresent. Digitale dei Testi

3 Ecosistema XML

4 Fondamenti XML

5 Conclusioni XML

Markup language e XML

soluzione corrente per la codifica dei testi

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

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 sul (meta)linguaggio standard **W3C 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).

Impiego di XML

Benefici

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

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

Impiego di XML

Benefici

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

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

Ecosistema XML

Le tecnologie XML per la definizione ed elaborazione di documenti XML

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

- 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

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

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.

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

Progress status

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

1 Introduzione e Strumenti di lavoro

2 Rappresent. Digitale dei Testi

3 Ecosistema XML

4 Fondamenti XML

5 Conclusioni XML

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML come meta-linguaggio

XML, *eXtensible Markup Language*, è un insieme di regole per definire linguaggi di marcatura personalizzati e personalizzabili (*custom-built vocabularies*).

Applicazioni XML

XML è nato per **strutturare, conservare e trasportare** informazioni.

I linguaggi di marcatura derivati da XML per strutturare e descrivere specifiche informazioni vengono chiamati *XML applications* oltre che *vocabolario XML*.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML: eXtensible

XML è **estensibile**: è pensato per essere *modificato* ed *esteso* al fine di soddisfare le varie necessità di rappresentazione dell'informazione. **XML non contempla un vocabolario predefinito!**

XML: standard W3C

XML è sviluppato e **manutenuto dal W3C** (World Wide Web Consortium), il quale sviluppa *protocolli* e *standard* riconosciuti dalla comunità scientifica e tecnica al fine di **condividere informazioni sul Web**.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML: riassumendo

XML, eXtensible Markup Language, deriva da SGML ed è una **specificazione**, un **formalismo**, per *strutturare, conservare e scambiare* informazioni in formato machine readable (*digitale*).

XML: riassumendo

XML è anche una specificazione per **descrivere la struttura dell'informazione** seguendo un **modello dei dati gerarchico**. XML è simile ad HTML, ma a differenza di questo non ha etichette predefinite.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

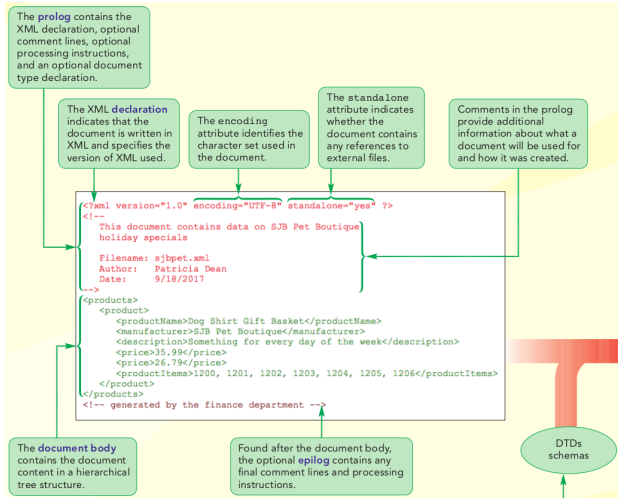


immagine dal libro *New Perspectives on XML, 3rd Edition*

Fondamenti XML

eXtensible Markup Language: regole sintattiche

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

- Ciascun elemento XML deve avere un tag di chiusura.
- I tag XML sono *case sensitive*.
- Gli elementi XML devono essere annidati in modo rigoroso.
- Tutti i documenti XML devono avere un elemento radice (root) che contiene tutti gli altri elementi opportunamente annidati.
- Gli elementi XML possono avere attributi con stile nome-valore.

Fondamenti XML

eXtensible Markup Language: regole sintattiche cont.

