

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Presentazione del Corso Codifica di Testi a.a. 2024-2025

Angelo Mario Del Grosso

angelo.delgrossos@ilc.cnr.it

CNR-ILC

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

Outline della lezione

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

1 Presentazione personale

2 Introduzione Corso Codifica di Testi

3 Materiale Didattico e Strumenti di lavoro

4 Progetto Esame

5 Codifica dei Testi

6 Ecosistema XML

Progress status

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

1 Presentazione personale

2 Introduzione Corso Codifica di Testi

3 Materiale Didattico e Strumenti di lavoro

4 Progetto Esame

5 Codifica dei Testi

6 Ecosistema XML

Di cosa mi occupo

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Filologia Digitale e Computazionale

Attività di ricerca scientifica e tecnologica volta allo sviluppo di sistemi computazionali per la filologia digitale finalizzati alla produzione, rappresentazione, analisi, fruizione e interrogazione di testi d'interesse storico e letterario.

Presentazione del Corso

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML



Progress status

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

1 Presentazione personale

2 Introduzione Corso Codifica di Testi

3 Materiale Didattico e Strumenti di lavoro

4 Progetto Esame

5 Codifica dei Testi

6 Ecosistema XML

Introduzione al Corso di Codifica di Testi

Obiettivi, competenze e conoscenze

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Obiettivo

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

Rationale

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).

Argomenti trattati

Obiettivi, competenze e conoscenze

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Alla fine del corso sapremo

- valutare il metodo di codifica più appropriato allo scenario d'interesse
- creare uno schema di codifica conforme alle linee guida TEI
- usare gli strumenti più idonei per la codifica di una risorsa testuale
- elaborare il testo codificato tramite opportune tecnologie
- presentare il testo codificato in ambiente web

Principali Argomenti

Obiettivi, competenze e conoscenze

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

- Codifica dei 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, XQuery)
- Esempi, esercitazioni e seminari

Calendario e Orario del Corso

Mercoledì 12.00-13.30 (Fib M-Lab) e Giovedì 12.00-13.30 (Fib M-LAB)

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025
A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

17 – 21 feb 2025		
mar 18/2	mer 19/2	gio 20/2
	08.30 - 10.00 PROGETTAZIONE E PROGRAMMAZIONE WEB - PROGETTAZIONE WEB (1) Fib PS1	08.30 - 10.00 LETTERATURA ITALIANA CONTEMPORANEA (3) Fib PS1
10:15 - 11:45 BASI DI DATI E LABORATORIO LINGUISTICA LABORATORIO WEB (GENERALE (1) Fib M-Lab	10:15 - 11:45 ISTITUZIONI DI LOGICA (3) Ricci R4 E. MORICONI	10:15 - 11:45 ALGORITMICA (2) Fib PS1 F. LEVI
12:00 - 13:30 ALGORITMICA (2) Fib PS1 F. LEVI	12:00 - 13:30 CODIFICA DI TESTI (3) Fib M-Lab A. DEL GROSSO	12:00 - 13:30 BASI DI DATI E CODIFICA DI TESTI LABORATORIO (Fib M-Lab inglese LABORATORIO A. DEL GROSSO (Leone) per
14:15 - 15:45 GEOGRAFIA CULTURALE (1) Fib PS1 A. ROMANO	14:15 - 15:45 PROGETTAZIONE E PROGRAMMAZIONE WEB - PROGRAMMAZIONE (1) Fib M-Lab, Fib I-Lab	14:15 - 15:45 LETTERATURA ITALIANA (2) Fib PS1 M. RICCUCCI
16:00 - 17:30 LETTERATURA ITALIANA (2) Fib PS1 M. RICCUCCI	16:00 - 17:30 LINGUA INGLESE E COMUNICAZIONE MULTIMEDIALE (1) Fib PS1	16:00 - 19:00 PROGETTAZIONI, ISTITUZIONI E LINGUA GRAFICA (3) (3) Fib H-Lab Ricci R1 B. RAPISARDA
		16:00 - 17:30 LATINA (2) Ricci R3



Bozza piano del Corso: Calendario 2024/2025

Mercoledì 12.00-13.30 (Fib M-Lab) e Giovedì 12.00-13.30 (Fib M-LAB)

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

- Introduzione al corso e strumenti software utilizzati
- Elementi teorici relativi alla rappresentazione digitale del testo
- Introduzione ai Linguaggi di Markup e fondamenti di XML
- Introduzione alla definizione degli Schemi XML (DTD, XSD)
- Elementi di Codifica XML-TEI

Bozza piano del Corso: Calendario 2024/2025

Mercoledì 12.00-13.30 (Fib M-Lab) e Giovedì 12.00-13.30 (Fib M-LAB)

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

- Moduli Editoriali TEI
- Introduzione ai Fogli di Stile XSL
- Elaborazione documenti XML-TEI (XPath, XSLT e XQuery)
- Approfondimenti Codifica TEI ed Elaborazione XPath, XSLT, XQuery
- Esercitazione e discussione sui progetti d'esame

Progress status

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

1 Presentazione personale

2 Introduzione Corso Codifica di Testi

3 Materiale Didattico e Strumenti di lavoro

4 Progetto Esame

5 Codifica dei Testi

6 Ecosistema XML

Materiale Didattico e strumenti di lavoro

Studio e approfondimento

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Studio

Sono disponibili numerose risorse, anche gratuite e online, per studiare ed approfondire i principali temi trattati dal corso

Pratica e strumenti di lavoro

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

Materiale Didattico

Studio e approfondimento

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Supporto allo studio XML-TEI

- Slide delle lezioni (Moodle, GitHub)
- Libri Burnard, Pierazzo e Ciotti
- Linee guida TEI (online)
- Riviste scientifiche (jTEI, UD, Balisage)

- Libri tecnologie XML
- Risorse online e specifiche ufficiali (W3C)
- Esempi ed esercitazioni

Slide delle lezioni

Moodle, GitHub

Presentazione
del Corso
Codifica di

Testi

a.a.

