

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021
A.M. Del
Grosso

Presentazione del Corso Codifica di Testi a.a. 2020-2021

Angelo Mario Del Grosso

angelo.delgrosso@ilc.cnr.it

CNR-ILC

Istituto di Linguistica Computazionale “A. Zampolli”,
22nd February 2021

Piano della presentazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

1 Presentazione

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.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

1 Presentazione

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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Filologia Digitale e Computazionale

Attività di ricerca scientifica e tecnologica volta allo sviluppo di sistemi di linguistica e filologia digitale e computazionale finalizzati alla produzione, rappresentazione, analisi, fruizione e interrogazione di testi di tradizione medievale, a stampa e di autori moderni e contemporanei.

Presentazione del Corso

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

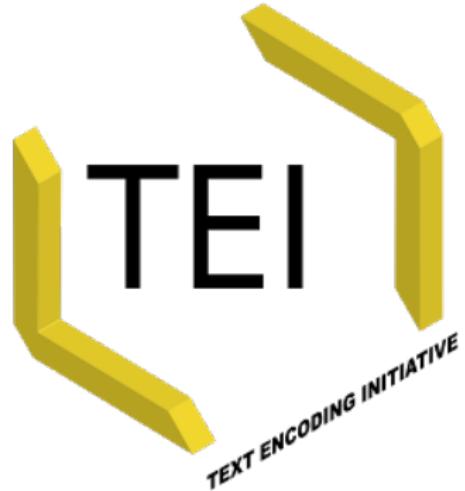
Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia



Progress status

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

1 Presentazione

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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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 criticamente una codifica digitale di testi complessi, in particolare testi letterari e di interesse storico-culturale, usando le linee guida della Text Encoding Initiative (TEI).

Argomenti trattati

Obiettivi, competenze e conoscenze

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Cosa ci aspettiamo alla fine del corso

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

Principali Argomenti

Obiettivi, competenze e conoscenze

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

- 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, DOM)
- Esempi, esercitazioni e seminari

Calendario e Orario del Corso

Lunedì dalle 14.15 alle 17.30

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

| | | | |
|-------------------|---|----------------------------|--|
| Codifica di testi | 6 | ANGELO MARIO DEL GROSSO | 14:15-15:45 Aula IFU/3 16:00-17:30 Aula IFU/3 |
|-------------------|---|----------------------------|--|

Bozza piano del Corso: Calendario 2020/2021

Lunedì dalle 14.15 alle 17.30

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

- Lezione 1 (22/02): Introduzione al corso e strumenti software utilizzati
- Lezione 2 (01/03): Elementi teorici relativi alla rappresentazione del testo
- Lezione 3 (08/03): Introduzione Linguaggi di Markup ed elementi di XML
- Lezione 4 (15/03): Introduzione alla definizione degli Schemi XML (DTD, XSD)
- Lezione 5 (22/03): Elementi di Codifica XML-TEI

Bozza piano del Corso: Calendario 2020/2021

Lunedì dalle 14.15 alle 17.30

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

- Lezione 6 (29/03): Moduli Editoriali TEI
- Lezione 7 (19/04): Introduzione ai Fogli di Stile
- Lezione 8 (26/04): Elaborazione documenti XML-TEI
(XSLT e DOM)
- Lezione 9 (03/05): Approfondimenti Codifica TEI e XSLT
- Lezione 10 (10/05): Esercitazione e discussione sui progetti d'esame
- Recupero (17/05): Esercitazione e Recupero Argomenti

Progress status

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

1 Presentazione

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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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, liberi da licenze a pagamento, per implementare le risorse oggetto di studio

Materiale Didattico

Studio e approfondimento

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021
A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

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 'Italiano (it)'. On the right, there's a user profile for 'ANGELO MARIO DEL GROSSO' with a notification icon. Below the header, the course title 'Codifica di Testi (IU, LINTRA, 2020/21)' is displayed, along with a breadcrumb trail: 'Home / I miei corsi / DISCIPLINE UMANISTICHE - Corsi ordinari / A.A. 2020 - 2021 / II semestre / 299LL_20'. There's also a 'Attiva modifica' button. The main content area has a sidebar on the left with 'Navigazione' containing links like 'Home', 'Dashboard', 'E-LEARNING AREA UMANISTICA', 'I miei corsi', 'DISCIPLINE UMANISTICHE', 'AA. 2020 - 2021', 'II semestre', '299LL_20', 'Partecipanti', 'Badge', 'Competenze', 'Valutazioni', 'Introduzione', and 'Argomento 1'. The main content area shows sections for 'Bacheca', 'Programma d'esame', and 'Risultati delle prove in itinere e degli esami'. It also features 'Argomento 1' and 'Argomento 2'. On the right, there are boxes for 'Ricerca nel forum', 'Annunci recenti', and 'Prossimi eventi'.