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

- Un attributo all'interno dell'elemento può apparire una sola volta
- Il valore degli attributi è una stringa e deve essere inserita tra apici
- Esistono alcuni caratteri speciali che non possono essere usati.
- I commenti non possono essere inseriti prima della dichiarazione XML e non possono essere annidati.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Manutenibilità

Data la semplicità delle regole e della sintassi XML incentrata sulla memorizzazione e scambio dei dati, la struttura generale di un documento XML è semplice sia dal punto di vista della progettazione sia dal punto di vista della manutenibilità.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML vista ad albero

XML ha un **modello dei dati gerarchico** e può quindi essere visto come un **albero etichettato ordinato**.

Per questo motivo le informazioni sono rappresentate in modo ottimale se sono gerarchiche e sequenziali.

Fondamenti XML

eXtensible Markup Language: vista ad albero

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML document

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!--
  This document contains data on SJB Pet Boutique
  holiday specials

  Filename: sjbpets.xml
  Author:   Patricia Dean
  Date:    9/18/2017
-->
<products>
  <product>
    <productName>Dog Shirt Gift Basket</productName>
    <manufacturer>SJB Pet Boutique</manufacturer>
    <description>Something for every day of the week</description>
    <price>35.99</price>
    <price>26.79</price>
    <productItems>1200, 1201, 1202, 1203, 1204, 1205, 1206</productItems>
  </product>
</products>
<!-- generated by the finance department -->
```

Hierarchy tree structure

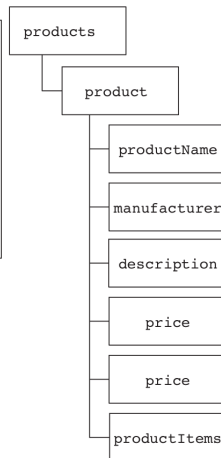


immagine dal libro New Perspectives on XML, 3rd Edition

Fondamenti XML

eXtensible Markup Language: vista ad albero

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

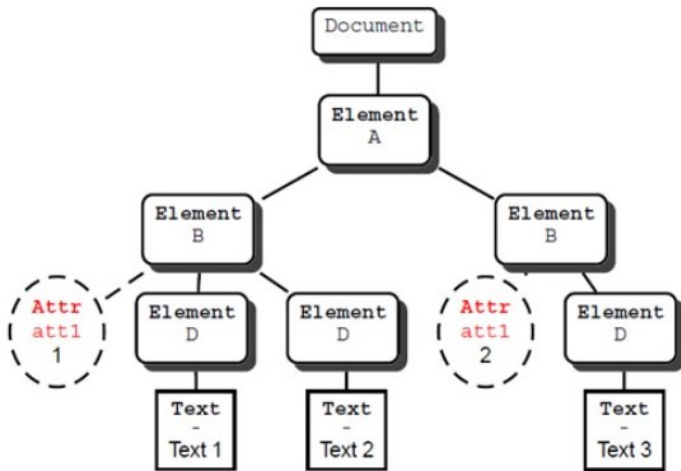
Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML



Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

TEI-XML vocabulary

Al fine di soddisfare i **requisiti degli studiosi del testo** il *vocabolario TEI-XML* è stato sviluppato nel corso degli ultimi decenni con l'obiettivo di *permettere la codifica di qualsiasi informazione testuale*.

Un vocabolario XML è un insieme di tag XML sviluppato per una particolare esigenza di codifica

Fondamenti XML

eXtensible Markup Language: Esempio TEI

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