2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

A.M. Del Grosso

Presentazione del Corso Codifica di Testi a.a. 2024-2025

15 / 73

The screenshot shows a Moodle course page. At the top, there's a navigation bar with links for 'Miei corsi', 'Accesso', 'Assistenza', 'Attivazione corsi', and a language switcher. A user profile for 'ANGELO MARIO DEL GROSSO' is also at the top right. Below the header, the course title 'Codifica di Testi (IU, LINTRA, 2020/21)' is displayed, along with the breadcrumb trail 'Home / I miei corsi / DISCIPLINE UMANISTICHE - Corsi ordinari / AA 2020 - 2021 / Il semestre / 299IU_20'. On the right, there's a button labeled 'Attiva modifica'. The main content area has a sidebar titled 'Navigazione' containing a tree view of course sections: Home, Dashboard, E-LEARNING AREA UMANISTICA, I miei corsi, DISCIPLINE UMANISTICHE (with sub-sections for AA 2020-2021, Il semestre, and 299IU_20), and various activity modules like Bacheca, Programma d'esame, and Risultati delle prove in itinere e degli esami. A note about access is present: 'Accesso visualizzato Accesso consentito fino al 30 aprile 2023, 23:55 (Altrimenti nascosto)'. To the right of the sidebar, there are three boxes: 'Ricerca nei forum' with a search field and 'Val' button, 'Annunci recenti' with a note about no recent announcements, and 'Prossimi eventi' with a note about no upcoming events. The bottom of the page features standard Moodle navigation icons.

Slide delle lezioni

Moodle, GitHub

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

The screenshot shows a GitHub repository page for 'angelodel80/corsoCodifica'. The repository is public and contains one branch, 'master'. The commit history shows four commits from 'angelodel80' adding README.md files to subdirectories 'progetto-esame', 'slide', and 'src'. A fifth commit from 'angelodel80' adds a README.md file to the root directory, starting the course description.

Commit	Message	Date
angelodel80 add first README.md files in subdirs	now	fbdb5949 now
..		
progetto-esame	add first README.md files in subdirs	now
slide	add first README.md files in subdirs	now
src	add first README.md files in subdirs	now
README.md	start CdT aa 22-23	2 hours ago

The README.md file content is:

```
Corso Codifica di Testi (299LL, 6 CFU) aa 2022 - 2023
```

<https://github.com/angelodel80/corsoCodifica>

Materiale di studio

Linee Guida della TEI - Online

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

A.M. Del Grosso

Presentazione del Corso Codifica di Testi a.a. 2024-2025

17 / 73

The screenshot shows the homepage of the Text Encoding Initiative (TEI) at <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>. The page title is "P5: Guidelines for Electronic Text Encoding and Interchange". The main content area includes sections for "Front Matter", "Text Body", and "TEI sourcecode", each listing various chapters and appendices. The "Text Body" section lists chapters 1 through 18. The "Front Matter" section lists chapters I through VI. The "Back Matter" section lists appendices A through H. The "TEI sourcecode" section lists links to GitHub, bug reports, and feature requests.

P5: Guidelines for Electronic Text Encoding and Interchange

Version 4.1.0. Last updated on 19th August 2020, revision b414be550

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

[Front Matter](#)

- 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 [Dictionaries](#)
- 10 [Manuscript Description](#)
- 11 [Representation of Primary Sources](#)
- 12 [Critical Apparatus](#)
- 13 [Names, Dates, People, and Places](#)
- 14 [Tables, Formulae, Graphics and Notated Music](#)
- 15 [Language Corpora](#)
- 16 [Linking, Segmentation, and Alignment](#)
- 17 [Simple Analytic Mechanisms](#)
- 18 [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

Open Edition Book - Lou Burnard

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

The screenshot shows the OpenEdition Press website. At the top, there's a logo for 'OpenEdition press' and the text 'OpenEdition Press'. On the right side of the header, there are links for 'HOME', 'ENCYCLOPÉDIE NUMÉRIQUE', 'READ/WRITE BOOK', 'SCIENCES PO | MÉDIALAB', and 'BRÉSIL / FRANCE | BRASIL / FRANÇA'. Below the header, there's a large image of the book cover for 'WHAT IS THE TEXT ENCODING INITIATIVE?'. The book cover has a colorful, abstract background with the title in white. To the left of the book image is a search bar with the placeholder 'CERCA NEL LIBRO' and an 'OK' button. Below the search bar are buttons for 'INDICE DEI CONTENUTI', 'CITA', 'CONDIVIDI', and 'CITATO DA'. To the right of the book image, the title 'WHAT IS THE TEXT ENCODING INITIATIVE?' is displayed in large bold letters, followed by a subtitle 'How to add intelligent markup to digital resources'. The author 'Lou Burnard' is mentioned, along with the publisher 'Encyclopédie numérique'. There are three ellipsis dots, followed by a detailed description of the Text Encoding Initiative (TEI) Guidelines. A call-to-action button 'Leggi il seguito' is shown. At the bottom, there's a table with book details: Editore: OpenEdition Press; Collana: Encyclopédie numérique | 3; Luogo di pubblicazione: Marseille; Anno di pubblicazione: 2014; Pubblicato su OpenEdition Books: 11 avril 2014; EAN (edizione cartacea): 9782821834590; EAN digitale: 9782821834606; DOI: 10.4000/books.oep.426; Numero di pagine: 114 p. On the far right, there's a sidebar with icons for 'LEGGI' (reading mode), 'ACCESSO APERTO' (open access), 'LETTURA CONFORTEVOLE' (comfortable reading), 'EPUB' (EPUB format), and 'PDF DEL LIBRO' (PDF of the book).