<https://elearning.humnet.unipi.it/course/view.php?id=3381>

Slide delle lezioni

Moodle, GitHub

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

The screenshot shows a GitHub repository interface. At the top, there's a navigation bar with a dropdown for 'master', the repository name 'corsoCodifica / CTaa20-21 /', and buttons for 'Go to file', 'Add file', and three dots. Below this is a list of files and their details:

| File | Description | Time Ago |
|----------------------------|----------------------------|---------------------|
| angelodel80 add shell code | | 29d4237 2 hours ago |
| .. | | |
| others | add saxon and xerces tools | 2 hours ago |
| progetto-esame | init repo CdT aa20-21 | 7 hours ago |
| slides | init repo CdT aa20-21 | 7 hours ago |
| src | add shell code | 2 hours ago |
| README.md | add saxon and xerces tools | 2 hours ago |

Below the file list is a 'README.md' file content box. It contains the text 'Corso Codifica di Testi aa 2020 / 2021' and a small edit icon.

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

Materiale di studio

Linee Guida della TEI - Online

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso



The screenshot shows the homepage of the Text Encoding Initiative (TEI). The header features the TEI logo and the text "<Text Encoding Initiative>". Below the header is a navigation bar with links to Home, Guidelines, Activities, Tools, Membership, Support, About, and News. A search bar is also present. The main content area is titled "P5: Guidelines for Electronic Text Encoding and Interchange" and includes a note about the version being 4.1.0, last updated on 19th August 2020. Below this are links to various language versions: English, Deutsch, Español, Italiano, Français, 日本語, 한국어, and 中文. There are also icons for user, GitHub, and a magnifying glass. The page is divided into sections: "Front Matter", "Text Body", and "TEI sourcecode". The "Front Matter" section lists chapters I through VI. The "Text Body" section lists chapters 1 through 18. The "TEI sourcecode" section lists links to GitHub repositories and bug reports.

P5: Guidelines for Electronic Text Encoding and Interchange
Version 4.1.0. Last updated on 19th August 2020, revision b414ba550

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

Front Matter

- I. [Release 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 [Descriptions](#)
- 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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML
Bibliografia

The screenshot shows the OpenEdition Press website. At the top left is the logo 'OpenEdition press'. To its right is the title 'OpenEdition Press' in a large, bold, pink font. On the far right of the header are links to 'HOME', 'ENCYCLOPÉDIE NUMÉRIQUE', 'READ/WRITE BOOK', 'SCIENCES PO | MÉDIALAB', and 'BRESIL / FRANCE | BRASIL / FRANÇA'. Below the header, there's a large image of the book cover for 'WHAT IS THE TEXT ENCODING INITIATIVE?'. The cover has a colorful, abstract background and the title in white text. To the right of the book image, the title 'WHAT IS THE TEXT ENCODING INITIATIVE?' is displayed in large, bold, black capital letters. Below it is the subtitle 'How to add intelligent markup to digital resources'. Underneath the subtitle is the author's name, 'Lou Burnard', and the word 'Encyclopédie numérique'. There are three small red dots followed by a paragraph of text about the Text Encoding Initiative (TEI) Guidelines. Further down, there's a section titled 'CERCA NEL LIBRO' with a search bar and an 'OK' button. Below that is a 'INDICE DEI CONTENUTI' button. At the bottom of this sidebar are buttons for 'CITA' and 'CONDIVIDI', and another for 'CITATO DA'. To the right of the book image, there's a section titled 'WHAT IS THE TEXT ENCODING INITIATIVE?' with a brief description and a link to 'Leggi il seguito'. At the bottom of this section is a horizontal line with two columns of book details. The left column includes 'Editore : OpenEdition Press', 'Luogo di pubblicazione : Marseille', 'Pubblicato su OpenEdition Books : 11 avril 2014', 'EAN digitale : 9782821834606', and 'Numero di pagine : 114 p.'. The right column includes 'Collana : Encyclopédie numérique | 3', 'Anno di pubblicazione : 2014', 'EAN (edizione cartacea) : 9782821834590', and 'DOI : 10.4000/books.oep.426'. On the far right of the page is a vertical sidebar with four icons: 'LEGGI' (with a computer monitor icon), 'ACCESSO APERTO' (with a smartphone icon), 'LETTURA CONFORTEVOLI' (with a laptop icon), 'EPUB' (with a smartphone icon), and 'PDF DEL LIBRO' (with a document icon). A red dot is visible above the 'LETTURA CONFORTEVOLI' and 'PDF DEL LIBRO' icons.