```
<div type="narrative" n="6">
  <head>Sixth Narrative</head>
  <head>contributed by Sergeant Cuff</head>
  <div type="fragment" n="6.1">
    <opener>
      <dateline>
        <name type="place">Dorking, Surrey,</name>
        <date>July 30th, 1849</date>
      </dateline>
      <salute>To <name>Franklin Blake, Esq.</name> Sir, -</salute>
    </opener>
    <p>I beg to apologize for the delay that has occurred in the
      production of the Report, with which I engaged to furnish you.
      I have waited to make it a complete Report ...</p>
    <closer>
      <salute>I have the honour to remain, dear sir, your
        obedient servant </salute>
      <signed>
        <name>RICHARD CUFF</name> (late sergeant in the
          Detective Force, Scotland Yard, London). </signed>
      </closer>
    </div>
  </div>
```

immagine dal sito TEI Guide Lines

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Documento ben formato (well-formed)

Un documento XML deve essere **ben formato** (*well-formed*), cioè non deve contenere **errori sintattici** e deve soddisfare le **regole generali della specifica**.

Un documento non ben formato non può essere letto dalle applicazioni che elaborano codice XML.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Parti principali di un documento XML

Un documento XML consiste di tre parti:

- il prologo
- il corpo (body)
- l'epilogo

Fondamenti XML

eXtensible Markup Language: Esempio TEI

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-stylesheet type="text/css" href="customStyle.css"?>
<!--The following document is made online by the Perseus Project -->
<!--Added the TEI-lite DTD and a processing instruction -->
<!DOCTYPE TEI.2 SYSTEM "teixbaby.dtd">

<TEI.2>
  <text lang="en">
    <body>
      <div1 type="book" n="1" org="uniform" sample="complete">
        <div2 type="section" n="327A" org="uniform" sample="complete">
          <p>
            327A - 328B Socrates describes how he visited the Piraeus in company with Glauco, and
            was induced by Polemarchus and others to defer his return to Athens.
          </p>
          <p>
            <lemma lang="greek" targOrder="U" from="ROOT" to="DITTO">κατέβην κτλ.</lemma>
            Dionys. Hal.
            <title lang="la">de comp. verb.</title>
            p. 208 (Reiske)
            <foreign lang="greek">
              ὁ δὲ Πλάτων, τοὺς
              ἑαυτοῦ διαλόγους κτενίζων καὶ βοστρυχίζων, καὶ πάντα τρόπον ἀναπλέκων, οὐ
              διέλιπεν ὀγδοήκοντα γεγονόσ' ἔτη. πᾶσι γὰρ δὴ πού τοις φιλολόγοις γινώριμα
              τὰ περὶ τῆς φιλοπονίας τάνδρὸς ιστορούμενα, τὰ τ' ἄλλα, καὶ δὴ καὶ τὰ
              περὶ τὴν δέλτον ἦν τελευτήσαντος αὐτοῦ λέγουσιν εὐρεθῆναι ποικίλως
              μετακειμένην τὴν ἀρχὴν τῆς πολιτείας ἔχουσαν τήνδε "κατέβην χθές
              εἰς Πειραιᾶ μετὰ Γλαύκωνος τοῦ Ἀριστάνου
            </foreign>
            ."
          </p>
        </div2>
      </div1>
    </body>
  </text>
</TEI.2>

<!-- This document is not completed and was cut without a special meaning -->
```


Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Documento XML: prologo

- XML declaration (obbligatorio)
- Processing instructions (opzionale)
- Commenti (opzionale)
- Document type declaration (opzionale)

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Documento XML: corpo

Il corpo del documento XML segue immediatamente il prologo. Questa parte del documento contiene il contenuto vero e proprio in una **struttura ad albero ordinata**.

Documento XML: epilogo

Opzionalmente, al corpo del documento XML segue un epilogo il quale può contenere commenti finali e processing instructions.

Fondamenti XML

eXtensible Markup Language: Prologo

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML declaration

```
<?xml version="version number" encoding="encoding  
      type" standalone="yes|no" ?>
```

Fondamenti XML

eXtensible Markup Language: Prologo

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML declaration: ERRORI

```
<?XML VERSION="1.0" ENCODING="ISO-8859-1"  
      STANDALONE="YES" ?>  
<?xml version=1.0 encoding=ISO-8859-1  
      standalone=yes ?>  
<?xml version="1.0" standalone="yes"  
      encoding="ISO-8859-1" ?>
```

Fondamenti XML

eXtensible Markup Language: Prologo

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML comments

I commenti XML vengono ignorati dai programmi che elaborano il documento.

I commenti quindi non influenzano i contenuti e la struttura del documento.