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



Materiale di studio

Libro Digital Scholarling Editing - E. Pierazzo

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

A.M. Del Grosso

Presentazione del Corso Codifica di Testi a.a. 2024-2025

19 / 73

The screenshot shows the HAL (archives-ouvertes.fr) website interface. At the top left is the HAL logo. To its right is a stylized illustration of a figure lying on a globe, looking up at floating geometric shapes. The main navigation menu includes Accueil, Dépot, Consultation, Recherche, and Documentation. Below the menu, a red button indicates the document is a 'Pré-publication, Document de travail'. The title of the document is 'Digital Scholarly Editing: Theories, Models and Methods' by Elena Pierazzo. A note below the title specifies it is a 'GERCi - Groupe d'étude et de recherche sur la culture italienne' document. The abstract discusses the book's treatment of digital scholarly editing across the timeline from choice of object to publication. The document is categorized under 'Digital Humanities', 'Philology', 'Digital Editing', and 'Scholarly editing'. It is identified as a 'Pré-publication, Document de travail'. The domain is listed as 'Sciences de l'Homme et Société / Littératures', 'Sciences de l'Homme et Société / Héritage culturel et muséologie', and 'Sciences de l'Homme et Société / Sciences de l'information et de la communication'. A link to a complete metadata list is provided. On the right side of the page, there are several sidebar sections: 'FICHIER' (containing the PDF file 'DSE-revised-29Sept2014.pdf' and information about the author(s)), 'LICENCE' (mentioning 'Copyright (Tous droits réservés)'), 'IDENTIFIANTS' (listing the HAL ID 'hal-01182162, version 1'), 'COLLECTIONS' (listing 'GERCi | UGA'), and 'CITATION' (providing the citation details: 'Elena Pierazzo. Digital Scholarly Editing: Theories, Models and Methods. 2014. (hal-01182162)').

<https://hal.archives-ouvertes.fr/hal-01182162>

Materiale di studio

Rivista jTEI

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

The screenshot shows the homepage of the **JOURNAL OF THE TEXT ENCODING INITIATIVE**. At the top, there is a navigation bar with links to **HOME**, **CATALOGUE OF 554 JOURNALS**, and **OPENEDITION SEARCH**. On the right, there are links for **All OpenEdition** and a share icon. The main content area features the **jTEI** logo and the journal's name. Below this, there is a sidebar with a search bar, an index section listing **Authors** and **Keywords**, and a section for **FULL TEXT ISSUES** with links to **Issue 11 | July 2019 - June 2020 Selected Papers from the 2016 TEI Conference** and **Issue 12 | July 2019 - May 2020 Selected Papers from the 2017 TEI Conference**. The main text area describes the journal as the official journal of the Text Encoding Initiative Consortium, publishing proceedings of the annual TEI Conference and Member's Meeting, and providing a forum for articles on the discussion of the interface between the TEI and other communities.

<https://journals.openedition.org/jtei/>

Materiale di studio

Ulteriori risorse

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

- Ciotti F. (2007). Il testo e l'automa. Saggi di teoria e critica computazionale dei testi letterari. Aracne.
- Driscoll, M. J., and Pierazzo, E. (Eds.). (2016). Digital Scholarly Editing: Theories and Practices. Open Book.
- Lenci, A., Montemagni S., Pirrelli V. (2016). Testo e Computer. Elementi Di Linguistica Computazionale. Aulamagna. Carocci.
- Mancinelli, T., Elena Pierazzo. (2020). Che cos'è un'edizione scientifica digitale. Roma: Carocci.
- Orlando, T. (2010). Informatica testuale: teoria e prassi. Laterza.

Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

<https://www.w3.org/standards/xml/>

XML TECHNOLOGY

On this page → technology topics • news • upcoming events and talks



XML Technologies including XML, XML Namespaces, XML Schema, XSLT, Efficient XML Interchange (EXI), and other related standards.

XML Essentials

XML is shouldered by a set of essential technologies such as the infoset and namespaces. They address issues when using XML in specific applications contexts.

Efficient Interchange

XML standards are omnipresent in enterprise computing, and are part of the foundation of the Web. Because the standards are highly interoperable and affordable, people have wanted to use them in a wide variety of applications. However, in some settings (on devices with low memory or low bandwidth, or where performance is critical) experience has shown that a more efficient form of XML is required.

Schema

Formal descriptions of vocabularies create flexibility in authoring environments and quality control chains. W3C's XML Schema, SML, and data binding technologies provide the tools for quality control of XML data.

Security

Manipulating data with XML requires sometimes integrity, authentication and privacy. XML signature, encryption, and xkms can help create a secure environment for XML.

Transformation

Very frequently one wants to transform XML content into other formats (including other XML formats). XSLT and XPath are very powerful tools for creating different representations of XML content.

Query

XQuery (supported by XPath) is a query language for XML to extract data, similar to the role of SQL for databases, or SPARQL for the Semantic Web.

Components

The XML ecosystem is using additional tools to create a richer environment for using and manipulating XML documents. These components include style sheets, xlink xml:id, xinclude, xpointer, xforms, xml fragments, and events.

Processing

A processing model defines what operations should be performed in what order on an XML document.

Internationalization

W3C has worked with the community on the Internationalization of XML, for instance for specifying the language of XML content.

Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

XML.com

Home Articles Authors News Job Board About Search

News

Articles

Foundations

Popular tags: articles

XFORMS XSLT VOCABULARY

UBL XPROC NISO RELAX-NG

XSD CONFERENCE XQUERY

Popular tags: news

XML XSLT DITA DOCBOOK

XQUERY XHTML DOCK

XMLMIND HTML EPUB

Announcing the XML slack community

Markup UK 2021 Conference Dates Announced

Markup UK is pleased to announce that the next edition of the conference will take ...

18 December 2020 [Read](#)

Release of XMLmind Word To XML v1.8.3

Release of XMLmind Word To XML v1.8.3 (<http://www.xmlmind.com/w2x/changes.html>)

Enhancements:

- Upgraded XMLmind Web Help ...

2 December 2020 [Read](#)

3 Useful Things DITA XML Can Do for Your Content

Tim Ludwig
You know what DITA is, but what can you actually do with it?

29 December 2020 [Read](#)

XProc 3.0 - Strategies for merging documents

Erik Siegel
Erik Siegel continues his series on XProc 3.0 with an analysis of potential strategies for merging documents.