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



Materiale di studio

Libro Digital Scholarling Editing - E. Pierazzo

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

The screenshot shows a digital scholarly editing book titled "Digital Scholarly Editing: Theories, Models and Methods" by Elena Pierazzo. The page includes a navigation bar with links to Accueil, Dépôt, Consultation, Recherche, and Documentation. A red button highlights the document identifier "hal-01182162, version 1". Below the title, the author's name "Elena Pierazzo" is listed with a link to her profile. A small thumbnail image of the book cover is visible. The right side of the page contains detailed metadata sections: FICHIER (file), LICENCE (license), IDENTIFIANTS (identifiers), COLLECTIONS (collections), and CITATION (citation). The URL of the document is provided at the bottom: <https://hal.archives-ouvertes.fr/hal-01182162>.

Materiale di studio

Rivista jTEI

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

The screenshot shows the homepage of the OpenEdition Journals website. At the top, there is a logo for "OpenEdition Journals" featuring a stylized blue flower-like icon. To the right of the logo are links for "English" (with a dropdown arrow), "Share" (with a link icon), and "All OpenEdition" (with a grid icon). Below the header, there are three main navigation links: "HOME", "CATALOGUE OF 554 JOURNALS", and "OPENEDITION SEARCH Q". The main content area is titled "JOURNAL OF THE TEXT ENCODING INITIATIVE" and features a large yellow "TEI" logo. To the left of the main text, there is a sidebar with a search bar, an "INDEX" section, and a "FULL TEXT ISSUES" section listing issues 11 and 12. The main text discusses the journal's role as the official journal of the Text Encoding Initiative Consortium, publishing proceedings from annual conferences and special thematic issues. It also mentions state-of-the-art reports on electronic textual editing, current trends in TEI encoding, and new use cases for TEI. The text further provides a forum for articles on the discussion of the interface between the TEI and other communities, and more generally on the role of technological standards in the digital humanities, including digital scholarly editing, linguistic analysis, corpora creation, and newer areas such as mass digitization, semantic web research, and editing within virtual worlds. Below this, there is a section for "ISSUE 11 | JULY 2019 - JUNE 2020" titled "Selected Papers from the 2016 TEI Conference", edited by Resch Claudia, Wissik Tanja and Hannesschlaeger Vanessa.

Materiale di studio

Ulteriori risorse

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

- 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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Gross

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, xp-pointer, 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.

Publishing ☐

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



Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia



Home Articles Authors News Job Board About Search

News

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

Articles

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

Foundations

Annotated XML



What is XSLT?



Introduction to XProc 3.0



Introduction to XForms

Popular tags: articles

XFORMS XSLT VOCABULARY

UBL XPROC NISO RELAX-NG

XSD CONFERENCE XQUERY

Popular tags: news

XML XSLT DITA DOCBOOK

XQUERY XHTML DOCX

XMLMIND HTML EPUB

Announcing the XML slack community

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

Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

Saxonica > Saxon > Reference: XSLT Elements

Reference: XSLT Elements

This section of the Saxon documentation lists all of the standard XSLT elements in the <http://www.w3.org/1999/XSL/Transform> namespace, providing brief descriptions of their function, details about their structure and usage, and any information specific to the Saxon implementation. Many also have further detailed descriptions and examples. Almost all of the elements are supported in Saxon stylesheets, and any that are not supported are clearly labelled as such.