XML comments: sintassi

```
<!-- il parser XML qui non entra -->
```

Un commento può occupare anche più righe

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

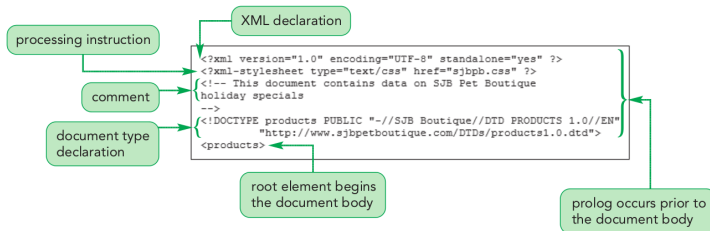


immagine dal libro New Perspectives on XML, 3rd Edition

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML parser

Un programma che legge ed interpreta un documento XML è chiamato XML parser (o processor).

Cosa fa un XML parser

- Verifica che il documento rispetti la sintassi XML
- Interpreta i dati con tipo PCDATA (*Parsed*)
- Risolve character or entity references
- Gestisce le processing instructions per interpretare i dati

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML body

Un documento XML è composto da elementi e attributi.
Gli elementi sono la base, le unità fondamentali di qualsiasi documento XML.

Elementi: Sintassi

```
<element>content</element>  
opening tag:  <element>;  
closing tag:  </element>
```

Un elemento può contenere testo e/o ulteriori elementi

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Element

Gli elementi XML possono avere diversi tipi di contenuto:

- contenuto strutturale: solo altri elementi, non testo
- contenuto misto: testo e anche altri elementi
- contenuto testuale: solo testo, non altri elementi

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Element: note importanti sul nome

- Gli elementi sono case sensitive.
- Gli elementi possono iniziare con una lettera o con un “_”.
- Un elemento non può iniziare con la stringa *xml*.
- Il tag di apertura e di chiusura devono avere lo stesso nome.
- Un tag può essere usato più di una volta.
- Un insieme di elementi costituiscono un vocabolario

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Element: empty e nested

- Un elemento vuoto (*empty*) è un elemento senza contenuto.
- Un elemento può contenere altri elementi opportunamente annidati (*nested element*).

XML esempi: empty e nested element

- `<element /><element></element>`
- `<choice><sic>testo con errore</sic><corr>
testo corretto</corr></choice>`

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Element: hierarchical relationship

- Un elemento annidato (*nested*) è un elemento *figlio*, cioè contenuto (annidato) in un ulteriore elemento detto padre/genitore (*parent*).
- Gli elementi che sono presenti su uno stesso livello gerarchico (*side by side*) sono detti *sibling element*.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

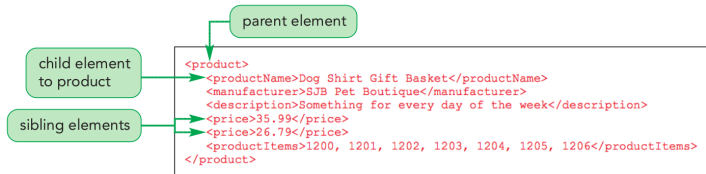


immagine dal libro New Perspectives on XML, 3rd Edition

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Element: hierarchical relationship

- Tutti gli elementi nel body del documento sono figli/discendenti di uno stesso elemento, chiamato radice (*root*).
- Un documento XML deve contenere un elemento root per essere considerato ben formato.
- Una gerarchia XML può essere rappresentata tramite un diagramma ad albero.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Element: hierarchical relationship cont.

- Il prologo e i commenti non fanno parte dell'albero del body.
- Elementi non annidati correttamente implicano un errore di sintassi nei parser.
- Le specifiche XML non consentono di sovrapporre i tag di apertura e di chiusura degli elementi annidati (*no overlap*).

