

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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. 2023-2024

Angelo Mario Del Grosso

angelo.delgrossos@ilc.cnr.it

CNR-ILC

Istituto di Linguistica Computazionale “A. Zampolli”,
21st February 2024

Outline della lezione

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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 14.15-15.45 (Fib M-Lab) e Venerdì 12.00-13.30 (Fib M-LAB)

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

Progetto
Esame

Codifica dei
Testi

Ecosistema
XML

mer 21/2	gio 22/2	ven 23/2
	08.30 - 10.00 LETTERATURA ITALIANA CONTEMPORANEA INFORMATICA UMANISTICA (3) Fib PS1	
10.15 - 11.45 LETTERATURA ITALIANA CONTEMPORANEA INFORMATICA UMANISTICA (3) Fib PS1		
		12.00 - 13.30 CODIFICA DI TESTI INFORMATICA UMANISTICA (3) Fib M-Lab
14.15 - 15.45 CODIFICA DI TESTI INFORMATICA UMANISTICA (3) Fib M-Lab		
16.00 - 19.15 ISTITUZIONI DI LOGICA INFORMATICA UMANISTICA () Fib PS4 G. VENTURI	16.00 - 19.15 PROGETTAZIONE GRAFICA INFORMATICA UMANISTICA (3) Fib H-Lab B. RAPISARDA	

Bozza piano del Corso: Calendario 2023/2024

Mercoledì 14.15-15.45 (Fib M-Lab) e Venerdì 12.00-13.30 (Fib M-LAB)

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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 2023/2024

Mercoledì 14.15-15.45 (Fib M-Lab) e Venerdì 12.00-13.30 (Fib M-LAB)

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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 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 a breadcrumb trail: 'Home / I miei corsi / DISCIPLINE UMANISTICHE - Corsi ordinari / AA 2020 - 2021 / Il semestre / 299IU_20'. There's a 'Attiva modifica' button on the right.

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
 - Corsi ordinari
 - AA 2020 - 2021
 - Il semestre
 - 299IU_20
 - Argomento 1
 - Argomento 2

<https://elearning22.humnet.unipi.it/Materiale>

Slide delle lezioni

Moodle, GitHub

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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 has 32 watchers and 4 forks. The 'Code' tab is selected. A commit from 'angelodel80' is shown, adding README.md files to subdirectories. The commit message is 'add first README.md files in subdirs'. The commit was made 'now' and has a commit ID of fdb5949. Below the commit, there is a file named 'README.md' containing the text '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.
2023-2024

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. 2023-2024

17 / 73

The screenshot shows the homepage of the Text Encoding Initiative (TEI). The header features the TEI logo and navigation links for Home, Guidelines, Activities, Tools, Membership, Support, About, and News. A search bar is present at the top right. Below the header, a banner for "P5: Guidelines for Electronic Text Encoding and Interchange" is displayed, along with a note about the version (4.1.0, last updated on 19th August 2020). The main content area is divided into three columns: "Front Matter" (containing sections like Title, Back Matter, and Appendices), "Text Body" (listing chapters from 1 to 18), and "TEI sourcecode" (with links to GitHub and bug reports). At the bottom, there are language selection buttons for English, Deutsch, Español, Italiano, Français, 日本語, 한국어, and 中文.

<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.
2023-2024

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 left is the logo 'OpenEdition press'. To its right is the text 'OpenEdition Press' and a navigation menu with links to 'HOME', 'ENCYCLOPÉDIE NUMÉRIQUE', 'READ/WRITE BOOK', 'SCIENCES PO | MÉDIALAB', and 'BRÉSIL / FRANCE | BRASIL / FRANÇA'. Below this is a large image of the book cover for 'WHAT IS THE TEXT ENCODING INITIATIVE?'. The cover has a colorful, abstract background with the title in white. Below the image is a search bar with placeholder 'CERCA NEL LIBRO' and an 'OK' button. Underneath 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's name 'Lou Burnard' and the series 'Encyclopédie numérique' are listed. Below the author information are three ellipsis dots. A detailed description of the Text Encoding Initiative follows, mentioning it as a de facto standard for digital textual resources in scholarly research. Further down, there is a section with examples of TEI-encoded text. On the far right, there is a sidebar with icons for reading options: 'LEGGI' (with a computer monitor icon), 'ACCESO APERTO' (with a lamp icon), 'LETTURA CONFORTEVOLE' (with a lamp icon), 'EPUB' (with a smartphone icon), and 'PDF DEL LIBRO' (with a document icon). At the bottom of the page is a URL: <https://books.openedition.org/oep/426>.

Materiale di studio

Libro Digital Scholarling Editing - E. Pierazzo

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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. 2023-2024

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 person's head and hands emerging from a globe, with geometric shapes floating around. Below the logo are navigation links: Accueil, Dépôt, Consultation ▾, Recherche, and Documentation. A search bar contains the identifier "hal-01182162, version 1". Next to it is a button labeled "Pré-publication, Document de travail" with a small icon. The main title of the document is "Digital Scholarly Editing: Theories, Models and Methods" by Elena Pierazzo. Below the title, it says "GERCi - Groupe d'étude et de recherche sur la culture italienne". The abstract discusses the book's treatment of digital scholarly editing across the timeline of an edition, from choice of object to publication, and its significance. The document is identified as a "Pré-publication, Document de travail". It includes sections for "Keywords" (Digital Humanities, Philology, Digital Editing, Scholarly editing), "Type de document" (Pré-publication, Document de travail), and "Domaine" (Sciences de l'Homme et Société / Littératures, Sciences de l'Homme et Société / Héritage culturel et muséologie, Sciences de l'Homme et Société / Sciences de l'information et de la communication). A link "Liste complète des métadonnées" leads to a detailed view. On the right side of the page, there are several sidebar boxes: "FICHIER" containing the PDF file "DSE-revised-29Sept2014.pdf" and information about the author(s); "LICENCE" showing "Copyright (Tous droits réservés)"; "IDENTIFIANTS" listing the HAL ID; "COLLECTIONS" showing "GERCi | UGA"; and "CITATION" providing the bibliographic 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.
2023-2024

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 a brief description of the journal. Below this, there is a sidebar with a search bar, an index section listing authors and keywords, and a full-text issues section listing issues 11 and 12. The main text area describes the journal as the official journal of the Text Encoding Initiative Consortium, publishing proceedings from annual conferences and special thematic issues. It also mentions current trends in TEI encoding and new use cases for TEI.

JOURNAL OF THE TEXT ENCODING INITIATIVE

The *Journal of the Text Encoding Initiative* is the official journal of the Text Encoding Initiative Consortium. It publishes the proceedings of the annual *TEI Conference* and Members' Meeting and special thematic issues: state-of-the-art reports on electronic textual editing, current trends in TEI encoding, and new use cases for TEI. It furthermore provides a forum for articles on the discussion of the interface between the TEI and other communities, and more generally of 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.

LATEST ISSUE

ISSUE 11 | JULY 2019 - JUNE 2020
Selected Papers from the 2016 TEI Conference
Edited by Resch Claudia, Wissik Tanja and Hanneschlaeger Vanessa

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

Materiale di studio

Ulteriori risorse

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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.
2023-2024

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.
2023-2024

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](#)

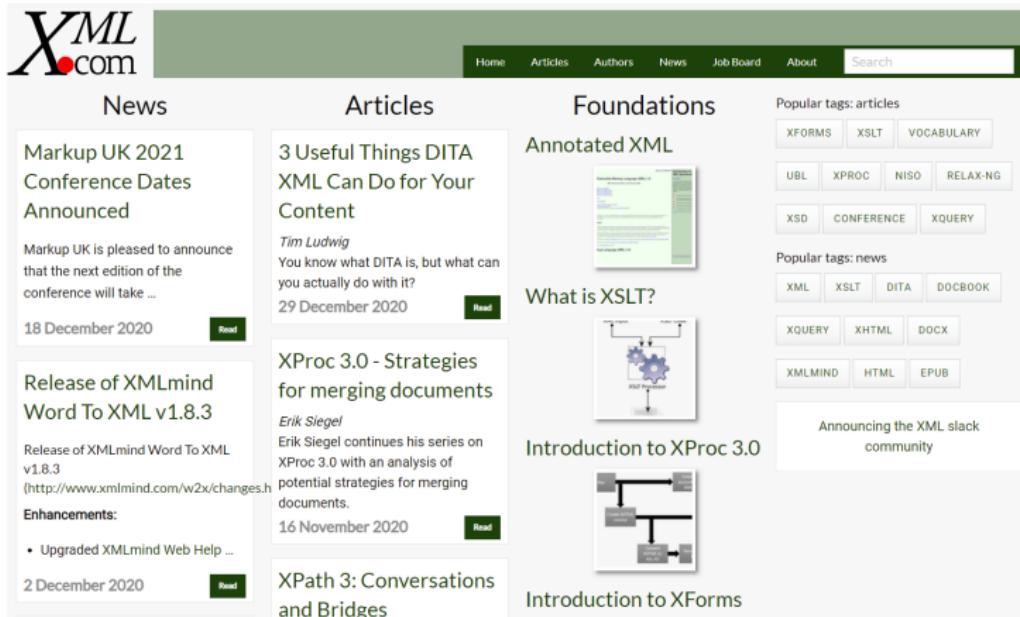
XPath 3: Conversations and Bridges

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.
2023-2024

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 provides a brief description of the standard XSLT elements. It mentions the W3C XSLT 3.0 Recommendation and links to the XSLT 3.0 Conformance page. Below this is a "Conformance" section, followed by "XSLT Patterns", and finally "Standard attributes". A "Section index" at the bottom lists all the XSLT elements from the sidebar.

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

Materiale di studio

Tecnologia di rappresentazione e di elaborazione

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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.
2023-2024

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. 2023-2024

26 / 73

XML
3RD EDITION



COMPREHENSIVE

Carey :: Vodník

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.
2023-2024

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.
2023-2024

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. 2023-2024

29 / 73

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

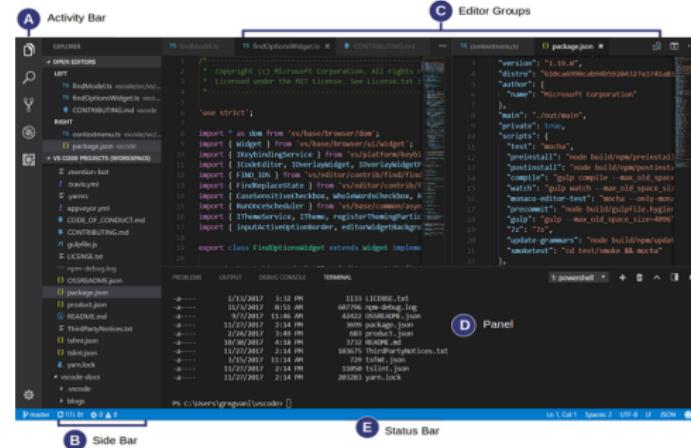
User Interface



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.

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.
2023-2024

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 file 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 path to "C:\Users\empire\myProject\samples\xml\personal-schema.xml". The schema defines three "person" elements with an "id" attribute. The right pane shows the XSLT Result Preview in HTML format, displaying two results: "one.worker" and "two.worker".

Editor XML

Cosa è consigliato usare

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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.
2023-2024

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'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. The main title "JAVA + YOU, DOWNLOAD TODAY!" is prominently displayed in large, bold, grey letters. 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's a photograph of a smiling woman with curly hair lying on a white surface, looking at a silver laptop. At the bottom, there's 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.
2023-2024

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. 2023-2024

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" 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.
2023-2024

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.
2023-2024

A.M. Del
Grosso

Presentazione
personale

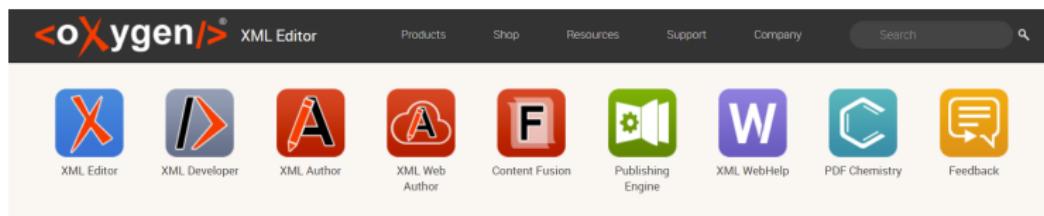
Introduzione
Corso Codifica
di Testi

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.
2023-2024

A.M. Del
Grosso

Presentazione
personale

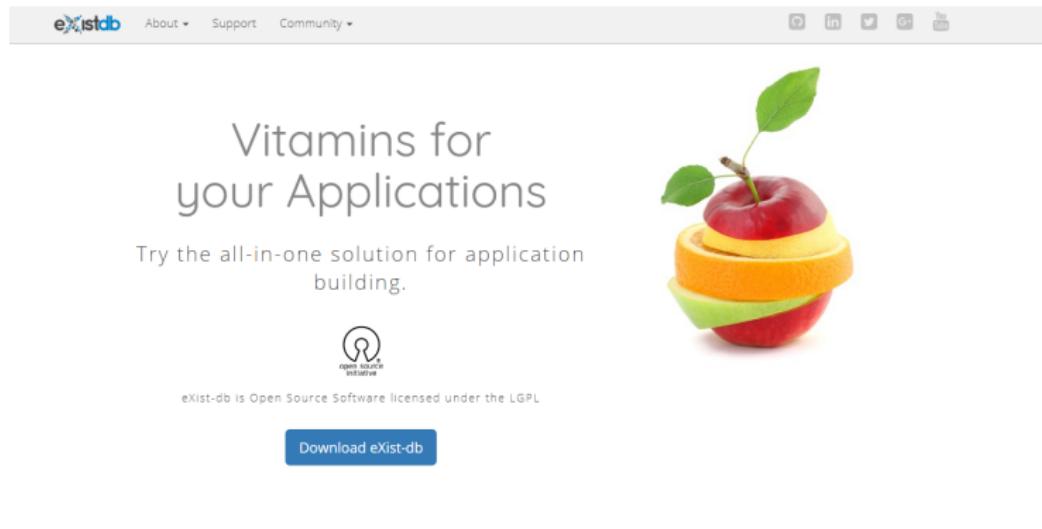
Introduzione
Corso Codifica
di Testi

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 is 'Vitamins for your Applications', followed by a subtext: 'Try the all-in-one solution for application building.' Below this is a logo for 'open source initiative'. A note states 'eXist-db is Open Source Software licensed under the LGPL'. A blue button says 'Download eXist-db'. On the right side, there's a decorative graphic of a red apple, an orange slice, and a green lime slice stacked together. At the bottom left, there's a red box titled 'Latest News' containing the text 'February 05 2023 eXist-db 6.2.0'. On the bottom right, there's a small screenshot of a database interface with some text and a timeline indicator.

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

Progress status

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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 software interface with several windows open:

- Paleographic Apparatus:** Shows Col. 64 of a manuscript. The text is in two columns. The left column contains lines 1-15, and the right column contains lines 16-20. Lines 1-15 are in Greek script, while lines 16-20 are in Latin script.
- Diacritic Transcription:** A window showing the same text as the paleographic apparatus, but with diacritical marks added to the Greek script.
- Literary Transcription:** A window showing the text in a simplified literary transcription.
- Palaeographical Apparatus:** Shows the same text as the paleographic apparatus, but with additional annotations and markings.
- Translation:** A window showing the translation of the text into modern Italian.
- Infrared image at 950nm:** A window showing the infrared image of the manuscript page.
- Copper engraving for the Collectio aliorum:** A window showing a copper engraving of the manuscript page.
- Transcription by S. Crillo:** A window showing a transcription of the text by S. Crillo.
- Notes:** A window showing various notes and annotations.
- Normalised paragraph:** A window showing the text in a normalised paragraph format.

Laboratorio CNR-ILC CoPhiLab

Tirocini, Tesi, EVT

Pronti a far parte della comunità?

Presentazione
del Corso
Codifica di
Testi
a.a.
2023-2024

A.M. Del
Grosso

Presentazione
personale

Introduzione
Corso Codifica
di Testi

Materiale
Didattico e
Strumenti di
lavoro

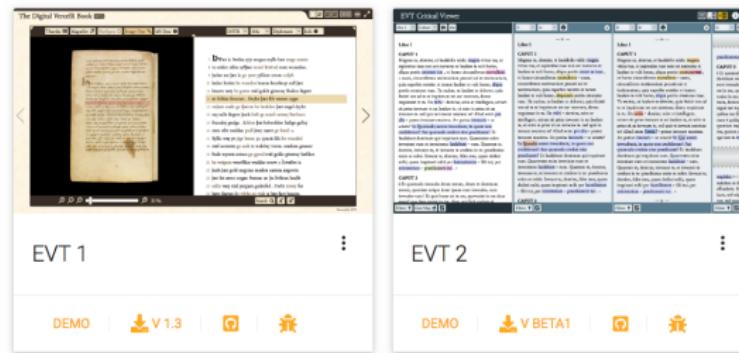
Progetto
Esame

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.
2023-2024

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.
2023-2024

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.
2023-2024

A.M. Del
Grosso

Presentazione
personale

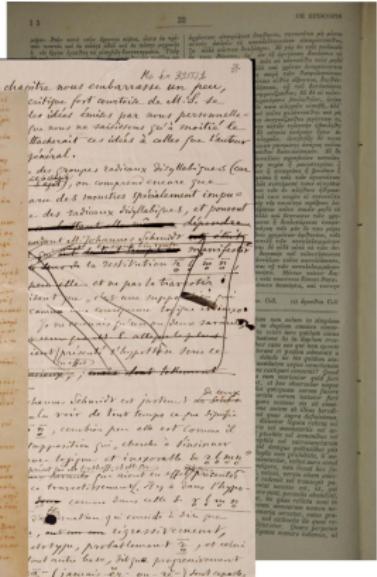
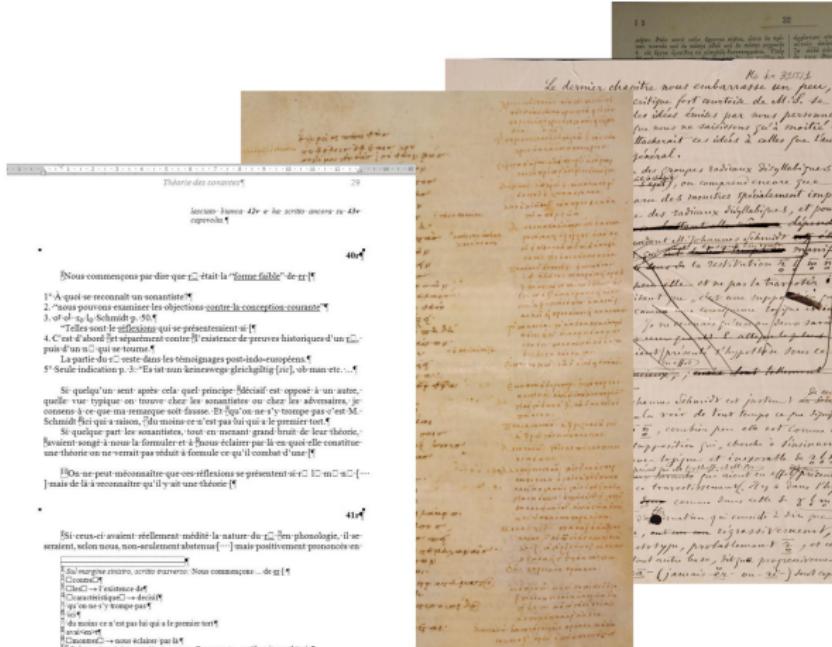
Introduzione
Corso Codifica
di Testi

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.
2023-2024

A.M. Del
Grosso

Presentazione
personale

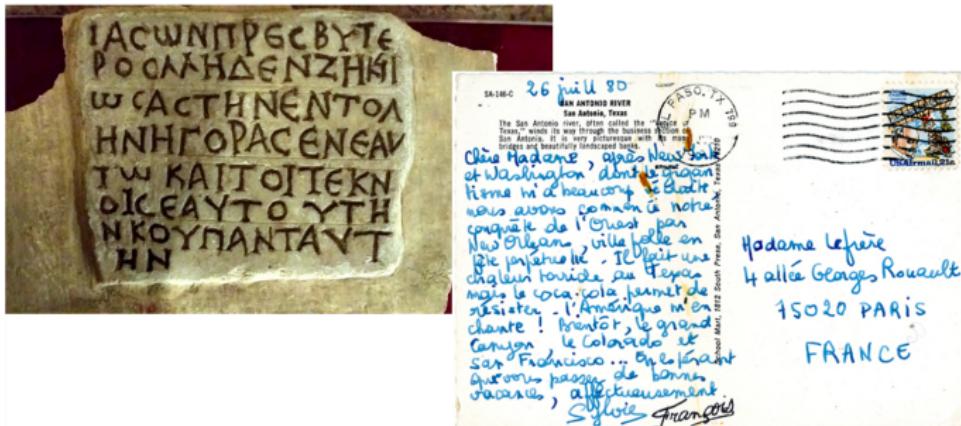
Introduzione
Corso Codifica
di Testi

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.
2023-2024

A.M. Del
Grosso

Presentazione
personale

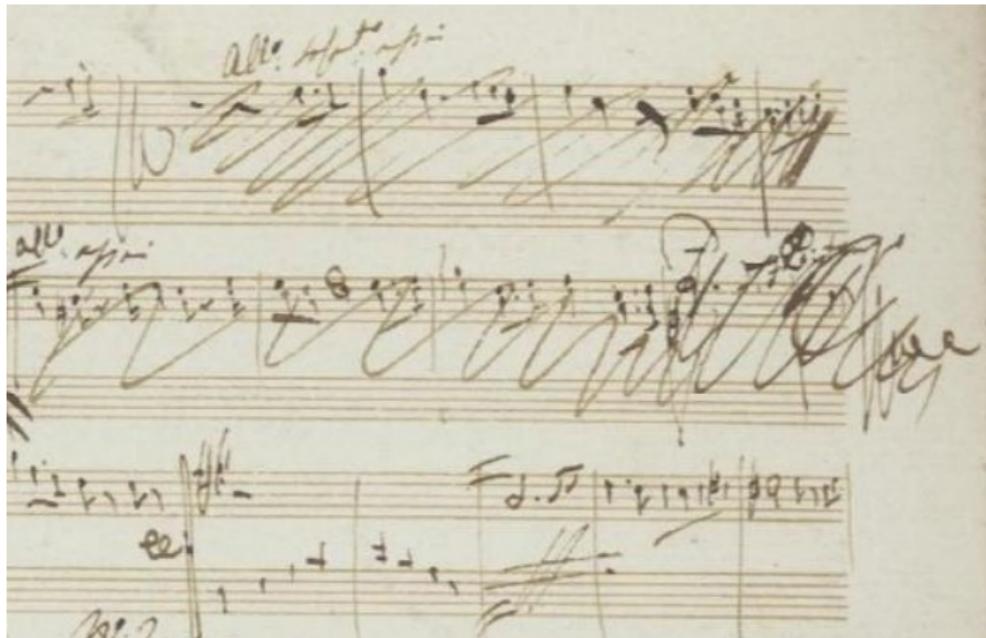
Introduzione
Corso Codifica
di Testi

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.
2023-2024

A.M. Del
Grosso

Presentazione
personale

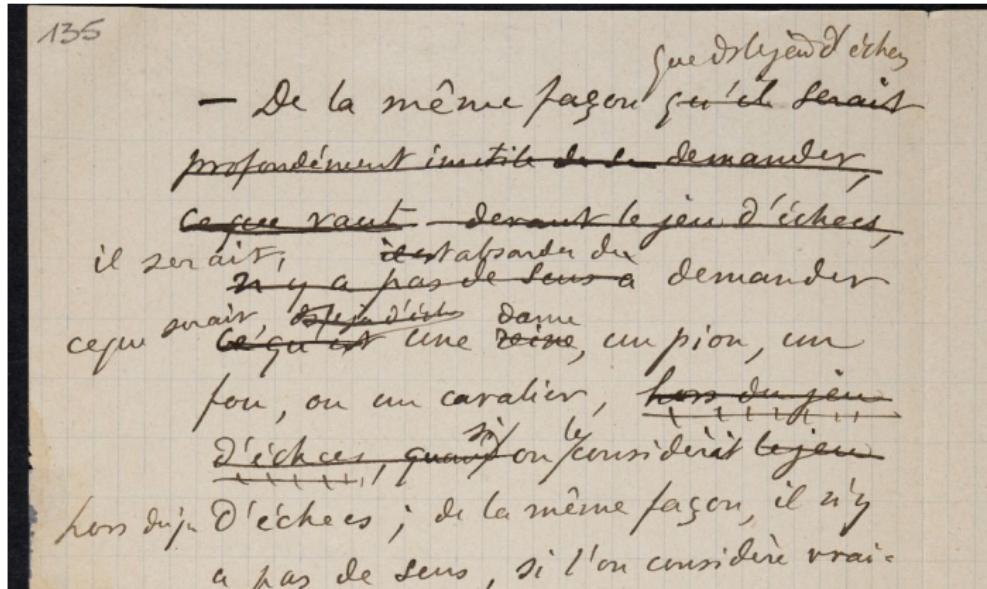
Introduzione
Corso Codifica
di Testi

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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		
.	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": "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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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.
2023-2024

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>

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Codifica di Testi - Introduzione Teorica a.a. 2023-2024

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Angelo Mario Del Grosso

angelo.delgrossos@ilc.cnr.it

CNR-ILC

Istituto di Linguistica Computazionale “A. Zampolli”,
1st March 2024

Contenuto della lezione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Codifica e Modellizzazione

Problemi e opportunità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo, oggetto di studio

Le discipline umanistiche hanno come *oggetto di studio* privilegiato: **i testi**.

L'avvento delle tecnologie informatiche ha *cambiato la produzione, la trasmissione e lo studio del sapere* e della conoscenza (quindi anche il testo, in senso lato).

Il testo in ambiente digitale

Il nuovo formato digitale ha determinato **profonde trasformazioni nelle pratiche scientifiche** che hanno per oggetto o per fine lo studio, la produzione, la pubblicazione di testi.

Codifica e Modellizzazione

Problemi e opportunità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Vero e proprio cambio di paradigma

Gli studiosi si sono interrogati sul *nuovo rapporto che le discipline umanistiche devono avere con il testo* che diventa elettronico o digitale.

Nuove sfide

La natura del formato digitale ha favorito *molte ricerche volte alla risoluzione di nuovi problemi* connessi alla **rappresentazione formale dei testi.**

Codifica e Modellizzazione

Problemi e opportunità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Discussione nella comunità

L'informatica allo stato attuale è in grado di fornire allo studioso umanista **un ampio ecosistema di tecnologie** finalizzato all'implementazione di *efficaci sistemi di rappresentazione formale del testo* (anche conseguenti sistemi di elaborazione).

Nuove sfide

La discussione sulla rappresentazione digitale dei testi ha rivelato un intreccio di complesse **questioni teoriche, metodologiche e pratiche.**

Perché è importante la codifica dei testi

Motivazioni teoriche

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

La codifica come metodologia

L'attività di codifica diviene **metodo di analisi della conoscenza** nell'ambito delle discipline che si occupano della cura scientifica 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

Definizione di codifica informatica dei testi

Fabio Ciotti 2007

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Definizione

Per **codifica** informatica dei testi intendiamo la *rappresentazione formale* di un **testo** ad un qualche livello descrittivo, su di un **supporto digitale**, in un formato utilizzabile da un elaboratore (*Machine Readable Form*) mediante un opportuno **linguaggio informatico** (*formalismo*).

Concetti alla base della codifica di testi

Gli elementi

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

- Cosa significa Codifica
- Cosa significa Testo
- Cosa significa Codifica del Testo
- Cosa significa Linguaggio Informatico

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Codifica e Decodifica

- Il termine codifica (in inglese *encoding*) può essere associato alla creazione di un testo
- Il termine decodifica (in inglese *decoding*) può essere associato alla comprensione e interpretazione di un testo
- Il messaggio può essere visto come contenitore di significato (~ *testo*)

Codifica del testo

definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codice

Sistema di regole e di segni (*convenzione*) per convertire informazioni da una forma (anche astratta) in un'altra forma o rappresentazione per la comunicazione tramite un canale.

La comunicazione Umana

Modello della Comunicazione - C. Shannon e W. Weaver (1949)

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

The Mathematical Theory of Communication

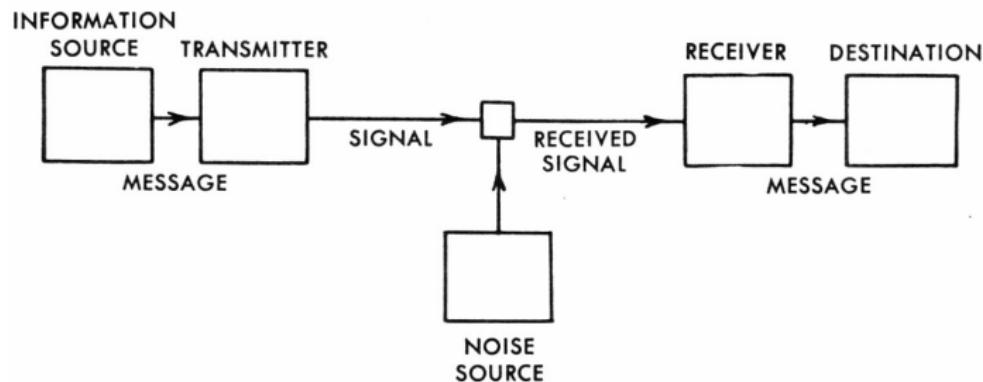


Fig. 1.— Schematic diagram of a general communication system.

La comunicazione Umana

Diasistema

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Messaggio

$M_e \neq M_d$ Il **Messaggio codificato è diverso dal messaggio decodificato**

La comunicazione Umana

Diasistema

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Messaggio

Il sistema di codifica e il sistema di decodifica sono due sottosistemi di uno stesso **diasistema**.

La comunicazione Umana

Diasistema

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Comunicazione come diasistema

Comunicazione = Sistema dell'emittente (S1) ~ Sistema del destinatario (S2).

*Il risultato di un compromesso tra il sistema S1 e il sistema S2**

* Segre, 1985

Codifica del testo

Sinteticamente

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

La codifica digitale di un testo

la codifica testuale è la rappresentazione formale di un testo e delle sue caratteristiche mediante un linguaggio artificiale.
(Ciotti)

Grammatica formale

Questo linguaggio consente di **descrivere esplicitamente in modo non ambiguo** le caratteristiche di un testo applicando **etichette meta-testuali** la cui funzione e struttura è descritta da **regole definite in una grammatica formale** (formalismo e meta-linguaggio).

Codifica del testo

Linguaggio formale che diventa linguaggio teorico

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

La codifica del testo come modello formale

La codifica digitale del testo diviene a questo punto un **linguaggio teorico formale** attraverso il quale lo studioso **costruisce modelli formali** del testo.

Linguaggi di Markup

Ai fini della modellizzazione di testi, tra i linguaggi formali si sono imposti quei sistemi basati sui cosiddetti **linguaggi di marcatura descrittivi** (*markup language*).

Codifica del testo

Sistema linguistico formalizzato

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Secondo Lou Burnard 1995

La codifica informatica di un testo impone l'**adozione di un sistema linguistico formalizzato** che permetta a uno studioso di **rendere esplicita una interpretazione di un testo** e le varie operazioni che la hanno prodotta.

I sistemi dichiarativi

I sistemi dichiarativi/descrittivi forniscono un potente dispositivo metalinguistico.

Codifica del testo

Finalità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Capacità rappresentazionale

**Il principale requisito di uno schema di codifica, pertanto,
è la capacità rappresentazionale che esso offre allo
studioso**

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Modellare il testo

Il testo come oggetto del dominio di studio

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Informatica nelle scienze umane

L'informatica umanistica si occupa, tra le altre cose, **della rappresentazione e dell'elaborazione degli oggetti** d'interesse e di studio per le discipline umanistiche.

Testo come oggetto di studio

Il testo è, tra questi, l'oggetto più ricorrente e privilegiato

Modellare il testo

qual è la natura del testo?

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Modellare il testo

Per trattare e rappresentare il testo in ambiente digitale,
bisogna formulare un suo modello

Modello del testo

Un efficace modello del testo non può prescindere da una determinazione di **cosa si intenda per testo e quale sia la sua natura.**

Modellare il testo

Testo: oggetto complesso

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è il testo

Il testo è un oggetto complesso in quanto è in grado di veicolare **significato su più livelli strutturali** (logico, ontologico, linguistico, autoriale, editoriale, fisico, ecc.), anche attraverso *l'instaurazione di molteplici relazioni tra più dimensioni*.

Modellare il testo

Testo: oggetto complesso

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Ma..

**Non abbiamo a disposizione nessuna teoria completa e
condivisa del testo**

Modellare il testo

Tentativo..

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è il testo

Quale concezione o modello ontologico del testo è implicata nella rappresentazione informatica di esso?

A quali teorie possiamo rivolgerci?

Molteplici teorie del testo, quasi tutte sbilanciate sul livello verbale-semantico (*linguistica testuale*).

Modellare il testo

Tentativo..

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è il testo

Il termine “testo” si riferisce a un oggetto plurale, in cui co-esistono:

- **Un livello astratto:** la *sequenza verbale*, la quale a sua volta genera una serie di *livelli di contenuti semantici*.
- **Un livello materiale:** il supporto e le tracce d'inchiostro.
- **Un livello dinamico:** il testo viene **creato** da un autore e **ricreato** dal lettore.

Cos'è il testo

- Un documento materiale composto da fogli che contengono inchiostro variamente disposti?
- Un discorso linguistico fissato tramite la scrittura su un documento materiale?
- Un'opera dell'ingegno che viene costituita e fissata tramite un insieme di simboli/segni?
- Il contenuto di un messaggio?

Modellare il testo

Il testo: Cercando una definizione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è il testo - uno sguardo più specialistico

- Lo stato linguistico di un singolo testimone materiale di un'opera?
- Lo stato linguistico di un medesimo testimone di un'opera che presenta diverse lezioni identificabili?
- Una versione edita di un'opera?
- Una sequenza coerente di enunciati in una lingua naturale?
- Uno scritto che può essere trasmesso anche oralmente?

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo

Dal latino *textum*, intreccio di unità di significato (**monemi**) veicolato da simboli (**grafemi**).

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo secondo Segre 1985

Il testo è dunque una successione fissa di significati grafici.
Questi significati sono poi portatori di significati semantici.