16 November 2020 [Read](#)

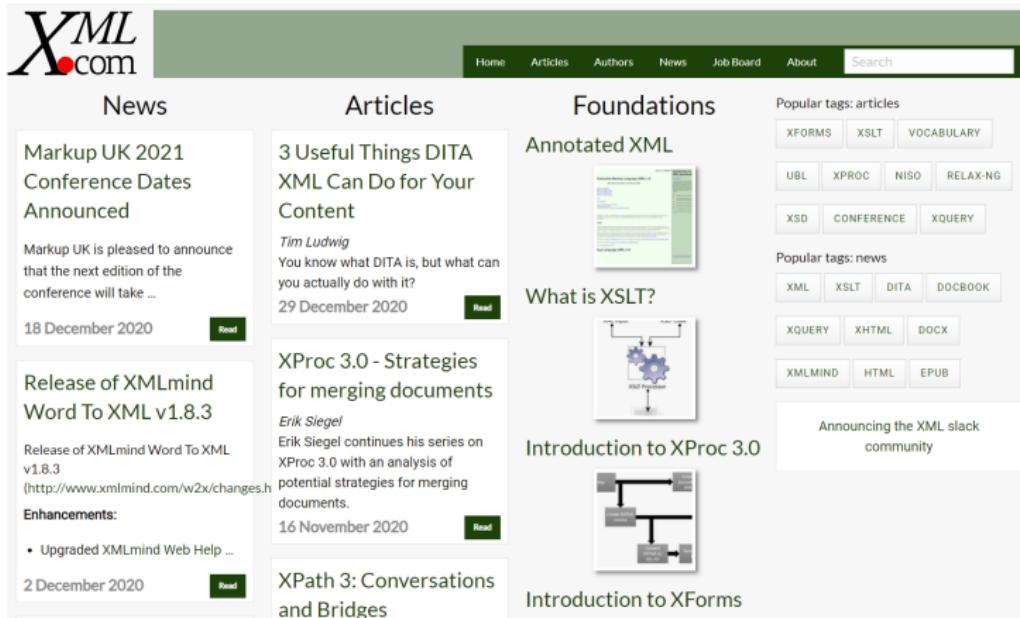
XPath 3: Conversations and Bridges

Annotated XML

What is XSLT?

Introduction to XProc 3.0

Introduction to XForms



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



Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

The screenshot shows a web browser displaying the Saxonica documentation. The URL in the address bar is <https://www.saxonica.com/documentation/>. The page title is "Reference: XSLT Elements". The left sidebar contains a navigation tree with categories like "Using XML Schema (XSD)", "The Gizmo Utility", "Saxon on .NET", "Extensibility Mechanisms", "Saxon Extensions", "Sample Saxon Applications", "Saxon SQL Extension", and "Reference: XSLT Elements". Under "Reference: XSLT Elements", there is a list of XSLT elements: xsl:accept, xsl:accumulator, xsl:accumulator-rule, xsl:analyze-string, xsl:apply-imports, xsl:apply-templates, xsl:assert, xsl:attribute, xsl:attribute-set, xsl:break, xsl:call-template, xsl:catch, xsl:character-map, xsl:choose, xsl:comment, xsl:context-item, xsl:copy, xsl:copy-of, xsl:decimal-format, xsl:document, and xsl:element. The main content area starts with a section titled "Reference: XSLT Elements" which describes the standard XSLT elements. It mentions the W3C XSLT 3.0 Recommendation and provides links to the official standard and extension instructions. Below this is a "Conformance" section, followed by "XSLT Patterns" and "Standard attributes". A "Section index" at the bottom lists the XSLT elements again.

<https://www.saxonica.com/documentation/>

Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

mdn web docs References Guides MDN Plus

Theme Log in Get MDN Plus

English (US)

References > XSLT: Extensible Stylesheet Language Transformations

XSLT

EXSLT

XPath

Guides

▼ XSLT Elements

- <xsl:apply-imports>
- <xsl:apply-templates>
- <xsl:attribute-set>
- <xsl:attribute>
- <xsl:call-template>
- <xsl:choose>
- <xsl:comment>
- <xsl:copy-of>
- <xsl:copy>
- <xsl:decimal-format>
- <xsl:element>
- <xsl:fallback>
- <xsl:for-each>
- <xsl:if>
- <xsl:import>
- <xsl:include>

XSLT: Extensible Stylesheet Language Transformations

XSLT/XPath Reference: XSLT elements, EXSLT functions, XPath functions, XPath axes

Extensible Stylesheet Language Transformations (XSLT) is an [XML](#)-based language used, in conjunction with specialized processing software, for the transformation of XML documents.

Although the process is referred to as "transformation," the original document is not changed; rather, a new XML document is created based on the content of an existing document. Then, the new document may be serialized (output) by the processor in standard XML syntax or in another format, such as [HTML](#) or plain text.

XSLT is most often used to convert data between different XML schemas or to convert XML data into web pages or PDF documents.

Documentation

[XSLT Element Reference](#)

Reference.

[Transforming XML with XSLT](#)

XSLT allows a stylesheet author to transform a primary XML document in two significant ways: manipulating and sorting the content, including a wholesale reordering of it if so desired, and transforming the content into a different format.



Modernize your database with MongoDB Atlas, the leading developer data platform.

More about

Don't want to see ads?

<https://developer.mozilla.org/en-US/docs/Web/XSLT>

Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

A.M. Del Grosso



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

Materiale di studio

Ulteriori risorse

- 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.
- Kelly, D. J. (2015). *XSLT Jumpstarter: Level the Learning Curve and Put Your XML to Work*. Pragmatic Bookshelf.
- XML Guild Team. (2006). *Advanced XML Applications from the Experts at The XML Guild*. Course Technology Press.
- Fawcett, J., Ayers D., Quin L. R. E. (2012). *Beginning XML 5th Edition*. Wrox.