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Element: hierarchical relationship as tree structure

Un modo rapido e comodo per visualizzare la struttura completa di un documento XML è quello di disegnare attraverso un diagramma ad albero ordinato gli elementi del documento XML.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

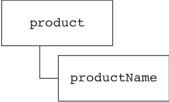
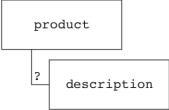
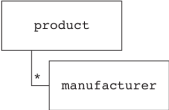
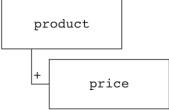
Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Symbol	Description	Chart	Interpretation
[none]	The parent contains a single occurrence of the child element.	 <pre> graph TD product[product] --- productName[productName] </pre>	A product element must contain a single productName element.
?	The parent contains zero or one of the child elements.	 <pre> graph TD product[product] --- ?[?] --- description[description] </pre>	A product element may contain a description element.
*	The parent contains zero or more of the child elements.	 <pre> graph TD product[product] --- *[] --- manufacturer[manufacturer] </pre>	A product element can contain zero or more manufacturer elements.
+	The parent contains at least one of the child elements.	 <pre> graph TD product[product] --- +[+] --- price[price] </pre>	A product element must contain one or more price elements.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

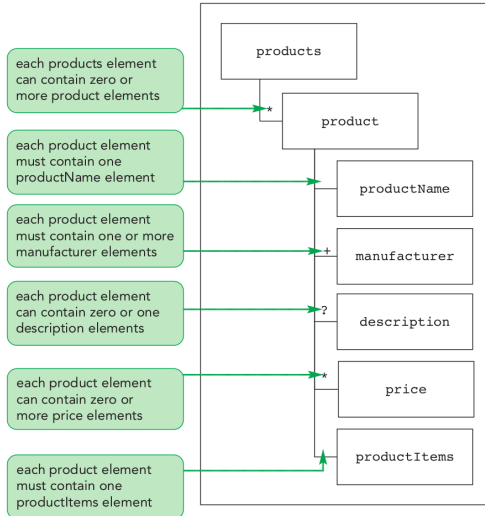
Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML



Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Element: Mixed Content

Un elemento può contenere contemporaneamente sia testo sia altri elementi.

Questo modello di contenuto si chiama Mixed Content ed è ideale per descrivere informazioni text-based (**dati semi-strutturati**).

XML Element: Mixed Content

```
<p><salute>Salve</salute> il mio nome è  
<persName>Angelo</persName></p>
```

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Attributi

Gli elementi in un documento XML possono avere uno o più attributi.

Un attributo descrive una caratteristica dell'elemento in cui appare.

XML Attributi

Un attributo ha senso solo all'interno del proprio elemento e non è possibile separarlo da esso in alcun modo.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Attributi: valore

Un attributo ha due componenti: nome - valore. Il valore di un attributo è una stringa e deve essere sempre racchiusa tra apici (singoli o doppi).

XML Attributi: valore

```
<element attribute='value'> ... </element>
<element attribute='value' />
<element attribute='value',
    attribute2='value2' />
```

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Attributi: restrizioni ai nomi

- Il nome di un attributo può iniziare con una lettera oppure underscore.
- Gli spazi non sono consentiti in un nome di un attributo.
- Il nome di un attributo non può iniziare con la stringa *xml*.

XML Attributi

- Il nome degli attributi è *case sensitive*.
- L'ordine degli attributi non è significativo.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

XML Character and Entity References

- numeric character reference: `&#nnn;`
- hexadecimal character reference: `&#xhhhh`
- character entity reference: `&entity;`

XML References

- `A` (*carattere A*)
- `&` (*carattere &*)

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Symbol	Character Reference	Entity Reference	Description
>	>	>	Greater than
<	<	<	Less than
'		'	Apostrophe (single quote)
"		"	Double quote
&	&	&	Ampersand
©	©	©	Copyright
®	®	®	Registered trademark
™	™		Trademark
°	°		Degree
£	£		Pound
€	€	€	Euro
¥	¥	¥	Yen

immagine dal libro New Perspectives on XML, 3rd Edition

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Text Character Parsing

Il contenuto testuale di un elemento XML può essere diviso in tre categorie: parsed character data, character data, and white space.