For information about the additional extension elements provided with the Saxon product, see [Extension instructions](#).

Conformance

Saxon implements the [XSLT 3.0 Recommendation](#) from the World Wide Web Consortium: see [XSLT 3.0 Conformance](#). The information here is designed to give a comprehensive summary of the features, but for the full specification, consult the official standard.

XSLT Patterns

This section of the documentation also contains an informal description of the syntax of [Patterns](#) in XSLT 3.0.

Standard attributes

There are a number of standard attributes that may appear on any XSLT element (or any [literal result element](#), in which case the names of the attributes must be in the XSLT namespace): `default-collation`, `default-mode`, `default-validation`, `exclude-result-prefixes`, `expand-text`, `extension-element-prefixes`, `use-when`, `version`, and `xpath-default-namespace`. For information about these, see the [xsl:stylesheet](#) element.

Section index

- [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](#)
- [xsl:element](#)
- [xsl:if](#)
- [xsl:import](#)
- [xsl:import-schema](#)
- [xsl:include](#)
- [xsl:iterate](#)
- [xsl:function](#)
- [xsl:global-context-item](#)
- [xsl:output](#)
- [xsl:output-character](#)
- [xsl:override](#)
- [xsl:package](#)
- [xsl:param](#)
- [xsl:perform-sort](#)
- [xsl:preserve-space](#)

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

Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso



MDN Web Docs
mozilla.org

▶ Technologies ▶ References & Guides ▶ Feedback

Search MDN

Sign in

XSLT: Extensible Stylesheet Language Transformations

Web technology for developers ▶ XSLT: Extensible Stylesheet Language Transformations

English (US) Change language

Table of contents

Documentation

Related Topics

[XSLT](#)

[EXSLT](#)

[XPath](#)

▶ Guides

▶ XSLT Elements

Documentation

[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.

<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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso



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



Materiale di studio

Ulteriori risorse

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

- 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 Giult 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.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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

Strumenti di lavoro

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

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



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

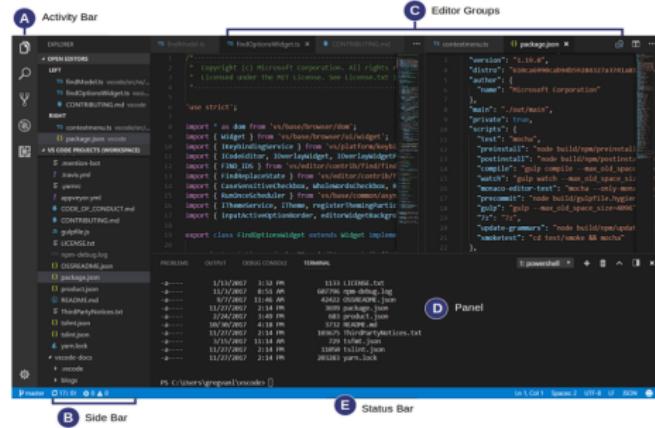
Ask questions

Follow @code

Request features

Report issues

Watch videos



Basic Layout

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

Strumenti di lavoro

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

Presentazione del Corso Codifica di Testi a.a. 2020-2021

30 / 71

The screenshot shows the EditiX XML Editor 2015 interface. The main window displays an XSLT file named "personal.xsl" with a Document Type of "XSLT version 1.0". The code editor contains the following XSLT code:

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

Below the code editor, the "XML Data source" pane shows a tree structure with nodes for "personnel" and "person". The "person" node has three children: "id", "name", and "value-of". The "value-of" node is expanded, showing its "select" attribute set to "@id". The "XSLT Result Preview" pane on the right shows the output for two "person" nodes: "one.worker" and "two.worker", both displaying the value "one.worker" as the result of the transformation.