Strumenti di lavoro

requisiti di base

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

- 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 shell da terminale per lavorare sui documenti XML

Strumenti di lavoro

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Overview

SETUP

GET STARTED

Intro Videos

Tips and Tricks

User Interface

Themes

Settings

Key Bindings

Display Language

USER GUIDE

LANGUAGES

NODEJS /
JAVASCRIPT

PYTHON

JAVA

AZURE

EXTENSION
AUTHORING

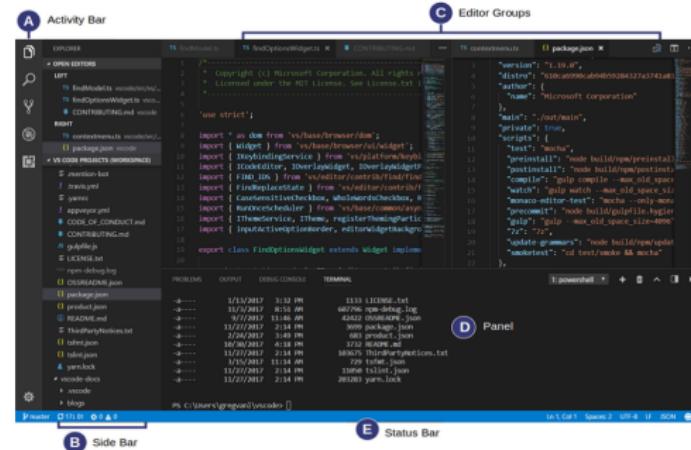
EXTENSIBILITY
REFERENCE

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

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



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
- [Tweet](#)
- [Subscribe](#)
- [Ask questions](#)
- [Follow @code](#)
- [Request features](#)
- [Report issues](#)
- [Watch videos](#)

Strumenti di lavoro

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

<http://www.editix.com>

The screenshot shows the Editix XML Editor 2015 interface. The main window displays an XSLT document named "personal.xsl". The code is as follows:

```
<stylesheet template="person" h1 value-of="id"/>
<xsl:output method="html"/>
<xsl:template match="person">
  <h1>
    <xsl:value-of select="@id"/>
  </h1>
</xsl:template>
</xsl:stylesheet>
```

The left pane shows the XML Data source with a tree view of "personnel" nodes. The right pane shows the XSLT Result Preview in HTML format, displaying "one.worker" and "two.worker". A status bar at the bottom indicates "10:12" and "Num".

Editor XML

Cosa è consigliato usare

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

- 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

Strumenti di sviluppo

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML



The image shows the Java download page. At the top, there is a red header with the Java logo (a steaming coffee cup) and the word "Java". Below the header, there are "Download" and "Help" buttons. The main title "JAVA + YOU, DOWNLOAD TODAY!" is displayed prominently in large, bold, grey font. Below the title is a red button labeled "Java Download". Underneath the button, there are links for "What is Java?", "Need Help?", and "Uninstall". To the right of the text, there is a photograph of a smiling woman with curly hair lying on a white surface, looking at a silver laptop. At the bottom, there is a horizontal row of six icons representing different Java-related resources: "About Java", "Go Java", "Alice", "Java + Alice", "Greenfoot", "Java + Greenfoot", "JavaOne", "Java Developer Conference", "Oracle Academy", "Java Magazine", and "Java Get it now".

<https://www.java.com/en/download/manual.jsp>

Strumenti di sviluppo

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

A.M. Del Grosso

Presentazione del Corso Codifica di Testi a.a. 2024-2025

33 / 73

The screenshot shows the Apache Xerces™ Project website. At the top, there's a logo featuring a colorful butterfly above the word "Xerces" in blue, with "apache" in smaller letters to the right. Below the logo is a navigation bar with links for "Home", "Wiki", and "Mailing List Archives". The "Home" link is currently selected. A sidebar on the left contains a navigation tree:

- Xerces
 - Overview
 - Charter
 - News
 - Downloads
 - Get Involved
 - Who We Are
 - License
 - Legal Stuff
 - Mailing Lists
 - Code Repositories
 - Reference
 - Security
 - Sponsorship
 - Thanks
- Local Projects
- Hibernated Projects
- Related Projects

Below the sidebar, a banner says "Built with Apache Forrest". To the right of the sidebar, a breadcrumb trail shows the path: apache.org > XML Federation > xerces.apache.org >. The main content area has a header "The Apache Xerces™ Project - xerces.apache.org". It includes a welcome message, a section on the project's history ("Over a Decade at the Apache Software Foundation"), and a section on current projects. At the bottom of the page is a banner for "10 YEARS OF The ASF" with the subtitle "Celebrating a Decade of Open Source Leadership" and a link to "http://www.apache.org".

<http://xerces.apache.org/index.html>

Strumenti di sviluppo

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

SAXONICA
XSLT AND XQUERY PROCESSING

Our Products Technology Download Buy Documentation Support About us Site map

SAXONICA > Download > Java

Download | Java | .NET | C/C++, PHP & Python | JavaScript | Evaluation License | Downloads Info

Java Platform Downloads

To use Saxon-PE or Saxon-EE you need a license key. If you don't have a current license you can request a 30-day evaluation license.

Saxon-HE is open source software and can be downloaded from SourceForge or Maven.

Use the buttons below to download Saxon-PE and Saxon-EE directly. The Saxon-HE buttons redirect to SourceForge.

Java	Saxon-EE	Saxon-PE	Saxon-HE
	10.3 (19.2Mb)	10.3 (17.0Mb)	10 (SourceForge)
	9.9.1.8 (19.8Mb)	9.9.1.8 (17.6Mb)	9.9 (SourceForge)

The latest release for Java is Saxon 10.3, released on 28 October 2020. **Saxon 10 is now considered the most stable and reliable release.** For more information see [Latest Releases](#).

The latest maintenance release of 9.9 is Saxon 9.9.1.8, released on 22 October 2020.