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo secondo Tito Orlandi 2010

La produzione di un'attività linguistica intesa in senso stretto,
cioè in una delle lingue storicamente date

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Ancora Tito Orlando

Il Testo si presenta al lettore sotto la forma di **rappresentazione di un'espressione linguistica**, e poiché l'attività linguistica è multiforme, potrà essere una rappresentazione puramente mentale, *astratta*, ovvero *fonica* (verbale), ovvero *grafica* (scrittura).

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

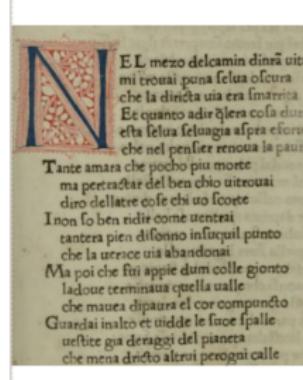
Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



NEL mezo del cammin di nostra vita,
Mi ritrovai per una selva oscura,
Che la dritta via era smarrita.
E quanto a dir qual era cosa dura!
Questa selva selvaggia, ed aspra e forte,
Che nel pensier rinnova la paura! 5
Tanto è amara che poco è più morte;
Ma, per trattar del ben ch' i' vi trovai,
Dirò dell' altre cose ch' i' v'ho scorte.
I' non so ben ridir com' i' v'entrai;
Tant'era pien di sonno in su quel punto, 10
Che la verace via abbandonai.
Ma po' ch' i' fui appiè d'un colle giunto,
Là ove terminava quella valle,
Che m'avea di paura il cor compunto; 15
Guardai' in alto, e vidi le sue spalle
Vestite già de' raggi del pianeta
Che mena dritto altri per ogne calle.

Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura,
ché la dritta via era smarrita.
Ah quanto a dir qual era cosa dura
esta selva selvaggia e aspra e forte
che nel pensier rinnova la paura!
Tant' è amara che poco è più morte;
ma per trattar del ben ch' i' vi trovai,
dirò de l' altre cose ch' i' v'ho scorte.
Io non so ben ridir com' i' v'intrai,
tant' era pien di sonno a quel punto
che la verace via abbandonai.
Ma poi ch' i' fui al piè d'un colle giunto,
là dove terminava quella valle
che m' avea di paura il cor compunto,
guardai in alto e vidi le sue spalle
vestite già de' raggi del pianeta
che mena dritto altri per ogne calle.

Possiamo concludere che

Il testo è l'invariante, la successione di valori invariabili rispetto alle variabili materiali dei caratteri, della scrittura (*principio di identità*).

Modellare il testo

Complessità del testo

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo oggetto complesso multidimensionale/pluralista

- analisi grafematica (caratteri, ideogrammi)
- analisi strutturale (struttura editoriale, interna)
- analisi metrica (piedi, versi, stanze, strofe)
- analisi stilistica e retorica
- analisi tematica
- analisi semantica
- ecc..

Modellare il testo

Il documento

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il Testo non è il Documento

La parola **documento** deriva dalla parola latina “*documentum*” che ha la stessa radice di “*doceo*” che significa “*insegnare*” / “*dimostrare*” con l’aggiunta del suffisso “*-umentum*” che intende **lo strumento per fare**.

Il Documento

La parola “*documento*” designa quindi uno **strumento** per **insegnare** o uno strumento atto a **dimostrare qualcosa**.

Modellare il testo

Il documento: Cercando una definizione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è un documento

- Una prova o una evidenza?
- Un atto cartaceo originale e/o ufficiale attraverso il quale dimostrare o provare qualcosa?
- Qualsiasi cosa materiale che serva come evidenza o prova?
- Un qualcosa di scritto contenente informazioni?
- Un supporto materiale sul quale sono stati rappresentati fatti o pensieri attraverso un qualche tipo di convenzione fatta di simboli e segni?

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Testo

Il testo è una entità astratta invariante, che in ogni operazione di realizzazione materiale della sequenza di simboli grafici determina la struttura fisica di un oggetto sensibilmente concreto (ovvero capace di attivare uno dei canali recettivi dell'uomo verso stimoli esterni).

Documento

L'oggetto materiale, sensibile, concreto che **costituisce il supporto**, stabile e riproducibile dell'informazione testuale che determina una modalità d'interazione (lettura, decifrazione).

Modellare il testo

Codifica del documento

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo non è il documento

La codifica di un documento è diversa dalla codifica di un testo
(testo != documento).

Documento digitalizzato

Corrispondenza perfetta con l'originale: **documento
digitalizzato**

Modellare il testo

Struttura del testo

Codifica di
Testi -
Introduzione

Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Elementi di un testo

Oltre alla sequenza di grafemi, **un testo si presenta strutturato in segmentazioni logiche** e partizioni interne di interi blocchi

Struttura di un testo

Il testo ha una certa struttura, i cui **elementi sono determinati dalla struttura logico-semanticà del discorso** con *finalità* e *funzioni* ben distinte (titolo di un capitolo, corpo di un paragrafo)

Modellare il testo

Struttura del documento

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Struttura di un documento

Un documento, ugualmente, è costituito da **elementi significativi, non solo verbali, da esplicitare formalmente** (layout, segmentazione, foliazione, aspetti paratestuali, aspetti tipografici, aspetti topografici).

Modellare il testo

Ordered Hierarchy of Content Objects (OHCO)

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

La natura del testo come OHCO

Gli oggetti di contenuto testuale sono perlopiù le **strutture editoriali astratte** di cui si compone un testo (divisioni logiche, segmentazioni e partizioni di un libro).

La natura del testo come OHCO

Questi oggetti sono **gerarchici** e di **contenuto** in quanto tali elementi ne possono contenere altri, ma sempre entità che veicolano contenuto. Sono **ordinati** in quanto esiste una *relazione lineare tra due oggetti posti sul medesimo livello gerarchico* (pensiamo alle parole di una frase all'interno di un paragrafo).

Modellare il testo

OHCO come primo modello concettuale TEI

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

OHCO e TEI

Il primo modello implementato dalla *Text Encoding Initiative* si è basato sull'**impianto teorico OHCO**.

OHCO e TEI

Sono però stati riscontrati una serie di *problem di rappresentazione ed espressività*: esistono diversi insiemi di **elementi di contenuto che non possono essere ricondotti a una struttura gerarchica unitaria** (pensiamo alla rappresentazione di parole e di righe).

Modellare il testo

Prospettive analitiche

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Strutture sovrapposte

Gli elementi di contenuto che si sovrappongono si comportano come se appartenessero a **diverse gerarchie di oggetti testuali.**

Prospettiva analitica

Ogni prospettiva analitica su un testo determina una struttura gerarchica di oggetti di contenuto. (F. Ciotti)

Modellare il testo

Sottoprospective

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Sottoprospective analitiche

In questa ottica **il testo diventa un sistema multi-dimensionale, ciascuna dimensione corrisponde a specifici punti di vista** dell'osservatore. (F. Ciotti)

Modellare il testo

La struttura gerarchica del testo

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Le gerarchie del testo

Le strutture del testo sono strutture gerarchiche: le sottoprospective, comunque esse siano definite, danno luogo a gerarchie.

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Digitalizzare e codificare un testo

Digitalizzare e codificare per favorire la **conservazione, l'accesso, elaborazione e il trattamento automatico** dei testi

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Trattamento dei testi

- analisi di tipo linguistico (linguistica computazionale, database testuali, corpora linguistics)
- analisi di altro tipo (metrica, stilistica, ecc.)
- ricerca testuale avanzata
- pubblicazione in vari formati (sul Web, come ebook, a stampa)
- didattica

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Digitalizzare un testo

Per facilitare e garantire una **universalità di accesso** al loro contenuto

Vantaggi della digitalizzazione

- Edizioni elettroniche garantiscono **diffusione** capillare (via Web) e nuove **funzionalità** (ipertesti, ricerca, ecc.)
- Permettono anche di **preservare** i documenti più **antichi** (e fragili) riducendone la *consultazione diretta*

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Superare i problemi dei documenti digitali

- disponibilità di hardware e software
- sistemi proprietari chiusi
- elevata obsolescenza e limitata manutenibilità
- difficile portabilità su piattaforme diverse

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Massimizzare le seguenti proprietà: Portabilità

- **indipendenza dall'hardware:** processore, supporto, output
- **indipendenza dal software:** sistemi operativi, applicazioni di authoring, applicazioni di visualizzazione
- **indipendenza dai sistemi di codifica dei caratteri**
- **indipendenza logica:** da un particolare processo applicativo

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Portabilità

Un documento digitale (come qualsiasi risorsa informativa digitale) è portatile se è **interscambiabile tra sistemi diversi**, **riutilizzabile** in molteplici processi computazionali anche a distanza di tempo, e **integrabile** da ulteriori risorse informative omogenee

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Standard

Le modalità di codifica devono divenire uno standard per **favorire la portabilità** del documento digitale.

Standard

Insieme di norme relative a una particolare tecnologia emesse da un **ente istituzionale** nazionale o internazionale deputato a tale scopo (*ISO - International Standardization Organization*).

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Standardizzare: vantaggi

- produrre **risultati confrontabili e condivisibili** all'interno di una comunità che adotta lo stesso tipo di standard
- lavorare in modo tale che i **prodotti** di oggi siano **usufruibili** anche in **futuro**
- codificare testi che possano essere usufruiti anche da chi non ha il tipo di **piattaforma hardware e software** che ha prodotto la codifica

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Linguaggio di codifica del contenuto

Il sistema di codifica deve assistere nel processo di **rappresentazione del testo** focalizzando l'attenzione sul **contenuto** (o sulla struttura del contenuto) piuttosto che sulla sua *forma grafica*.

Linguaggio di codifica dichiarativo

I sistemi di codifica dichiarativa si sono rivelati estremamente efficaci nel rappresentare strutture complesse come **riferimenti incrociati** e **collegamenti tra elementi** all'interno di un testo, ma anche tra più testi.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Linguaggio di codifica dichiarativo - Markup

I sistemi di markup dichiarativo introducono consistenti vantaggi poiché un medesimo **schema di codifica** dichiarativo che corrisponde ad un **modello di codifica** può essere utilizzato in *molteplici forme di trattamento informatico*.

markup dichiarativo: annotano la struttura e il significato degli elementi costitutivi del testo, tralasciandone l'aspetto.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Descrizione formale del testo

Così è possibile descrivere formalmente le caratteristiche di un testo in modo **indipendente** da particolari **finalità** di trattamento da contingenti **forme di presentazione** grafica su un qualsivoglia **supporto fisico**.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Linguaggio di codifica come linguaggio teorico

In questo modo la codifica permette allo studioso di **esPLICITARE** le sue ipotesi interpretative.

Linguaggio di codifica dichiarativo flessibile

I linguaggi di markup dichiarativi permettono di **predicare** l'appartenenza di un dato segmento testuale a una **classe di strutture testuali** definita dall'utente.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Gerarchia ordinata di oggetti di contenuto (OHCO)

I linguaggi di markup dichiarativi sono particolarmente **efficienti** dal punto di vista computazionale se modellano una **struttura gerarchica**.

Linguaggi di codifica gerarchici

La **sintassi** del linguaggio di codifica gerarchico può essere usata per **rappresentare le relazioni tra gli elementi** strutturali di un testo, a un determinato livello di descrizione.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Gerarchia ordinata di oggetti di contenuto (OHCO)

I linguaggi di markup dichiarativi gerarchici, in particolare **SGML** prima e **XML** poi, si sono rivelati ottimi strumenti di supporto alla **codifica e all'analisi computazionale dei testi**.

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Elementi di Codifica dei Caratteri

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Rappresentare il testo in formato digitale

L'adozione di metodologie informatiche per il trattamento dei testi richiede in primo luogo la disponibilità di un'adeguata **rappresentazione dei dati testuali in formato digitale.**

Elementi di Codifica dei Caratteri

Problemi di rappresentazione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

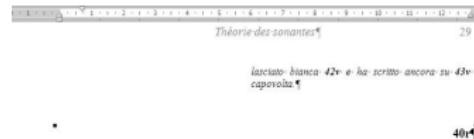
Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



«Nous commençons par dire que t était la "forme faible" de er.»

1. À quoi se reconnaît un sonantiste?
2. «nous pouvons examiner les objections contre la conception courante»
3. cf. ol 70, 40 Schmidt p. 50.¶
"Telles sont les réflexions qui se présentent si [¶]
4. C'est d'abord l'opposition contre l'existence de preuves historiques d'un t, puis d'un n qui se touche.¶
La partie du t se trouve dans les témoignages post-indo-européens.¶
5. Seule indication p. 3: "Es ist nur keineswegs gleichgültig [sic], ob man etc. ..."

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

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

¶On ne peut néanmoins que ces réflexions se présentent si t l m n [...] mais de là à reconnaître qu'il y ait une théorie.¶

41r

¶Si ceux-ci avaient réellement médité la nature du t [en phonologie, il se seraient, selon nous, non-seulement abstenus [...] mais positivement prononcés en-

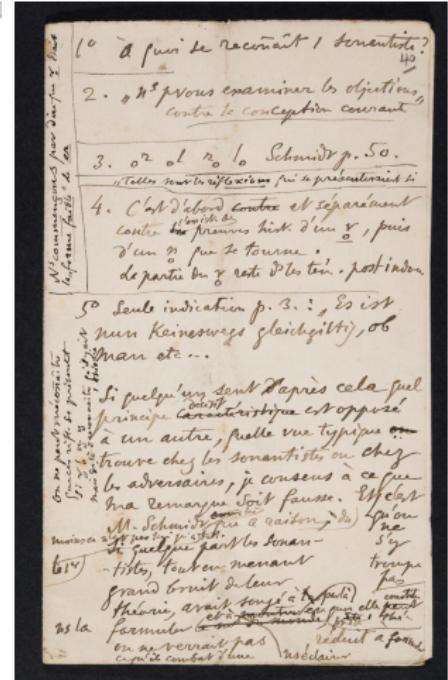
Sur margine sinistro, scritto traverso: Nous commençons... de gr [¶]

- control
- les t → l'existence de
- [caractéristique] → décisif
- que ce n'est y trompe pas
- et
- du moins ce n'est pas lui qui a le premier tort

gras [¶]

□ montrer → nous éclairer par là¶

¶Sur margine sinistro, scritto traverso: On ne peut... qu'il y ait une théorie



Elementi di Codifica dei Caratteri

Definizioni

Codifica di
Testi -
Introduzione

Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Perché è importante la codifica dei caratteri

La codifica dei caratteri costituisce il **grado zero** (basso livello) della rappresentazione di testi su supporto digitale.

Le codifiche dei caratteri sono la base di qualsiasi schema di codifica testuale.

Rappresentazione digitale dei caratteri

I caratteri vengono rappresentati all'interno di un elaboratore mediante una sequenza di codici binari costituiti da opportune disposizioni di cifre composte da 0 e 1: 01100001 *lettera a*

Elementi di Codifica dei Caratteri

American Standard Code for Information Interchange

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Tabella Code Page ASCII 7 bit

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
10	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{	}	~	DEL	

7 bit = 128 possibili caratteri; 32 caratteri di controllo; 96 caratteri effettivi

Elementi di Codifica dei Caratteri

Esempio codifica binaria

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

codifica *ciao mondo!* 7 bit ASCII

6369 616F 206D 6F6E 646F 210A

codifica *ciao è mondo!* 8 bit ASCII

6369 616f 20e8 206d 6f6e 646f 210a

codifica **ciao è mondo!** UNICODE UTF-8

6369 616f 20c3 a820 6d6f 6e64 6f21 0a

Elementi di Codifica dei Caratteri

Definizioni

Codifica di
Testi -
Introduzione

Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Character Set Per le discipline che studiano i sistemi di scrittura e l'analisi del linguaggio naturale, un **insieme di caratteri astratti** è detto Character set (unità alfabetiche). Astratto perché non riguarda la rappresentazione materiale della forma sul supporto, ma è relativo alla **forma mentale**, fatta di simboli di codifica (referenti).

Coded Char Set Per poter trattare un insieme di unità alfabetiche in formato digitale bisogna assegnare a ciascun carattere **un numero intero non negativo detto code point**.

Elementi di Codifica dei Caratteri

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Character Encoding Il fine ultimo della codifica è quello di rappresentare una sequenza di caratteri in una **sequenza di byte**. La codifica di un carattere utilizza uno “encoding schema” che a sua volta mappa o *trasforma ciascun code point in una sequenza di byte* e quindi in ultima istanza in una sequenza di bit.

Tabella del code page Generalmente i code points sono espressi attraverso un sistema numerico esadecimale e disposti in una tabella di associazione.

Elementi di Codifica dei Caratteri

In sintesi

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica dei caratteri

Quindi trasformare una sequenza di caratteri appartenenti ad un char set in una sequenza di byte (bit) significa prima di tutto trasformare/mappare ciascun carattere nel proprio corrispettivo code point e successivamente codificare/serializzare questo code point nella relativa sequenza di byte (bit).

Elementi di Codifica dei Caratteri

Complessità e rappresentazione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Complessità di rappresentazione universale dei caratteri

Se si considerano tutti i possibili alfabeti del mondo e le molteplici esigenze poste dalla scrittura delle fonti manoscritte antiche e medievali, ci si accorge che la realizzazione di un sistema universale per la codifica dei caratteri è un progetto molto complesso con svariate sfide da affrontare.

Complessità della Codifica dei Caratteri

Un Esempio

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Complessità della Codifica dei Caratteri

Un Esempio

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

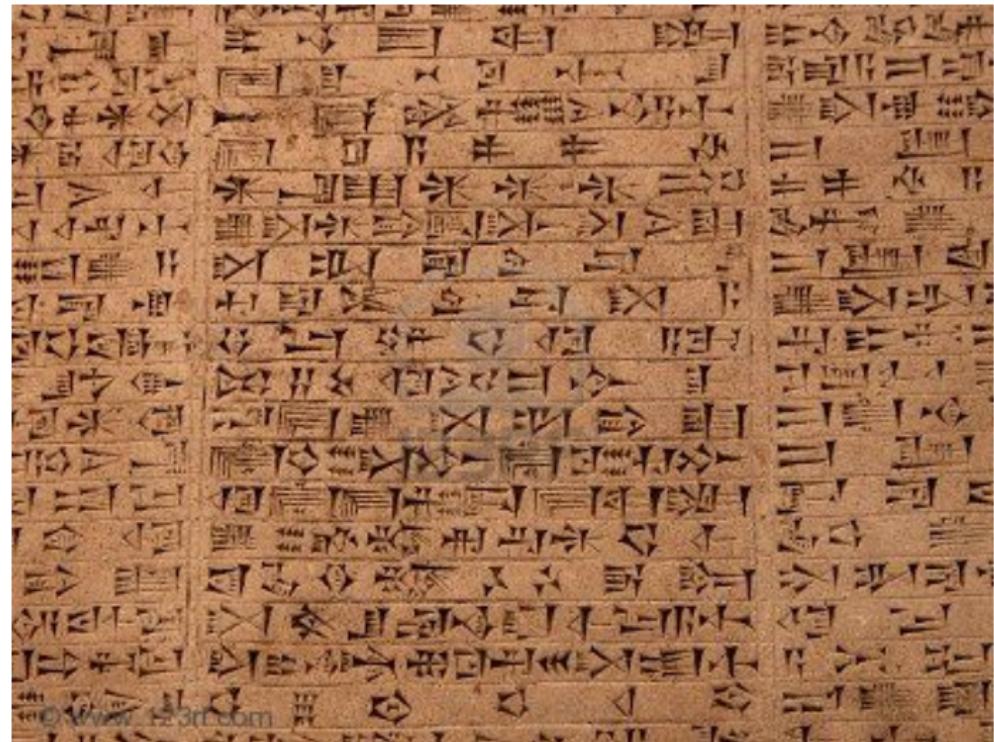
Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



Elementi di Codifica dei Caratteri

Unicode

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Complessità di rappresentazione universale

Ad oggi, lo standard de facto per la codifica dei caratteri è lo **UNICODE**. Esso è in grado di codificare **più di un milione di differenti unità alfabetiche**, segni di interpunkzione e diacritici, appartenenti a centinaia di diverse lingue.

Complessità di rappresentazione universale

Unicode assegna i propri code point in un range che va da 0x0 a 0x10FFFF. In Unicode il code point viene indicato con una "U" seguita da un segno "+" seguito a sua volta dall'esadecimale con padding del codice (es: U+0041 lettera A).

Elementi di Codifica dei Caratteri

Unicode

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Unicode Transformation Format

Lo Unicode è un Coded Char Set e per essere concretamente serializzato su un supporto elettronico deve essere trasformato attraverso qualche tipo di schema di codifica. L'UTF (Unicode Transformation Format) mappa i code point Unicode in sequenze di byte (bit).

UTF standards

Esistono tre tipi di schemi di codifica che vanno sotto il nome di UTF, ciascuno è identificato dal minimo numero di bit necessario a codificare ciascun code point: UTF-8; UTF-16; UTF-32.

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Markup language e XML

soluzione corrente per la codifica dei testi

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

```
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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

- 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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Vantaggi

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

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo

L'applicazione di metodologie computazionali nell'ambito della ricerca umanistica comporta due aspetti di formalizzazione

Due elementi

Formalizzazione dei dati e formalizzazione dell'elaborazione

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Formalizzazione dei dati

Definizione e implementazione di strutture dati adeguate alla cattura dei fenomeni di interesse dell'umanista, e in particolare alla rappresentazione formale dei testi;

Formalizzazione dell'elaborazione

Specificazione di algoritmi che, applicati alle strutture dati, siano in grado di simulare i processi di manipolazione dei testi tipici della ricerca umanistica o in generale delle pratiche sociali che hanno a che fare in vario modo con i testi.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo

Il problema della codifica testuale rientra in generale nel primo tipo di formalizzazione (Dati).

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Problemi teorici

La codifica è un processo assai più complesso della semplice e meccanica correlazione biunivoca di strutture rappresentazionali.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Norme TEI

Le linee guida di codifica TEI impongono a chi codifica di effettuare delle scelte teoriche e interpretative che non sono imputabili alla semplice trascrizione.

Processo di codifica

La codifica è un processo interpretativo non solo un processo riproduttivo.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica come interpretazione

Conseguentemente ogni processo di codifica (inclusi quelli di codifica informatica del testo) è il risultato di una interpretazione da esplicitare e formalizzare.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Esempio

L'assunzione che una data traccia grafica "A" sia una istanza di una data classe astratta di tracce che identifichiamo come il carattere "a" nella sua forma maiuscola richiede molti sforzi interpretativi e quindi cognitivi.

Rappresentazione e interpretazione

In linea di principio non è sempre possibile dire in modo non ambiguo che una traccia su un supporto fisico appartiene a una certa classe di iscrizioni che chiamiamo carattere.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Certezza e soggettività

Ogni interpretazione può godere di diversi gradi di certezza e di soggettività. In ogni caso non esiste nessun genere di rappresentazione di un testo che si possa definire libera da processi interpretativi.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Linguaggio teorico

Lo schema di codifica è un linguaggio teorico usato per costruire teorie o modelli di fenomeni testuali

Linguaggio teorico

La stessa costruzione di un linguaggio teorico riflette un determinato modello del mondo (soggettivo o condiviso).
Si presuppone una teoria ontologica del testo.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Obiettivo

- Sviluppare teorie e modelli formali del testo (o di alcuni suoi livelli descrittivi)
- Individuare formalismi atti a esprimerli in modo computazionalmente accettabile

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

OHCO

Il modello di struttura gerarchica del testo si sposa perfettamente con XML, il quale può essere considerato sia un formalismo sia un modello di dati espresso da quel formalismo, e tale meta-modello è appunto un albero ordinato etichettato.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione

Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

OHCO: modello di dati così vincolante

Il modello OHCO presenta dei vincoli, che vanno sotto il nome di “problema delle gerarchie sovrapposte” (overlapping hierarchies).

Gerarchie Sovrapposte

Dati due oggetti logici presenti in un testo, le coppie di tag bilanciati che li rappresentano non si annidano propriamente ma si sovrappongono. *Tale situazione è sintatticamente e semanticamente vietata in XML.*

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Modelli gerarchici

Vantaggio di poter validare in modo automatico ogni istanza di dati rispetto al modello mediante algoritmi generali ben conosciuti e computazionalmente trattabili, ciò che a sua volta consente di costruire sistemi di elaborazione degli stessi dati consistenti ed efficaci (al netto dei costrutti ID/IDREF).

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Modelli non gerarchici

Negli ultimi anni la ricerca teorica sullo sviluppo di sistemi di markup non gerarchici ha avuto un notevole stimolo

Modelli non gerarchici

Tutte le soluzioni proposte finora hanno grossi limiti per quel che concerne la facilità di gestione e manutenzione

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Ricapitolando

Il trasferimento del testo su supporto informatico pone allo studioso una serie di **quesiti teorici** (oltre che a numerosi problemi pratici) a partire dalla determinazione di quale particolare oggetto del mondo considerare come "**fonte**" della codifica e memorizzazione.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

In conclusione

Secondo la teoria OHCO un testo è una struttura gerarchica di oggetti logici, e la codifica non fa altro che esplicitare questa sua struttura essenziale.

In conclusione

Individuazione di diverse unità o elementi di contenuto testuale a partire da un punto di vista interno alle discipline che studiano ed analizzano il testo.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Punti di vista

TEI ha una struttura modulare, in cui ogni modulo corrisponde alla rappresentazione di un determinato punto di vista metodologico sul testo.

TEI e Gerarchie sovrapposte

L'impianto finale dello schema di codifica della Text Encoding Initiative ha accolto in parte la nozione di sottoprospettiva ed ha sviluppato una serie di costrutti sintattici e semantici in grado di rappresentare adeguatamente fenomeni di sovrapposizione e di parallelismo tra elementi testuali in XML.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

TEI e Gerarchie sovrapposte

La TEI ha sviluppato un sistema di codifica sufficientemente potente da permettere di rappresentare le caratteristiche testuali che interessano e di poter esplicitare le sue interpretazioni sul ruolo di tali caratteristiche.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Scopo applicativo

Occorre tenere presente nella rappresentazione del testo anche i possibili processi applicativi a cui esso può essere sottoposto.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

La codifica informatica di un testo

La codifica informatica di un testo può essere vista come il prodotto di un insieme di inferenze che vengono espresse mediante un linguaggio formalizzato (Ciotti).

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo: riassumendo

- elaborazione di un quadro teorico di riferimento entro cui collocare i procedimenti analitici
- definizione di un modello di rappresentazione informatica o codifica del testo e delle strutture rilevanti in relazione al testo di riferimento
- individuazione di metodi e processi di analisi testuale applicabili al modello del testo e loro definizione sottoforma di procedure formali o algoritmi

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo: riassumendo

- implementazione del modello di rappresentazione e dei processi di analisi mediante adeguati linguaggi informatici
- applicazione delle procedure informatiche al testo digitalizzato
- analisi e interpretazione critica dei risultati

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo: riassumendo

Si tratta di costruire un modello informatico del testo nel quadro di un determinato contesto teorico, per poi interrogare opportunamente tale modello e avanzare ipotesi interpretative sul testo.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Non è codifica di un testo

- fare una scansione di un documento e diffonderne le immagini (digitalizzazione)
- usare un software OCR per ricavarne una versione in formato testo
- creare una pagina html di un testo
- creare un documento con software di word processing
- creare banche dati con testo destrutturato

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Fare la codifica di un testo significa

- convertirlo in un formato comprensibile al calcolatore elettronico
- usare, a tal fine, un linguaggio di codifica di tipo formale
- stabilire le caratteristiche del testo che si intendono esplicitare (modello di codifica)
- definire e seguire uno schema di codifica ben preciso

References

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



Ciotti, F. (2007). Il testo e l'automa: saggi di teoria e critica computazionale dei testi letterari. Aracne.



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.



Segre, C. (1985). Avviamento all'analisi del testo letterario. Einaudi.



Segre, C. (1979). Critica testuale, teoria degli insiemi e diasistema, in Semiotica filologica. Einaudi.

References

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



DeRose, S. J., Durand, D. G., Mylonas, E., and Renear, A. H. (1990). What is text, really? *Journal of Computing in Higher Education*, 1(2), 3–26.



Renear, Allen H.; Mylonas, Elli; Durand, David (1995). Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies.



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>

References

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



TEI Guide Reference. <http://www.tei-c.org/>

Codifica di Testi - Introduzione XML Markup

a.a. 2023-2024

Angelo Mario Del Grosso

angelo.delgrosso@ilc.cnr.it

CNR-ILC

Istituto di Linguistica Computazionale “A. Zampolli”,
8th March 2024

Contenuto della lezione

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

I linguaggi di codifica

introduzione

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Definizione di codifica digitale del testo

Per **codifica** digitale dei testi intendiamo la *rappresentazione formale* di un **testo** ad un qualche livello descrittivo, su di un supporto digitale, in un formato utilizzabile da un elaboratore (*Machine Readable Form*) mediante un opportuno **linguaggio informatico** (F. Ciotti).

I linguaggi di codifica

Riassumendo

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Impostazione teorico-pratica

- un testo è molto di **più della sequenza di caratteri** che lo compongono
- per mezzo della codifica vogliamo **rendere esplicite le caratteristiche** che vogliamo analizzare
- solo quello che è esplicito può essere **interpretato ed elaborato dal computer**
- vogliamo codificare il **testo per quello che è**, non per quello che sembra
- codifica da effettuare mediante **linguaggio di markup**

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Il markup

Il termine **markup** è stato utilizzato in passato per denotare i **segni grafici** che accompagnavano un testo apposti sul documento per **indicare correzioni o modalità grafiche di stampa**.

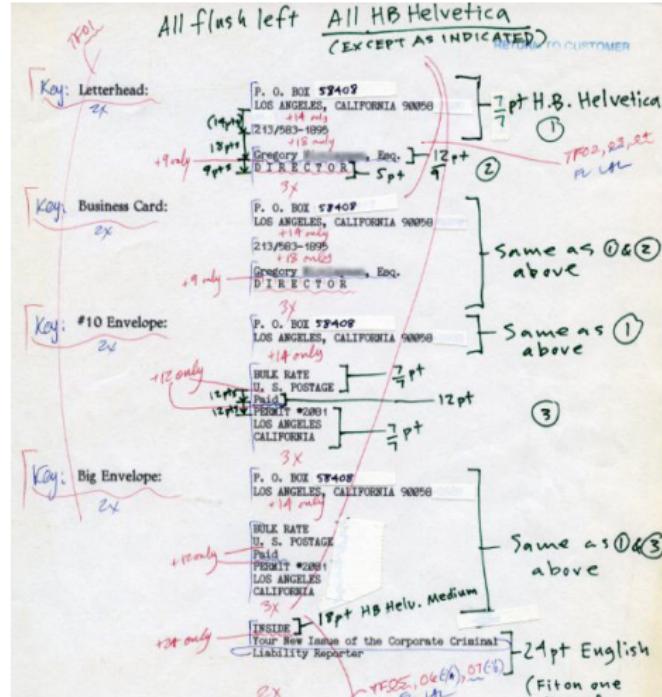
I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di codifica



I linguaggi di codifica

Linguaggi di marcatura

Codifica di Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

<u>g</u> <u>e</u>	Delete	<u>fL</u>	Flush Left	<u>bFital</u>	Set in Bold Face Italic
<u>x</u>	Insert	<u>fR</u>	Flush Right	<u>lf</u>	Set in Light Face
<u>c</u>	Join	<u>JC</u>	Center Horizontally	<u>wF</u>	Wrong Font
<u>-</u>	Move closer	<u>CV</u>	Center Vertically	<u>-/</u> <u>=/</u>	Hyphen
<u>#</u>	Space	<u>[and]</u>	Move to the next line	<u>1/</u> <u>en</u>	En Dash
<u>#</u>	Add Space	<u>end.</u>	Move to the preceding line	<u>1/</u> <u>em</u>	Em Dash
<u>#</u> <u>de</u>	Delete Space	<u>□</u>	Indent 1 em	<u>S</u>	Superscript
<u>rw</u>	Transpose Word	<u>□□</u>	Indent 2 ems	<u>^</u>	Subscript
<u>rl</u>	Transpose Letters	<u>P</u>	Paragraph	<u>,</u>	Comma
<u>/</u>	To separate two or more marks	<u>≡</u>	All Caps	<u>'</u>	Apostrophe
<u>Stet</u>	Let it Stand (ignore correction)	<u>==</u>	Small Caps	<u>○</u>	Period
<u>l</u>	Move Left	<u>=====</u>	Caps & Small Caps	<u>:</u> <u>;/</u>	Semicolon
<u>r</u>	Move Right	<u>Caps</u>	Capital Letter	<u>:</u> <u>①</u>	Colon
<u>u</u>	Move Up	<u>lc</u>	Lower Case	<u>“”</u> <u>‘’</u>	Quotation Marks
<u>d</u>	Move Down	<u>Rom</u>	Set in Roman	<u>(/)</u>	Parentheses
<u> </u>	Align Vertically	<u>ital</u>	Set in Italic	<u>[/]</u>	Brackets
<u>—</u>	Align Horizontally	<u>bF</u>	Set in Bold Face		

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Il markup

La codifica con linguaggi di marcatura (markup) è in sostanza **un insieme di convenzioni**, rese attraverso specifiche **sequenze di caratteri, etichette, codici**, (detti *tags*) **intercalati nel testo** per permettere agli elaboratori elettronici di distinguere le varie parti di un documento.

Il markup formale

Un linguaggio di markup è un **sistema formale** per *scambiare e pubblicare* informazioni in **formato testo in modo strutturato**.

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Il markup formale

Markup formale: costituito da un **sistema non ambiguo** di istruzioni, ognuna delle quali è **dotata di una specifica semantica e sintassi**.

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Diversi tipi di markup

Esistono diversi linguaggi di markup, per rappresentare diversi tipi di documenti.

- **Linguaggi procedurali** (specific markup languages)
- **Linguaggi dichiarativi** (generic markup languages)

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Linguaggi procedurali

- **Orientati al documento**, indicando come deve essere elaborato e *disposto il testo in resa grafica*
- Istruzioni da inserire nel testo per connotarne specifiche *caratteristiche di visualizzazione*
- Font, dimensione, spaziatura del carattere, posizionamento nella pagina, colore, etc.

Esempi: *TeX e LaTeX, RTF*

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Esempio RTF