<http://www.editix.com>

Editor XML

Cosa è consigliato usare

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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

Strumenti di sviluppo

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia



The screenshot shows the Java download page. At the top, there's a red header with the Java logo (a steaming coffee cup) and the word "Java". Below the header, there are "Download" and "Help" buttons. A search bar with a magnifying glass icon is positioned in the top right corner. The main title "JAVA + YOU, DOWNLOAD TODAY!" is prominently displayed in large, bold, grey font. Below the title is a red button labeled "Java Download". Underneath the button, there are three small links: "What is Java?", "Need Help?", and "Uninstall". To the right of the text, there's a photograph of a smiling woman with curly hair lying on a white surface, looking at a silver laptop. At the bottom of the page, there's a horizontal row of six links: "About Java", "Go Java", "Alice", "Greenfoot", "JavaOne", "Oracle Academy", and "Java Magazine". Each link has a small thumbnail image next to its name. The URL <https://www.java.com/en/download/manual.jsp> is displayed at the bottom left.

Strumenti di sviluppo

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

Presentazione del Corso Codifica di Testi a.a. 2020-2021

33 / 71

The screenshot shows the Apache Xerces™ Project homepage. At the top, there's a logo featuring a stylized orange and yellow butterfly above the word "Xerces" in blue, with "apache" written in smaller letters to the right. Below the logo is a navigation bar with links for "Home", "Wiki", and "Mailing List Archives". The main content area has a dark blue header bar with the text "apache.org > XML Federation > xerces.apache.org >". The main title is "The Apache Xerces™ Project - xerces.apache.org". A sidebar on the left contains a tree menu with nodes like "Xerces" (selected), "Overview", "Charter", "News", "Downloads", "Get Involved", "Who We Are", "License", "Legal Stuff", "Mailing Lists", "Code Repositories", "Reference", "Security", "Sponsorship", "Thanks", "Local Projects", "Hibernated Projects", and "Related Projects". Below the sidebar is a link "Built with Apache Forrest". The main content includes sections for "Welcome to the Apache Xerces™ Project", "Over a Decade at the Apache Software Foundation", and "Apache Xerces Project - Current Projects" which lists "Apache Xerces C++", "Apache Xerces2 Java", "Apache Xerces Perl", and "Apache XML Commons". There are also sections for "Apache Xerces C++ API", "Apache Xerces2 Java API", "Apache Xerces Perl API", and "Apache XML Commons API". At the bottom, there's a banner for "APACHE EVENTS" with a "LEARN MORE" button, and another banner for "10 YEARS OF The ASF" with the text "Celebrating a Decade of Open Source Leadership" and a link "http://www.apache.org/".

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

Strumenti di sviluppo

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

Presentazione del Corso Codifica di Testi a.a. 2020-2021

34 / 71

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 |
|--|---------------------|---------------------|----------------------|
|  Java | 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

<https://www.saxonica.com/documentation/index.html#!using-xsl/commandline>

Strumenti di sviluppo

esercitazioni, esempi, progetto

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

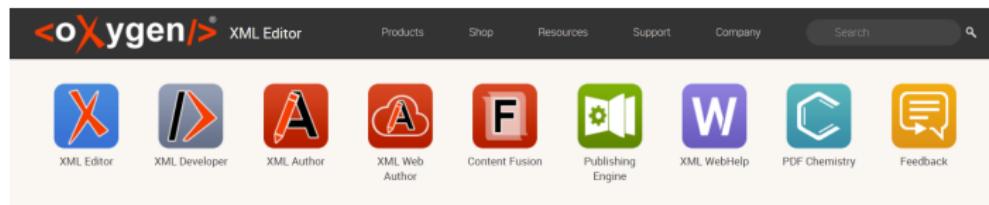
Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia



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/>

Progress status

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

1 Presentazione

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.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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

Modalità d'esame

In cosa consiste l'esame

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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

Tirocini, Tesi

Pronti a far parte della comunità?

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

A.M. Del Grosso

Daniele Gambarara: SGL
Beyond the Course. A Knowledge Site for Rethinking Saussure's General Linguistics

SIDE BAR:
[view the source in ext software](#)

ENGLISH TRANSLATION ^

Interpretative Edition of the Source

Facsimile
(BGE AdS 372/4 f. 135)

Explore Saussure Thoughts

Dive Into **échec** from BGE AdS 372/4, f. 135

Diplomatic
(BGE AdS 372/4 f. 135)

Interpretative
(Science of the language, p. 231)

— De la même façon <que dans le jeu d'échecs> qu'il serait profondément inutile de demander;
2 ce que vaut devant le jeu d'échecs;
3 <il serait> n'y a pas de sens à <il est absurde de> demander
5 <ce que serait> ce qu'est <dans le jeu d'échecs> une reine <dame>, un
fou, ou un cavalier, here du jeu
7 d'échecs quand <so> considérait le jeu
8 <hors du jeu> D'échecs; de la même façon, il n'y
9 a pas de sens, si l'on considère vrai;
10 ment le langage, à demander cher=
11 cher ce qu'est chaque élément par
12 lui-même. Il n'est rien d'autre
13 qu'une pièce valant par son
14 opposition avec d'autres selon
15 certaines conventions.