On the download info page:

- Release Notes
- Installing the Software
- Earlier releases

See also:

- Download .NET Saxon products
- Download C/C++, PHP and Python Saxon/C products
- Download saxon-resources

Strumenti di sviluppo

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

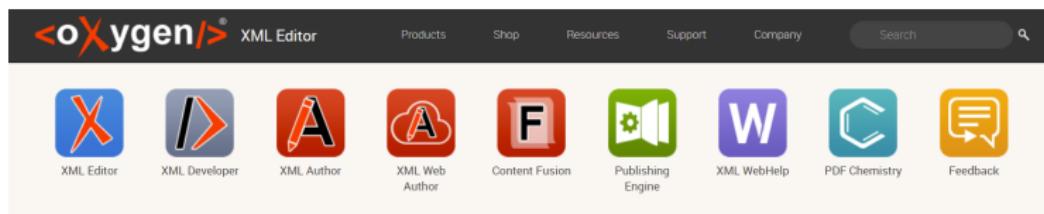
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
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 plugin. 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

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

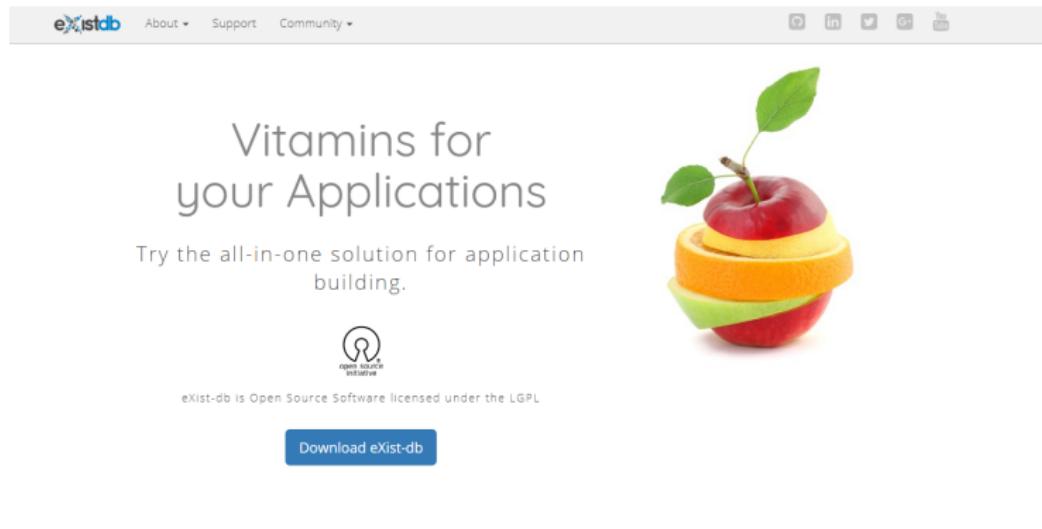
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

The screenshot shows the official website for eXist-db. At the top, there's a navigation bar with the eXist-db logo, 'About', 'Support', 'Community', and social media links. The main heading 'Vitamins for your Applications' is displayed prominently. Below it, a sub-headline reads 'Try the all-in-one solution for application building.' A small circular icon with a stylized 'R' is shown, followed by the text 'eXist-db is Open Source Software licensed under the LGPL'. A blue button labeled 'Download eXist-db' is visible. In the bottom right corner, there's a thumbnail image of a presentation slide titled 'July 29 2014' and 'The 3 amigos'.

A red box highlights the 'Latest News' section. It features a news item from February 05, 2023, about the release of eXist-db 6.2.0.

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

Progress status

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

1 Presentazione personale

2 Introduzione Corso Codifica di Testi

3 Materiale Didattico e Strumenti di lavoro

4 Progetto Esame

5 Codifica dei Testi

6 Ecosistema XML

Programma d'esame

Cosa bisogna fare per superare l'esame

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

- 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
 - *seguiranno indicazioni più dettagliate!*

Modalità d'esame

In cosa consiste l'esame

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

- 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)

Tirocini, Tesi

Pronti a far parte della comunità?

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

The screenshot displays the CoPhiEditor application window with several open panels:

- Paleographic Apparatus**: Shows the original manuscript image at 950mm resolution, with a red box highlighting a specific area of text.
- Diplomatic Transcription**: Shows the diplomatic transcription of the highlighted text from the manuscript image.
- Literary Transcription**: Shows the literary transcription of the same text.
- Palaeographical Apparatus**: Shows the palaeographical transcription of the text.
- Translation**: Shows the translation of the text into modern Italian.
- Infrared image at 950mm**: Shows the infrared image of the manuscript.
- Copper engraving for the Collectio aliorum**: Shows a copper engraving of the manuscript text.
- Transcription by S. Crifo**: Shows a handwritten transcription of the text by S. Crifo.
- Notes**: Shows a panel with various notes and annotations.
- Normalised paragraph**: Shows a panel with a normalised version of the text.

Laboratorio CNR-ILC CoPhiLab

Tirocini, Tesi, EVT

Pronti a far parte della comunità?

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

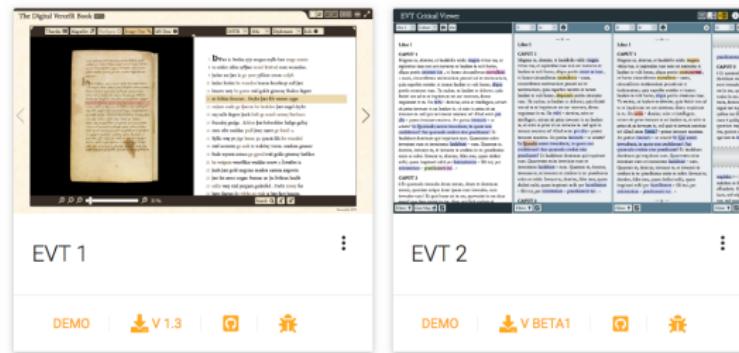
Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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.



EVT tool <http://evt.labcd.unipi.it/>

Progress status

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

1 Presentazione personale

2 Introduzione Corso Codifica di Testi

3 Materiale Didattico e Strumenti di lavoro

4 Progetto Esame

5 Codifica dei Testi

6 Ecosistema XML

Perché è importante la codifica dei testi

Motivazioni pratiche

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Dove troviamo i testi

Nella nostra cultura tradizionale la quasi totalità dei testi è “registrata”, “trasmessa” e quindi “conservata” mediante supporti e materiali fisici di varia natura e forma (iscrizioni su pietra, manoscritti di pergamena, papiri, carta, libri a stampa, incunabula, cinquecentine, etc).