```
{\rtf1\ansi\deff0\adeflang1025
{\fonttbl{\f0\froman\fprq2\fcharset0 Times New Roman;}}
{\f1\froman\fprq2\fcharset0 Times New Roman;}}
{\f2\fnil\fprq2\fcharset0 Lucida Sans Unicode;}}
{\colortbl;\red0\green0\blue0;\red128\green128\blue128;}
{\stylesheet{\s1\cf0{*\hyphen2\hyphlead2\hyphtrail2\hyphmax0}
\rtlch\af5\afs24\lang255\ltrch\dbch\af2\afs24\langfe255
\loch\f0\fs24\lang1040\snext1 Standard;}}
```

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Esempio LaTex

```
\documentclass [a4paper,10pt] {article} \usepackage[utf8] {inputenc}
\usepackage[T1] {fontenc}
\usepackage[italian] {babel}
\title {Il mio primo documento}
\author {Angelo Mario Del Grosso}
\begin {document}
\maketitle
\begin {abstract}
Primo tentativo di scrivere in \LaTeX .
\end {abstract}
\section {titolo della sezione}
Questo documento è vuoto.
\footnote {nota a piè di pagina.}

\end {document}
```

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Il markup procedurale

L'unico utilizzo di un testo codificato tramite un linguaggio procedurale è la **creazione di un output orientato alla visualizzazione.**

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Scrivere la tesi di laurea con L^AT_EX 2 _{ε}

Dipartimento di Ingegneria Meccanica,
Nucleare e della Produzione
Università di Pisa
56126 Pisa PI

Sommario

Lo scopo del presente articolo è fornire gli strumenti per scrivere una tesi di laurea utilizzando L^AT_EX 2 _{ε} . Tale obiettivo è conseguito analizzando i problemi tipici incontrati durante la stesura della tesi e le possibili soluzioni; si pone particolare attenzione ai pacchetti da usare nelle varie circostanze. I singoli argomenti non vengono approfonditi nei dettagli ma si rimanda alla letteratura specifica o ad i manuali dei pacchetti suggeriti, ove necessario.

*Ringrazio in primo luogo Fabiano Busdraghi che ha collaborato alla scrittura delle sezioni riguardanti le figure e gli oggetti flottanti. Ringrazio inoltre tutti coloro che mi hanno consigliato durante la stesura e la revisione di questo documento ed in particolare Claudio Beccari, Gustavo Cevolani, Massimo Guiggiani, Maurizio Himmelmann, Lorenzo Pantieri e Emiliano Vavassori.

I linguaggi di codifica

Linguaggi di marcatura dichiarativi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Linguaggi dichiarativi

Orientati al testo, annotano la *struttura*, la *funzione* ed il *significato* degli elementi costitutivi del testo, **tralasciandone l'aspetto**.

- La posizione che il brano in questione occupa all'interno del documento (**markup strutturale**)
- Peculiarità del testo stesso (**markup semantico**)
- I fogli di stile definiscono la formattazione dell'output
- *Molteplici usi del medesimo testo*

Esempio: famiglia SGML, XML

I linguaggi di codifica

Linguaggi di marcatura dichiarativi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Markup dichiarativi: contenuto e presentazione

La **separazione tra contenuto e presentazione** non solo è intenzionale, ma è la **caratteristica principale** di questi sistemi di marcatura: essa permette di concentrarsi sull'**annotazione logica-semantica** per funzioni di *ricerca e di analisi*, lasciando ad altro (ai fogli di stile) la resa grafica.

Unico testo più usi

In questo modo si ha inoltre la possibilità di utilizzare uno **stesso testo codificato con finalità o formattazioni differenti**, a seconda delle varie esigenze.

I linguaggi di codifica

Markup dichiarativi: esempio SGML

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Standard Generalized Markup Language

```
1 <!DOCTYPE testo [  
2   <!ELEMENT testo (titolo?, paragrafo+)>  
3   <!ELEMENT titolo (#PCDATA)>  
4   <!ELEMENT paragrafo (#PCDATA)>  
5 ]>  
6 <testo>  
7   <titolo> Questo è il titolo del documento</titolo>  
8   <paragrafo> Questo è un paragrafo </paragrafo>  
9 </testo>  
10
```

I linguaggi di codifica

Markup dichiarativi vs Markup procedurali

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

resa a video della frase

Le *Guidelines for Electronic Text Encoding and Interchange* sono molto complete e descrivono uno standard di *markup* del testo basato su XML.

Le \textit{Guidelines for Electronic Text Encoding and Interchange} sono \textit{molto} complete e descrivono uno standard di \textit{markup} del testo basato su XML.

<titolo>Le Guidelines for Electronic Text Encoding and Interchange</titolo> sono <enfasi>molto</enfasi> complete e descrivono uno standard di <linguastraniera> markup</linguastraniera> del testo basato su XML.

LaTeX vs SGML

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

linguaggi semi-dichiarativi e/o semi-procedurali

Esistono anche linguaggi che possono essere definiti **semi-procedurali**, o **semi-dichiarativi**, che come si intuisce utilizzano le istruzioni sia per una codifica di tipo procedurale, sia per una codifica di tipo descrittivo o dichiarativo.

HTML

HTML ha tra le sue etichette istruzioni di tipo procedurale per indicare come devono essere rese determinate porzioni di testo, e istruzioni di tipo dichiarativo che hanno una base semantica.

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML origini

XML affonda le proprie origini nel linguaggio **Standard Generalized Markup Language (SGML)**.

SGML è stato introdotto negli anni ottanta con il fine di **descrivere la struttura e il contenuto di qualsiasi informazione** “machine readable”.

XML è una semplificazione di SGML

XML può essere pensato come una **versione semplificata di SGML**. Infatti, come SGML, XML è *un meta-linguaggio*, usato per *create linguaggi di marcatura* (detti **vocabolari**).

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML come meta-linguaggio

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

Applicazioni XML

Allo stesso modo di SGML, XML è nato per **strutturare, conservare e trasportare** informazioni.

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML: eXtensible

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

XML: standard W3C

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML: riassumendo

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

XML: riassumendo

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

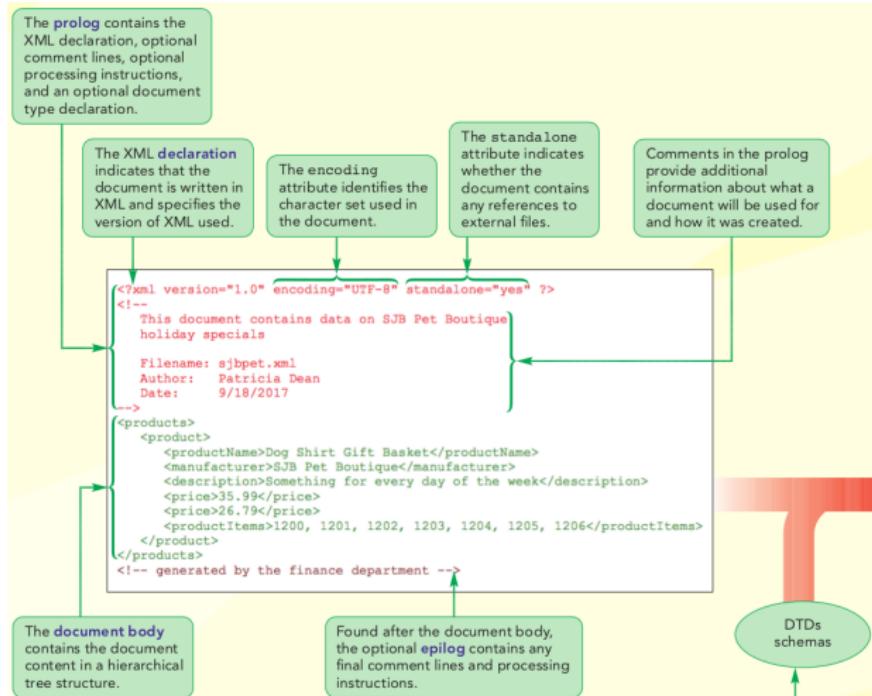


immagine dal libro *New Perspectives on XML, 3rd Edition*

Fondamenti XML

eXtensible Markup Language: regole sintattiche

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

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

Fondamenti XML

eXtensible Markup Language: regole sintattiche cont.

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Manutenibilità

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML vista ad albero

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

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

Fondamenti XML

eXtensible Markup Language: vista ad albero

Codifica di Testi - Introduzione XML Markup a.a. 2023-2024

A.M. Del Grosso

I linguaggi di codifica

Fondamenti del linguaggio XML

Validare un documento XML e Definire uno Schema

Conclusioni

XML document

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

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

Hierarchy tree structure

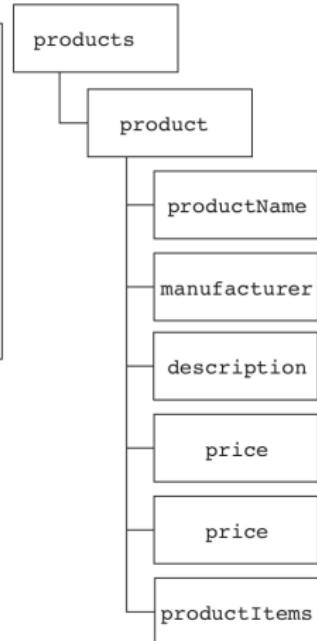


immagine dal libro New Perspectives on XML, 3rd Edition

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

TEI-XML vocabulary

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

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

Fondamenti XML

eXtensible Markup Language: Esempio TEI

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

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

immagine dal sito TEI Guide Lines

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Documento ben formato (well-formed)

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

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Parti principali di un documento XML

Un documento XML consiste di tre parti:

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

Fondamenti XML

eXtensible Markup Language: Esempio TEI

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

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

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

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Documento XML: prologo

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Documento XML: corpo

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

Documento XML: epilogo

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

Fondamenti XML

eXtensible Markup Language: Prologo

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML declaration

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

Fondamenti XML

eXtensible Markup Language: Prologo

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML declaration: ERRORI

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

Fondamenti XML

eXtensible Markup Language: Prologo

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML comments

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

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

XML comments: sintassi

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

Un commento può occupare anche più righe

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

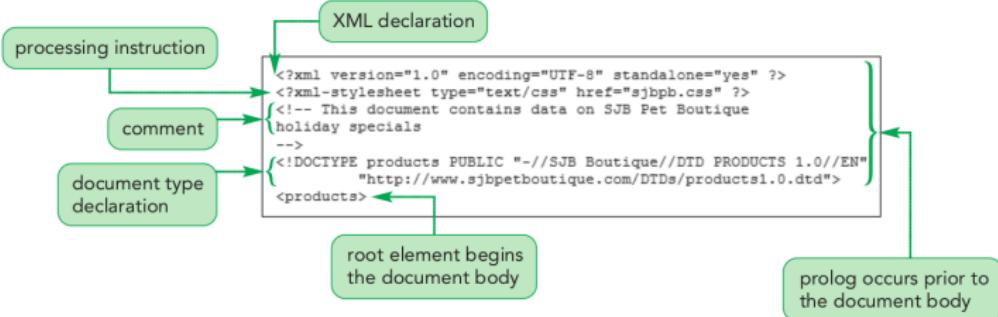


immagine dal libro New Perspectives on XML, 3rd Edition

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Esercizio prologo

Creare un file `.xml` ed inserire un prologo con la dichiarazione XML e un commento con le vostre informazioni.

Esercizio prologo

```
<!-- This document contains data on Codifica di  
Testi.
```

Filename: project.xml

Author: your name

Date: today's date -->

Salvare il file su github nel repository del progetto

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML parser

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

Cosa fa un XML parser

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XMLlint

XMLINT(1)

xmllint Manual

XMLINT(1)

NAME

xmllint - command line XML tool

SYNOPSIS

```
xmllint [--version | --debug | --shell | --xpath "XPath_expression" | --debugent | --copy |
--recover | --noent | --noout | --nonet | --path "PATH(S)" | --load-trace | --htmlout |
--nowrap | --valid | --postvalid | --dtdvalidfpi URL | --dtdvalidfpi FPI | --timing |
--output FILE | --repeat | --insert | --compress | --html | --xmlout | --push | --memory |
--maxmem NBBYTES | --nowarning | --noblanks | --nocdata | --format | --encode ENCODING |
--dropdtd | --nsclean | --testIO | --catalogs | --nocatalogs | --auto | --xinclude |
--noxincludenode | --loaddtd | --ddattr | --stream | --walker | --pattern PATTERNVALUE |
--chkregister | --relaxng SCHEMA | --schema SCHEMA | --c14n] {XML-FILE(S)... | -}
```

xmllint --help

DESCRIPTION

The **xmllint** program parses one or more XML files, specified on the command line as **XML-FILE** (or the standard input if the filename provided is **-**). It prints various types of output, depending upon the options selected. It is useful for detecting errors both in XML code and in the XML parser itself.

xmllint is included in **libxml(3)**.

OPTIONS

xmllint accepts the following options (in alphabetical order):

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML body

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

Elementi: Sintassi

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

Un elemento può contenere testo e/o ulteriori elementi

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element

Gli elementi XML possono avere diversi tipi di contenuto:

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element: note importanti sul nome

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element: empty e nested

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

XML esempi: empty e nested element

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element: hierarchical relationship

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

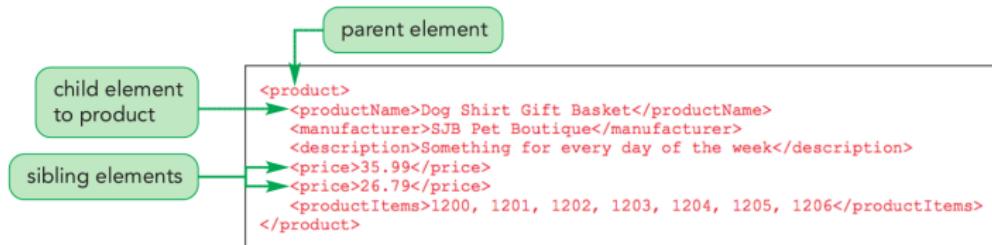


immagine dal libro New Perspectives on XML, 3rd Edition

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element: hierarchical relationship

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element: hierarchical relationship cont.

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element: hierarchical relationship - Esercizio

Scrivere e fare il check di un xml non opportunamente annidato

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element: hierarchical relationship as tree structure

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

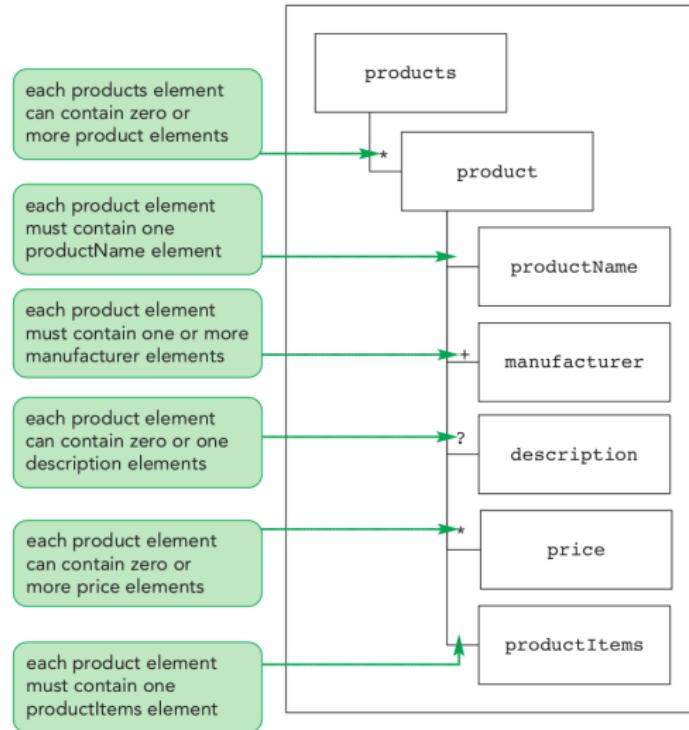
A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni



Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element: Mixed Content

Un elemento può contenere contemporaneamente sia testo sia altri elementi.

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

XML Element: Mixed Content

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Element: Esercizio

Aprire il file XML non ben formato presente nel repository
github:

- validarlo con un parser XML
- correggerlo (commentando gli errori e le modifiche)
- aggiungere un figlio (child) ad un elemento
- aggiungere un fratello (sibling) ad un elemento

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Attributi

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

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

XML Attributi

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Attributi: valore

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

XML Attributi: valore

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Attributi: restrizioni ai nomi

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

XML Attributi

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML Character and Entity References

- numeric character reference: &#nnn;
- hexadecimal character reference: &#xhhh
- character entity reference: &entity;

XML References

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

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

immagine dal libro New Perspectives on XML, 3rd Edition

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Text Character Parsing

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

Text Character Parsing

- PCDATA
- CDATA
- White Space

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Parsed Character Data

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

PCDATA

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Parsed Character Data

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

XML PCDATA

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Character Data

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

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

XML CDATA: sintassi

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Character Data

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

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

CDATA: qualche vincolo

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Esempio ed esercizio

Inserire all'interno di un tag un frammento di codice HTML

CDATA: esempio

```
<htmlCode> <! [CDATA[ <h1>Capitolo Primo</h1>
    <h2>Sezione Seconda</h2> ]]> </htmlCode>
```

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

White Space: esempio

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

I white space sono caratteri non stampabili

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Processing Instruction

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

Processing Instruction: sintassi

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

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Processing Instruction: sintassi

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

Processing Instruction

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

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Namespaces

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

Namespaces

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

Fondamenti XML

eXtensible Markup Language

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

Namespaces

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

Namespaces

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

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

XML per rappresentare il testo

- I markup language per supportare la rappresentazione, memorizzazione, pubblicazione di un testo.
- XML è un markup language flessibile e potente.
- le istruzioni dei markup language sono per lo più dichiarazioni indicando particolari funzioni del dato.
- le istruzioni sono etichette visibili.

Elementi XML

Conclusioni

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML per rappresentare il testo

- Una sintassi e una grammatica regolano l'applicabilità del linguaggio di marcatura
- Sintassi: documento well formed (ben formato)
- Grammatica: documento valido

Elementi XML

Conclusioni

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Conclusioni

XML per rappresentare il testo

- XML deriva dal linguaggio SGML.
- XML è una specifica del consorzio W3C.
- XML è un meta-linguaggio.
- XML è plain text.
- XML è portable.

XML per rappresentare il testo

- XML definisce markup dichiarativi e descrittivi.
- XML ha un modello dati ad albero ordinato.
- XML può avere associato un tipo di documento (DTD) o uno schema (XSD).

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Codifica di Testi - Introduzione XML Markup

a.a. 2023-2024

Angelo Mario Del Grosso

angelo.delgrosso@ilc.cnr.it

CNR-ILC

Istituto di Linguistica Computazionale “A. Zampolli”,
13th March 2024

Contenuto della lezione

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

- Document Type Definition (DTD)
- XML Schema Definition (XSD)
- RELAX NG

4 Conclusioni

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

I linguaggi di codifica

introduzione

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Definizione di codifica digitale del testo

Per **codifica** digitale dei testi intendiamo la *rappresentazione formale* di un **testo** ad un qualche livello descrittivo, su di un supporto digitale, in un formato utilizzabile da un elaboratore (*Machine Readable Form*) mediante un opportuno **linguaggio informatico** (F. Ciotti).

I linguaggi di codifica

Riassumendo

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Impostazione teorico-pratica

- un testo è molto di **più della sequenza di caratteri** che lo compongono
- per mezzo della codifica vogliamo **rendere esplicite le caratteristiche** che vogliamo analizzare
- solo quello che è esplicito può essere **interpretato ed elaborato dal computer**
- vogliamo codificare il **testo per quello che è**, non per quello che sembra
- codifica da effettuare mediante **linguaggio di markup**

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Il markup

Il termine **markup** è stato utilizzato in passato per denotare i **segni grafici** che accompagnavano un testo apposti sul documento per **indicare correzioni o modalità grafiche di stampa**.

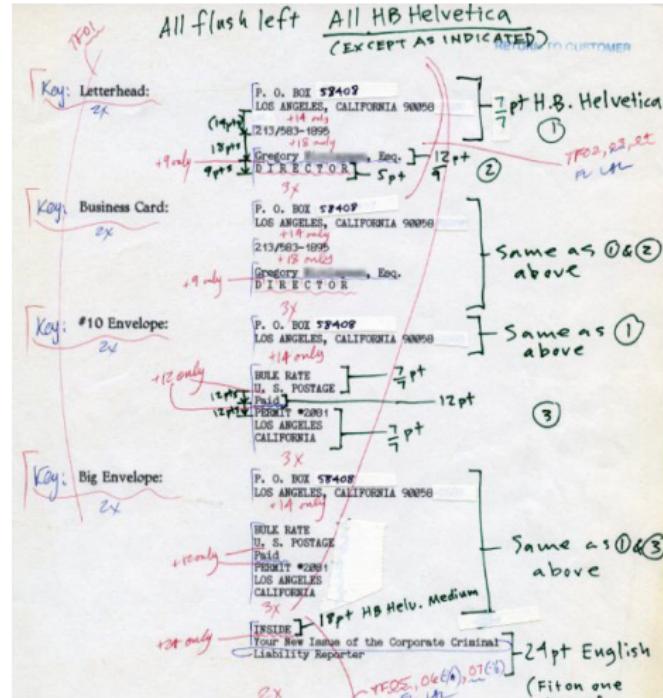
I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di codifica



I linguaggi di codifica

Linguaggi di marcatura

Codifica di Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

<u>g</u> <u>e</u>	Delete	<u>fL</u>	Flush Left	<u>bFital</u>	Set in Bold Face Italic
<u>x</u>	Insert	<u>fR</u>	Flush Right	<u>lf</u>	Set in Light Face
<u>c</u>	Join	<u>jc</u>	Center Horizontally	<u>wF</u>	Wrong Font
<u>-</u>	Move closer	<u>cv</u>	Center Vertically	<u>-/</u> <u>=/</u>	Hyphen
<u>#</u>	Space	<u>[and]</u>	Move to the next line	<u>en</u>	En Dash
<u>#</u>	Add Space	<u>end.</u>	Move to the preceding line	<u>em</u>	Em Dash
<u>#</u>	Delete Space	<u>□</u>	Indent 1 em	<u>^</u>	Superscript
<u>tr</u>	Transpose Word	<u>□□</u>	Indent 2 ems	<u>^</u>	Subscript
<u>tr</u>	Transpose Letters	<u>P</u>	Paragraph	<u>,</u>	Comma
<u>/</u>	To separate two or more marks	<u>≡</u>	All Caps	<u>'</u>	Apostrophe
<u>Stet</u>	Let it Stand (ignore correction)	<u>==</u>	Small Caps	<u>○</u>	Period
<u>[</u>	Move Left	<u>==</u>	Caps & Small Caps	<u>:</u> <u>;</u>	Semicolon
<u>]</u>	Move Right	<u>Caps</u>	Capital Letter	<u>○</u>	Colon
<u>↑</u>	Move Up	<u>lc</u>	Lower Case	<u>“”</u> <u>‘’</u>	Quotation Marks
<u>↓</u>	Move Down	<u>Rom</u>	Set in Roman	<u>(/)</u>	Parentheses
<u> </u>	Align Vertically	<u>ital</u>	Set in Italic	<u>[/]</u>	Brackets
<u>—</u>	Align Horizontally	<u>bF</u>	Set in Bold Face		

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Il markup

La codifica con linguaggi di marcatura (markup) è in sostanza **un insieme di convenzioni**, rese attraverso specifiche **sequenze di caratteri, etichette, codici**, (detti *tags*) **intercalati nel testo** per permettere agli elaboratori elettronici di distinguere le varie parti di un documento.

Il markup formale

Un linguaggio di markup è un **sistema formale** per *scambiare e pubblicare informazioni in formato testo in modo strutturato*.

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Il markup formale

Markup formale: costituito da un **sistema non ambiguo** di istruzioni, ognuna delle quali è **dotata di una specifica semantica e sintassi**.

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Diversi tipi di markup

Esistono diversi linguaggi di markup, per rappresentare diversi tipi di documenti.

- **Linguaggi procedurali** (specific markup languages)
- **Linguaggi dichiarativi** (generic markup languages)

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Linguaggi procedurali

- **Orientati al documento**, indicando come deve essere elaborato e *disposto il testo in resa grafica*
- Istruzioni da inserire nel testo per connotarne specifiche *caratteristiche di visualizzazione*
- Font, dimensione, spaziatura del carattere, posizionamento nella pagina, colore, etc.

Esempi: *TeX e LaTeX, RTF*

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Esempio RTF

```
{\rtf1\ansi\deff0\adeflang1025
{\fonttbl{\f0\froman\fprq2\fcharset0 Times New Roman;}}
{\f1\froman\fprq2\fcharset0 Times New Roman;}}
{\f2\fnil\fprq2\fcharset0 Lucida Sans Unicode;}}
{\colortbl;\red0\green0\blue0;\red128\green128\blue128;}
{\stylesheet{\s1\cf0{*\hyphen2\hyphlead2\hyphtrail2\hyphmax0}
\rtlch\af5\afs24\lang255\ltrch\dbch\af2\afs24\langfe255
\loch\f0\fs24\lang1040\snext1 Standard;}}
```

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Esempio LaTex

```
\documentclass [a4paper,10pt] {article} \usepackage[utf8] {inputenc}
\usepackage[T1] {fontenc}
\usepackage[italian] {babel}
\title {Il mio primo documento}
\author {Angelo Mario Del Grosso}
\begin {document}
\maketitle
\begin {abstract}
Primo tentativo di scrivere in \LaTeX .
\end {abstract}
\section {titolo della sezione}
Questo documento è vuoto.
\footnote {nota a piè di pagina.}

\end {document}
```

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Il markup procedurale

L'unico utilizzo di un testo codificato tramite un linguaggio procedurale è la **creazione di un output orientato alla visualizzazione.**

I linguaggi di codifica

Linguaggi di marcatura procedurale

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Scrivere la tesi di laurea con L^AT_EX 2 _{ε}

Dipartimento di Ingegneria Meccanica,
Nucleare e della Produzione
Università di Pisa
56126 Pisa PI

Sommario

Lo scopo del presente articolo è fornire gli strumenti per scrivere una tesi di laurea utilizzando L^AT_EX 2 _{ε} . Tale obiettivo è conseguito analizzando i problemi tipici incontrati durante la stesura della tesi e le possibili soluzioni; si pone particolare attenzione ai pacchetti da usare nelle varie circostanze. I singoli argomenti non vengono approfonditi nei dettagli ma si rimanda alla letteratura specifica o ad i manuali dei pacchetti suggeriti, ove necessario.

*Ringrazio in primo luogo Fabiano Busdraghi che ha collaborato alla scrittura delle sezioni riguardanti le figure e gli oggetti flottanti. Ringrazio inoltre tutti coloro che mi hanno consigliato durante la stesura e la revisione di questo documento ed in particolare Claudio Beccari, Gustavo Cevolani, Massimo Guiggiani, Maurizio Himmelmann, Lorenzo Pantieri e Emiliano Vavassori.

I linguaggi di codifica

Linguaggi di marcatura dichiarativi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Linguaggi dichiarativi

Orientati al testo, annotano la *struttura*, la *funzione* ed il *significato* degli elementi costitutivi del testo, **tralasciandone l'aspetto**.

- La posizione che il brano in questione occupa all'interno del documento (**markup strutturale**)
- Peculiarità del testo stesso (**markup semantico**)
- I fogli di stile definiscono la formattazione dell'output
- *Molteplici usi del medesimo testo*

Esempio: famiglia SGML, XML

I linguaggi di codifica

Linguaggi di marcatura dichiarativi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Markup dichiarativi: contenuto e presentazione

La **separazione tra contenuto e presentazione** non solo è intenzionale, ma è la **caratteristica principale** di questi sistemi di marcatura: essa permette di concentrarsi sull'**annotazione logica-semantica** per funzioni di *ricerca e di analisi*, lasciando ad altro (ai fogli di stile) la resa grafica.

Unico testo più usi

In questo modo si ha inoltre la possibilità di utilizzare uno **stesso testo codificato con finalità o formattazioni differenti**, a seconda delle varie esigenze.

I linguaggi di codifica

Markup dichiarativi: esempio SGML

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Standard Generalized Markup Language

```
1 <!DOCTYPE testo [  
2   <!ELEMENT testo (titolo?, paragrafo+)>  
3   <!ELEMENT titolo (#PCDATA)>  
4   <!ELEMENT paragrafo (#PCDATA)>  
5 ]>  
6 <testo>  
7   <titolo> Questo è il titolo del documento</titolo>  
8   <paragrafo> Questo è un paragrafo </paragrafo>  
9 </testo>  
10
```

I linguaggi di codifica

Markup dichiarativi vs Markup procedurali

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

resa a video della frase

Le *Guidelines for Electronic Text Encoding and Interchange* sono molto complete e descrivono uno standard di *markup* del testo basato su XML.

Le **\textit{Guidelines for Electronic Text Encoding and Interchange}** sono **\textit{molto}** complete e descrivono uno standard di **\textit{markup}** del testo basato su XML.

<titolo>Le Guidelines for Electronic Text Encoding and Interchange**</titolo>** sono **<enfasi>**molto**</enfasi>** complete e descrivono uno standard di **<linguastraniera>** markup**</linguastraniera>** del testo basato su XML.

LaTeX vs SGML

I linguaggi di codifica

Linguaggi di marcatura

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

linguaggi semi-dichiarativi e/o semi-procedurali

Esistono anche linguaggi che possono essere definiti **semi-procedurali**, o **semi-dichiarativi**, che come si intuisce utilizzano le istruzioni sia per una codifica di tipo procedurale, sia per una codifica di tipo descrittivo o dichiarativo.

HTML

HTML ha tra le sue etichette istruzioni di tipo procedurale per indicare come devono essere rese determinate porzioni di testo, e istruzioni di tipo dichiarativo che hanno una base semantica.

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

- Document Type Definition (DTD)
- XML Schema Definition (XSD)
- RELAX NG

4 Conclusioni

Elementi per la definizione degli schemi xml

principi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Validare il contenuto di un documento XML

Se vogliamo condividere efficacemente informazioni bisogna avere dei meccanismi per controllare che i dati trasmessi rispettino una ben precisa struttura e abbiano un ben preciso e coerente modello dei contenuti (*rispettino una grammatica*).

Validare il contenuto di un documento XML

Per essere sicuri che un documento XML sia corretto da un punto di vista della struttura e del contenuto, cioè sia **valido**, bisogna riferirsi ad uno *schema*.

Elementi per la definizione degli schemi xml

principi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Schema XML

Per condividere quindi efficacemente un vocabolario XML bisogna definire delle regole che controllino come utilizzare correttamente gli elementi e gli attributi del vocabolario.

Schema XML

Gli strumenti per descrivere le regole relative ad una corretta compilazione di un documento XML sono principalmente:
Document Type Definition (DTD) oppure *XML Schema Definition (XSD)*.

Elementi per la definizione degli schemi xml

principi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

XML schema come contratto

Gli schemi XML possono essere visti come un contratto (formale) condiviso tra chi codifica i dati e chi deve consumarli. In questo modo può avvenire in modo rigoroso la comunicazione per lo scambio delle informazioni codificate attraverso il formato XML.

XML schema come contratto

Validare un documento XML vuol dire verificare che il documento sia aderente al formato definito nel contratto (schema).

Elementi per la definizione degli schemi xml

principi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Documento XML valido

Un documento XML si dice **valido** se esso è ben formato (well formed) e se soddisfa anche le regole specificate all'interno di uno schema XML associato.

Elementi per la definizione degli schemi xml

principi

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

schemi XML come contratti

- definiscono la struttura di un documento XML
- definiscono le regole per validare il contenuto degli elementi e degli attributi
- permette ad un programma (*validator, checker*) di verificare la validità di un documento rispetto allo schema prescelto.

Elementi per la definizione degli schemi xml

Tipi di formalismi per definire schemi XML

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Molti tipi di schemi XML

Nel corso degli anni sono stati proposti molti formalismi per codificare gli schemi XML

- DTD, XSD, RELAX NG
- Schematron
- XDR, SOX, DSD, DCD, DDML

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Document Type Definition (DTD)

Una document type definition (DTD) descrive le regole relative alla struttura e al contenuto di un documento XML.

Document Type Definition (DTD)

Una DTD dichiara gli elementi, gli attributi, le entità e le notazioni ammesse in un documento XML.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

well-formed document != valid document

Se un documento XML manca di riferirsi ad una DTD oppure non rispetta le regole di una DTD, esso può essere tutt'al più ben formato, ma sicuramente non può essere valido.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Document Type Definition (DTD)

La validazione dei documenti XML è alla base della condivisione e scambio dati in quanto è possibile confidare sulla natura dei dati trasmessi.

Attenzione: non possiamo validare la correttezza della semantica dei dati!

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Document Type Definition (DTD)

L'associazione tra documento XML e DTD viene realizzata tramite una dichiarazione inclusa nel prologo all'inizio del documento.

Root element and content

La DTD dichiara l'elemento radice del vocabolario e il suo *content model* (children elements).

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Element Declaration (con figli)

```
<!ELEMENT element-name (child-element1,  
child-element-2 ...)>
```

Element Declaration (solo testo)

```
<!ELEMENT element-name (#PCDATA)>
```

*Il Parsed Character Content designa contenuto testuale piano
senza figli*

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Child Element Declaration

La dichiarazione di un elemento figlio, è analoga in tutto e per tutto alla dichiarazione dell'elemento radice. Cioè utilizzando l'etichetta `<!ELEMENT >`

Element Declaration (root)

La dichiarazione dell'elemento radice deve sempre essere la prima

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Modificatori

Nella dichiarazione di un elemento possono essere inclusi optionalmente dei modificatori, per stabilire il numero di occorrenze degli elementi figli.

Modificatori

- + Una o più occorrenze
- ? Zero o una occorrenza
- * Zero o più occorrenze

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Modificatori

```
<!ELEMENT element-name (B, C)+ >
<!ELEMENT element-name (B+, C) >
<!ELEMENT element-name (B, C+) >
<!ELEMENT element-name (B+, C+) >
```

Se un elemento figlio deve presentarsi solo una volta, allora non c'è bisogno di modificatori.

Attenzione l'ordine dei figli nella dichiarazione è significativa

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Esercizio

Definire i seguenti elementi:

- elemento root: **TEI**
- elementi figli:
 - header (obbligatorio una occorrenza)
 - facsimile (opzionale una occorrenza)
 - text (obbligatorio almeno una occorrenza)

Gli elementi header, facsimile e text hanno tutti un content model testuale

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

choice declaration

```
<!ELEMENT element-name (child-a | child-b) >
```

Dichiarazione di Choice

La sintassi della DTD consente di dichiarare una scelta (*choice*) tra due o più elementi come content model di un elemento.

Il choice indica una scelta tra una lista di possibilità

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Attributi degli elementi XML

Un attributo è dichiarato sfruttando l'elemento `<!ATTLIST >`

Cos'è un attributo di un elemento XML

Un attributo è una proprietà, una caratteristica di un elemento
e descrive il contenuto dell'elemento stesso.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Attributi: sintassi

```
<!ATTLIST Element-name Attr-name Attr-type  
          Attr-state? default-value?>
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Attributi: sintassi

- “Element-name” è il nome dell’elemento a cui l’attributo si riferisce
- “Attr-name” è il nome dell’attributo dichiarato
- “Attr-type” è il tipo di dato atteso dell’attributo
- “Attr-state” indica uno tra i tre stati possibili di un attributo
- “default-value” indica il valore di default per quell’attributo, se non fornito.