— De la même façon que dans le jeu d'échecs il serait absurde de demander / ce que serait une dame, un pion / un fou, ou un cavalier, / si on le considérait / hors du jeu d'échecs; de la même façon, il n'y a pas de sens, si l'on considère vrai / ment la langue, à chercher ce qu'est chaque élément par / lui-même. Il n'est rien d'autre / qu'une pièce valant par son / opposition avec d'autres selon / certaines conventions.

show similar texts

ERC-2020-ADV Grant

© 2020 Beyond the Course. A Knowledge Site for Rethinking Saussure's General Linguistics

SGL project

Open Sandbox

Tirocini, Tesi, EVT

Pronti a far parte della comunità?

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

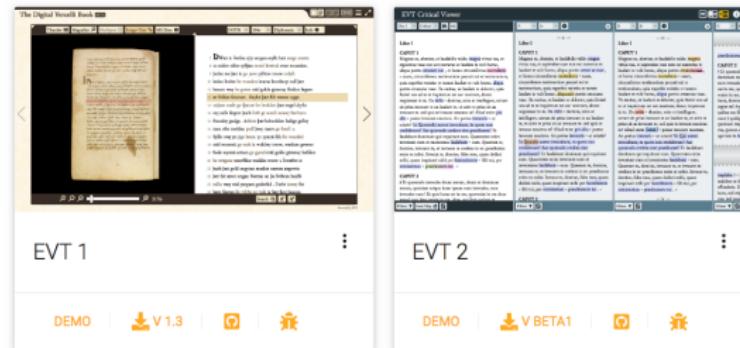
Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

1 Presentazione

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.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

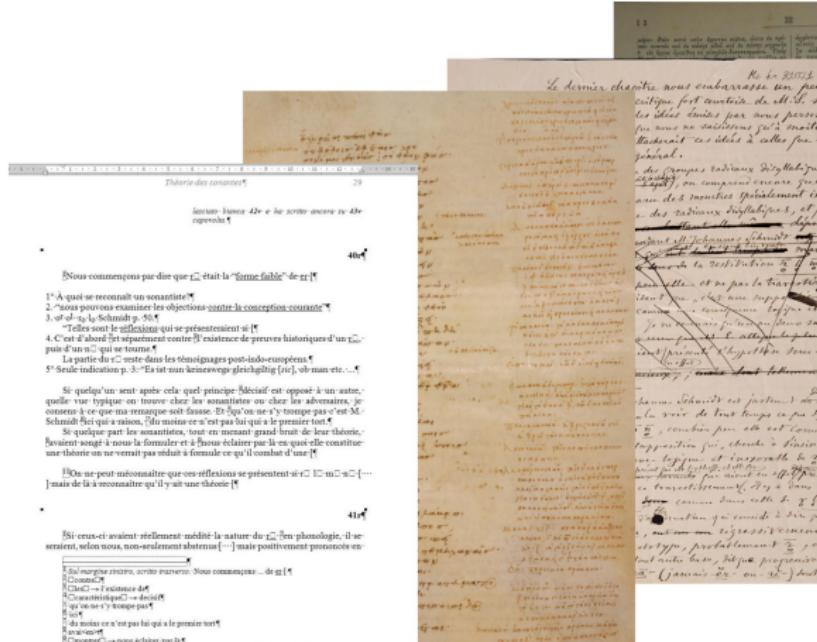
Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia



Perché è importante la codifica dei testi

Dove troviamo i testi

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

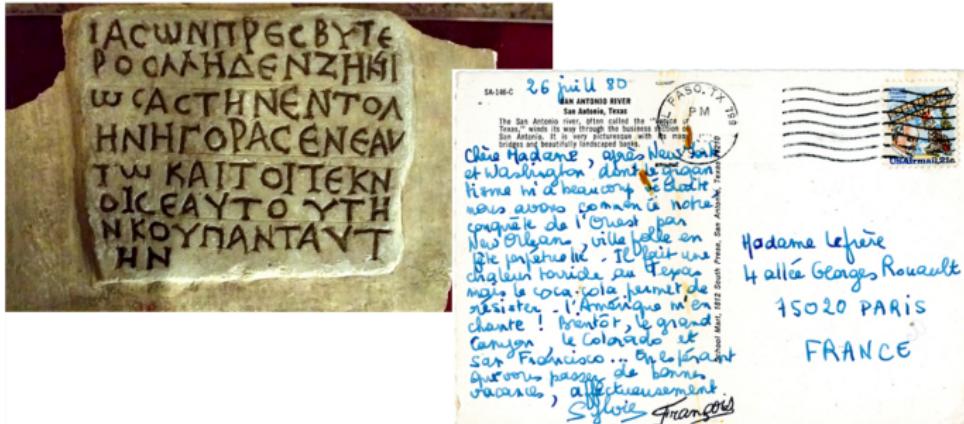
Ecosistema
XML

Bibliografia

A.M. Del Grosso

Presentazione del Corso Codifica di Testi a.a. 2020-2021

44 / 71



Perché è importante la codifica dei testi

Dove troviamo i testi

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

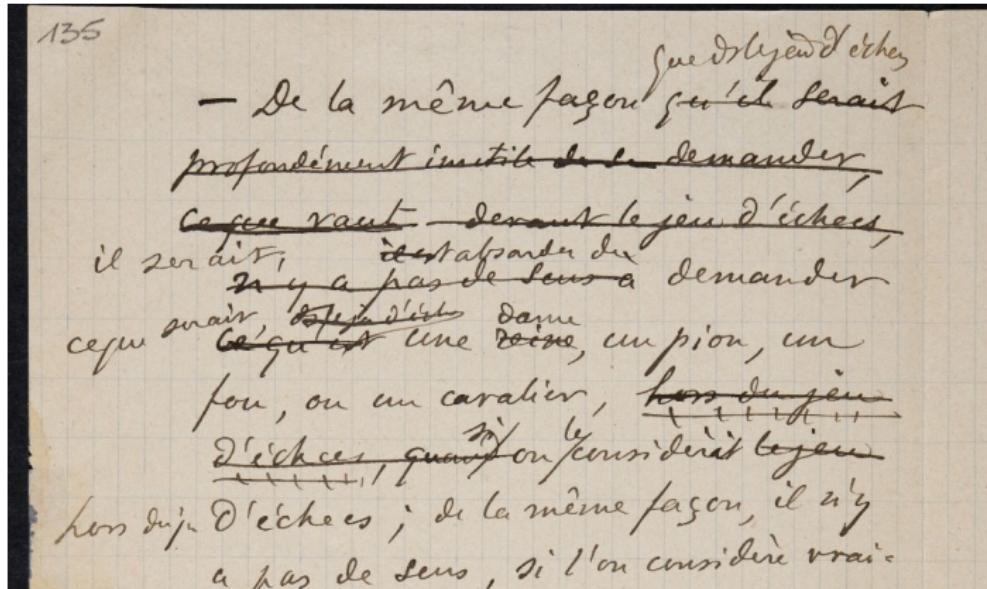
Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia



Perché è importante la codifica dei testi

Motivazioni pratiche

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021
A.M. Del
Grosso

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.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Codificare un testo

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

Codificare un testo

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

Rappresentazione digitale dei testi

basso e alto livello di codifica

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Rappresentare un testo

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

Rappresentazione digitale dei testi

Modello dati di un testo

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Un testo non ha una struttura rigida, predefinita:

- Non è rappresentabile solo come un insieme di record di un archivio elettronico.
- Non è rappresentabile solo come un insieme di tabelle di una banca dati.
- Non è rappresentabile solo come un albero o un insieme di sotto-alberi (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.
2020-2021
A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi
Materiale
Didattico e
Strumenti di
lavoro
Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