Perché è importante la codifica dei testi

Dove troviamo i testi

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

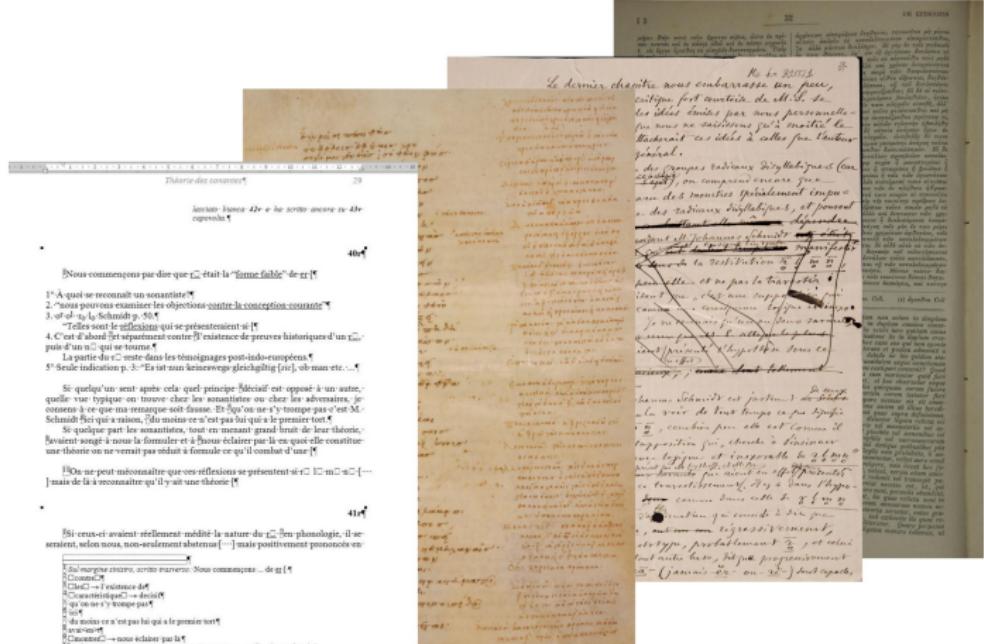
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML



Perché è importante la codifica dei testi

Dove troviamo i testi

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

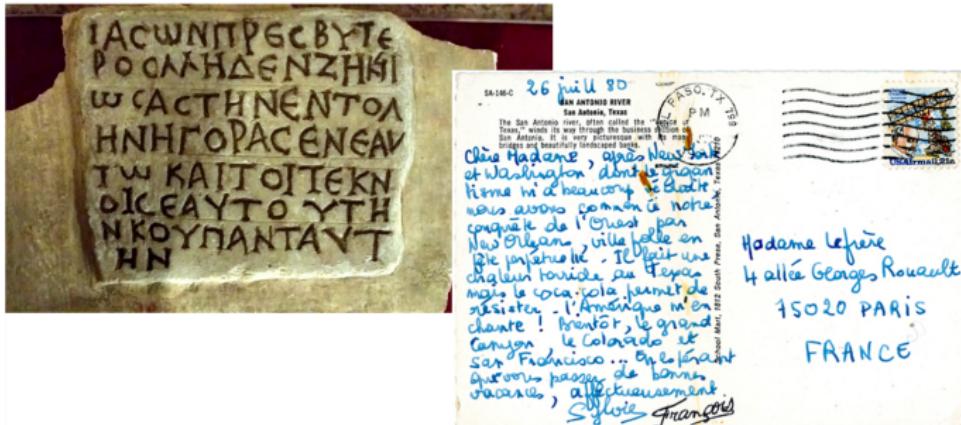
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML



Perché è importante la codifica dei testi

Dove troviamo i testi

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

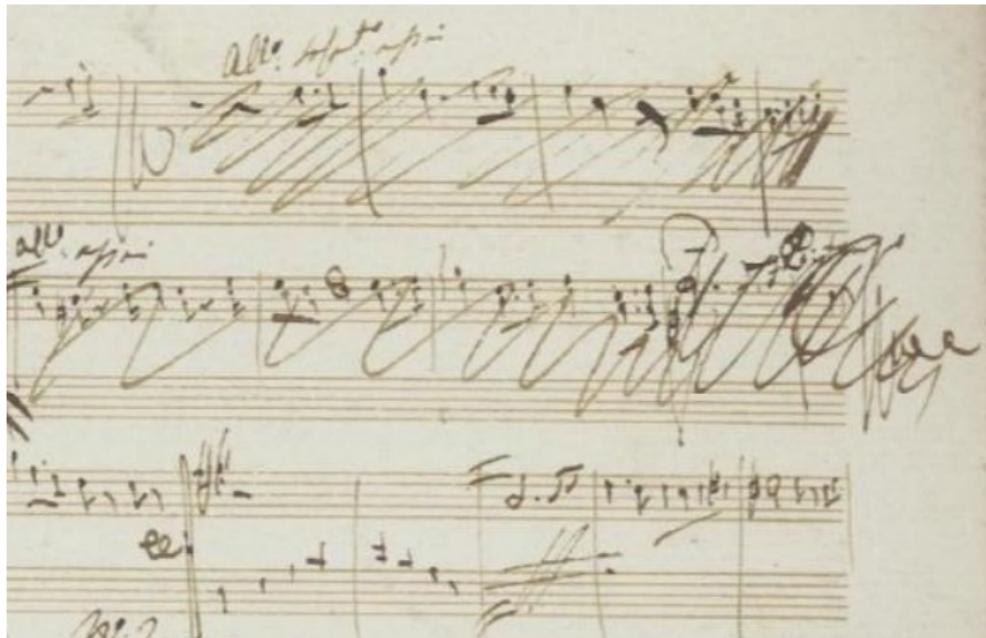
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML



Perché è importante la codifica dei testi

Dove troviamo i testi

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

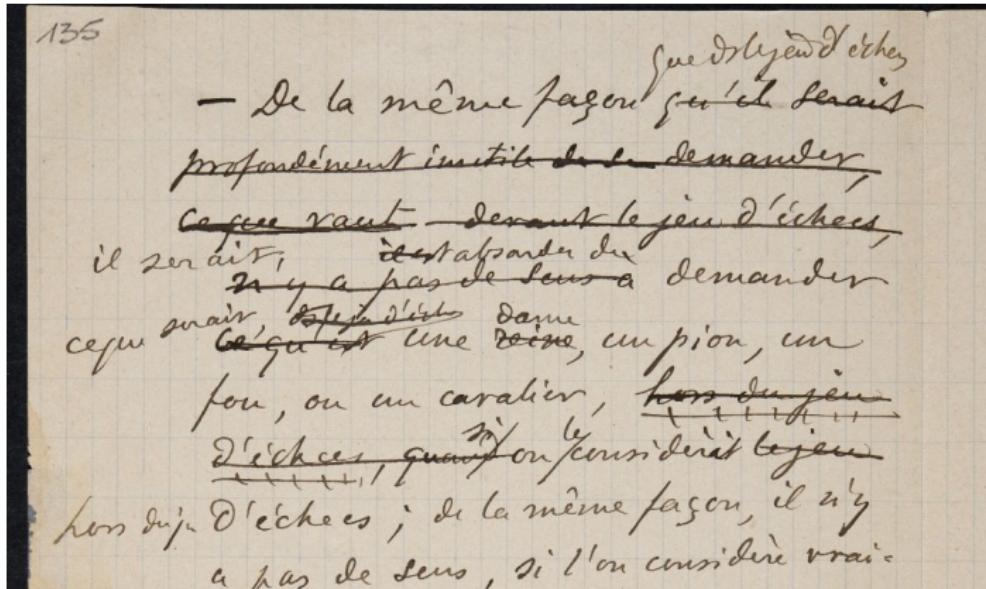
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML



Perché è importante la codifica dei testi

Motivazioni pratiche

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Perché codificare i testi

Per rendere disponibile questo patrimonio attraverso i sistemi per la gestione dell'informazione digitali e computazionali è necessario effettuare una trasposizione/transcodifica* dei testi dal loro supporto originario verso il nuovo supporto elettronico (*Machine Readable Form*).

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

Perché è importante la codifica dei testi

Motivazioni teoriche

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Scienze umane vs Scienze informatiche

Il rapporto tra sapere umanistico e informatica non è solo una questione meramente strumentale:

l'informatica non è solo un utensile dalle notevoli capacità.

Salto di paradigma sia da un punto di vista teorico sia
metodologico.

Perché è importante la codifica dei testi

Motivazioni teoriche

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

La codifica come metodologia

L'attività di codifica diviene metodo di analisi della conoscenza nell'ambito delle discipline che si occupano del testo.

La codifica come analisi teorica

Il linguaggio di codifica adottato può essere considerato come un linguaggio teorico:

Esplicitare e formalizzare le ipotesi interpretative su un certo oggetto di studio

Perché è importante la codifica dei testi

In sintesi

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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.

Rappresentazione digitale dei testi

basso e alto livello di codifica

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Rappresentare un testo

La rappresentazione digitale di un testo è una operazione intrinsecamente assai difficile perché coinvolge una pletora 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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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 (struttura gerarchica)
- Non è rappresentabile solo come un grafo o come un insieme di sotto grafi (struttura reticolare)

Molteplici modelli per diverse esigenze

Strutture dato e testo

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
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

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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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 NN phenomenon
has VBN have
been VBN be
measured VBN measure
,

but CC but
the DT the
data NN data
is VBN be
corrupted VBN corrupt
by IN by
errors NNS error
.

```
<? s_id="1"?>
<x surface="Cathy" lemma="Cathy" pos="N" syn="su:2" />
<x surface="zag" lemma="zile" pos="V" syn="ROOT:0" />
<x surface="hen" lemma="hen" pos="Pron" syn="obj1:2" />
<x surface="wild" lemma="wild" pos="Adj" syn="mod:5" />
<x surface="zwaaien" lemma="zwaai" pos="N" syn="vc:2" />
<x 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_shai": "58bad08927902ff9307b621c54716dcc5083e339"
  },
  "selector": {
    "type": "TextPositionSelector",
    "start": 15,
    "end": 25
  }
}
```

<<http://example1.com/citingwork>>
cito:cites <<http://example2.com/citedwork>>;
cito:inTextCitationFrequency [
 a cito:IntTextCitationCount ;
 cito:inTextCountValue "10"^^xsd:integer ;

Elementi di Codifica del testo

Formalismi

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Formati come formalismi

Data l'importanza metodologica accanto ad un formato di rappresentazione del dato testuale, abbiamo la necessità di adottare un vero e proprio formalismo. Si parla cioè di linguaggi di codifica basati su un insieme di istruzioni rigorose definite da grammatiche formali.

Elementi di Codifica del testo

Formalismi

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Formati e formalismi di codifica

Quindi ogni pezzo di informazione aggiunta ad un testo grezzo attraverso l'inserimento di dati metatestuali (markup, annotazione, codifica), constituisce 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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

1 Presentazione personale

2 Introduzione Corso Codifica di Testi

3 Materiale Didattico e Strumenti di lavoro

4 Progetto Esame

5 Codifica dei Testi

6 Ecosistema XML

Codifica ad alto livello

Sistemi di marcatura

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
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 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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

```
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.

Impiego di XML

Benefici

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
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

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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
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

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
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.

XML come linguaggio per la codifica di testi

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

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.

References

Presentazione
del Corso
Codifica di
Testi
a.a.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML



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.
2024-2025

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML



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/xmllintro/xmllintro.html>



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