Elementi per la definizione degli schemi xml

TABELLA dei tipi

Attribute Value	Description
CDATA	Any character data except characters reserved by XML
enumerated list	A list of possible attribute values
ID	A unique text string
IDREF	A reference to an ID value
IDREFS	A list of ID values separated by white space
ENTITY	A reference to an external unparsed entity
ENTITIES	A list of entities separated by white space
NMTOKEN	An accepted XML name
NMTOKENS	A list of XML names separated by white space
NOTATION	The name of a notation defined in the DTD

Elementi per la definizione degli schemi xml

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Stato di attributi

Lo stato di un attributo può essere uno tra:

- #IMPLIED (attributo opzionale)
- #REQUIRED (attributo obbligatorio)
- #FIXED (valore fisso dell'attributo)

Il valore di un attributo fisso viene fornito come valore di default

Elementi per la definizione degli schemi xml

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Choice attributi

```
<!ATTLIST element attribute (value1 | value2 |  
value3 | ...) type default >  
<!ATTLIST name title (Mr. | Mrs. | Ms.)  
#IMPLIED ‘‘Mr.’’>
```

Attenzione non è possibile avere un valore di default se un attributo è marcato #REQUIRED

Elementi per la definizione degli schemi xml

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Attributi ID IDREF IDREFS

```
<!ATTLIST element attribute ID >
<!ATTLIST order orderID ID #REQUIRED>
<!ATTLIST customer orders IDREFS>
<!ATTLIST package orderRef IDREF>
```

Elementi per la definizione degli schemi xml

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Mixed content - DTD

```
<!ELEMENT element-name (#PCDATA|child-element)* >
```

Mixed content XML

```
<p>Ieri pomeriggio sono andato a  
<placeName>Pisa</placeName>, per un giro</p>
```

Elementi per la definizione degli schemi xml

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Esercizio

root: TEI

Figli:

- header(obbligatorio una volta sola)
- facsimile(opzionale una volta sola)
- testo(obbligatorio una o più volte)

* testo è un mixed content con possibile elemento <seg>

Attributi:

- header: type:(fixed, CDATA “intestazione”); lang(opzionale, NMTOKEN)
- facsimile: source:(obbligatorio); ref(optionale, IDREFS)
- testo: id(obbligatorio, ID) type(opzionale contenuto testuale)

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Empty elements

Un elemento XML può essere vuoto (empty). La dichiarazione di un elemento vuoto si realizza con la parola chiave **EMPTY**.

Empty content

```
<!ELEMENT element-name EMPTY>
```

Empty content

```
<!ELEMENT lb EMPTY>
```

```
<lb />
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Any elements

E' possibile dichiarare anche elementi che hanno qualsiasi tipo di content model.

A tal proposito viene impiegata la parola chiave "ANY".

Any content

```
<!ELEMENT element-name ANY >
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Dichiarare il tipo del documento XML

La dichiarazione della DTD viene inserita attraverso una URL nel prologo del documento XML, tra la dichiarazione del documento XML e l'elemento radice.

Dichiarare il tipo del documento XML

Grazie al sistema di dichiarazione della DTD è possibile massimizzare il riuso e collegare lo schema a tutti i documenti che si vuole validare.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

DOCTYPE

```
<!DOCTYPE root-element SYSTEM “External DTD’s  
URL” [Internal DTD ]>  
<!DOCTYPE root-element [Internal DTD] >  
<!DOCTYPE root-element SYSTEM “Ext-DTD URL” >
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

DOCTYPE PUBLIC

```
<!DOCTYPE root PUBLIC \id" \uri">  
standard//owner//description//language  
-//W3C//DTD XHTML 1.0 Strict//EN
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Esercizi

- includere all'interno di un documento XML la dichiarazione del tipo, definire internamente gli elementi e gli attributi e validare.
- inserire nel prologo di un documento XML la dichiarazione del tipo di documento e validare.

Creare un file esterno con estensione .dtd prima di includerlo nel prologo XML.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Entity

Per includere dati da diverse fonti, DTD prevede l'uso di entità. Due tipologie di entità sono state definite: general entities e parameter entities.

Entity: generiche e parametriche

- le general entities vengono espanso nel documento XML
- le parameter entities vengono espanso nel documento DTD

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Entity

- Le general entities si possono classificare in interne ed esterne, a loro volta possono essere parsed oppure unparsed.
- Le parameter entities si possono classificare in interne ed esterne; che possono essere solo parsed.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Entity

Le entità generiche interne aiutano ad includere nel documento XML quei caratteri speciali che altrimenti causerebbero errori al passaggio del parser.

Internal General Entity: Sintassi

```
<!ENTITY entity-name ‘replacement-string’ or  
‘hexadecimal-code’ >
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Internal General Entity: Sintassi

Per usare le entità all'interno del documento XML basta prefissare al nome dell'entità una “e commerciale” (**&**) e aggiungere alla fine come suffisso un “punto e virgola” (**;**):
`&entity-name;`

Una entità può contenere un frammento XML ben formato

Internal General Entity: Sintassi

```
<!ENTITY firma '<i>Angelo Mario Del Grosso</i>'>
<p><salutation>&firma;<salutation></p>
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Internal General Entity: Esempio con UNICODE

Spesso le entità vengono utilizzate per dare un nome ai riferimenti a carattere.

Internal General Entity: Sintassi

```
<!ENTITY amaiuscola '&#65;''>
<!ENTITY amaiuscola '&x0041;''>
<p><salutation>&amaiuscola;ddio<salutation></p>
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

External General Entity

Un documento XML può essere composto da altri pezzi di XML distribuiti in diversi luoghi.

Grazie alle entità generiche esterne è stata implementata questa caratteristica.

External General Entity: Sintassi

```
<!ENTITY entity-name SYSTEM ''URL'' >  
<!ENTITY salutation SYSTEM ''salut.xml.ent'' >
```

L'URL punta al luogo dove risiede la external entity

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

External General Entity

Le entità esterne non possono contenere una DTD per ovvi motivi di gestione dei potenziali conflitti.

E' possibile utilizzare altre entità all'interno delle entità esterne.

General Entity

Le entità generiche vengono dichiarate nella DTD, ma possono essere utilizzate solo all'interno di un documento XML e non nella DTD stessa.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Parameter Entity

Le parameter entity sono entità sfruttabili all'interno del documento DTD. Ma non possono essere utilizzate all'interno del documento XML.

Esestono due tipi di entità parametriche:

1) **internal** parameter entities 2) the **external** parameter entities.

Parameter Entity: Sintassi

```
<!ENTITY % entity-name ‘‘replacement-string’’>  
<!ENTITY % parameter-name SYSTEM ‘‘URL’’ >
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

Parameter Entity: Impiego

```
<!ENTITY % biblinfo '<(title,author?,cost?)'>
<!ELEMENT biblInfo %biblinfo;*>
```

Parameter Entity: Impiego

```
<!ENTITY % biblInfo SYSTEM '<biblInfo.dtd'>
<!ELEMENT listBibl (bib+)>
%biblInfo;
```

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Parameter Entity: utilità

Quando una entità parametrica viene inserita in una DTD, essa viene rimpiazzata dal suo contenuto a tempo di esecuzione.

Parameter Entity: utilità

Ciò permette di facilitare lo sviluppo della DTD e di ottimizzarne la manutenibilità.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Parameter Entity: utilità

Le entità parametriche esterne facilitano la modularità di grandi DTD e permettono un linking dinamico ai vari documenti di definizione.

Parameter Entity: utilità

Grazie a questi tipi di entità è possibile includere pezzi di DTD residenti in posizioni remote e formare un completo e unico documento DTD a tempo di esecuzione.

Elementi per la definizione degli schemi xml

Principi Document Type Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

DTD: Pros

- sono compatte e facilmente comprensibili
- sono definibili inline all'interno del documento XML
- possono definire entità
- sono utilizzate da quasi tutti i vocabolari esistenti
- sono supportate da quasi tutti i parser esistenti

DTD: Cons

- non sono scritte con una sintassi XML
- richiedono parser specifici
- non supportano i namespaces
- non hanno un vero meccanismo per i tipi di dati
- la validazione del contenuto di un elemento è molto limitato e limitante
- non ci sono meccanismi per indicare esattamente il numero di figli che può contenere un elemento.

Elementi per la definizione degli schemi xml

principi XML Schema Definition

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Cos'è uno schema XML

Uno schema XML è un documento XML standard che descrive come deve essere realizzato un altro documento XML.

Ci riferiamo a questa tecnologia con l'acronimo XSD.

A cosa serve uno Schema XML

I documenti XSD sono usati per validare documenti XML. Tuttavia un documento XSD viene realizzato tramite l'uso di un vocabolario predefinito riferibile attraverso un namespace ad un URL standard.

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XSD Schema

Il termine XSD o XML Schema denota un documento XML che descrive e valida la struttura e il contenuto di un altro documento XML.

XSD Schema

Dichiarazione del documento (declaration) e istanza del documento (instance).

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XSD elemento root

L'elemento radice di uno schema XSD è sempre l'elemento “`<schema>`”.

Tutte le definizioni devono seguire quindi l'elemento “`<schema>`”.

XSD Schema

Tutti gli elementi e gli attributi dello schema sono dichiarati all'interno del namespace

“<http://www.w3.org/2001/XMLSchema>.”.

Tutti i documenti XSD contengono la dichiarazione a questo namespace con prefisso convenzionale **xsd** oppure **xs**.

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XSD componenti di base

I componenti di base di uno Schema XSD sono le dichiarazioni degli elementi e le dichiarazioni degli attributi.

XSD Schema

Le dichiarazioni più complesse si poggiano su queste unità: elementi e attributi.

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XSD dichiarazioni

Scrivere un pezzo di codice XSD per descrivere e validare un elemento per un documento XML è detto *element declaration*.

XSD dichiarazioni di base

XSD permette di dichiarare elementi, attributi e di specificare il numero di figli, le occorrenze, l'ordine di apparizione, e i tipi di dati del *content model*.

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Element Types: simple and complex

La dichiarazione di un elemento può avere un tipo semplice (*simple type*) oppure un tipo complesso (*complex type*) a seconda della sua struttura e del suo contenuto.

Simple Type e Complex Type

La dichiarazione di un elemento ha un tipo semplice se non possiede **né figli né attributi**.

La dichiarazione di un elemento ha un tipo complesso in tutti gli altri casi.

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XSD esempio

```
<xsd:schema  
xmlns:xsd='http://www.w3.org/2001/XMLSchema'>  
<xsd:element name='text'/>  
</xsd:schema>
```

XSD esempio elemento di tipo semplice

```
<text>Il primo documento XML Validato</text>
```

Elementi per la definizione degli schemi xml

principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

XML XSD esempio

Il documento XML istanza dello schema XSD per essere valido deve contenere un elemento radice. Validare il documento XML con il relativo XSD con XMLlint.

XMLlint

```
xmlint xmlfirst.xml --schema  
./schema/xsd/xsdfirst.xsd
```

Elementi per la definizione degli schemi xml principi XSD

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)

XML Schema
Definition (XSD)

RELAX NG

Conclusioni

Element Complex Types: esempio

```
<xsd:schema  
xmlns:xsd='http://www.w3.org/2001/XMLSchema'>  
<xsd:element name='Employee'> <xsd:complexType>  
<xsd:attribute name='FirstName' />  
</xsd:complexType> </xsd:element> </xsd:schema>
```

Element Complex Types: esempio

Il documento XML istanza dello schema:

```
<Employee FirstName="Jacob"/>
```

Progress status

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

1 I linguaggi di codifica

2 Fondamenti del linguaggio XML

3 Validare un documento XML e Definire uno Schema

4 Conclusioni

XML per rappresentare il testo

- I markup language per supportare la rappresentazione, memorizzazione, pubblicazione di un testo.
- XML è un markup language flessibile e potente.
- le istruzioni dei markup language sono per lo più dichiarazioni indicando particolari funzioni del dato.
- le istruzioni sono etichette visibili.

Elementi XML

Conclusioni

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

XML per rappresentare il testo

- Una sintassi e una grammatica regolano l'applicabilità del linguaggio di marcatura
- Sintassi: documento well formed (ben formato)
- Grammatica: documento valido

Elementi XML

Conclusioni

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

XML per rappresentare il testo

- XML deriva dal linguaggio SGML.
- XML è una specifica del consorzio W3C.
- XML è un meta-linguaggio.
- XML è plain text.
- XML è portable.

Elementi XML

Conclusioni

Codifica di
Testi -
Introduzione
XML Markup
a.a.
2023-2024

A.M. Del
Grosso

I linguaggi di
codifica

Fondamenti
del linguaggio
XML

Validare un
documento
XML e
Definire uno
Schema

Document Type
Definition (DTD)
XML Schema
Definition (XSD)
RELAX NG

Conclusioni

XML per rappresentare il testo

- XML definisce markup dichiarativi e descrittivi.
- XML ha un modello dati ad albero ordinato.
- XML può avere associato un tipo di documento (DTD) o uno schema (XSD).

Elementi
Codifica TEI
a.a.
2023-2024
A.M. Del
Grosso

Elementi Codifica TEI a.a. 2023-2024

Angelo Mario Del Grosso

angelo.delgrossos@ilc.cnr.it

CNR-ILC

<http://ilc.cnr.it/>

Istituto di Linguistica Computazionale “A. Zampolli”,
15th March 2024

Sommario della Lezione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Introduzione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Cos'è la TEI

la TEI - *acronimo di Text Encoding Initiative* - rappresenta un **punto di riferimento** per tutte le iniziative il cui scopo principale è quello di **digitalizzare risorse testuali** in ambito umanistico per fini di *ricerca* e di *conservazione*.

Introduzione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Qual è l'obiettivo della TEI

L'obiettivo della TEI è quello di **fornire linee guida** per la creazione e la gestione in **forma digitale** di qualsiasi tipo di dato creato e usato in **ambito umanistico**.

Per questo motivo il consorzio investe molte risorse per garantire la massima **accessibilità e divulgazione** delle tecnologie che da anni sviluppa.

Modularità della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

- parleremo degli elementi basilari
 - Intestazione TEI (TEIHeader)
 - Elementi e attributi presenti in tutti i documenti TEI
 - Esempi di codifica
- parleremo del sistema Modulare della TEI
 - Moduli
 - Classi
 - Macro
 - Datatype

Modularità della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

- parleremo degli elementi basilari
 - Intestazione TEI (TEIHeader)
 - Elementi e attributi presenti in tutti i documenti TEI
 - Esempi di codifica
- parleremo del sistema Modulare della TEI
 - Moduli
 - Classi
 - Macro
 - Datatype

I principi fondamentali della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI



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

I principi fondamentali della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI



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

I principi fondamentali della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI



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

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Motto

TEI: Yesterday's information tomorrow

Dal sito TEI

an international and interdisciplinary standard that enables libraries, museums, publishers, and individual scholars to represent a variety of literary and linguistic texts for online research, teaching, and preservation

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Testo di riferimento

Guidelines for Electronic Text Encoding and Interchange
(<http://www.tei-c.org/Guidelines/>)

Testo di ausilio

BURNARD, Lou. What is the Text Encoding Initiative? How to add intelligent markup to digital resources. Nouva edizione [online]. Marseille: OpenEdition Press, 2014 (creato il 13 octobre 2018). Disponibile su Internet:
<http://books.openedition.org/oep/426>. ISBN: 9782821834606. DOI: 10.4000/books.oep.426.

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

un po' di storia

- **1987:** necessità di standard che permetta la creazione e l'interscambio di documenti per mezzo di archivi informatici (convegno NY)
- **1990:** prima versione delle Guidelines (TEI P1)
- **1990-94:** fondi garantiti da enti quali NEH, Mellon Foundation, la Comunità Europea; supporto di ACH, ACL, ALLC

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

un po' di storia

- **2000:** nascita del TEI Consortium, associazione non profit per lo sviluppo dello standard TEI
- **2002:** passaggio da SGML a XML con la versione P4
- **2007:** nuova versione TEI P5, continuamente aggiornata

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Guidelines

Versione TEI P4

- standard precedente, ancora impiegata
- basata su XML, DTD tradizionale
- pubblicata in forma definitiva nel 2002
- <http://www.tei-c.org/Guidelines/P4/>

versioni P1 e P3 basate su SGML

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Guidelines: versione P5

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

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Guidelines: Obiettivi (dalle guide lines)

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

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Guidelines: Obiettivi

These apparently incompatible goals result in a **highly flexible, modular, environment for DTD customization.**

*Lou Burnard, TEI and XML: a marriage made in heaven?
(<http://www.tei-c.org/Talks/marriage.xml?style=printable>)*

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Che cosa offre la TEI

- un ricco e complesso **manuale di codifica**,
(le Guidelines for Electronic Text Encoding and Interchange)
- un **vocabolario** con un numero elevato di elementi (sia strutturale sia semantico)
- **schemi di codifica**
- infrastruttua **modulare e personalizzabile**

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Che cosa offre la TEI

- **è possibile scegliere soltanto i moduli necessari**
- **è possibile modificare le definizioni degli elementi**

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Supporto per gli utenti

- il sito del consorzio (<http://www.tei-c.org/>)
- pagine relative alle varie versioni delle Guidelines
- software, tutorial, ecc.
- il wiki:
http://www.tei-c.org/wiki/index.php/Main_Page
- la mailing list TEIL
- Github: <https://github.com/TEIC>
- TEI by example: <http://teibyexample.org/>

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Novità della versione P5

- modulo di descrizione dei manoscritti
- grafica e multimedia
- standoff markup
- supporto per i namespace XML
- miglioramenti nel modulo feature structure

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI: Struttura modulare

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

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI: Struttura modulare - Moduli essenziali

- **tei**: definisce le classi di elementi, le macro e i datatype che verranno usati per tutti i moduli
- **header**: l'intestazione contenente i metadati relativi al documento TEI XML
- **textstructure**: elementi strutturali per qualsiasi tipo di testo
- **core**: elementi utili in qualsiasi tipo di documento

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI: Struttura modulare - Moduli facoltativi

- **analysis**: strumenti per analisi (linguistica etc.) del testo
- **corpus**: gestione di corpora linguistici
- **drama**: elementi per testi teatrali e drammatici
- **gaiji**: rappresentazione di caratteri e glifi non standard
- **msdescription**: metadati relativi a manoscritti

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI: Struttura modulare - Moduli facoltativi

- **spoken**: trascrizione del parlato
- **textcrit**: apparato critico
- **transcr**: trascrizione di fonti primarie (manoscritti)
- **verse**: elementi supplementari per testi poetici

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Lite

una specifica **personalizzazione** della TEI versione P4/5

A simple demonstration of how the TEI encoding scheme might be adopted to meet 90% of the needs of 90% of the TEI user community

(dalla Prefatory Note: <http://www.tei-c.org/Lite/>)

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Lite

La **versione P4** è stata tradotta anche in italiano: TEI Lite: introduzione alla codifica dei testi, a cura di *F. Ciotti* (http://www.tei-c.org/Lite/teiu5_it.html).

Molto usata, ma presenta varie limitazioni, con la versione P5 è più semplice produrre una versione semplificata o personalizzata

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Pizza chef

la TEI P4 poteva essere modificata usando un programma su web chiamato Pizza chef

<http://www.tei-c.org.uk/pizza.html>.

Metafora della base e dei condimenti (toppings), corrispondenti ai moduli indispensabili e a quelli facoltativi.

Meccanismo efficace, soprattutto considerando l'alternativa (modifica manuale delle DTD TEI).

Tool obsoleto. Il tool ad oggi funzionante per la personalizzazione dello schema si chiama TEI ROMA

<https://roma.tei-c.org/>

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Roma

Per la P5 si è deciso di proporre una modularizzazione più efficace, e uno strumento di personalizzazione più potente.

Strumento basato sul nuovo formato ODD (*One Document Does it all*).

TEI Roma

Il metodo da seguire per la versione P5

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Roma

- Possibilità di **scegliere, escludere, modificare** sia gli *elementi* (e le classi di elementi), sia gli *attributi* (e le classi di attributi)
- Possibilità di **aggiungere elementi** (eventualmente inserendoli nelle classi preesistenti)
- Possibilità di **salvare lo schema** in tre *formati diversi*: DTD tradizionale, W3C e RelaxNG (anche in forma compatta)

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Roma: Il formato ODD

le versioni P1 – P4 delle DTD TEI erano nel formato DTD language e dipendevano dalla sintassi SGML per molti aspetti

TEI Roma: Il formato ODD

One Document Does it all (ODD): set di specifiche in base alle quali un semplice documento TEI XML “*defines a schema in terms of the modules it requires, together with any possible modifications, such as the desired root element*”.

Risultato finale: lo schema nel linguaggio desiderato e la relativa documentazione

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI Roma: Il formato ODD

tutorial: Getting Started with P5 ODDs

www.tei-c.org/Guidelines/Customization/odds.xml

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Il futuro della TEI

**negli ultimi anni l'approccio alla codifica TEI è stato
oggetto di alcune critiche**

Il futuro della TEI

- la TEI è troppo grande / complicata / piccola
- la TEI è basata su XML e questo formato è in declino
- la TEI/XML non supporta le gerarchie multiple
- la TEI non supporta nativamente il markup di tipo stand-off (implementato un metodo ad-hoc)

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Il futuro della TEI

Importante: il formato XML non è l'unico formato per realizzare documenti TEI (in passato SGML).

TEI: Modello astratto

Se c'è una discrepanza fra Guidelines e schemi di codifica, la precedenza va alle Guidelines (**abstraction level**).

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

text encoding con la TEI

E' caldamente raccomandato usare direttamente la **versione più recente della P5**.

La flessibilità della P5 permette di definire uno schema di codifica che corrisponda precisamente al modello

La comunità di utenti e sviluppatori TEI offre un buon supporto.

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Un documento TEI P5 'minimo'

- prologo XML
- intestazione TEI
- elementi strutturali
- elementi semanticci dei moduli base

Intro Text Encoding Initiative

TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Moduli di base: tei, header, textstructure, core

un documento TEI P5 ‘minimo’

Anche usando soltanto i moduli essenziali si ha a disposizione uno schema adatto alla marcatura di numerosi tipi di testi.

Schemi “leggieri” consigliati: la TEI Lite, o se necessario una versione più ridotta della P5 (TEI Absolutely Bare)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Caratteristiche del vocabolario TEI-XML

- gli **elementi TEI** rientrano nelle categorie generali di **elementi XML**
- elementi che possono contenere solo altri elementi (*elementi strutturali*)
- elementi che possono contenere altri elementi e testo
- elementi che possono contenere solo testo
- elementi vuoti (es. <pb/>, <milestone />)
- gli elementi vuoti marcano una gerarchia differente

Intro Text Encoding Initiative

Gerarchie multiple

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<?xml version="1.0" encoding="utf-8"?>
<text>
  <titolo>Gli assassini della Rue Morgue</titolo>
  <intestazione> I </intestazione>

  <pagina n="5">
    <p>Le facolt`a mentali che si sogliono chiamare analitiche sono, di per se stesse, poco suscettibili di analisi [...]</p>
    <p>La facolta` di risolvere `e probabilmente molto rinfor- <u>za</u></p>
  </pagina>
  <pagina n="6"> zata dallo studio delle matematiche e in modo particolare dell'altissimo ramo di questa scienza che[...]
    </p>
  </pagina>
</text>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Gerarchie multiple

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<?xml version="1.0" encoding="utf-8"?>
<text>
  <titolo>Gli assassini della Rue Morgue</titolo>
  <intestazione> I </intestazione>

  <pagina n="5"/>
  <p>
    Le facoltà mentali che si sogliono chiamare analitiche sono, di per se stesse [...]
  </p>
  <p>
    La facoltà di risolvere e probabilmente molto rinforzata dallo studio delle matematiche e in modo particolare dell'altissimo ramo di questa scienza che [...]
  </p>
</text>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Struttura di un documento TEI

- *struttura fondamentale all'interno della radice (<TEI>)*
- una intestazione TEI (<teiHeader>)
- un testo: <text> (o più testi, cfr. infra)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Contenuto del TEI header

- **metadati** relativi al documento (utili per collezioni di testi codificati)
- descrizione del **file** usando `<fileDesc>` (*obbligatoria*)
- descrizioni relative al tipo di **codifica**, al contenuto del documento, alle sue revisioni (*facoltative*)

E' possibile includere testi introduttivi e spiegazioni relative alla codifica effettuata (preziosi per l'interscambio!)

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

```
<?xml version="1.0" encoding="utf8"?>
<!DOCTYPE TEI SYSTEM "tei_lite.dtd">
<TEI xmlns="http://www.teic.org/ns/1.0">
    <teiHeader></teiHeader>
    <text>
        <div><p></p></div>
    </text>
</TEI>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-model href="tei-lite.rng"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
...
    <teiHeader>
        |    <!-- metadati -->
    </teiHeader>
    <text>
        |    <div> <p> <!-- contenuto testuale --> </p> </div>
    </text>
</TEI>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

documento TEI - schema di intestazione TEI minima

Metadati essenziali riguardano il **titolo**, la **modalità di diffusione** e la **fonte originaria** di un testo codificato. Permettono classificazione, archiviazione ed elaborazione bibliografica

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione TEI minima

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<teiHeader>
  ...
    <fileDesc>
      <titleStmt>
        <!--
          | Sezione per la codifica del titolo
          | dell'edizione digitale
        -->
      </titleStmt>
      <publicationStmt>
        <!--
          | Sezione per la codifica dei
          | dati di diffusione e distribuzione dell'edizione
        -->
      </publicationStmt>
      <sourceDesc>
        <!--
          | Descrizione della fonte primaria
        -->
      </sourceDesc>
    </fileDesc>
  </teiHeader>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>La Divina Commedia: versione elettronica</title>
      <respStmt>
        <resp>Conversione TEI P5 a cura di</resp>
        <name>M. Rossi</name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <publisher>Università di Pisa</publisher>
      <date>2002-11-07</date>
      <availability status="restricted"><p></p>
      </availability>
    </publicationStmt>
    <sourceDesc><bibl>
      <title>La Divina Commedia</title>
      <author>Dante Alighieri </author>
      <publisher>Mondadori</publisher>
      <date>1988</date>
    </bibl></sourceDesc>
  </fileDesc>
</teiHeader>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Le altre componenti dell'intestazione TEI

- <encodingDesc> informazioni riguardo lo schema (e il modello di codifica) utilizzato
- <profileDesc> descrizione del testo: quando è stato creato, da chi, usando quali lingue etc.
- <revisionDesc> informazioni sulle versioni del file

I metadati sono una componente essenziale di qualsiasi progetto di digitalizzazione

Lo schema TEI permette di includere anche metadati registrati con altri schemi di codifica usando il descrittore <xenodata>

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.

2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Elementi strutturali

- <text> un singolo testo di qualsiasi tipo (punto di partenza della gerarchia).
- <facsimile> riproduzione della fonte primaria, può affiancare o sostituire <text>
- <front> figlio di <text> materiale che precede il testo
- <body> figlio di <text> rappresenta il testo stesso
- <back> figlio di <text> materiale che segue il testo
- <group> figlio di <text> alternativo a <body>, raggruppa testi diversi

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

```
<TEI>
  ...
  <teiHeader>
    |   <!-- informazioni del TEI Header -->
  </teiHeader>
  <text>
    <front>
      |   <!-- premessa, dedica ... -->
    </front>

    <body>
      |   <!-- corpo del testo ... -->
    </body>

    <back> <!-- postfazione, appendice ... --></back>
  </text>
</TEI>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Costruire documenti compositi

- rimpiazzando il <body> con un gruppo (<group>) di testi si ottiene un documento composito
- ciascuno di questi testi è rappresentato secondo una struttura standard
- un'altra possibilità è creare un corpus con <teiCorpus>
- intestazioni (<teiHeader>) separate per il corpus e per ciascun gruppo di testi
- struttura più complessa, su più livelli

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

```
<TEI>
  ...
  <teiHeader> <!-- intestazione --> </teiHeader>
  <text>
    <front> <!-- frontespizio testo composito --> </front>
    <group>
      <text>
        <front> <!-- frontespizio del primo testo --> </front>
        <body> <!-- body del primo testo --> </body>
        <back> <!-- appendice del primo testo --> </back>
      </text>
      <text>
        <front> <!-- frontespizio del secondo testo --> </front>
        <body> <!-- body del secondo testo --> </body>
        <back> <!-- appendice del secondo testo --> </back>
      </text> <!-- altri testi o gruppi di testi -->
    </group>
    <back> <!-- appendice del testo composito -->
  </back>
</text>
</TEI>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<teiCorpus>
  ...
    <teiHeader> <!-- metadati per il corpus --></teiHeader>
    <TEI>
      <teiHeader> <!-- metadati relativi al I testo--></teiHeader>
      <text> <!-- primo testo del corpus--> </text>
    </TEI>
    <TEI>
      <teiHeader> <!-- metadati relativi al II testo--></teiHeader>
      <text> <!-- secondo testo del corpus--> </text>
    </TEI>
  </teiCorpus>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Altri elementi strutturali fondamentali

- suddivisioni del testo, **non numerati**: <div> (nessun limite di nidificazione)
- suddivisioni del testo, **numerati**: <div1> ... <div7> (massimo 7 livelli)
- **paragrafi**: <p>
- **testo riferito**: <q> (discorso diretto, citazioni, etc.)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Altri elementi strutturali fondamentali

- **versi**: strofe `<lg>` e singoli versi `<l>`
- **testi teatrali**: discorsi `<sp>` che possono contenere paragrafi `<p>` o versi `<l>`, oltre a direzioni di scena `<stage>`
- **milestone tags**: `<pb/>`, `<lb/>`, `<cb/>`, `< milestone/>`
- notare che un `<div>` può contenere un `<floatingText>` (possibilità di introdurre **gerarchie complesse**).

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Apertura e chiusura di un <div>

- <head>: qualunque tipo di **intestazione**: il titolo di un opera, l'intestazione di un paragrafo, di una sezione, ecc.
- l'attributo *type* permette di **classificare** in base a una tipologia
- <epigraph> citazione all'**inizio del testo**, o nella pagina del titolo, eventualmente con riferimento bibliografico
- <opener> raggruppa un serie di elementi (data, luogo, saluti, ecc.) all'**inizio** del <div>, specie di una lettera

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Apertura e chiusura di un <div>

- <argument>: lista degli **argomenti trattati** nel <div>
- <trailer> frase che compare alla **fine** del <div> (ad esempio “Fine del capitolo 1”)
- <closer> raggruppa un serie di elementi (data, luogo, saluti, etc.) alla **fine** del <div>, specie di una lettera

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi

Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<text><body>
...
<div type="lettera">
    <opener>
        <dateline>
            <name type="place">Pisa</name>
            <date>20 marzo 2015</date>
        </dateline>
        <salute>Gentilissima Prof.ssa Scannagatti,</salute>
    </opener>
    <p>sono spiacente di doverle comunicare che un'invasione di cavallette si
è abbattuta sui miei quaderni incautamente lasciati in giardino, e li ha
divorati interamente.</p>
    <p>Questo purtroppo significa che non posso mostrare i compiti svolti,
come sempre, con solerzia e assiduo impegno.</p>
    <closer>
        <salute>Certo di poter contare sulla sua comprensione le porgo i miei
migliori saluti,</salute>
        <signed>Pierino Rossi</signed>
    </closer>
</div>
</body>
</text>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Errori frequenti

Si faintende il significato dell'elemento <fileDesc>

- serve in primo luogo a dare informazioni sul file stesso, non sul testo originale
- il riferimento alla fonte dalla quale è tratto il testo codificato deve essere inserito nel <sourceDesc>

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Errori frequenti

I titoli sono codificati con `<title>` soltanto nel caso di riferimenti bibliografici.

I titoli del testo, dei capitoli etc. si marcano con `<head>`

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Nota sugli errori possibili

Tre categorie:

- **errori sintattici:** un elemento inserito in un punto sbagliato della gerarchia, o che non può contenere testo etc.
- **errori di marcatura semantica:** usare un elemento inadatto allo scopo, ad esempio marcare un titolo con `<emph>`
- **errori di interpretazione** del testo (che portano al tipo d'errore precedente o all'assenza del markup che andrebbe inserito)

Gli errori del primo tipo sono i più facili da individuare e correggere, quelli del terzo i più difficili

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024
A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Alcuni attributi possono essere usati con qualsiasi elemento (v. la classe att.global)

Attributi globali

- **n** un numero o un nome non univoco, possibilmente breve, per identificare un elemento
- **rend** informazioni relative all'aspetto (*originale!*) del testo
- **rendition** simile a *@rend*, ma fa riferimento a elementi `<rendition>` inseriti nell'`<encodingDesc>` (annidato in `<tagsDecl>`)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Attributi globali

- **xml:lang** la lingua del testo contenuto da un elemento
- **xml:id** un identificatore univoco per l'elemento

NOTA: in base ai moduli usati nello schema sono disponibili ulteriori attributi globali

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi

Codifica TEI

a.a.

2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<text>
  ...
    <body>
      <div n="ch1" type="chapter">
        <pb n="1" />
        <p> [...] risulta chiaro se avete letto
          <title rend="underline" xml:lang="fra">
            | Lesfleurs du mal
          </title> [...]
        </p>
        <p> [...] un grande esempio di
          | <foreign xml:lang="fra">savoir faire</foreign> [...]
        </p>
      </div>
      <!-- altri elementi <div> -->
    </body>
  </text>
```

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<text>
  ...
  <body>
    <div n="ch1" type="chapter">
      <pb n="1" />
      <p n="1">[...] descritto altrove (si veda ad esempio
      |   <ref target="#Rossi94">Rossi 1994</ref>) [...]
      </p>
    </div>
    <div n="bib" type="bibliography"> [...]
      <bibl xml:id="Rossi94">
        <author>Rossi, M.</author> [...]
      </bibl>
    </div>
  </body>
</text>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Errori frequenti

<div> non può essere usato allo stesso livello gerarchico di <p>, in altre parole non si può alternare <div> con <p>

Errore!

```
<div> [...] </div> <p> [...] </p> <div> [...]  
</div>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Errori frequenti

<div> e tutti gli altri elementi strutturali *puri* non possono contenere testo.

Errore!

<div>Pippo</div> <person>Pippo</person>

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Enfasi e termini particolari

- **<emph>** parole o frasi enfatizzate nel testo. (Questo è il **<emph>mio</emph>** computer!)
- **<foreign>** parola o frase in una lingua diversa. (In quel punto entrò il bidello a dare il **<foreign xml:lang='lat'>finis</foreign>**).

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Enfasi e termini particolari

- <distinct> “diverso” dal testo perché arcaico, gergale, ecc. (Saltò in groppa al <distinct>fido destriero</distinct>)
- <hi> elemento generico. (<hi rend='‘double’’>N</hi>el mezzo del cammin di nostra vita. Il suo nome è <hi rend='‘italic’’>Mario Rossi</hi>)

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Enfasi e termini particolari

- <mentioned> parola o frase menzionata ma non usata.
(Il termine corretto è
<mentioned>epigrafe</mentioned>)
- <soCalled> parola o espressione da cui ci si distanzia (il cosiddetto <soCalled>darwinismo
sociale</soCalled>)

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Enfasi e termini particolari

- <term> una o più parole considerate termine tecnico.
(Possiamo definire il <term xml id="NPL" rend='italic'>neopositivismo logico</term>)
- <gloss> una spiegazione o glossa riguardo il testo.
(<gloss target="#NPL">una corrente filosofica basata sul principio che la filosofia debba aspirare al rigore proprio della scienza </gloss>)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Esercizio

Marcare un testo plain text di circa 3000 caratteri a piacere.

- inserire prologo XML
- marcare la struttura usando gli elementi fin qui descritti in particolare marcare tutti i paragrafi usando `<p>` e la struttura editoriale usando `<div>`
- verificare che sia ben formato con `xmllint`
- salvare il file XML su github

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Contenuto del TEI header

Qualsiasi documento TEI deve includere una intestazione (TEIHeader).

TEI header: componenti

file description, encoding description, profile description,
revision description, no TEI metadata

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Contenuto del TEI header

- metadati relativi al documento (utili per collezioni di testi codificati)
- descrizione del file usando <fileDesc> (obbligatoria)
- descrizioni relative al tipo di codifica, al contenuto del documento, alle sue revisioni (facoltative)

E' possibile includere testi introduttivi e spiegazioni relative alla codifica effettuata (preziosi per l'interscambio!)

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Testo in prosa

Molti elementi dell'intestazione contengono semplicemente del testo in prosa (giunti a qualche livello dell'albero del documento).

Testo in prosa

Alcuni elementi possono contenere testo in prosa suddivisa da **elementi di paragrafo (*p*)**, oppure elementi più specifici strutturati, che a loro volta contengono testo in prosa.

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Grouping elements

Gli elementi che terminano con suffisso **Stmt** (esempio *titleStmt*), racchiudono un gruppo di elementi specifici e specializzati per la codifica di informazioni strutturate.

Grouping elements

In molti casi questi elementi possono contenere anche descrizioni con testo in prosa. Ciò permette al curatore della codifica di decidere se utilizzare una strategia strutturata, semi-strutturata o completamente destrutturata.

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Dichiarazioni

Gli elementi che terminano con il suffisso **Decl** (per esempio *refsDecl*) racchiudono informazioni inerenti specifiche prassi di codifica che sono state usate all'interno del testo digitale.

Dichiarazioni

Spesso queste prassi sono descritte in un linguaggio formale, attraverso una serie di dichiarazioni che identificano strutture complesse e descrizioni dettagliate.

*Le dichiarazioni possono essere riferite con l'attributo **decls** per ciascuna parte di testo a cui si applica.*

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Descrizioni

Gli elementi che terminano con il suffisso **Desc** (esempio *projectDesc*) contengono testo in prosa, che può essere anche strutturata usando specifici elementi.

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

documento TEI - schema di intestazione TEI minima

```
<teiHeader>
  <fileDesc>
    <titleStmt>...</titleStmt>
    <publicationStmt>...</publicationStmt>
    <sourceDesc>...</sourceDesc>
  </fileDesc>
</teiHeader>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

documento TEI - schema di intestazione TEI minima

Metadati essenziali riguardano il titolo, la modalità di diffusione e la fonte originaria di un testo codificato.