Text Character Parsing

- PCDATA
- CDATA
- White Space

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Parsed Character Data

Parsed character data (PCDATA) si riferisce a tutti quei caratteri che XML tratta come parte del codice e quindi vengono interpretati dai parser.

PCDATA

- XML declaration
- Opening tag e closing tag
- Character or entity references
- Commenti

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Parsed Character Data

La presenza di contenuti di tipo PCDATA può causare errori inaspettati.

XML PCDATA

Caratteri speciali che sono utilizzati dalla specifica XML come &, <, > non possono essere utilizzati come contenuto testuale.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Character Data

I dati di tipo “Character Data” non vengono interpretati dal parser XML.

La sequenza di caratteri viene trattata come puro contenuto. In definitiva una sezione *CDATA* è un blocco di testo.

XML CDATA: sintassi

```
<![CDATA [  
character data  
]]>
```

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Character Data

Le sezioni di testo CDATA possono essere inserite in qualsiasi parte del documento XML.

Utile per inserire una sezione di testo con molti caratteri speciali.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

CDATA: qualche vincolo

- Non è possibile inserire commenti in una sezione CDATA.
- Non è possibile annidare sezioni CDATA.
- Non possono essere vuote.
- i simboli “]]” non sono ammessi.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

White Space: esempio

- Gli spazi bianchi sono ignorati quando sono tra i tag.
- Gli spazi bianchi sono ignorati all'interno del prologo e dell'epilogo e all'interno dei tag.
- Gli spazi bianchi inseriti nel valore di un attributo sono trattati come parte del contenuto.
- Non vengono strippati gli spazi all'interno del contenuto testuale degli elementi.

I white space sono caratteri non stampabili

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Processing Instruction

Una *processing instruction* è un comando, una direttiva, che indica al parser XML in che modo elaborare e trattare tutto o parte del documento XML.

Processing Instruction: sintassi

```
<?target instruction ?>  
  
<?xml-stylesheet type="text/css" href="main.css"  
media="all" ?>
```

Molteplici processing instruction possono co-esistere all'interno di un unico documento XML.

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Processing Instruction: sintassi

```
<?target instruction ?>  
  
<?xml-stylesheet type="text/css" href="main.css"  
media="all" ?>
```

Processing Instruction

Target: identifica il tool al quale la processing instruction è diretta.

Instruction: identifica le informazioni che il documento passa al parser per essere elaborate. Le istruzioni hanno la forma degli attributi (nome-valore).

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Namespaces

Un namespace può essere visto come una collezione di elementi e attributi e un insieme di regole che ne determinano la struttura e il contenuto.

Namespaces

```
<element xmlns:prefix="uri"> ... </element>
    <element xmlns="uri"> ... </element>
        <tei:TEI
xmlns:tei='“http://www.tei-c.org/ns/1.0”'>
<TEI xmlns='“http://www.tei-c.org/ns/1.0”'>
```

Fondamenti XML

eXtensible Markup Language

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

Namespaces

Un namespace viene ereditato da tutti gli elementi discendenti dell'elemento in cui esso è stato dichiarato.

Namespaces

Generalmente si dichiarano tutti i namespace nell'elemento root così da avere a disposizione tutti gli elementi dei vari namespace in tutto il documento XML

Progress status

Seminario
XML/TEI -
FeDHLab

A.M. Del
Grosso

Introduzione e
Strumenti di
lavoro

Rappresent.
Digitale dei
Testi

Ecosistema
XML

Fondamenti
XML

Conclusioni
XML

1 Introduzione e Strumenti di lavoro

2 Rappresent. Digitale dei Testi

3 Ecosistema XML

4 Fondamenti XML

5 Conclusioni XML

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 che indicano 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.

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).