La rappresentazione di un testo

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

Elementi di Codifica del testo

Tabella Formalismi

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Formalismi

| | Data | Text | Hierarchies | Presentation | Validation | References | Annotations | Overlapping |
|---------------------|-------------|-------------|-------------|--------------|------------|-------------|-------------|-------------|
| CSV | Light Green | Light Red | Light Red | Light Red | Dark Green | Light Green | Light Green | Light Red |
| JSON | Dark Green | Light Green | Dark Green | Light Red | Light Red | Light Red | Light Green | Light Red |
| RDF | Dark Green | Light Red | Light Red | Light Red | Dark Green | Dark Green | Light Green | Light Red |
| Markdown | Light Red | Dark Green | Light Red | Light Green | Light Red | Light Green | Light Red | Light Red |
| HTML | Light Red | Dark Green | Dark Green | Dark Green | Light Red | Light Green | Light Green | Light Red |
| HTML+RDFa | Dark Green | Dark Green | Dark Green | Dark Green | Light Red | Dark Green | Dark Green | Light Red |
| XML | Dark Green | Dark Green | Dark Green | Dark Green | Dark Green | Dark Green | Dark Green | Light Red |
| Overlapping formats | | | | Light Red | Light Red | Dark Green | Dark Green | Dark Green |

courtesy of *Fabio Vitali*

Elementi di Codifica del testo

Formalismi

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Formati di rappresentazione

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

Importanza dei formati

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

Elementi di Codifica del testo

Esempio di codifica del testo utilizzando CSV

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Why it's worth giving up .
The human_body is remarkably resilient and will start to repair the damage caused by smoking almost as soon as you start giving_up .
The benefits to your health can be seen in as little as 20 minutes from the time you put_down your last cigarette .
The whole process remains only marginally_safe on average for women_and_children .
Both men_and_women want the process to be safer .
Both men_and_women are involved in developing and exploiting the sciences and technologies to make_this_happen .
The gender_dynamics of public_perceptions on DNA_paternity_testing are likely to become more important as the tests become more widespread .
Some perceptions of paternity_testing will obviously be dependent on the inherently_different reproductive_roles of men_and_women .

Assume VB assume ''
that IN that
a DT a
natural JJ natural
cyclic JJ cyclic
phenomenon 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
SENT .

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

```
{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "http://574heritago.com/annotations/2/",
  "type": "Annotation",
  "created": "2017-03-12T12:03:45Z",
  "body": {
    "type": "TextualBody",
    "purpose": "tagging",
    "format": "text/plain",
    "language": "en",
    "value": "Where is the origin of this statue?"
  },
  "creator": {
    "id": "http://heritago.com/users/1",
    "type": "Person",
    "email_shai": "5bad08927902ff9307b621c54716dcc5083e339"
  },
  "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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Formati e formalismi di codifica

Quindi ogni pezzo di informazione aggiunta ad un testo grezzo attraverso l'inserimento di dati metatestuali (markup, annotazione, codifica), 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.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

1 Presentazione

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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

XML per la descrizione e la codifica

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

TEI-XML

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

Lo schema TEI-XML

Estratto di documento TEI-XML

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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

Perché TEI

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

Impiego di XML

Benefici

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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

Impiego di XML

Benefici

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

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

Linguaggio di marcatura XML

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Perché XML

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

XML come linguaggio per la codifica di testi

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia

Vantaggi

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

References

Presentazione
del Corso
Codifica di
Testi
a.a.
2020-2021
A.M. Del
Grosso

Presentazione
Introduzione
Corso Codifica
di Testi
Materiale
Didattico e
Strumenti di
lavoro
Progetto
Esame
Codifica dei
Testi
Ecosistema
XML
Bibliografia

-  Lenci, A., Montemagni S., and Pirrelli V. (2016). *Testo e Computer. Elementi Di Linguistica Computazionale*. Aulamagna. Carocci.
-  Pierazzo, E. (2015). *Digital Scholarly Editing : Theories, Models and Methods*. Farnham Surrey: Ashgate.
-  Orlando, 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.
2020-2021

A.M. Del
Grosso

Presentazione

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

Bibliografia



XML Standards Reference, MSDN.

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



IBM XML Tutorial,

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



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