Permettono classificazione, archiviazione ed elaborazione bibliografica

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<teiHeader>
  ...
    <fileDesc>
      <titleStmt>
        <title>La Divina Commedia: versione elettronica</title>
        <respStmt>
          <resp>Conversione TEI P5 a cura di</resp>
          <name>M. Rossi</name>
        </respStmt>
      </titleStmt>
      <publicationStmt>
        <publisher>Università di Pisa</publisher>
        <date>2002-11-07</date>
        <availability status="restricted"><p></p>
      </availability>
      </publicationStmt>
      <sourceDesc><bibl>
        <title>La Divina Commedia</title>
        <author>Dante Alighieri </author>
        <publisher>Mondadori</publisher>
        <date>1988</date>
      </bibl></sourceDesc>
    </fileDesc>
  </teiHeader>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Le altre componenti dell'intestazione TEI

- `<encodingDesc>` informazioni riguardo lo schema (e il modello di codifica) utilizzato
- `<profileDesc>` descrizione del testo: quando è stato creato, da chi, usando quali lingue etc.
- `<revisionDesc>` informazioni sulle versioni del file

I metadati sono una componente essenziale di qualsiasi progetto di digitalizzazione

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Le altre componenti dell'intestazione TEI

■ elemento **xenoData**

Gli elementi di struttura specifici e specializzati che occorrono all'interno di questo elemento non sono parte del vocabolario TEI. Provengono da altri schemi di metadati come ad esempio il *Dublin Core*.

I metadati sono una componente essenziale di qualsiasi progetto di digitalizzazione

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Errori frequenti

Si faintende il significato dell'elemento <fileDesc>

- serve in primo luogo a dare informazioni sul file stesso, non sul testo originale
- il riferimento alla fonte dalla quale è tratto il testo codificato deve essere inserito nel <sourceDesc>

Intro Text Encoding Initiative

Schemi di codifica TEI – Intestazione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Esempio Intestazione TEI raccomandata *Tratto dalle linee
Guida della TEI. Capitolo 2 paragrafo 2.7*

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Elementi
Codifica TEI
a.a.
2023-2024
A.M. Del
Grosso

Elementi Codifica TEI a.a. 2023-2024

Angelo Mario Del Grosso

angelo.delgrossos@ilc.cnr.it

CNR-ILC

<http://ilc.cnr.it/>

Istituto di Linguistica Computazionale “A. Zampolli”,
20th March 2024

Sommario della Lezione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Introduzione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Cos'è la TEI

la TEI - *acronimo di Text Encoding Initiative* - rappresenta un **punto di riferimento** per tutte le iniziative il cui scopo principale è quello di **digitalizzare risorse testuali** in ambito umanistico per fini di *ricerca* e di *conservazione*.

Introduzione

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Qual è l'obiettivo della TEI

L'obiettivo della TEI è quello di **fornire linee guida** per la creazione e la gestione in **forma digitale** di qualsiasi tipo di dato creato e usato in **ambito umanistico**.

Per questo motivo il consorzio investe molte risorse per garantire la massima **accessibilità e divulgazione** delle tecnologie che da anni sviluppa.

Modularità della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

- parleremo degli elementi basilari
 - Intestazione TEI (TEIHeader)
 - Elementi e attributi presenti in tutti i documenti TEI
 - Esempi di codifica
- parleremo del sistema Modulare della TEI

■ Moduli

■ Classi

■ Macro

■ Datatype

Modularità della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

- parleremo degli elementi basilari
 - Intestazione TEI (TEIHeader)
 - Elementi e attributi presenti in tutti i documenti TEI
 - Esempi di codifica
- parleremo del sistema Modulare della TEI
 - Moduli
 - Classi
 - Macro
 - Datatype

I principi fondamentali della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI



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

I principi fondamentali della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI



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

I principi fondamentali della TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI



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

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Citazioni

- <q> testo citato da altre fonti: discorso diretto, esempi (nei dizionari), ecc.

Citazioni - Esempio

(La mia maestra della prima superiore mi salutò sulla porta della classe e mi disse: <q rend='‘PRE mdash’’>Enrico, tu vai al piano di sopra, quest’anno; non ti vedrò nemmen più passare!</q>)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Citazioni

- <quote> frase o brano attribuito a fonte esterna

Citazioni - Esempio

(<p>E allora disse: <q rend='PRE lsquo POST rsquo'>Ecco come comincia la Divina Commedia:
<quote>Nel mezzo del cammin di nostra vita / Mi ritrovai per una selva oscura</quote>.</q></p>)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Citazioni

- < said > testo pronunciato ad alta voce o pensato
- < cit > citazione con riferimento bibliografico

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.

2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Citazioni - Esempio

Lexicography has shown little sign of being affected by the work of followers of J.R. Firth, probably best summarized in his slogan, <cit> <quote>You shall know a word by the company it keeps.</quote> <ref>(Firth, 1957)</ref> </cit>

Semplice riferimento bibliografico nell'esempio, possibile aggiungere un collegamento (a capitolo/sezione o a una specifica entrata bibliografica) usando l'attributo @target

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI: elementi di citazione

A proposito di <q> e <quote> possono contenere non solo altri elementi simili (<q> e <quote>) ma anche elementi come <p>, <l>, ecc.:

Possibili problemi relativi alla gerarchia (tag overlap, gerarchie sovrapposte): <p>Allora disse: <q>Sarò breve.</p><p>Ho finito.</q></p>

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024
A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<p>
    <q>The Lord! The Lord! It is Sakya Muni
himself,</q> the lama half sobbed; and under his
breath began the wonderful Buddhist invocation:
    <q>
        <quote>
            <l>To Him the Way the Law Apart </l>
            <l>Whom Maya held beneath her heart</l>
            <l>Ananda's Lord the Bodhisat</l>
        </quote>
        And He is here! The Most Excellent Law is here
also. My pilgrimage is well begun. And what
work! What work!
    </q>
</p>
```

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Nomi, numeri e date

- <rs> letteralmente “referring string”, nome o etichetta generica

<q>Mio caro <rs type='‘person’’>Filippo</rs></q>, gli disse <rs type='‘person’’>sua moglie</rs>.)
- <name> nome proprio (di persona, luogo, ecc.)

<q>Mio caro <name type='‘person’’>Filippo</name></q>

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Nomi, numeri e date

- <num> un numero in qualsiasi formato <num value='‘23’’>XXIII</num>
- <date> una data in qualsiasi formato nato il <date when='‘18680210’’>10 febb. 1868</date>
- <time> l'ora in qualsiasi formato alle <time when='‘8.00’’>otto del mattino</time>

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Nomi, numeri e date

possibile usare un modulo specifico:

13 Names, Dates, People, and Places (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>)
(`<persName>`, `<forename>`, `<surname>`, `<roleName>`,
`<addName>`, ecc.)

Questo modulo permette una granularità molto maggiore grazie agli elementi specifici che mette a disposizione.

Possibile creare sistema prosopografico.

Ricchezza degli attributi (anche nella versione base)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

La pagina del titolo

elemento `<titlePage>` può contenere:

- `<docTitle>` e (o direttamente) `<titlePart>`: titolo anche in più parti
- `<docEdition>`: informazioni riguardo l'edizione
- `<byline>` e (o direttamente) `<docAuthor>`: autore
- `<docImprint>`: informazioni di stampa, a sua volta include

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

La pagina del titolo

elemento `<docImprint>` può contenere:

- `<publisher>`: editore
- `<pubPlace>`: luogo di stampa
- `<date>`: data
- `<docDate>`: data

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<front>
  ...
    <titlePage>
      <titlePart>Cuore</titlePart>
      <byline>di <docAuthor>E. de Amicis</docAuthor>
      </byline>
      <docEdition>Edizione integrale</docEdition>
      <docImprint>
        <publisher>Newton Compton editori</publisher>
        <pubPlace>Roma</pubPlace>
        <date>1994</date>
      </docImprint>
    </titlePage>
  </front>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<titlePage>
  ...
    <docAuthor>E. DE AMICIS</docAuthor>
    <docTitle>
      <titlePart type="main">CUORE</titlePart>
      <titlePart>Libro per i ragazzi</titlePart>
    </docTitle>
    <docEdition>98.a edizione</docEdition>
    <graphic url="publisher.png" />
    <docImprint>
      <pubPlace>MILANO</pubPlace>
      <publisher>FRATELLI TREVES, EDITORI</publisher>
      <date>1889</date>
    </docImprint>
  </titlePage>
```

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024
A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Collegamenti interni ed esterni

- <ptr/> specifica un puntatore a un altro *luogo* (un altro punto dello stesso testo o di un altro testo)

<p>Il sistema di puntatori è basato sul meccanismo W3C Xpointer. Per maggiori informazioni si veda il capitolo <ptr target="#SAid"/>; per le specifiche si veda <ptr target='‘<http://www.w3.org/TR/xptrxpointer/>’’/’>; anche <ptr type="image" target="#fig22"/>. </p>

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Collegamenti interni ed esterni

- `<ref>` specifica un puntatore a un altro *luogo*, può includere del testo.

Si veda il `<ref>terzo capitolo, p. 24</ref>`. Si veda il `<ref target='#cap3.24'>terzo capitolo, par. 24</ref>`

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Liste e tavole

- **<list>** qualsiasi tipo di lista (type per specificare)
- **<head>** intestazione (titolo) della lista
- **<item>** un elemento della lista
- **<label>** numero o esponente associato all'<item>
- **<headLabel>** intestazione per gli esponenti della lista
- **<headItem>** intestazione per gli elementi della lista

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.

2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Agli item può essere associata o meno una etichetta (label), ma in caso affermativo deve essere presente per tutti.

Liste e tavole

```
<list><head>Ingredienti:</head> <item>un cucchiaio  
di zucchero;</item> <item>mezzo chilo di  
farina;</item> <item>due uova.</item></list>
```

Usare l'attributo type per specificare il tipo di lista.

*Il testo non deve essere necessariamente ordinato come lista
per essere marcato come tale*

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

```
<list type="simple">
  ...
    <head>Lista della spesa:</head>
    <item>pane;</item>
    <item>frutta;</item>
    <item>verdura;</item>
    <item>latte;</item>
    <item>farina;</item>
    <item>uova;</item>
    <item>tovaglioli;</item>
    <item>bicchieri;</item>
    <item>piatti.</item>
</list>
```

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

```
<list type="gloss">
  ...
    <head>Tecnologie XML:</head>
    <label>XSL</label>
    <item>eXtensible Stylesheet Language</item>
    <label>XSLT</label>
    <item>XSL Transformations</item>
    <label>XSLF0</label>
    <item>XSL Formatting Objects</item>
    <label>XQuery</label>
    <item>XML Query Language</item>
    <label>XPath</label>
    <item>XML Path Language</item>
</list>
```

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Tabelle

gli elementi per tavole sono <table>, <row>, <cell> disponibili con il modulo figures, v. il cap. 14 Tables, Formulae and Graphics (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FT.html>)

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

```
<table rows="3" cols="2">
...
<row>
    <cell>HTML</cell>
    <cell>
        Derivato da SGML (come applicazione),
        il linguaggio del World Wide
        Web.</cell>
    </row>
    <row>
        <cell>XML</cell>
        <cell>Derivato da SGML (per semplificazione).</cell>
    </row>
    <row>
        <cell>XSLT</cell>
        <cell>Fogli di stile per XML.</cell>
    </row>
</table>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Note

- l'elemento `<note>` permette di inserire una nota di qualsiasi tipo (attributi *type* e *place*)
- la nota può appartenere al testo codificato, o può essere opera di chi lo codifica (attributo *resp*)
- se non nel flusso del testo usare gli attributi *target* e *targetEnd* per stabilire collegamento preciso

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Note - Esempio

Le <title>Guidelines</title><note n="2" place="foot" resp='amdg'>C.M. SperbergMcQueen and Lou Burnard, <title>Guidelines for Electronic Text Encoding and Interchange</title> (Chicago, Oxford: Text Encoding initiative, 2002).</note> sono nate con lo scopo di ...

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Esercizio

Utilizzare puntatori, note, lista, glossario marcare nomi, rs, term.

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Bibliografia

- <bibl> citazione bibliografica di tipo *flessibile*: possono essere presenti solo alcuni elementi; usata anche nel corpo del testo
- <biblStruct> citazione bibliografica di tipo strutturato: gli elementi devono seguire uno schema fisso
- <biblFull> citazione bibliografica di tipo strutturato completo: gli elementi devono seguire uno schema fisso e devono essere tutti presenti
- <listBibl> lista di citazioni bibliografiche in uno qualsiasi dei formati citati sopra

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi

Codifica TEI

a.a.

2023-2024

A.M. Del

Grosso

Introduzione

al progetto

TEI

Introduzione

Text Encoding

Initiative

Intestazione

TEI

Elementi di

base TEI

Infrastruttura

TEI

Personalizzare

TEI

Una voce bibliografica può essere codificata impiegando tre possibili tipi:

Tipi di pubblicazione

- <analytic> “analitico”: informazioni su un titolo che non costituisce una pubblicazione autonoma (articolo di rivista o in una miscellanea)
- <monogr> “monografico”: informazioni su una pubblicazione, anche di tipo periodico (rivista)
- <series> “serie”: informazioni su di una serie editoriale

tipicamente impiegati con gli elementi <biblStruct> e <biblFull>, più flessibilità con <bibl>

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Titolo, autore, curatore

- <title> il titolo della pubblicazione; più livelli (attr. level)
- *level="a"* articolo (article)
- *level="j"* rivista (journal)
- *level="m"* monografia (monograph)
- *level="s"* serie (series)
- *level="u"* non pubblicato (unpublished)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Titolo, autore, curatore

- <author> autore della pubblicazione
- <editor> curatore della pubblicazione

Questi elementi possono contenere direttamente il testo con il nome oppure elementi più complessi (<persName> e sub-elementi)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Elementi descrittivi

- <imprint> raggruppa informazioni relative alla stampa o diffusione di una pubblicazione
- <publisher> l'editore responsabile della produzione a stampa o diffusione
- <pubPlace> il luogo di pubblicazione
- <date> la data di pubblicazione

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Elementi descrittivi

- <biblScope> l'estensione (in pagine, volumi) della pubblicazione;
- con attributo *unit* permette di specificare il tipo di estensione ad es. se si tratta di pagine
- <citedRange> porzione di lavoro citato.

<biblScope> può essere usato anche all'interno di <series>

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

```
<biblStruct xml:id="Robinson1993d">
  ...
    <monogr>
      <author>Robinson, P.</author>
      <title level="m">The Digitization of Primary
      Textual Sources</title>
      <imprint>
        <pubPlace>Oxford</pubPlace>
        <publisher>Office for Humanities
        Communication</publisher>
        <date>1993</date>
      </imprint>
    </monogr>
  </biblStruct>
```

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

```
<biblStruct xml:id="Mohler2001">
...
<analytic>
    <author>Mohler, Peter Ph.</author>
    <author>Zuell, Cornelia</author>
    <title level="a">Applied Text Theory:
        | Qualitative Analysis of Answers to Open Ended Questions
    </title>
</analytic>
<monogr>
    <editor>Mark D. West</editor>
    <title level="m">Application of Computer Content Analysis</title>
    <imprint>
        <pubPlace>Westport CN</pubPlace>
        <publisher>Ablex</publisher>
        <date>2001</date>
        <biblScope unit="pages">116</biblScope>
    </imprint>
</monogr>
</biblStruct>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

```
<biblStruct xml:id="Ester1994">
...
<analytic>
    <author>Ester, Michael</author>
    <title level="a">Digital Images in the Context of
Visual Collections and Scholarship</title>
</analytic>
<monogr>
    <title level="j">Visual Resources</title>
    <imprint><date>1994</date></imprint>
    <biblScope unit="vol">X,1</biblScope>
    <biblScope unit="pages">23</biblScope>
</monogr>
<note type="key">iproc</note>
</biblStruct>
```

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Componenti non testuali: immagini

- in formato XML (*SVG, il formato TEI*)
- in formato binario (*TIFF, JPEG, PNG*)

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Componenti non testuali: immagini

- <graphic/> contiene un puntatore a una immagine
- attributo **url** può essere locale o esterno
- <binaryObject> dati binari che costituiscono una immagine
- <media> riferimento a file audio, video ecc.

Intro Text Encoding Initiative

Schemi di codifica TEI: Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Componenti non testuali: immagini

- <figure> raggruppa informazioni relative a un'immagine
- <head> titolo o didascalia relativa all'immagine
- <figDesc> descrizione dell'immagine
- <graphic> puntatore a un'immagine

<figure> nella P5 fa parte del modulo figures
(<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FT.html#FTGRA>)

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi

Codifica TEI

a.a.

2023-2024

A.M. Del

Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

*Nella TEI P4 bisogna inserire dichiarazioni per le immagini
nell'intestazione*

```
<!DOCTYPE figEx [
    <!NOTATION JPEG PUBLIC 'ISO DIS 10918//NOTATION
JPEG Graphics Format//EN'>
    <!ENTITY PDatabase
        SYSTEM 'support/PDatabase.jpeg' NDATA JPEG> ]>
<figEx>
    <p>The system is build around a relational
        database, called the <soCalled>palaeographical
        database</soCalled> ... </p>
    <figure entity="PDatabase">
        <figDesc>Schematic of the SPI system.</figDesc>
    </figure>
</figEx>
```

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Immagine in un documento P5

```
<p>Questo fenomeno è chiaramente visibile nella  
figura che segue: <graphic url="diagramma.png"/>  
Altre variazioni delle crescute sono  
disponibili...</p>  
<head> <graphic  
url="http://www.name.org/hpral02.gif"/> </head>
```

immagine impiegata come intestazione in un documento P5

Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Esercizi documenti di tipo e genere diversi

- codificare usando gli opportuni elementi TEI un articolo di rivista
- codificare usando gli opportuni elementi TEI una cartolina

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Infrastruttura TEI

Tabella Moduli TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

TEI framework

La tecnologia TEI ha un framework concettuale diviso in

- Moduli
- Classi
- Macro
- Tipi di Dato

Infrastruttura TEI

Tabella Moduli TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Moduli TEI

Un modulo è semplicemente un contenitore per una serie di dichiarazioni uniformi e coerenti per gli elementi TEI e le relative classi.

Infrastruttura TEI

Tabella Moduli TEI

Module name	Formal public identifier	Where defined
analysis	Analysis and Interpretation	<i>17. Simple Analytic Mechanisms</i>
certainty	Certainty and Uncertainty	<i>21. Certainty, Precision, and Responsibility</i>
core	Common Core	<i>3. Elements Available in All TEI Documents</i>
corpus	Metadata for Language Corpora	<i>15. Language Corpora</i>
dictionaries	Print Dictionaries	<i>9. Dictionaries</i>
drama	Performance Texts	<i>7. Performance Texts</i>
figures	Tables, Formulae, Figures	<i>14. Tables, Formulae, Graphics and Notated Music</i>
gaiji	Character and Glyph Documentation	<i>5. Characters, Glyphs, and Writing Modes</i>
header	Common Metadata	<i>2. The TEI Header</i>
iso-fs	Feature Structures	<i>18. Feature Structures</i>
linking	Linking, Segmentation, and Alignment	<i>16. Linking, Segmentation, and Alignment</i>
msdescription	Manuscript Description	<i>10. Manuscript Description</i>
namesdates	Names, Dates, People, and Places	<i>13. Names, Dates, People, and Places</i>
nets	Graphs, Networks, and Trees	<i>19. Graphs, Networks, and Trees</i>
spoken	Transcribed Speech	<i>8. Transcriptions of Speech</i>
tagdocs	Documentation Elements	<i>22. Documentation Elements</i>
tei	TEI Infrastructure	<i>1. The TEI Infrastructure</i>
textcrit	Text Criticism	<i>12. Critical Apparatus</i>
textstructure	Default Text Structure	<i>4. Default Text Structure</i>
transcr	Transcription of Primary Sources	<i>11. Representation of Primary Sources</i>
verse	Verse	<i>6. Verse</i>

Infrastruttura TEI

Tabella Moduli TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Component	Name	Example
Attribute Classes	att.*	att.global
Model Classes	model.*	model.biblPart
Macros	macro.*	macro paraContent
Datatypes	data.*	data.pointer

Infrastruttura TEI

Tabella Moduli TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Classi TEI

Le classi sono usate per esprimere due distinti tipi di **caratteristiche comuni** ad un insieme di elementi.

Classi TEI

Gli elementi di una classe possono *condividere un insieme di attributi* oppure possono far parte di uno stesso *content model*.

Infrastruttura TEI

Tabella Moduli TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Classi TEI

- Un elemento appartenente ad una classe attributo condivide gli attributi con tutti gli altri elementi membri della stessa classe.
- Un elemento appartenente alla classe modello condivide il luogo del content model dove appare con gli altri elementi membri della stessa classe.

*In entrambi i casi un **elemento eredita** proprietà dalle classi di cui è membro.*

Infrastruttura TEI

Tabella Moduli TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

class name	module name	see further
att.global.linking	linking	<i>16. Linking, Segmentation, and Alignment</i>
att.global.analytic	analysis	<i>17. Simple Analytic Mechanisms</i>
att.global.facs	transcr	<i>11.1. Digital Facsimiles</i>
att.global.change	transcr	<i>11.7. Identifying Changes and Revisions</i>

Infrastruttura TEI

Tabella Moduli TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Macro TEI

Le Macro sono shortcut per dichiarazioni che occorrono frequentemente.

Le Macro sono utilizzate in due modi diversi:

- per content model o parti di content model
frequently-encountered
- per datatype di attributi

Infrastruttura TEI

Tabella Moduli TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Data Type TEI

I valori che possono assumere gli attributi sono definiti da tipi di dato all'interno delle *TEI datatype specification*.

Data Type TEI

Le specifiche TEI definiscono i propri tipi di dato sfruttando altri tipi di dato primitivi e quelli derivati dalle specifiche W3C.

Attributi Globali

Esempio @xml:lang

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

xml:lang indica la lingua e il sistema di scrittura usato

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
    <teiHeader xml:lang="en">
        <!-- ... -->
    </teiHeader>
    <text xml:lang="fr">
        <body>
            <div>
                <!-- chapter one is in French -->
            </div>
            <div xml:lang="de">
                <!-- chapter two is in German -->
            </div>
            <div>
                <!-- chapter three is French -->
            </div>
            <!-- ... -->
        </body>
    </text>
</TEI>
```

Infrastruttura TEI

Classificazione degli elementi

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

*Quasi tutti gli elementi TEI possono essere **classificati informalmente** come appartenenti alle seguenti categorie:*

TEI element classification

- divisions
- chunks
- phrase-level elements
- inter-level elements
- components

Infrastruttura TEI

Classificazione degli elementi

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

■ **divisions**

Divisioni ad alto livello dei testi, molto spesso elementi annidati.

■ **chunks**

Elementi come i paragrafi e altri elementi simili i quali sono posizionati all'interno dei testi e divisioni.

Soltanamente non sono elementi che possono annidarsi o apparire all'interno di altri elementi di livello chunk.

Infrastruttura TEI

Classificazione degli elementi

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

■ **phrase-level elements**

Elementi che occorrono solo all'interno di elementi di livello chunk.

■ **inter-level elements**

Elementi che possono occorrere sia tra chunks all'interno di division, sia all'interno di essi.

■ **components**

Elementi che possono occorrere direttamente all'interno dei testi o delle divisioni di testo. E' una combinazione di elementi di livello inter e chunk.

Progress status

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

1 Introduzione al progetto TEI

2 Introduzione Text Encoding Initiative

3 Intestazione TEI

4 Elementi di base TEI

5 Infrastruttura TEI

6 Personalizzare TEI

Intro Text Encoding Initiative

Schemi di codifica TEI – Personalizzare TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Personalizzare TEI

Nessun progetto di codifica richiede di utilizzare tutte le specifiche definite dalle linee guida della TEI.

Personalizzare TEI

La TEI fornisce un insieme specifico di elementi che può essere usato per creare uno schema TEI puntuale e ritagliato sulle specifiche del progetto di codifica in corso.

Intro Text Encoding Initiative

Schemi di codifica TEI – Personalizzare TEI

```
<schemaSpec ident="tei-custom"
...
    start="TEI teiCorpus" >
    <moduleRef key="analysis"
        include="interp interpGrp pc s w" />
    <moduleRef key="linking"
        include="anchor seg" />
    <moduleRef key="tei " />
</schemaSpec>
```

*Si selezionano o si escudono gli elementi attraverso gli attributi
@include e @exclude*

Intro Text Encoding Initiative

Schemi di codifica TEI – Personalizzare TEI

Elementi
Codifica TEI
a.a.
2023-2024

A.M. Del
Grosso

Introduzione
al progetto
TEI

Introduzione
Text Encoding
Initiative

Intestazione
TEI

Elementi di
base TEI

Infrastruttura
TEI

Personalizzare
TEI

Modifica agli elementi e attributi dei Moduli

```
<classSpec type="atts" ident="att.datable.w3c"
...
    module="tei" mode="change" >
    <attList>
        <attDef ident="notAfter" mode="delete" />
        <attDef ident="from" mode="delete" />
        <attDef ident="to" mode="delete" />
    </attList>
</classSpec>
```

Codifica TEI -
Elementi
editoriali e
facsimile
A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Codifica TEI - Elementi editoriali e facsimile

Angelo Mario Del Grosso
(Materiale tratto dalle lezioni di R. Rosselli Del Turco)

angelo.delgrosso@ilc.cnr.it

CNR-ILC
<http://ilc.cnr.it/>

Istituto di Linguistica Computazionale “A. Zampolli”,
26th April 2024

Sommario della Lezione

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

1 Introduzione agli elementi editoriali e facsimile TEI

2 Elementi editoriali TEI

3 Elementi facsimile TEI

4 Conclusioni

Progress status

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

1 Introduzione agli elementi editoriali e facsimile TEI

2 Elementi editoriali TEI

3 Elementi facsimile TEI

4 Conclusioni

Introduzione Elementi Editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

In parte già disponibili nei moduli TEI di base

Elementi per interventi editoriali

Per la critica testuale indispensabili i moduli

- **msdescription** *descrizione del manoscritto*
- **trans** *trascrizione di fonti primarie*
- **textcrit** *apparato critico*
- **gaiji** *caratteri non standard*

Introduzione Elementi Editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Altri Moduli Utili

- **analysis** modulo per l'inserimento di semplici analisi e interpretazioni ad elementi con contenuto testuale
- **certainty** modulo per la registrazione del grado di "certezza/incertezza" relativo al testo e alla corrispondente marcatura
- **figures** modulo per l'inserimento di immagini con relative informazioni
- **namesdates** modulo per la codifica dei nomi: persone, luoghi, organizzazioni
- **verse** modulo per la codifica di testi poetici

Progress status

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

1 Introduzione agli elementi editoriali e facsimile TEI

2 Elementi editoriali TEI

3 Elementi facsimile TEI

4 Conclusioni

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Testo aggiunto, cancellato, lacunoso

- **<add> una o più parole aggiunte nel testo**
questa parola è `<add place="supralinear">stata</add>` aggiunta in un secondo momento

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Testo aggiunto, cancellato, lacunoso

- ** una o più parole cancellate nel testo**
questa invece era <del rend="overstrike">
>era di troppo e l'ho cancellata

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Testo aggiunto, cancellato, lacunoso

- <gap /> **parte di testo omessa, mancante o illeggibile**
questa <gap reason="illegible" quantity="6"
unit="chars"/> è illeggibile (forse "parola"?)

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Testo danneggiato, poco chiaro, inserito dal curatore

- <damage> **testo danneggiato nel documento originale**
per qualche goccia d'acqua questa parola si è
<damage agent="water" >scolorita</damage>
molto

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Testo danneggiato, poco chiaro, inserito dal curatore

- <unclear> **parte di testo interpretabile con difficoltà**
<unclear reason="faded" >questa</unclear> si
legge ancora ma con difficoltà

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Testo danneggiato, poco chiaro, inserito dal curatore

- **<supplied> testo inserito dal curatore perché illeggibile nell'originale o assente**
qui <supplied>mancava qualcosa</supplied> nel testo

*<supplied> (fa parte del modulo *transcr*)*

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Dal Vercelli Book

```
<w>ge<lb n="24"/>bund<g ref="#eenl"/>nne</w>
<damage type="faded"><g ref="#inodotenl"/>n</damage>
<unclear><g ref="#sins"/>umne</unclear>
<gap reason="illegible" extent="unknown"/>
<supplied resp="scragg">cafertun
    þa eode <name type="person">Petrus</name>
    ond þes ilca</supplied>
```

*Manoscritto in inglese antico, dialetto tardo sassone
occidentale, X secolo ca.*

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento sostituzione

Può succedere che una parola non sia semplicemente cancellata, ma che sia anche sostituita da un altro termine

Elemento sostituzione

In questo caso è possibile usare l'elemento <subst> per collegare la sequenza cancellazione - nuovo testo

<subst> fa parte del modulo *transcr*, per ulteriori informazioni (@seq) v. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHSU>

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento sostituzione: parola

questa parola è stata <subst> <del
rend="overstrike" >scritta <add
place="supralinear" >aggiunta</add> </subst> in un
secondo momento

Elemento sostituzione: carattere

<subst><del rend="overtype"
>t<add>T</add></subst>i scrivo una mail
domani mattina

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Attributi e valori consigliati

- <add> **place** *inline, supralinear, margin* **hand** *author, scribe1, scribe2*
- **rend** *overstrike, subpunct, overtype, dotted* **hand** *author, scribe1, scribe2*
- <gap> **reason** *illegible, cancelled, irrelevant, missing, omissis, censored* **hand** **editor agent** *water, smoke, hole, missing page extent chars, words, cm*
- <unclear> **reason** *part erased, ink blot* **hand** *author, scribe1, scribe2* **agent** *water, smoke, ink*

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Correzioni e normalizzazioni

- **<sic> parola o frase ritenuta errata, ma riportata così come appare scritta**
questa parola è **<sic>statta</sic>** sbagliata

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Correzioni e normalizzazioni

- **<corr> correzione di una parola o frase errata**
questa parola è **<corr>stata</corr>** corretta in
un secondo momento

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Correzioni e normalizzazioni

- **<orig> parola o frase ritenuta non standard**
Allora, mi dici <orig>'ndo</orig> vai?

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Correzioni e normalizzazioni

- **<reg> parola o frase normalizzata (regularised)**
Allora, mi dici <reg>dove</reg> vai?

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Abbreviazioni ed espansioni

■ <abbr> parola abbreviata, brevigrafo

chiedi al <abbr>dott.</abbr> Rossi

in nomine Patris <abbr>7</abbr> Filii

<abbr>7</abbr> Spiritus Sancti

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Abbreviazioni ed espansioni

- **<expan> espansione di un'abbreviazione**
chiedi al <expan>dottor</expan> Rossi
in nomine Patris <expan>et</expan> Filii
<expan>et</expan> Spiritus Sancti

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Abbreviazioni ed espansioni

- **<abbr> usare l'attributo type per specificare il tipo**
chiedi al <abbr type="titolo" >dott.</abbr>
Rossi
in nomine Patris <abbr type="brevigrafo" >7</abbr> Filii <abbr type="brevigrafo" >7</abbr> Spiritus

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Disponibili ulteriori elementi aggiungendo il modulo transcr:

Abbreviazioni ed espansioni

- **<am> lettere o segni di abbreviazione**
- **<ex> lettere aggiunte espandendo un'abbreviazione**
`<abbr>sanct<am>u~</am></abbr>`
`<expan>sanctu<ex>m</ex></expan>`

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Alcuni elementi di intervento editoriale sono perfettamente speculari

Elementi speculari

- <sic> - <corr>
- <orig> - <reg>
- <abbr> - <expan>

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento choice

Nella *TEI P5* è stato introdotto un nuovo elemento <choice> che comprende ogni coppia

Elemento choice

La relazione fra le coppie è importante per essere sicuri che ciascun elemento sia collegato all'altro

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Disponibili ulteriori elementi aggiungendo il modulo transcr:

Coppie in choice

■ <sic> - <corr>

questa parola è <choice> <sic>statta</sic>
<corr>stata</corr> </choice> scritta sbagliata

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Disponibili ulteriori elementi aggiungendo il modulo transcr:

Coppie in choice

- <orig> - <reg>

Allora, mi dici <choice><orig>'ndo</orig>
<reg>dove</reg></choice> vai?

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Disponibili ulteriori elementi aggiungendo il modulo transcr:

Coppie in choice

- <abbr> - <expan>
chiedi al <choice><abbr>dott.</abbr>
<expan>dott<ex>or</ex></expan></choice> Rossi

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Attributi responsabilità

Per <add> e abbiamo l'attributo @hand per specificare, se necessario, **l'autore** della **modifica** all'originale

Attributi responsabilità

Per qualche intervento editoriale, invece, potrebbe essere importante specificare il **responsabile** (in particolare per l'elemento <supplied>!) e il grado di **certezza**

Per fare questo si usano gli attributi @resp e @cert

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Attributi responsabilità

questa parola è <choice> <sic>statta</sic> <corr
resp="#amdg" >stata</corr> </choice> scritta per
errore

*Il valore di @resp rimanda ad un <respStmt> <corr
resp="#amdg" >stata</corr>*

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Altri elementi utili

■ <gb/> **gathering begins**

*indica l'inizio di un nuovo fascicolo nella trascrizione di un
manoscritto*

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Altri elementi utili

■ <space> **modulo transl**

usato per marcare la presenza di spazio significativo (ad esempio spazio lasciato dallo scriba per inserire una iniziale minciata)

normalmente usato come elemento vuoto

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Altri elementi utili

■ <pc> **modulo analysis**

contiene uno o più caratteri che costituiscono una forma di punteggiatura nel testo
punctuation character(s)

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Altri elementi utili

■ **<ab> anonymous block**

*contenitore di testo simile a un paragrafo ma senza il
valore semantico di quest'ultimo
modulo linking*

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Altri elementi utili

■ <seg> **arbitrary segment**

*usato per marcare qualunque porzione di testo, usare
@type per specificare il contenuto semantico
modulo linking*

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Altri elementi utili

■ **<w> word**

*marka una parola a livello grammaticale, @lemma per indicare il lemma e @lemmaref per stabilire un link con un dizionario online
modulo analysis*

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Altri elementi utili

■ **<c> character**

*marka un singolo carattere nel testo
modulo analysis*

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

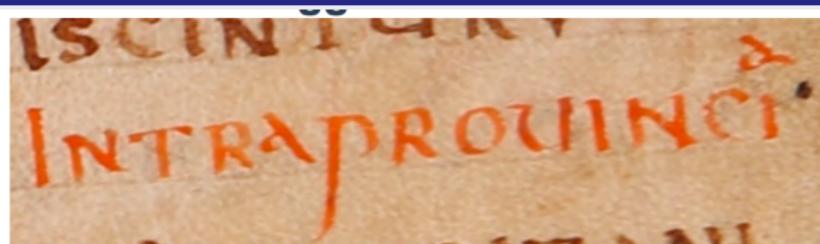
Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Aggiunte scribali



<w>intra</w>

<w>pro<c corresp="#v">u</c>inci<add place="above" rend="smaller">a</add></w>

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

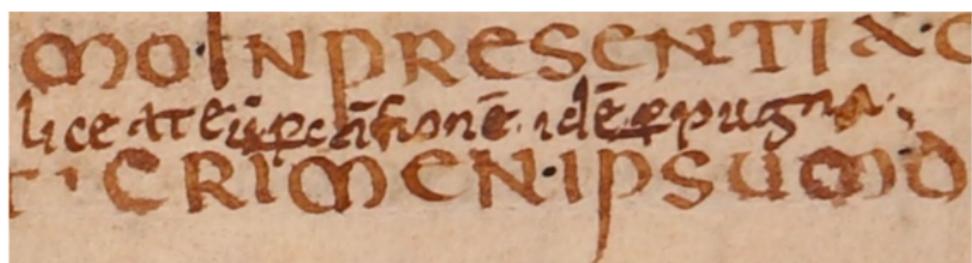
Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Aggiunte scribali



```
<add place="above"><w>liceat</w> <w>cum</w> <w>per</w> <w>camphionem</w><reg>, </reg>  
<w>id</w> <w>est</w> <w>per</w> <w>pugnam</w><reg>, </reg>  
</add>
```

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Cancellature



```
<w>d<subst><del rend="overwrite">o</del><add>u</add></subst>cem</w>
```

Cancellature scribali: Dal Vercelli Book

```
<del rend="erasure" >ff</del>fore wæron <del  
rend="dot" ><g ref="#aeligddot"/></del>
```

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Abbreviazioni e Espansioni



```
<w>
  <choice>
    <abbr>pericul<am><g ref="#umacr"/></am></abbr>
    <expan>periculu<ex>m</ex></expan>
  </choice></w>
```



```
<w><choice>
  <abbr>iussion<am><g ref="#emacr"/></am></abbr>
  <expan>iussione<ex>m</ex></expan>
</choice></w>
```



```
<w>
  <choice>
    <abbr>d<am><g ref="#imacr"/></am></abbr>
    <expan>d<ex>e</ex>i</expan>
  </choice></w>
```

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

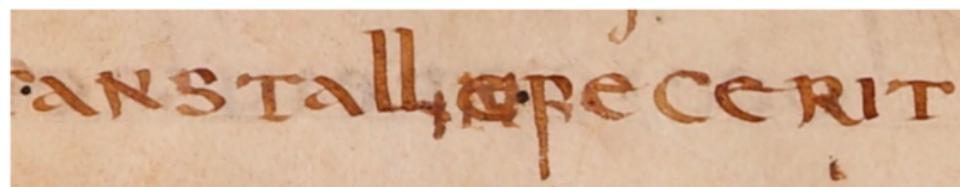
Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Codifica fenomeni annidati



```
<term xml:lang="Eng"><w>  
  <choice>  
    <orig>anstall<add place="below">i</add><subst><del rend="overwrite">c</del><add>n</add></subst></orig>  
    <reg>astalin</reg>  
  </choice></w></term>
```

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

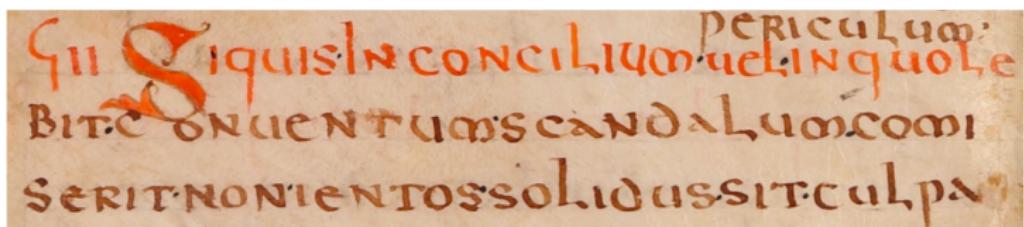
Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Errori Scribali



```
<w>
  <choice>
    <sic>quole<db n="11" xml:id="CV188_lb_019v_11_sic"/>bit</sic>
    <corr resp="RRDT">quoli<db n="11" xml:id="CV188_lb_019v_11_corr"/>bet</corr>
  </choice></w>
  <w>c<space extent="1"/>on<cc corresp="#v">u</c>entum</w><!-- conventu -->
```

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni



```
<p>
<lb xml:id="CV188_lb_018v_05" n="05"/>
<label n="1"><num><w>I</w></num></label>
<seg rend="cap">
<w><hi rend="decorated">S</hi>i</w>
<w>q<<c corresp="#u">v</c>is</w>
<w>homin<<c corresp="#u">v</c>m</w>
<lb xml:id="CV188_lb_018v_06" n="06"/><w>contra</w>
```

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Marcatura singole parole

```
<lb xml:id="CV188_lb_018v_10" n="10" type="rubric"/>
<label n="2"><num><w>II</w></num></label>
<w><hi rend="cap1.2">S</hi>i</w>
<w>quis</w>
<w>cum</w>
<w>rege</w>
<w>de</w>
<w>morte</w>
<w>alte<lb xml:id="CV188_lb_018v_11" n="11"/>rius</w>
<w>consilia<c corresp="#v">u</c>erit</w><reg>, </reg>
<w>aut</w>
<w>hominem</w>
<lb xml:id="CV188_lb_018v_12" n="12"/><w>per</w>
<w>ipsius</w>
<w><choice>
    <abbr>iussion<am><g ref="#emacr"/></am></abbr>
    <expan>iussione<ex>m</ex></expan>
</choice></w>
<choice><reg> </reg><choice><reg> </reg><choice><reg> </reg>
```

Elementi interventi editoriali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Esercizio

Trascrivere e codificare le prime righe della pagina seguente,
selezionata dal sito del CDEC

([http://digital-library.cdec.it/cdec-web/viewer/cdecxDamsHist003/IT-CDEC-ST0003-000032#page/3 mode/1up](http://digital-library.cdec.it/cdec-web/viewer/cdecxDamsHist003/IT-CDEC-ST0003-000032#page/3	mode/1up))

25 Aprile 1920 - 23 Luglio 1960 - Ventic'anni: (1)
appena il limitare della vita - ma breve è luminosa,
la veramente, come dice il racconto marmoreo
come forma di meditare che si volgono fra cosa e parola,
promessa, speranza, desirio. Quelle che
fanno, a sbocciare di un animo selenita, di un'intelligenza
che ha conosciuto nella prima infanzia, di solito,
se anche i ricordi di quei primi anni, dicono.
Quando si aspira giocondamente alla vita.

Progress status

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

1 Introduzione agli elementi editoriali e facsimile TEI

2 Elementi editoriali TEI

3 Elementi facsimile TEI

4 Conclusioni

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Tipi di edizione digitale

- **edizione ipertestuale**
- **facsimile digitale**
- **image-based**

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizione ipertestuale

Prime a essere prodotte, ancora oggi spesso in formato HTML
(derivate anche da elaborazioni di documenti TEI XML).

Formato ideale per edizioni critiche.

Distribuzione sul Web (es.: Biblioteca Digitale Italiana).

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Facsimile digitale

Riproduzione del manoscritto basata su scansione digitale.
Distribuzione sul Web (es.: Progetto e-codices)

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Image-based digital edition

Edizione basata su immagini.

Il testo dell'edizione (diplomatica, interpretativa, critica) con le
immagini del manoscritto

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizione basata sulle immagini

- Trascrizione collegata alle immagini del manoscritto
- Funzionalità principali collegate all'edizione digitale
 - immagini in formati e/o risoluzioni diverse
 - lente d'ingrandimento ed evidenziazione dettagli
 - restauro digitale
 - motore di ricerca testuale e bibliografia
 - commento interattivo al testo

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Esempio edizione digitale image-based

Electronic Beowulf

<http://ebeowulf.uky.edu/ebeo4.0/CD/main.html>

The screenshot shows a digital edition of the Anglo-Saxon epic Beowulf. On the left, there is a transcription of a manuscript page. Above the transcription are two small windows labeled "Left/Top" and "Right/Bottom" which show images of the manuscript pages. The transcription itself is in a large, clear font, with line numbers on the left and the text of the poem. The text is in Old English, with some words in modernized spelling. The manuscript image on the right shows the original parchment with handwritten text in two columns.

Left/Top

Goto Edition 129r, 1-21 << >> Fit frame

Right/Bottom

Goto MS 129r, BL 132r << >> Fit frame

129r | HWÆT WE GAR-DENA IN GEARDAGUM
þeodcyninga þrym gefrunon.
Hu ða æhelingas ellen fremdon!
Oft Scyld Scefing sceapena þreatum
5 monegum mæghum meodosetla ofteah,
egsode corl, syððan ærest wearð
feasceaf funden. He þes frofne gebad,
weox under wolcnum, weorðmyndum þah,
oð hæt him æghwylc þara ymbstittendra
10 ofer hronrade hyran scolde,

PYE TE GARDE
na mægha datum. heod cym
þrym se þunon. huða æhelingas ellen
fremdon. oft scyld scepung sceches
þær cum monegus mæghum meodo setla
oð ealh egsode ealh syððan ærest wearð
feasceaf funden. he þes frofne gebad,
weox under wolcnum, weorðmyndum þah,
oð hæt him æghwylc þara ymbstittendra
ofer hronrade hyran scolde. zonibar

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizione digitale di un manoscritto

- **immagini** del manoscritto
- **trascrizione** del/i testo/i
- creazione del **facsimile digitale**
- **collegamento** testo-immagine

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Livelli di edizione

- **edizione diplomatica**
- **edizione diplomatico-interpretativa**
- **edizione critica**

In caso di singolo manoscritto possiamo avere edizione diplomatica e/o interpretativa

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

livelli di edizione: diplomatica

Trascrizione del testo di un testimone rispettando la disposizione e la grafia originale, senza nessun tipo di correzione (errori manifesti) o altri interventi editoriali (espansione abbreviazioni).

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

livelli di edizione: diplomatico-interpretativa

Sempre rispettando il testo originale, vengono corretti gli errori più evidenti, regolarizzate certe particolarità ortografiche (suddivisione delle parole), espande le abbreviazioni, etc.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Livelli di edizione: critica

Sulla base della collazione di tutte le trascrizioni dei testimoni viene stabilito lo stemma codicum e si tenta di ricostruire il testo originale confrontando le varianti dei testimoni più validi

Facsimile ed edizione digitale

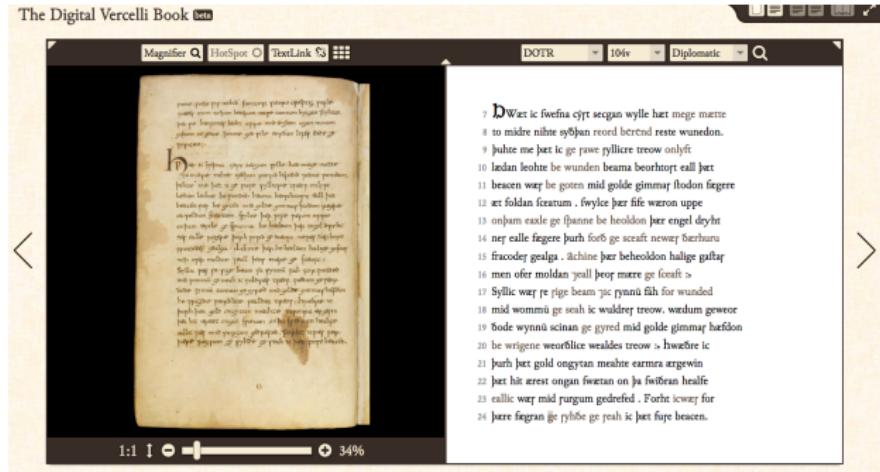
Codifica TEI -
Elementi
editoriali e
facsimile

Elementi
facsimile TEI

Esempio livelli di edizione

Vercelli Book

<http://vbd.humnet.unipi.it/beta/>



Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Rapporto testo-immagine

- **collegamento mirato (hot-spot)**
- **collegamento generalizzato**

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Collegamento mirato (hot-spot)

Una specifica area dell'immagine viene evidenziata in maniera tale che, interagendo con la stessa, vengono visualizzate delle informazioni quali note editoriali, versione migliorata di un dettaglio, commento al testo, etc.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Collegamento generalizzato

Tutto il testo dell'edizione viene messo in relazione diretta con le immagini, o parti di immagine, corrispondenti, in modo da poter accedere facilmente alla porzione di immagine corrispondente partendo dal testo, e viceversa.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Obiettivo

Realizzare un collegamento fra testo e immagine in maniera tale che cliccando sul testo viene visualizzata la parte di immagine corrispondente e viceversa

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizione digitale Facsimile

Gli schemi di codifica TEI versione P5 (2007) introducono numerosi miglioramenti per quanto riguarda la gestione e trascrizione di manoscritti

Tra queste la nuova sezione Digital facsimiles nel capitolo 11 Representation of Primary Sources: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizione digitale Facsimile

Includendo il modulo **transcr** nello schema di codifica TEI si rende disponibile un attributo globale @facs

Edizione digitale Facsimile

@facs (facsimile) collega tutta o parte di una immagine al corrispettivo contenuto di un elemento.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizione digitale Facsimile

Questo attributo può essere usato in qualsiasi elemento per associare il contenuto dello stesso a un'immagine

Edizione digitale Facsimile: esempi

```
<p n="1" facs="para1.jpg" >  
<head facs="head.jpg" >  
<pb facs="page1.jpg" />
```

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

oltre a @facs è necessario usare gli altri elementi del modulo transcription per collegare testo a immagine

Edizione image-based

- <facsimile>
- <surface>
- <zone>

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Facsimile

■ <facsimile>

costituisce una rappresentazione digitale di una fonte con testo scritto espressa attraverso un insieme di immagini, piuttosto che attraverso la trascrizione e la codifica del testo.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Facsimile

■ <surface>

definisce una superficie di scrittura in termini di uno spazio espresso attraverso un sistema di coordinate.

Opzionalmente può essere inserita una collezione di elementi *graphic* che rappresentano graficamente lo spazio indicato. E' possibile caratterizzare la superficie con successivi elementi *zone* che rappresentano specifiche regioni di interesse su di essa.

L'attributo @start punta ad un elemento di testo che codifica dove ha inizio il contenuto della superficie interessata.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Elementi facsimile TEI

Elemento Facsimile

■ <zone>

definisce un'area di interesse all'interno dell'elemento surface.



Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Facsimile

L'elemento <facsimile> è di tipo strutturale e si pone allo stesso livello di <text> o addirittura in alternativa a quest'ultimo

Elemento Facsimile

La TEI permette una **grande flessibilità**: facsimile digitale con le immagini del manoscritto; facsimile digitale con trascrizione del testo; facsimile digitale con trascrizione del testo e collegamento all'immagine.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Quando il modulo *transcr* viene aggiunto allo schema di codifica è possibile scegliere fra:

Elemento Facsimile

- un <teiHeader> e un <facsimile>
- un <teiHeader> e un <text>
- un <teiHeader>, un <facsimile> e un <text> .

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Facsimile

```
<TEI>
<teiHeader>
<!-- ... -->
</teiHeader>
<facsimile>
<graphic url="page1.png"/>
<graphic url="page2.png"/>
<graphic url="page3.png"/>
<graphic url="page4.png"/>
</facsimile>
</TEI>
```

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Facsimile

```
<TEI>
<teiHeader>
<!-- ... -->
</teiHeader>
<text>
  <pb facs="page1.png"/>
  <!-- inserire qui il testo di pagina 1 -->
  <pb facs="page2.png"/>
  <!-- inserire qui il testo di pagina 2 -->
</text>
</TEI>
```

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Facsimile

Grazie a un foglio di stile XSLT è possibile generare una pagina HMTL divisa in due riquadri (immagine e testo)

Elemento Facsimile

Nessun collegamento testo-immagine a livello diverso dalla pagina e non è possibile nemmeno individuare aree particolari sulle immagini.

Infine, i puntatori alle immagini sono sparsi per tutto il documento.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Facsimile

La soluzione più efficace è la **parallel transcription** basata su <facsimile> e <text> e all'uso di <surface> e <zone> all'interno di <facsimile>.

Elemento Facsimile

Possibilità di definire le aree delle immagini e collegare il testo della trascrizione a tali aree e/o immagini secondarie.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Facsimile

Le aree delle immagini sono individuate per mezzo di un sistema di coordinate cartesiane registrate come valori di attributi di `<surface>` e `<zone>`.

Coordinate area immagini

- **ulx, uly** coordinate x e y dell'angolo superiore sinistro
- **lrx, lry** coordinate x e y dell'angolo inferiore destro

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Surface

<surface> individua la superficie scritta di un'immagine

Elemento Surface: esempio

```
<surface ulx="0" uly="0" lrx="400" lry="280" >  
  <graphic url="page1.png"/>  
</surface>
```

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Surface

può contenere più di un elemento <graphic>

Elemento Surface: esempio

```
<surface> <graphic url="page1-highRes.png"/>  
<graphic url="page1-lowRes.png"/> </surface>
```

<surface> stesso si trova all'interno di <facsimile>

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

l'elemento surface invece di <graphic> può contenere una o più <zone>

Elemento Zone

<zone> definisce una specifica area dell'immagine usando lo stesso sistema di coordinate di <surface>

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elemento Zone: esempio

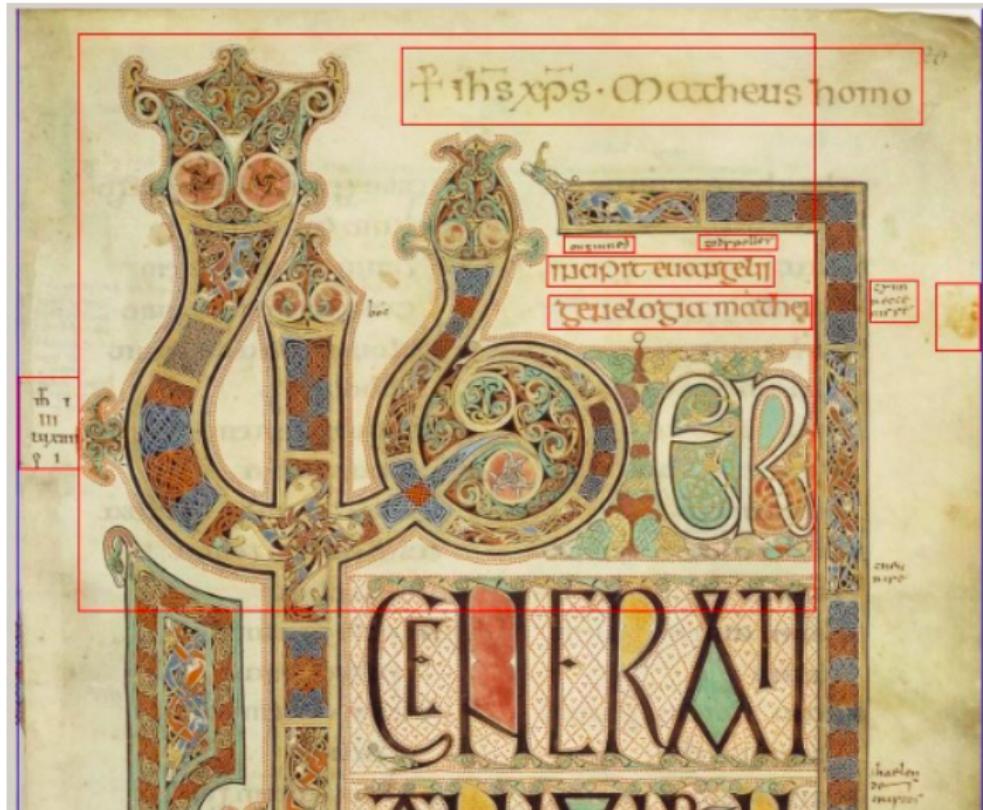
```
<surface ulx="0" uly="0" lrx="500" lry="321" >
  <zone ulx="50" uly="20" lrx="400" lry="280" >
    <graphic url="scrittura.png"/>
  </zone>
  <note>first page</note>
</surface>
```

<surface> stesso si trova all'interno di <facsimile>

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

Elementi facsimile TEI



Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Esempio Facsimile Zone

```
<facsimile xml:id="imtAnnotatedImage">
  ...
  <surface>
    <graphic height="1797px" url="IncipitMatt.jpg" width="1266px" />
    <zone lrx="1268" lry="1797" ulx="0" uly="4" xml:id="imtArea-0" />
    <zone lrx="1267" lry="450"  ulx="1202" uly="356" xml:id="imtArea-1" />
    <zone lrx="1050" lry="792"  ulx="81" uly="30" xml:id="imtArea-3" />
    <zone lrx="1190" lry="154"  ulx="503" uly="48" xml:id="imtArea-4" />
    <!-- altre zone -->
  </surface>
</facsimile>
```

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Collegare il testo alle immagini

Per collegare il testo della trascrizione alle aree corrispondenti dell'immagine bisogna **assegnare un identificatore univoco a ciascun elemento del facsimile** usando **l'attributo xml:id**.

Collegare il testo alle immagini

Usare l'attributo **facs** negli elementi testuali per **specificare l'id** degli elementi `<surface>` e `<zone>` corrispondenti

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Collegare le immagini al testo

Per collegare le aree delle immagini ai corrispondenti elementi di testo bisogna assegnare un **identificatore univoco** a *ciascun elemento della trascrizione* usando l'**attributo xml:id**

Collegare le immagini al testo

Usare l'attributo **start** oppure **corresp** negli elementi **<surface>** e **<zone>** per **specificare l'id** degli *elementi testuali corrispondenti*.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Esempio collegamento completo

```
<text>
    <body>
        <div>
            <pb facs="#page1" n="1" xml:id="page-1" />
            <p>Lorem ipsum ... <gloss facs="#det1">semper</gloss>
        </div>
    </body>
</text>
<facsimile>
    <surface xml:id="page1" start="#page-1" ulx="0" uly="0" lrx="500" lry="321">
        <graphic url="page1.png" />
        <zone xml:id="line1" ulx="50" uly="80" lrx="200" lry="321">
            <graphic url="line1.png" />
            <note>First page.</note>
        </zone>
        <zone xml:id="det1" ulx="120" uly="48" lrx="143" lry="56">
            <graphic url="gloss.png" />
            <note>Scribe gloss.</note>
        </zone>
    </surface>
</facsimile>
```

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Embedded transcription

Un metodo a metà fra facsimile digitale ed edizione basata su immagini è quello della **embedded transcription**.

Differenza con il metodo parallel transcription

Il testo è considerato di accompagnamento, il focus infatti è sulle immagini (ad esempio disposizione fisica delle parti).

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHZLAB>

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Embedded transcription

Per implementare il metodo si usa l'elemento `<sourceDoc>` sullo stesso livello gerarchico e in alternativa a `<facsimile>` e `<text>`.

Embedded transcription

All'interno di `<sourceDoc>` gli elementi `<surface>` e `<zone>` vengono utilizzati in maniera simile a quanto visto per l'elemento `<facsimile>`.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Embedded transcription

L'elemento `<zone>` contiene una serie di elementi `<line>` corrispondenti alle righe di testo

Embedded transcription: esempio

```
<zone uly="40" lry="180" ulx="20" lrx="120" >  
    <line>prima riga di trascrizione</line>  
    <line>seconda riga di trascrizione</line>  
</zone>
```

Il content model di `<line>` è limitato e con ci sono conflitti di gerarchie

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Embedded transcription

```
<sourceDoc>
  <surfaceGrp n="South_Wall">
    <surface xml:id="EETS.29">
      <zone>
        <graphic url="img_0479.JPG"/>
        <line><orig>Now <damage>in the</damage> name of ou<damage>r</damage>e
          <damage>or</damage>d ih<ex>esu</ex>s</orig></line>
        <line><orig>of right hole herte <ex>&amp;</ex> in our <damage>b</damage>est
          entent</orig></line>
        <line><orig>ou<damage>r</damage>
          <damage>l</damage>yf reme<ex>m</ex>bryng froward and
          vicious</orig></line>
        <line><orig>a<damage>y</damage> co<damage>n</damage>t<damage>ra</damage>rye to
          the commandement</orig></line>
        <line><orig>of crist i<ex>hes</ex><damage>u</damage> now wy<damage>th</damage>
          aviseme<damage>nt</damage></orig></line>
        <line><orig>the l<damage>ord beseching of mercy and pete</damage></orig></line>
        <line><orig>our you<damage>th</damage>e <ex>&amp;</ex> a<damage>g</damage>
          tha<damage>l</damage> we have <damage>myspent</damage></orig></line>
        <line><orig>wyt<damage>h this woord mercy knelyng o</damage>
          o<damage>ur</damage> kne</orig></line>
      </zone>
    </surface>
    <surface />
    <surface />
  </surfaceGrp>
</sourceDoc>
```

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Embedded Transcription

Esempio Taccuini di Proust

[http://research.cch.kcl.ac.uk/proust_prototype/
index.html](http://research.cch.kcl.ac.uk/proust_prototype/index.html)

The screenshot shows a digital edition interface for a manuscript. At the top, a navigation bar includes links for 'About', 'Writing sequence' (which is highlighted in red), 'Reading sequence', 'Contact', 'Go to opening', 'Previous', 'Next', 'Opening 46v-47r', and 'Hide Banner'. Below the bar, there are two pages of handwritten French text. The left page has a large red oval drawn around the bottom portion of the text. On the right page, there is a red box at the top containing some text. A vertical toolbar on the left side of the main area contains icons for zooming in and out, and numbered circles (1, 2, 3, 4) which likely correspond to specific annotations or sections of the text.

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Esistono numerosi programmi per calcolare le coordinate di aree su immagini facsimile.

Come inserire le coordinate

Software di disegno in formato bitmap strumenti per
programmatori di siti web.

Software progettati per creatori di edizioni digitali (*EPPT*).

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Strumenti per immagini facsimile

Un utile strumento per ottenere le coordinate di regioni di interesse in formato TEI è **TEIzoner**.

<http://teicat.huma-num.fr/zoner.php>

Facsimile ed edizione digitale

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizioni image-based: Esercizio

Codificare lettera Bellini utilizzando il tagset facsimile e il collegamento completo testo-immagine



Progress status

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

1 Introduzione agli elementi editoriali e facsimile TEI

2 Elementi editoriali TEI

3 Elementi facsimile TEI

4 Conclusioni

Edizioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Moduli TEI P5

Le nuove caratteristiche degli schemi TEI P5 offrono un'ottima base per edizioni digitali complesse con collegamento testo-immagine

Conclusioni: Editioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizioni digitali scientifiche

- elementi contenuti nei *moduli di base*
- elementi del modulo di *descrizione dei manoscritti*
- elementi del modulo di *trascrizione delle fonti primarie*
- elementi del modulo di *apparato critico*
- elementi del modulo di *gestione di caratteri non standard*

Conclusioni: Edizioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

elementi per interventi editoriali:

<abbr> <expan>, <orig> <reg>, <sic> <corr>,
<subst> <gap/>, <supplied>, <unclear>, <damage>

strutturali specifici:

<gb/>, <line>

Conclusioni: Edizioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elementi di intervento editoriali

- <damage> marca la parte di testo danneggiata non proprio “intervento editoriale” ma spesso usato contestualmente con <gap/>, <unclear> e <supplied>
- <supplied> testo inserito dal curatore perché l’originale è mancante o illeggibile

Conclusioni: Edizioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elementi di intervento editoriali: esempio

```
<l n="1" >Nel mezzo del cammin di nostra vita</l>
<l n="2" ><damage agent="fire" extent="1line"
><unclear>Mi ritrovai</unclear> <supplied
reason="illegible" resp="amdg" >per una selva
oscura,</supplied></damage></l>
<l n="3" >Ché la diritta via era smarrita</l>
```

Conclusioni: Edizioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elementi di intervento editoriale

- **<subst>** raggruppa una cancellazione e un'aggiunta
scribale per rendere evidente che si tratta di una
sostituzione
Stessa semantica funzionale di **<choice>**

Conclusioni: Edizioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elementi di intervento editoriale

```
<l n="1" >Nel mezzo del cammin di nostra vita</l>
<l n="2" >Mi ritrovai
<subst>
<del>pir</del>
<add>per</add>
</subst>
una selva oscura,</l>
<l n="3" >Ché la diritta via era smarrita</l>
```

Conclusioni: Edizioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizione image-based: Elementi strutturali

- **<gb/> gathering begins**
marca il punto in cui si presenta un nuovo fascicolo
all'interno di un manoscritto
- @type: classificazione in base al tipo
- @n: numero progressivo

Conclusioni: Edizioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Edizione image-based: Elementi strutturali

- **<line>**
trascrizione di una riga del foglio del manoscritto
può essere contenuto solo da <surface> e <zone>!
- per le righe di testo da inserire all'interno di <text> è
sempre necessario usare <lb/>.

Conclusioni: Edizioni Digitali

Codifica TEI -
Elementi
editoriali e
facsimile

A.M. Del
Grosso

Introduzione
agli elementi
editoriali e
facsimile TEI

Elementi
editoriali TEI

Elementi
facsimile TEI

Conclusioni

Elementi per interventi editoriali

A causa della relativa complessità di codifica di edizioni digitali image-based è preferibile usare strumenti software per facilitare la creazione di un facsimile digitale

strutturali specifici

Manca però ancora uno strumento/funzione per collegare le immagini annotate al testo della trascrizione

XPath - XSL

Angelo Mario Del Grosso

CNR-ILC

<http://ilc.cnr.it/>

angelo.delgrosso@ilc.cnr.it

**Selezione, Elaborazione e Presentazione di documenti XML-TEI mediante
i linguaggi XPath e XSL**

Istituto di Linguistica Computazionale “A. Zampolli”,
8th May 2024

Outline

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

1 Lesson path

2 Introduction

3 XPath

4 XSL Transformations

5 XSL in action

6 References

Progress status

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

1 Lesson path

2 Introduction

3 XPath

4 XSL Transformations

5 XSL in action

6 References

Lesson path

XPath - XSL

A.M. Del
Grosso

Lesson path

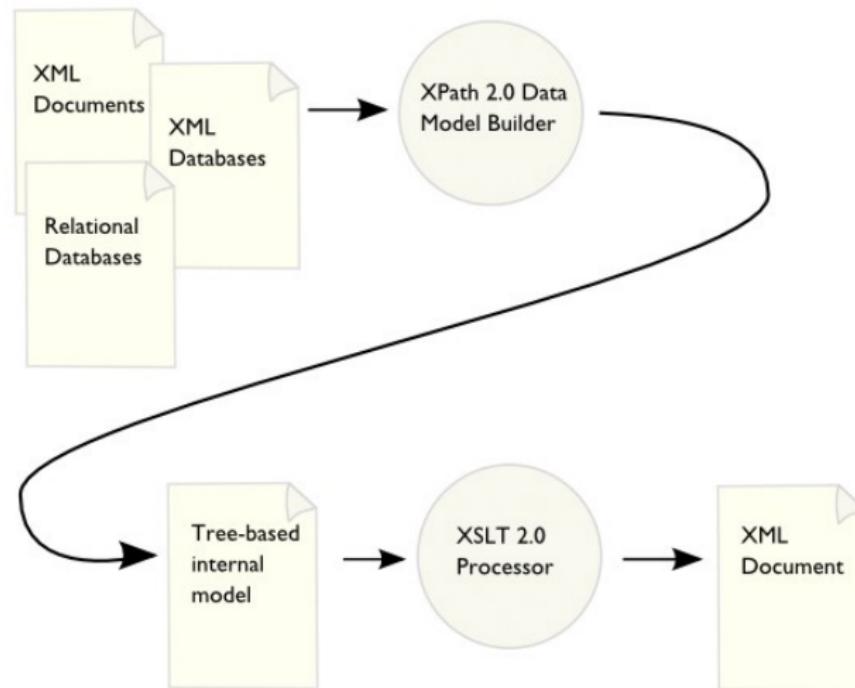
Introduction

XPath

XSL Trans-
formations

XSL in action

References



Progress status

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

1 Lesson path

2 Introduction

3 XPath

4 XSL Transformations

5 XSL in action

6 References

XSL XQuery

A family of languages

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XSL is a family of W3C recommendations for defining XML document transformation and presentation.

XSL

- **XSL Transformations (XSLT):**
a language for transforming XML
- **The XML Path Language (XPath):**
an expression language to refer to parts of an XML document;
- **XSL Formatting Objects (XSL-FO):**
an XML vocabulary for specifying formatting semantics.

XQuery

XQuery is a query language for XML to extract data.

XSL Transformations

Versioning

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

W3C > Standards > All Standards and Drafts

Skip

XSLT COVER PAGE

2021-03-30	Recommendation	XSL Transformations (XSLT) Version 2.0 (Second Edition) This specification defines the syntax and semantics of XSLT 2.0, a language for transforming XML documents into other XML documents. XSLT 2.0 is a revised version of the XSLT 1.0 Recommendation [XSLT 1.0] published on 16 November 1999. XSLT 2.0 is designed to be used in conjunction with XPath 2.0, which is defined in [XPath 2.0]. XSLT shares the same data model as XPath 2.0, which is defined in [Data Model], and it uses the library of functions and operators defined in [Functions and Operators]. XSLT 2.0 also includes optional facilities to serialize the results of a transformation, by means of an interface to the serialization component described in [XSLT and XQuery Serialization]. <i>This document contains hyperlinks to specific sections or definitions within other documents in this family of specifications. These links are indicated visually by a superscript identifying the target specification: for example XP for XPath, DM for the XDM data model, FO for Functions and Operators.</i>
2017-06-08	Recommendation	XSL Transformations (XSLT) Version 3.0 This specification defines the syntax and semantics of XSLT 3.0, a language for transforming XML documents into other XML documents.
1999-11-16	Recommendation	XSL Transformations (XSLT) Version 1.0 This specification defines the syntax and semantics of XSLT, which is a language for transforming XML documents into other XML documents. XSLT is designed for use as part of XSL, which is a stylesheet language for XML. In addition to XSLT, XSL includes an XML vocabulary for specifying formatting. XSL specifies the styling of an XML document by using XSLT to describe how the document is transformed into another XML document that uses the formatting vocabulary. XSLT is also designed to be used independently of XSL. However, XSLT is not intended as a completely general-purpose XML transformation language. Rather it is designed primarily for the kinds of transformations that are needed when XSLT is used as part of XSL.

RELATED RETIRED SPECIFICATIONS

2001-08-24	Retired	XSL Transformations (XSLT) Version 1.1
------------	---------	---

XSL Transformations

Versioning

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

W3C > Standards > All Standards and Drafts

> Skip <

XPATH COVER PAGE

2017-03-21	Recommendation	XML Path Language (XPath) 3.1 XPath is an expression language that allows the processing of values conforming to the data model defined in the XQuery and Xpath Data Model.
2014-04-08	Recommendation	XML Path Language (XPath) 3.0 XPath 3.0 (renamed from XPath 2.1 to align with the family of "3.0" specifications) is an expression language that allows the processing of values conforming to the data model defined in [XQuery and XPath Data Model (XDM) 3.0]. Some of the important new features since XPath 2.0 are: Literal function items, inline functions, dynamic function invocations, and function item coercion Clarification of rules associated with sequence type matching let expressions EQNames (QNames with a namespace URI instead of a namespace prefix) Support for union types in casting and function arguments
2010-12-14	Recommendation	XML Path Language (XPath) 2.0 (Second Edition) XPath is a way to refer to parts of an XML document. XPath 2.0 is based on the XQuery 1.0 and XPath 2.0 Data Model (XDM), and also introduces Schema awareness and data typing.
1999-11-16	Recommendation	XML Path Language (XPath) Version 1.0 XPath is a language for addressing parts of an XML document, designed to be used by both XSLT and XPointer.

XSL Transformations

elements

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

80 elements ca

xsl:accept	xsl:copy	xsl:include	xsl:non-matching-substring	xsl:sort
xsl:accumulator	xsl:copy-of	xsl:iterate	xsl:number	xsl:source-document
xsl:accumulator-rule	xsl:decimal-format	xsl:key	xsl:on-completion	xsl:stream
xsl:analyze-string	xsl:document	xsl:map	xsl:on-empty	xsl:strip-space
xsl:apply-imports	xsl:element	xsl:map-entry	xsl:on-non-empty	xsl:stylesheet
xsl:apply-templates	xsl:evaluate	xsl:matching-substring	xsl:otherwise	xsl:template
xsl:assert	xsl:expose	xsl:merge	xsl:output	xsl:text
xsl:attribute	xsl:fallback	xsl:merge-action	xsl:output-character	xsl:transform
xsl:attribute-set	xsl:for-each	xsl:merge-key	xsl:override	xsl:try
xsl:break	xsl:for-each-group	xsl:merge-source	xsl:package	xsl:use-package
xsl:call-template	xsl:fork	xsl:message	xsl:param	xsl:value-of
xsl:catch	xsl:function	xsl:mode	xsl:perform-sort	xsl:variable
xsl:character-map	xsl:global-context-item	xsl:namespace	xsl:preserve-space	xsl:when
xsl:choose	xsl;if	xsl:namespace-alias	xsl:processing-instruction	xsl:where-populated
xsl:comment	xsl:import	xsl:next-iteration	xsl:result-document	xsl:with-param
xsl:context-item	xsl:import-schema	xsl:next-match	xsl:sequence	

XSL Transformations

Documentation

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

xsl:template

Defines a processing rule for source elements or other nodes of a particular type.

Available in XSLT 1.0 and later versions. Available in all Saxon editions.

- **Category:** declaration
- **Content:** (`xsl:context-item?`, `xsl:param*`, *sequence-constructor*)
- **Permitted parent elements:** `xsl:package`; `xsl:stylesheet`; `xsl:transform`; `xsl:override`

Attributes

`match?`

`pattern`

Pattern to identify the type of node to be processed. The most common form of pattern is simply an element name. However, more complex patterns may also be used: the syntax of patterns is given in more detail in XSLT Pattern Syntax. The following examples show some of the possibilities:

Pattern

Meaning



XPath functions

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Number of functions: 177 of 177

[abs](#)
[accumulator-after](#)
[accumulator-before](#)
[adjust-date-to-timezone](#)
[adjustdateTime-to-timezone](#)
[adjust-time-to-timezone](#)
[analyze-string](#)
[apply](#)
[available-environment-variables](#)
[available-system-properties](#)
[avg](#)
[base-uri](#)
[boolean](#)
[ceiling](#)
[codepoint-equal](#)
[codepoints-to-string](#)
[collation-key](#)
[collection](#)
[compare](#)
[concat](#)
[contains](#)
[contains-token](#)
[copy-of](#)
[count](#)
[current](#)
[current-date](#)
[current-datetime](#)
[current-group](#)
[current-grouping-key](#)

[fold-left](#)
[fold-right](#)
[for-each](#)
[for-each-pair](#)
[format-date](#)
[format-datetime](#)
[format-integer](#)
[format-number](#)
[format-time](#)
[function-arity](#)
[function-available](#)
[function-lookup](#)
[function-name](#)
[generate-id](#)
[has-children](#)
[head](#)
[hours-from-datetime](#)
[hours-from-duration](#)
[hours-from-time](#)
[id](#)
[idref](#)
[implicit-timezone](#)
[in-scope-prefixes](#)
[index-of](#)
[innermost](#)
[insert-before](#)
[iri-to-uri](#)
[json-doc](#)
[json-to-xml](#)

[parse-xml](#)
[parse-xml-fragment](#)
[path](#)
[position](#)
[prefix-from-QName](#)
[put](#)
[QName](#)
[random-number-generator](#)
[regex-group](#)
[remove](#)
[replace](#)
[resolve-QName](#)
[resolve-uri](#)
[reverse](#)
[root](#)
[round](#)
[round-half-to-even](#)
[seconds-from-datetime](#)
[seconds-from-duration](#)
[seconds-from-time](#)
[serialize](#)
[snapshot](#)
[sort](#)
[starts-with](#)
[static-base-uri](#)
[stream-available](#)
[string](#)
[string-join](#)
[string-length](#)

XPath

Documentation

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

fn:substring

`substring($sourceString as xs:string?, $start as xs:double) → xs:string`

Returns a substring of a given string starting at the given starting position and continuing to the end of the string.

Arguments

\$sourceString	xs:string?	The input string
\$start	xs:double	The position of the first character of the input string to be included in the result

Result

xs:string

`substring($sourceString as xs:string?, $start as xs:double, $length as xs:double) → xs:string`

Returns a substring of a given string starting at the given starting position and continuing to the end of the string, or \$length characters if shorter.

Arguments

\$sourceString	xs:string?	The input string
\$start	xs:double	The position of the first character of the input string to be included in the result
\$length	xs:double	The number of characters to be included in the result

Result

xs:string

Namespace

<http://www.w3.org/2005/xpath-functions>

Links to W3C specifications

[XPath 2.0 Functions and Operators](#)

[XPath 3.0 Functions and Operators](#)

[XPath 3.1 Functions and Operators](#)

W3C XML standards

XML specs

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XML family specifications

[https:](https://www.w3.org/TR/?tag=xml&status=REC&version=latest)

[//www.w3.org/TR/?tag=xml&status=REC&version=latest](https://www.w3.org/TR/?tag=xml&status=REC&version=latest)

Tree data model (XDM)

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References



- The W3C specifications for XSLT, XQuery, and XPath **model an XML document as a tree**. This data model is known as **XDM**, and the nodes of an XDM tree are known as **XDM nodes**.
- XDM defines the information contained in the input to an XSLT processor as well as it defines all **permissible values of expressions** in the XSLT
- The **node-sets of XPath 1.0** are replaced in XPath 2.0 by **sequences of nodes**.

Tree data model (XDM)

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References



- The W3C specifications for XSLT, XQuery, and XPath **model an XML document as a tree**. This data model is known as **XDM**, and the nodes of an XDM tree are known as **XDM nodes**.
- XDM defines the information contained in the input to an XSLT processor as well as it defines all **permissible values of expressions** in the XSLT
- The **node-sets of XPath 1.0** are replaced in XPath 2.0 by **sequences of nodes**.

Tree data model (XDM)

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References



- The W3C specifications for XSLT, XQuery, and XPath **model an XML document as a tree**. This data model is known as **XDM**, and the nodes of an XDM tree are known as **XDM nodes**.
- XDM defines the information contained in the input to an XSLT processor as well as it defines all **permissible values of expressions** in the XSLT
- The **node-sets of XPath 1.0** are replaced in XPath 2.0 by **sequences of nodes**.

Tree data model (XDM)

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References



- The W3C specifications for XSLT, XQuery, and XPath **model an XML document as a tree**. This data model is known as **XDM**, and the nodes of an XDM tree are known as **XDM nodes**.
- XDM defines the information contained in the input to an XSLT processor as well as it defines all **permissible values of expressions** in the XSLT
- The **node-sets of XPath 1.0** are replaced in XPath 2.0 by **sequences of nodes**.

XML Trees

Hierarchical Ordered Nodes

XPath - XSL

A.M. Del
Grosso

Lesson path

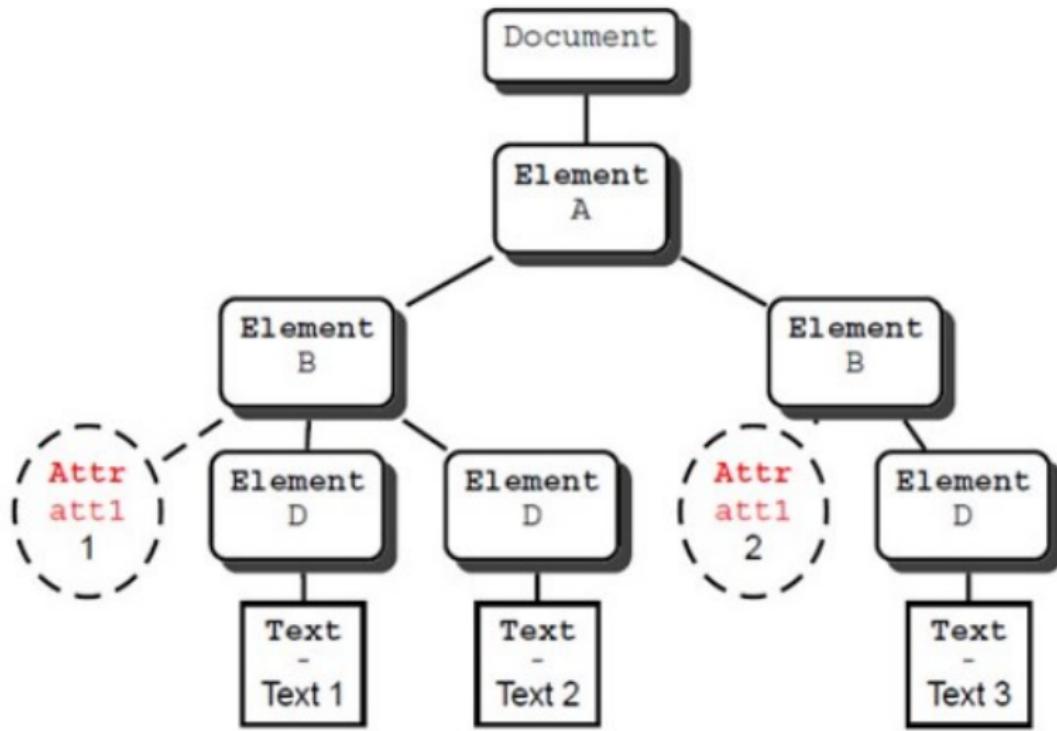
Introduction

XPath

XSL Trans-
formations

XSL in action

References



XPath Data Model

XDM

XPath - XSL

A.M. Del
Grosso

Lesson path

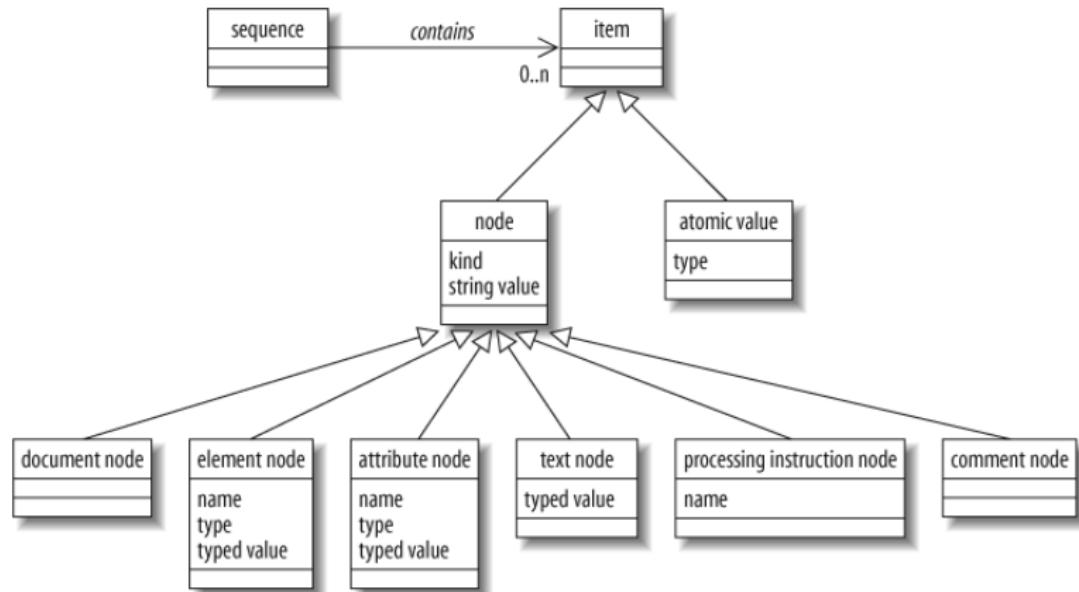
Introduction

XPath

XSL Trans-
formations

XSL in action

References



Fondamenti Extensible Stylesheet Language

XPath - XSL

A.M. Del
Grosso

Lesson path

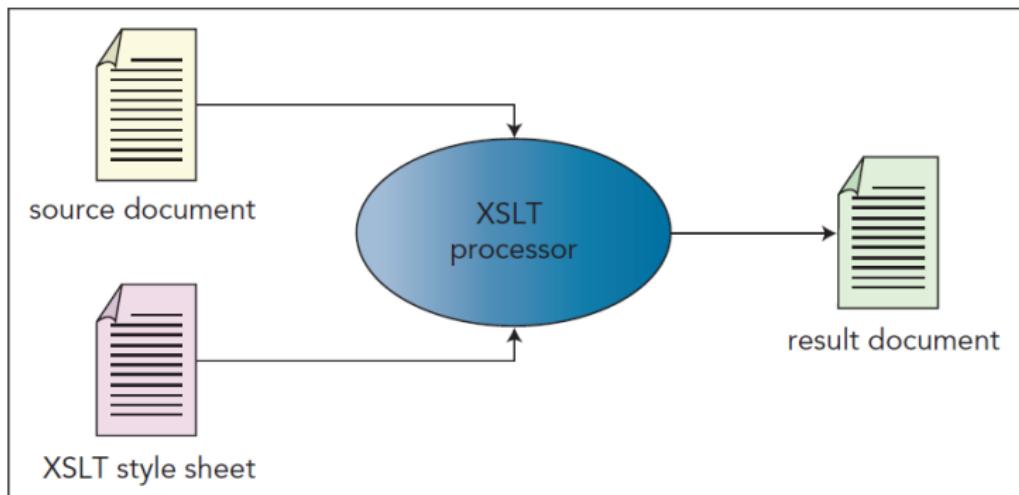
Introduction

XPath

XSL Transformations

XSL in action

References



XML Document

Exemple from The Inscriptions of Roman Tripolitania

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xml:lang="en">
  <teiHeader> ...
  </teiHeader>
  <text>
    <body>
      <div type="bibliography"> ...
      </div>
      <div subtype="text-constituted-from" type="history"> ...
      </div>
      <div type="edition" xml:lang="la"><head xml:lang="en">Text</head><ab> ...
      </ab></div>
      <div type="apparatus"> ...
      </div>
      <div type="translation"><head>Translation</head> ...
      </div>
      <div type="commentary"> ...
      </div>
      <div type="figure"> ...
      </div>
    </body>
  </text>
</TEI>
```

eXtensible Style Sheet

Example

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <xsl:stylesheet version="1.0">
3  --- xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
4  --- <xsl:output method="text" />
5  --- <xsl:template match="/">
6  -----<xsl:apply-templates select="current()/descendant::text" />
7  -----</xsl:template>
8  -----<xsl:template match="div[@type='edition']|div[@type='translation']">
9  -----<xsl:value-of select="normalize-space(.)" />
10 -----</xsl:template>
11 -----<xsl:template match="div" />
12 </xsl:stylesheet>
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

wsl + □ ☰ ...

```
angelodel80@LAPTOP-V8V3MLGO:/mnt/c/Users/angel/risorse/universita/corsoCodifica/tools$ java -jar SaxonHE10-3J/
saxon-he-10.3.jar -s:source/IRT030.xml -xsl:source/built-in.xsl
```

Text Imperator Caesar Marci Antonini Pii fili diui Pii nep diui Hadriani pronep diui Traiani Parthici
abnep diui Neruae adnep Luci Aeli Aureli Commod August

Translation Emperor Caesar case unknown son of Marcus Antonius Pius, grandson of deified Pius, great
grandson of deified Hadrian, great great grandson of deified Trajan victor in Parthia, great great great grand
son of deified Nerva, Lucius Aelius Aurelius Commodus Augustus



How Does it work

The XSLT processor **uses the instructions** in the Style Sheet to process the input XML document by **traversing the document's hierarchy**.

How Does it work

XSLT instructions indicate what **portion of the tree should be traversed**, how it should be inspected, and what **output fragment should be generated** at each point.

Ordered Hierarchy of Content Objects

OHCO

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

■ OHCO

The nodes of the tree are ordered. The child nodes of a parent node, which are siblings of one another, **occur in a particular order.**

This is why XML can be described as representing an ordered hierarchy of content objects.

Progress status

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

1 Lesson path

2 Introduction

3 XPath

4 XSL Transformations

5 XSL in action

6 References

Selecting and Processing XML Document Trees

Basic Concepts

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath

Path **expressions** will return **node sequences** whose nodes are in document order

Selecting and Processing XML Document Trees

Basic Concepts

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Sequence

The group of nodes that an XPath expression returns is a sequence, which is a technical term for an **ordered collection of items** that permits duplicates

Selecting and Processing XML Document Trees

XPath expression language

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath

- Select nodes sequence from XML tree
- Process data via functions

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath

XPath expressions are extremely accurate (selection of elements, attributes, texts etc.)

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <xsl:stylesheet version="2.0">
3      xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
4      <xsl:output method="text"/>
5
6      <xsl:template match="/">
7          ...
8              <xsl:value-of
9                  select="/TEI/text/body/div/@type" />
10             ...
11         </xsl:template>
12     </xsl:stylesheet>
13
14
15
```

out-text.txt U ×

1 bibliography history edition apparatus translation commentary figure

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath Basics

Path expressions are used to navigate from a current location
(called the context node) to other nodes in the tree.

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

XPath Basics

Steps in a path expression are indicated with slash characters
and the context node changes with each step.

XPath Basics

- "div" (*div child*)
- "div/head" (*child of div*)
- "div/*/persName" (*child of child*)
- "div//persName" (*descendant*)

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath

The XPath expression can be absolute or relative to the **context node**

XPath

The XPath expression encompasses three components: (**Axes, Test, Predicate**)

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath

XPath expressions navigate the XML tree by using the so called (*expression axes*).

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

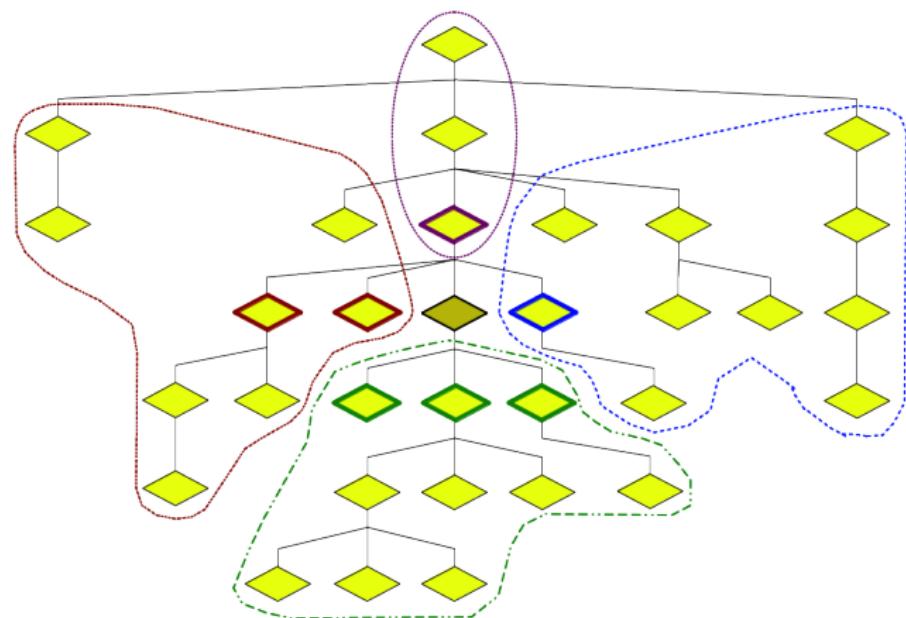
Introduction

XPath

XSL Trans-
formations

XSL in action

References



Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Axes	Depiction	Nodes
child	Dark green edges	The three nodes immediately below the current location
descendant	Dashed green line	The three child nodes mentioned above, plus the seven nodes below them, all the way down (their children and their children's children)
parent	Magenta edges	The node immediately above the current location
ancestor	Magenta dashed line	The parent plus its parent, and its parent's parent
preceding-sibling	Dark red edges	The two nodes to the left of the current location that have the same parent
preceding	Dark red dashed line	The preceding-sibling nodes plus the six other nodes that are entirely to the left of the current location
following-sibling	Blue edges	The node to the right of the current location that has the same parent
following	Blue dashed line	The following-sibling node plus the nine other nodes that entirely to the right of the current location

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Symbol	Meaning	Expanded version
.	current context node	<code>self::*</code> (for elements)
..	parent element	<code>parent::*</code>
//	descendant axis	<code>descendant::</code> . At the beginning of a path expression, it means that the path starts at the document node.
@	attribute axis	<code>attribute::</code>

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <xsl:stylesheet version="2.0"
3  .... xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
4  .... <xsl:output method="text" encoding="UTF-8" />
5  ....
6  .... <xsl:template match="/">
7  ....
8  .... |.... <xsl:value-of select="TEI/descendant::langUsage/language" />
9  ....
10 ....
11 </xsl:template>
12 </xsl:stylesheet>
```

out-text.txt U X

```
1 Arabic English French German Ancient Greek Transliterated Greek Modern Greek Hebrew Italian Latin Punic
```

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath Predicate

Predicates are **conditional expression** and are used to **filter the results** of the path expression.

XPath Predicate

Predicate expression are written in **square brackets after the step** in the path expression to which they apply

Any expression in square brackets that filters a step in a path expression is a predicate

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <xsl:stylesheet version="2.0">
3      xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
4      <xsl:output method="text" encoding="UTF-8" />
5
6      <xsl:template match="/">
7          <xsl:value-of select="//div[1]" />
8
9      </xsl:template>
10
11
12  </xsl:stylesheet>
```

out-text.txt U X

```
1
2      Bibliography
3          Not previously published.
4
```

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

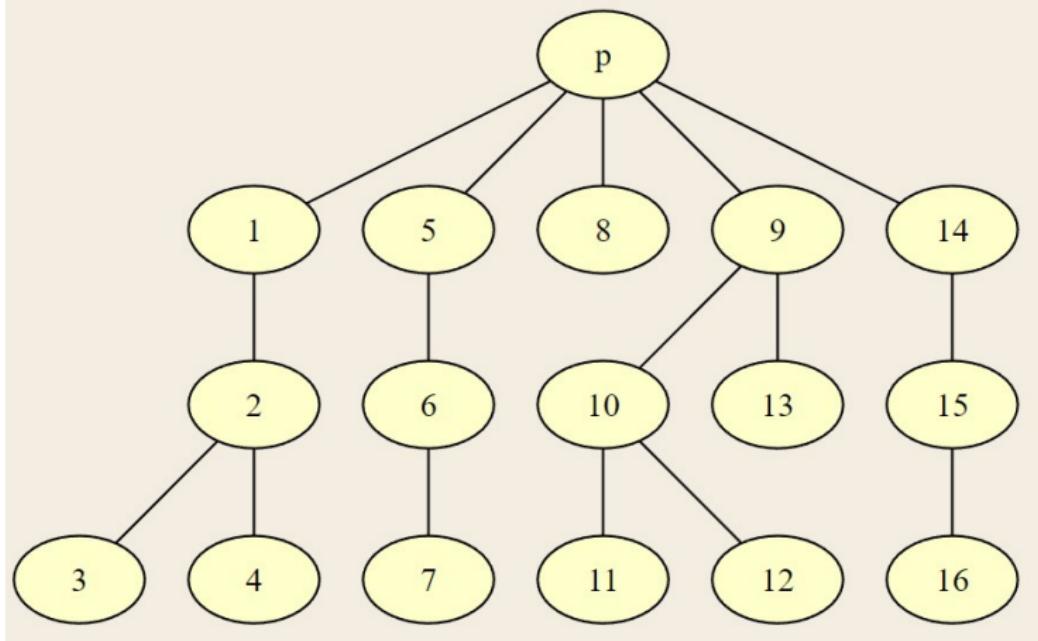
Introduction

XPath

XSL Trans-
formations

XSL in action

References



Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Description	Value	General
Equal to	eq	=
Not equal to	ne	!=
Greater than	gt	> (>)
Greater than or equal to (not less than)	ge	>= (>=)
Less than	lt	< (<)
Less than or equal to (not greater than)	le	<= (<=)

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Selecting and Processing XML Document Trees

Examples of Predicates

- `div[@type]`
- `div[@type="edition"]`

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath Basics

"jolly" element selection

- *
- *[@type]
- *[@type="book"]

Selecting and Processing XML Document Trees

XPath: functions

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Esempio predicati

- `//div[@type="edition"]`
- `//div[@type!="translation"]`
- `//div[@n > 2]`
- `//div[1]`
- `//div[last()]`
- `//div[position() = last() - 1]`
- `//div[position() mod 2 = 0]`

Selecting and Processing XML Document Trees

XPath location path

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath Basics

Use of OR to select more than one named elements and ID
function to select the element with the given ID

- "title | author"
- id("irt1952")

Selecting and Processing XML Document Trees

XPath location path

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <xsl:stylesheet version="2.0">
3  ....xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
4  ....<xsl:output method="text" />
5  ....
6  ....<xsl:template match="/">
7  ....
8  ....<xsl:value-of
9  .... select="descendant::langUsage/language[@ident='it']/preceding-sibling::*[1]/text()"
10 .... />
11 ....
12 ....</xsl:template>
13
14 </xsl:stylesheet>
15
```

out-text.txt U ×

1 Hebrew

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

XPath functions

- Functions operate on the information returned by a path expression or another function
- Functions can be nested
- functions can be used in predicates to filter expressions

Selecting and Processing XML Document Trees

XPath functions

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
2 <xsl:stylesheet version="2.0">
3   xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
4   <xsl:output method="text" encoding="UTF-8" />
5   <xsl:template match="/">
6     <xsl:value-of select="concat(
7       name(TEI/descendant::langUsage/language[1]), ',',
8       count(TEI/descendant::langUsage/language), ',',
9       //w[contains(., 'filii')]/@lemma, ',',
10      upper-case(
11        substring(TEI/descendant::langUsage/language[@ident eq 'it'], 0, 4)
12      ), ','
13    )"
14   </xsl:template>
15 </xsl:stylesheet>
```

out-text.txt U X

1 language 12 filius ITA

Selecting and Processing XML Document Trees

XPath functions

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

fn:count

Counts the number of items in a sequence.

`count($arg as item()*) → xs:integer`

Arguments

\$arg item()*

The sequence whose items are
to be counted

Result

xs:integer

Namespace

<http://www.w3.org/2005/xpath-functions>

Links to W3C specifications

[XPath 2.0 Functions and Operators](#)

[XPath 3.0 Functions and Operators](#)

[XPath 3.1 Functions and Operators](#)

Saxon availability

Available in XPath 2.0, XSLT 2.0, XQuery 1.0, and later versions. Available in all Saxon

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

fn:contains

Returns true if the second string is a substring of the first.

`contains($arg1 as xs:string?, $arg2 as xs:string?) → xs:boolean`

Arguments		
\$arg1	xs:string?	The containing string
\$arg2	xs:string?	The contained string
Result	xs:boolean	

`contains($arg1 as xs:string?, $arg2 as xs:string?, $collation as xs:string) → xs:boolean`

Arguments		
\$arg1	xs:string?	The containing string
\$arg2	xs:string?	The contained string
\$collation	xs:string	The collation to be used for comparing the strings
Result	xs:boolean	

Namespace

<http://www.w3.org/2005/xpath-functions>

Links to W3C specifications

[XPath 2.0 Functions and Operators](#)

[XPath 3.0 Functions and Operators](#)

[XPath 3.1 Functions and Operators](#)

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

fn:name

`name() → xs:string`

Returns the name of the context node, as a string in the lexical form of a QName.

There are no arguments

<i>Result</i>	xs:string
---------------	-----------

`name($arg as node())? → xs:string`

Returns the name of the supplied node, as a string in the lexical form of a QName.

Arguments

\$arg	node()?
-------	---------

The node whose name is
required

<i>Result</i>	xs:string
---------------	-----------

Namespace

<http://www.w3.org/2005/xpath-functions>

Links to W3C specifications

[XPath 2.0 Functions and Operators](#)

[XPath 3.0 Functions and Operators](#)

[XPath 3.1 Functions and Operators](#)

Saxon availability

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

fn:substring

`substring($sourceString as xs:string?, $start as xs:double) → xs:string`

Returns a substring of a given string starting at the given starting position and continuing to the end of the string.

Arguments

<code>\$sourceString</code>	<code>xs:string?</code>	The input string
<code>\$start</code>	<code>xs:double</code>	The position of the first character of the input string to be included in the result

Result

`xs:string`

`substring($sourceString as xs:string?, $start as xs:double, $length as xs:double) → xs:string`

Returns a substring of a given string starting at the given starting position and continuing to the end of the string, or `$length` characters if shorter.

Arguments

<code>\$sourceString</code>	<code>xs:string?</code>	The input string
<code>\$start</code>	<code>xs:double</code>	The position of the first character of the input string to be included in the result
<code>\$length</code>	<code>xs:double</code>	The number of characters to be included in the result

Result

`xs:string`

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

fn:upper-case

Converts a string to upper case.

`upper-case($arg as xs:string?) → xs:string`

Arguments

\$arg	xs:string?	The string to be converted to upper-case
-------	------------	---

Result

xs:string

Namespace

<http://www.w3.org/2005/xpath-functions>

Links to W3C specifications

[XPath 2.0 Functions and Operators](#)

[XPath 3.0 Functions and Operators](#)

[XPath 3.1 Functions and Operators](#)

Saxon availability

Available in XPath 2.0, XSLT 2.0, XQuery 1.0, and later versions. Available in all Saxon

Selecting and Processing XML Document Trees

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

fn:concat

Concatenates the string-values of the arguments into a single string. There must be at least two arguments.

`concat($arg1 as xs:anyAtomicType?, $arg2 as xs:anyAtomicType?, $etc... as xs:anyAtomicType?) → xs:string`

Arguments

\$arg1	xs:anyAtomicType?	The first string
\$arg2	xs:anyAtomicType?	The second string
\$etc...	xs:anyAtomicType?	The third and subsequent strings (as many as required)

Result

xs:string

Namespace

<http://www.w3.org/2005/xpath-functions>

Links to W3C specifications

[XPath 2.0 Functions and Operators](#)

[XPath 3.0 Functions and Operators](#)

[XPath 3.1 Functions and Operators](#)

Progress status

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

1 Lesson path

2 Introduction

3 XPath

4 XSL Transformations

5 XSL in action

6 References

eXtensible Style Sheet

XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

Basics

A transformation expressed in XSLT **describes rules for transforming input data into output data**. The inputs and outputs will all be instances of the XDM data model.

eXtensible Style Sheet

XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

Basics (cont.)

The input often is an XML document referred to as the **source tree**, and the output is a document referred to as the **result tree**.

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

Basics (cont.)

The transformation is achieved by **a set of template rules**. A template rule associates a pattern, which typically **matches nodes** in the source document, with a **sequence constructor**.

eXtensible Style Sheet

XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

Basics (cont.)

The structure of the result trees can be completely different from the structure of the source trees.

XSLT main elements

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

URL to Documentation

- <https://xsltdev.com/xslt/xsl-stylesheet/>
- <https://xsltdev.com/xslt/xsl-template/>
- <https://xsltdev.com/xslt/xsl-value-of/>
- <https://xsltdev.com/xslt/xsl-apply-templates/>
- <https://xsltdev.com/xslt/xsl-for-each/>
- <https://xsltdev.com/xslt/xsl-if/>
- <https://xsltdev.com/xslt/xsl-choose/>

XSLT main elements

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

URL to Documentation

- <https://xsltdev.com/xslt/xsl-sort/>
- <https://xsltdev.com/xslt/xsl-variable/>
- <https://xsltdev.com/xslt/xsl-element/>
- <https://xsltdev.com/xslt/xsl-attribute/>
- <https://xsltdev.com/xslt/xsl-key/>
- <https://xsltdev.com/xslt/xsl-preserve-space/>
- <https://xsltdev.com/xslt/xsl-strip-space/>
- <https://xsltdev.com/xslt/xsl-analyze-string/>

XSLT main elements

xsl:stylesheet

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

xsl:stylesheet

The `xsl:stylesheet` element is always the top-level element of an XSLT stylesheet. The name `xsl:transform` may be used as a synonym.

Available in XSLT 1.0 and later versions. Available in all Saxon editions.

- **Content:** (*declarations*)

Attributes

`id?`

id

Used to reference stylesheet modules embedded in a document.

`version`

decimal

Standard attribute that may appear on any XSLT element. Indicates the version of XSLT required by the stylesheet.

<https://xsltdev.com/xslt/xsl-stylesheet/>



XSLT main elements

xsl:stylesheet

```
<?xml version="1.0" encoding="UTF-8"?>

<xsl:stylesheet version="2.0">
    ...
        xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    ...
</xsl:stylesheet>
```

<https://xsltdev.com/xslt/xsl-stylesheet/>

XSLT main elements

xsl:template

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

xsl:template

Defines a processing rule for source elements or other nodes of a particular type.

Available in XSLT 1.0 and later versions. Available in all Saxon editions.

- **Category:** declaration
- **Content:** (`xsl:context-item?` , `xsl:param*` , *sequence-constructor*)
- **Permitted parent elements:** `xsl:package` ; `xsl:stylesheet` ; `xsl:transform` ;
`xsl:override`

Attributes

`match?`

`pattern`

Pattern to identify the type of node to be processed. The most common form of pattern is simply an element name. However, more complex patterns may also be used: the syntax of patterns is given in more detail in XSLT Pattern Syntax. The following examples show some of the possibilities:

<https://xsltdev.com/xslt/xsl-template/>

XSLT main elements

xsl:template

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
<?xml version="1.0" encoding="UTF-8"?>

<xsl:stylesheet version="2.0">
    <!-- xmlns:xsl="http://www.w3.org/1999/XSL/Transform" -->

        <!--<xsl:template match="/">

        </xsl:template>

        <!--<xsl:template match="div">
            ...
            </xsl:template>

        <!--<xsl:template match="head">
            ...
            </xsl:template>

        <!--<xsl:template match="lb">
            ...
            </xsl:template>

    </xsl:stylesheet>
```

<https://xsltdev.com/xslt/xsl-template/>

XSLT main elements

xsl:value-of

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

xsl:value-of

Evaluates an expression as a string, and outputs its value to the current result tree.

Available in XSLT 1.0 and later versions. Available in all Saxon editions.

- **Category:** instruction
- **Content:** sequence-constructor
- **Permitted parent elements:** any XSLT element whose content model is sequence-constructor; any literal result element

Attributes

`select?`

`expression`

Identifies the expression. If this is not specified, the value to be output is obtained by evaluating the sequence constructor contained within the `xsl:value-of` element. The full syntax of expressions is outlined in XPath Expression Syntax. Here are some examples of expressions that can be used in the `select` attribute:

<https://xsltdev.com/xslt/xsl-value-of/>

XSLT main elements

xsl:value-of

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
<xsl:stylesheet version="2.0">
  <!-- xmlns:xsl="http://www.w3.org/1999/XSL/Transform" -->

  <xsl:template match="/">
    <xsl:value-of select="concat('&#10;', normalize-space(descendant::titleStmt/title=>string()))" />
  </xsl:template>

  <xsl:template match="div">
    <xsl:value-of select="@type" />
  </xsl:template>

  <xsl:template match="head">
    <xsl:value-of select="." />
  </xsl:template>

  <xsl:template match="lb">
    <xsl:value-of select="concat('&#10;', current()/@n)" />
  </xsl:template>

</xsl:stylesheet>

txt.txt u ×
<?xml version="1.0" encoding="UTF-8"?>
Dedication to Commodus
```

<https://xsltdev.com/xslt/xsl-value-of/>

XSLT main elements

xsl:apply-templates

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

xsl:apply-templates

Causes navigation from the current element, usually but not necessarily to process its children.
Each selected node is processed using the best-match `xsl:template` defined for that node.

Available in XSLT 1.0 and later versions. Available in all Saxon editions.

- **Category:** instruction
- **Content:** (`xsl:sort` | `xsl:with-param`)*
- **Permitted parent elements:** any XSLT element whose content model is sequence-constructor; any literal result element

Attributes

`select?`

`expression`

Sequence of nodes to be processed. If this attribute is omitted, then all the immediate children of the current node are processed.

`mode?`

<https://xsltdev.com/xslt/xsl-apply-templates/>



XSLT main elements

xsl:apply-templates

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
<?xml version="1.0" encoding="UTF-8"?>

<xsl:stylesheet version="2.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

  <xsl:template match="/">
    <xsl:value-of select="concat(&#10;, normalize-space(descendant::titleStmt/
      title=>string()))" />
    <xsl:apply-templates select="//text//div" />
  </xsl:template>

  <xsl:template match="div">
    <xsl:text>&#10;</xsl:text>
    <xsl:value-of select="@type" />
    <xsl:text>&#10;</xsl:text>
  </xsl:template>

  <xsl:template match="head">
    <xsl:value-of select=".." />
  </xsl:template>

  <xsl:template match="lb">
    <xsl:value-of select="concat(&#10;, current()/
      @n)" />
  </xsl:template>

</xsl:stylesheet>
```

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  Dedication to Commodus
3  bibliography
4
5  history
6
7  edition
8
9  apparatus
10
11 translation
12
13 commentary
14
15 figure
16
```

<https://xsltdev.com/xslt/xsl-apply-templates/>

XSLT main elements

xsl:apply-templates

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="2.0">
    <!-- xmlns:xsl="http://www.w3.org/1999/XSL/Transform" -->
    <xsl:template match="/">
        <xsl:value-of select="concat('&#10;', normalize-space(descendant::titleStmt/title[string()]))"/>
        <xsl:apply-templates select="//text//div" />
    </xsl:template>

    <xsl:template match="div">
        <xsl:text>&#10;</xsl:text>
        <xsl:value-of select="name(.)=>concat(':::',@type)" />
        <xsl:text>&#10;</xsl:text>
        <xsl:apply-templates select="head" />
    </xsl:template>

    <xsl:template match="head">
        <xsl:value-of select="name(.)=>concat(':::',.,'.')"/>
    </xsl:template>

    <xsl:template match="lb">
        <xsl:value-of select="concat('&#10;', current()/@n)" />
    </xsl:template>
</xsl:stylesheet>
```

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  Dedication to Commodus
3  div : bibliography
4  head : Bibliography
5  div : history
6  head : Text constituted from
7  div : edition
8  head : Text
9  div : apparatus
10 head : Apparatus
11 div : translation
12 head : Translation
13 div : commentary
14 head : Commentary
15 div : figure
16 head : Photographs
```

<https://xsltdev.com/xslt/xsl-apply-templates/>

XSLT main elements

xsl:apply-templates

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="2.0">
  <!-- xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

  <!--xsl:strip-space elements="" /-->

  <xsl:template match="/">
    <xsl:value-of select="concat('&#10;', 'title :: ', 
      normalize-space(descendant::titleStmt/title=>string()))
    " />
    <xsl:apply-templates select="//text//div" />
  </xsl:template>

  <xsl:template match="div">
    <xsl:text>&#10;</xsl:text>
    <xsl:value-of select="name(.)=>concat(':: ', @type)" />
    <xsl:text>&#10;</xsl:text>
    <xsl:apply-templates select="head" />
  </xsl:template>

  <xsl:template match="head">
    <xsl:value-of select="name(.)=>concat(':: ', ., '&#10;')
    " />
    <xsl:apply-templates select="following-sibling::*" />
  </xsl:template>

  <xsl:template match="lb">
    <xsl:value-of select="concat('&#10;', 'riga ', current
      ()//n, ' :: ')"/>
  </xsl:template>

</xsl:stylesheet>
```

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  title : Dedication to Commodus
3  div : bibliography
4  head : Bibliography
5  Not previously published.
6  div : history
7  head : Text constituted from
8  Transcription (Reynolds, Ward-Perkins)
9  div : edition
10 head : Text
11 riga 1 : ImperatorCaesarMarciAntoniniPiifili
12 riga 2 : diuipinepediuHadrianiprompediu
13 riga 3 : TraianiParthiciabnepdiuNeruaeadnep
14 riga 4 : LuciAeliAureliCommmodAugust
15 div : apparatus
16 head : Apparatus
17 Line 4 is represented by the upper parts only of
two letters with rounded bowls.
18 div : translation
19 head : Translation
20 Emperor Caesar son of Marcus Antonius Pius,
grandson of deified Pius, great grandson of
deified Hadrian, great great grandson of deified
Trajan Victor in Parthia, great great great
grandson of deified Nerva, Lucius Aelius Aurelius
Commodus Augustus
21 div : commentary
22 head : Commentary
23 No comment
24 div : figure
25 head : Photographs
26 head : Ward-Perkins Archive, BSR (BSR 48.XXVII.13)
```

<https://xsltdev.com/xslt/xsl-apply-templates/>

XSLT main elements

xsl:for-each

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

xsl:for-each

Causes iteration over the nodes selected by a node-set expression.

Available in XSLT 1.0 and later versions. Available in all Saxon editions.

- **Category:** instruction
- **Content:** (`xsl:sort` * , sequence-constructor)
- **Permitted parent elements:** any XSLT element whose content model is sequence-constructor; any literal result element

Attributes

`select`

`expression`

Defines the nodes over which the statement will iterate. The XSLT statements subordinate to the `xsl:for-each` element are applied to each source node selected by the node-set expression in turn. The full syntax of expressions is outlined in XPath Expression Syntax.

<https://xsltdev.com/xslt/xsl-for-each/>



XSLT main elements

xsl:for-each xsl:sort

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <xsl:stylesheet version="2.0">
3  ...  xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
4  ...  <xsl:output method="text" encoding="UTF-8" />
5
6  ...  <xsl:template match="/">
7  ...    <xsl:for-each select="distinct-values(//w/@lemma)">
8  ...      <xsl:sort select=". " data-type="text" lang="la"/>
9  ...      <xsl:text>&#32;</xsl:text>
10     <xsl:value-of select=". " />
11     <xsl:text>&#32;</xsl:text>
12   </xsl:for-each>
13 </xsl:template>
14
15 </xsl:stylesheet>

out-text.txt U X
1 diuus filius imperator
```

<https://xsltdev.com/xslt/xsl-for-each/>

XSLT main elements

xsl:if xsl:key

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

xsl:if

Used for conditional processing. It takes a mandatory test attribute, whose value is a boolean expression. The contents of the `xsl:if` element are expanded only if the expression is true.

Available in XSLT 1.0 and later versions. Available in all Saxon editions.

- **Category:** instruction
- **Content:** sequence-constructor
- **Permitted parent elements:** any XSLT element whose content model is sequence-constructor; any literal result element

Attributes

`test`

expression

The boolean expression to be tested. The full syntax of boolean expressions is outlined in XPath Expression Syntax.

<https://xsltdev.com/xslt/xsl-if/>



XSLT main elements

xsl:key xsl:if

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
1   <?xml version="1.0" encoding="UTF-8"?>
2   <xsl:stylesheet version="2.0">
3   |   ... xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
4   |   ... <xsl:output method="text" encoding="UTF-8" />
5
6   |   ... <xsl:key name="lang" match="language" use="@ident"></xsl:key>
7
8   |   ... <xsl:template match="/">
9   |   |   ... <xsl:if test="key('lang','la')">
10  |   |   |   ... <xsl:value-of select="key('lang','la')"/>
11  |   |   ... </xsl:if>
12  |   ... </xsl:template>
13
14  </xsl:stylesheet>
15
```

out-text.txt U ×

1 Latin

<https://xsltdev.com/xslt/xsl-key/>

XSLT main elements

xsl:variable

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

xsl:variable

Used to declare a variable and give it a value. If it appears at the top level (immediately within `xsl:stylesheet`) it declares a global variable, otherwise it declares a local variable that is visible only within the stylesheet element containing the `xsl:variable` declaration. The value of a variable can be referenced within an expression using the syntax `$name`.

Available in XSLT 1.0 and later versions. Available in all Saxon editions.

- **Category:** declaration
- **Category:** instruction
- **Content:** sequence-constructor
- **Permitted parent elements:** `xsl:package` ; `xsl:stylesheet` ; `xsl:transform` ;
`xsl:override` ; `xsl:function` ; any XSLT element whose content model is sequence-constructor; any literal result element

Attributes

`name`

`eqname`

Defines the name of the variable.

XSLT main elements

xsl:variable

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <xsl:stylesheet version="2.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
3   <xsl:output method="text" encoding="UTF-8" />
4   <xsl:key name="lang" match="language" use="@ident"></xsl:key>
5
6   <xsl:template match="/">
7     <xsl:if test="key('lang', 'la')">
8       <xsl:variable name="latLang" select="key('lang', 'la')"/>
9       <xsl:value-of select="">
10      <xsl:with-param value="concat('Selected Language: ', $latLang/text(), ' (' , $latLang/@ident, ')')"/>
11    </xsl:if>
12  </xsl:template>
13
14 </xsl:stylesheet>
```

out-text.txt U ×

1 Selected Language: Latin (la)

<https://xsltdev.com/xslt/xsl-variable/>

Progress status

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

1 Lesson path

2 Introduction

3 XPath

4 XSL Transformations

5 XSL in action

6 References

eXtensible Style Sheet

Working with XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>
        <rs type="textType">Christian</rs>
        <rs type="textType">funerary</rs> inscription for Helladios</title>
        <editor>J. M. Reynolds</editor>
        <editor>J. B. Ward-Perkins</editor>
      </titleStmt>
      <publicationStmt>
        <authority>Centre for Computing in the Humanities, King's College London</authority>
        <idno type="filename">IRT256a</idno>
        <availability>
          <p>Creative Commons licence Attribution UK 2.0 (<ref>http://creativecommons.org/licenses/by/uk/2.0/</ref>)
          <p>All reuse or distribution of this work must contain somewhere a link back to the original source</p>
        </availability>
      </publicationStmt>
    <sourceDesc> ...
```

The Inscriptions of Roman Tripolitania -
<https://inslib.kcl.ac.uk/irt2009/>

eXtensible Style Sheet

Working with XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!-- $Id$ -->
3 <xsl:stylesheet>
4   ... xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
5   ... xmlns:tei="http://www.tei-c.org/ns/1.0"
6   ... xmlns:xs="http://www.w3.org/2001/XMLSchema"
7   ... exclude-result-prefixes="t" version="2.0">
8 ...
9 ...
10 ...
11 ...
12 ...
13 ...
14 ...
15 ...
16 ...
17 ...
18 ...
19 ...
20 ...
21 ...
22 ...
23 ...
24 ...
25 ...
26 ...
27 ...
28 ...
29 ...
30 ...
31 ...
32 ...
33 ...
34 ...
35 ...
36 ...
37 ...
38 ...
39 ...
40 ...
41 ...
42 ...
43 ...
44 ...
45 ...
46 ...
47 ...
48 ...
49 ...
50 ...
51 ...
52 ...
53 ...
54 ...
55 ...
56 ...
57 ...
58 ...
59 ...
60 ...
61 ...
62 ...
63 ...
64 ...
65 ...
66 ...
67 ...
68 ...
69 ...
70 ...
71 ...
72 ...
73 ...
74 ...
75 ...
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```
angelodel80@LAPTOP-V8V3MLG0:/mnt/c/Users/angel/risorse/universita/corsoCodifica/tools$ java -jar SaxonHE10-
3J/saxon-he-10.3.jar -s:source/IRT256a.xml -xsl:source/Stylesheets-9.4/start-edition.xsl -o:source/out/out-
-text-epidoc.html
```

```
1 <!DOCTYPE HTML><html xmlns:i18n="http://apache.org/cocoon/i18n/2.1">
2 ...
3 ...
4 ...
5 ...
6 ...
7 ...
8 ...
9 ...
10 ...
11 ...
12 ...
```

The Inscriptions of Roman Tripolitania -
<https://inslib.kcl.ac.uk/irt2009/>

eXtensible Style Sheet

Working with XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Bibliography

Not previously published.

Text constituted from

Transcription (Reynolds, Ward-Perkins)

Text

[Εύμοιριτο] Ἐλλάδιος ἔζησε ἔτι ν'
[πλέον ἔλατ(τ)ον μῆν[ες γ'] ἡμέρας
[δεκαπέντε ὁ Χριστὸς μετὰ τ(ο)ῦ πνεύ-
[ματός σ(ο)υ ἐτ]ελεύ[τισεν μ]ηνὶ Παυνὶ⁵
[έβδο]μι κατὰ τοὺς Ἀφρ<ο>υς

Apparatus

Apparatus

The supplied letters are no longer legible.

Translation

[May he be well off]. Helladios lived approximately 50 years, three months and fifteen days. Christ be with your spirit. He died on the seventh of the month of Payn among the Africans.

eXtensible Style Sheet

Working with XSLT

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
<parameter>
  <name>edition-type</name>
  <value>interpretive</value>
  <value on="yes">diplomatic</value>
</parameter>


<parameter>
  <name>line-inc</name>
  <value>1</value>
</parameter>
```

The Inscriptions of Roman Tripolitania -
<https://inslib.kcl.ac.uk/irt2009/>

eXtensible Style Sheet

Working with XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Text

- 1 [.....]ΕΛΛΑΔΙΟΣΕΖΗΣΕΕΤΙΝ
- 2 [.....]Ν[...]ΗΜΕΡΑΣ
- 3 [.....]ΠΙΝΕΥ
- 4 [.....]ΕΛΕΥ[.....]ΝΙΑΥΝΙ
- 5 [...]ΜΙΚΑΤΑΤΟΥΣΑΦΡΥΣ

Apparatus

Apparatus

The supplied letters are no longer legible.

Translation

[.....]. Helladios lived approximately 50 years, three months and fifteen days. Christ be with your spirit.
He died on the seventh of the month of Payn among the Africans.

The Inscriptions of Roman Tripolitania -
<https://inslib.kcl.ac.uk/irt2009/>

eXtensible Style Sheet

Working with XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
<parameter>
  <name>edn-structure</name>
  <value>default</value>
  <value>ddbdp</value>
  <value>hgv</value>
  <value>inslib</value>
  <value>iосре</value>
  <value>edh</value>
  <value>edh-db</value>
  <value>rib</value>
  <value>sammelbuch</value>
  <value on="yes">sample</value>
  <value>eagle</value>
  <value>igcyr</value>
</parameter>
```

eXtensible Style Sheet

Working with XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

Current repository: Unknown

Text type: Christianfunerary

Editor(s): J. M. Reynolds, J. B. Ward-Perkins

Changes history: 2010-08-18 GB Converted from TEI P4 (EpiDoc DTD v. 6) to P5 (EpiDoc RNG schema v. 8); 2008-09-09 ZA converted 2009-05-19 RV Added Figures; 2009-08-24 RV Added Figures

Publication details: Centre for Computing in the Humanities, King's College London;

Creative Commons licence Attribution UK 2.0 ()

All reuse or distribution of this work must contain somewhere a link back to the URL

Interpretive

Text

- 1 [Εύμοιριτο] Έλλαδιος ἔζησε ἐπι ν'
- 2 [πλέον ἔλατ(τ)ον μῆν[ες γ'] ἡμέρας
- 3 [δεκαπέντε ὁ Χριστός μετά τ(ο)ῦ πνεύ-
- 4 [ματός σ(ο)υ ἐτ]ελεύτισεν μι]νι Παυνί
- 5 [έβδο]μι κατὰ τοὺς Αφρο>υς

Diplomatic

Text

- 1 [.....]ΕΛΛΑΔΙΟΣΕΖΗΣΕΕΤΙΝ
- 2 [.....]Ν[...]ΗΜΕΡΑΣ
- 3 [.....]ΠΝΕΥ
- 4 [.....]ΕΛΕΥ[.....]ΝΙ.ΑΥΝΙ
- 5 [...]ΜΙΚΑΤΑΤΟΥΣΑΦΡΥΣ

eXtensible Style Sheet

Working with XSL-T

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Trans-
formations

XSL in action

References

```
<p><b>Text type:</b>
<xsl:choose>
  <xsl:when test="//t:textClass//t:keywords//t:term[@type='textType']">
    <xsl:apply-templates select="//t:textClass//t:keywords//t:term[@type='textType']"/>
  </xsl:when>

  <!--xsl:when test="//t:teiHeader//t:rs[@type='textType']"-->
  <xsl:apply-templates select="//t:teiHeader//t:rs[@type='textType']"/>
</xsl:when-->
  <xsl:when test="//t:teiHeader//t:rs[@type='textType']">
    <xsl:for-each select="//t:teiHeader//t:rs[@type='textType']">
      <xsl:apply-templates select="current()"/><xsl:text></xsl:text>
    </xsl:for-each>
  </xsl:when>
</xsl:choose>
```

The Inscriptions of Roman Tripolitania -
<https://inslib.kcl.ac.uk/irt2009/>

eXtensible Style Sheet

Working with XSLT

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

The screenshot shows a software interface for XSLT transformation. At the top, there are three tabs: "new-document 1", "transform.xq", and "htm-tpl-struct-sample.xsl". The "transform.xq" tab is currently active. Below the tabs is a code editor containing XQuery code:

```
1 xquery version "3.1";
2
3 let $xml := doc("IRT256a.xml")
4 let $xsl := doc("Stylesheets-9.4/start-edition.xsl")
5
6 return transform:transform($xml,$xsl,())
7
```

Below the code editor is a toolbar with the following buttons: a back arrow, a forward arrow, "XHTML Output" dropdown, "Indent" checkbox (which is checked), "Live Preview" checkbox (which is checked), "Highlight Index Matches" checkbox (which is checked), and a plus sign button.

The main pane displays the transformed XML output:

```
<html xmlns:i18n="http://apache.org/cocoon/i18n/2.1">
<head>
    <title>IRT256a. Christian funerary inscription for Helladios</title>
    <meta http-equiv="content-type" content="text/html; charset=UTF-8" />
    <link rel="stylesheet" type="text/css" media="screen, projection" href="global.css" />
</head>
<body>
    <h1>IRT256a. Christian funerary inscription for Helladios</h1>
    <div id="mainContent">
```

The Inscriptions of Roman Tripolitania -
<https://inslib.kcl.ac.uk/irt2009/>

Progress status

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

1 Lesson path

2 Introduction

3 XPath

4 XSL Transformations

5 XSL in action

6 References

Bibliography

deepen into XPath and XSLT

XPath - XSL

A.M. Del
Grosso

Lesson path

Introduction

XPath

XSL Transformations

XSL in action

References

Some References

- **XQuery and XPath Data Model 3.1**
<https://www.w3.org/TR/xpath-datamodel-31/>
- **XSLT Recommendations**
<https://www.w3.org/TR/xslt/>
- **XPath Recommendations**
<https://www.w3.org/TR/xpath/>
- Kay, M. (2011). XSLT 2.0 and XPath 2.0 Programmer's Reference. Wiley.
- Williams, I. (2009). Beginning XSLT and XPath: Transforming XML Documents and Data. Wiley.
- Walmsley, P. (2015) XQuery: Search Across a Variety of XML Data. O'Reilly.
- Saxonica documentation:
<https://www.saxonica.com/documentation11/documentation.xml>

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Codifica TEI - Visualizzazione ed Elaborazione: Fogli di Stile

Angelo Mario Del Grosso
(Slide tratte dal materiale di C. Di Pietro)

`angelo.delgrosso@ilc.cnr.it`
CNR-ILC
`http://www.ilc.cnr.it/`

Istituto di Linguistica Computazionale “A. Zampolli”,
8th May 2024

Sommario della Lezione

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

1 Introduzione

2 XSL Transformations: Caratteristiche Fondamentali

3 Template (rules/named) e principali istruzioni XSLT

4 XPath: selezione dei nodi ed expression axes

5 Conclusioni

Progress status

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

1 Introduzione

2 XSL Transformations: Caratteristiche Fondamentali

3 Template (rules/named) e principali istruzioni XSLT

4 XPath: selezione dei nodi ed expression axes

5 Conclusioni

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

I fogli di stile (style sheet)

- Describe il modo in cui un documento elettronico deve essere visualizzato
- Il mezzo di visualizzazione può variare: lo schermo di un computer, la stampa, i sintetizzatori vocali, ecc.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Scopo dei fogli di stile

- Separazione forma-contenuto: *la visualizzazione del documento è un processo indipendente (e successivo)*
- Gestione della resa grafica per molti documenti contemporaneamente: *massima uniformità dello stile*
- Gestione di mezzi diversi dal monitor: smartphone, sintetizzatore vocale, stampante braille, ecc.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Metodi e Tecnologie

**Quelli più noti e utilizzati sono standard internazionali
definiti dal consorzio W3**

(<http://www.w3.org/Style/CSS>).

- CSS: Cascading Style Sheets
- XSL: eXtensible Stylesheet Language

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

CSS: Cascading Style Sheets

- Nati per HTML, possono essere utilizzati anche con XML
- Mostrano cosa c'è nel file, nell'ordine in cui questo compare
- Difficile manipolare struttura e contenuto

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XSL: eXtensible Stylesheet Language

- Trasforma XML in altra forma: HTML, PDF, ODT, EPUB
- Manipola ed Elabora la struttura e il contenuto del documento
- Molto espressivo e completo, ma complesso

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

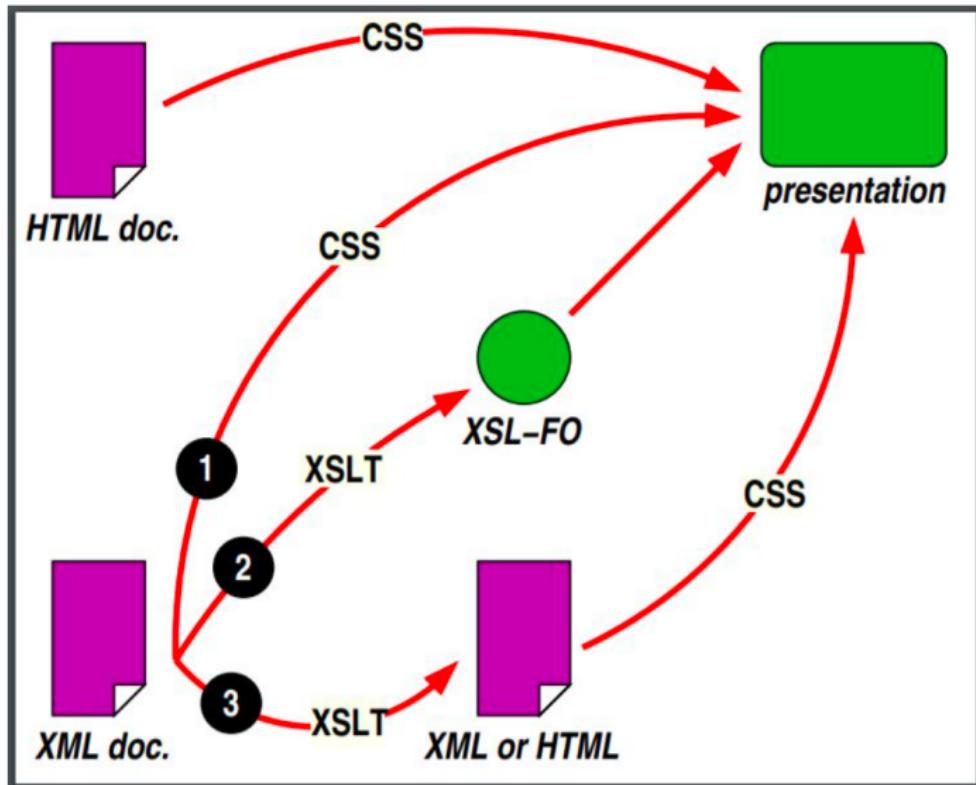
Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni



Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

CSS e XSL: caratteristiche a confronto

	CSS	XSL
Supporto HTML	✓	-
Supporto XML	✓	✓
Manipolazione testo	<i>limitato</i>	✓
Trasformazioni	-	✓
Supporto mezzi diversi	✓	✓
Tipo sintassi	CSS	XML

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Fogli di Stile CSS

- Offrono limitatissimi mezzi per modificare il documento al quale vengono applicati (in particolare aggiungere testo)
- Sono basati su una sintassi specifica piuttosto semplice

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

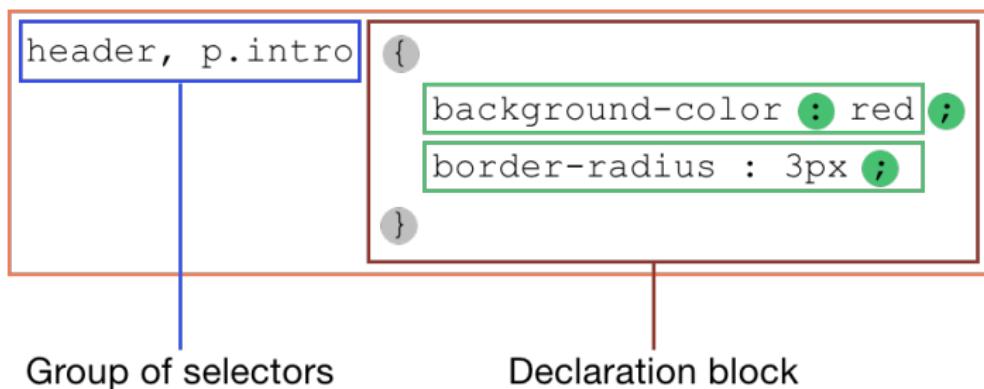
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Sintassi CSS



Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
table.header {  
    background-image: url('graphics/top_grad2.png');  
    width: 100%;  
    border-width:0px;  
    font-weight: normal;  
    color: #000000;  
    padding:0px;  
    white-space:nowrap;  
}  
  
table.header a{  
    display: block;  
}  
  
td.links{  
    padding-top:20px;  
    padding-right:20px;  
    border-width:0px;  
    width: 180px;  
}
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Fogli di Stile CSS: invocazione HTML

```
<link rel="stylesheet"  
      type="text/css" href="default.css" >
```

Fogli di Stile CSS: invocazione XML

```
<?xml-stylesheet  
    type="text/css" href="style.css"?>
```

Progress status

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

1 Introduzione

2 XSL Transformations: Caratteristiche Fondamentali

3 Template (rules/named) e principali istruzioni XSLT

4 XPath: selezione dei nodi ed expression axes

5 Conclusioni

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

eXtensible Stylesheet Language (XSL)

- Specifica del W3C che descrive un metodo per la visualizzazione e manipolazione dei documenti XML
- Maggiore controllo sulla presentazione dei dati XML
- Generazione di layout complessi e compositi

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XSL incorpora tre linguaggi

- **XSL Transformations (XSL-T)**: *trasformazione di un documento XML in un altro tipo di documento (es.: HTML)*
- **XSL Formatting Objects (XSL-FO)**: *applicazione degli stili e della resa grafica di un documento XML*
- **XML Path (XPath)**: *usato nei fogli di stile XSLT per selezionare le parti di un documento XML*

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

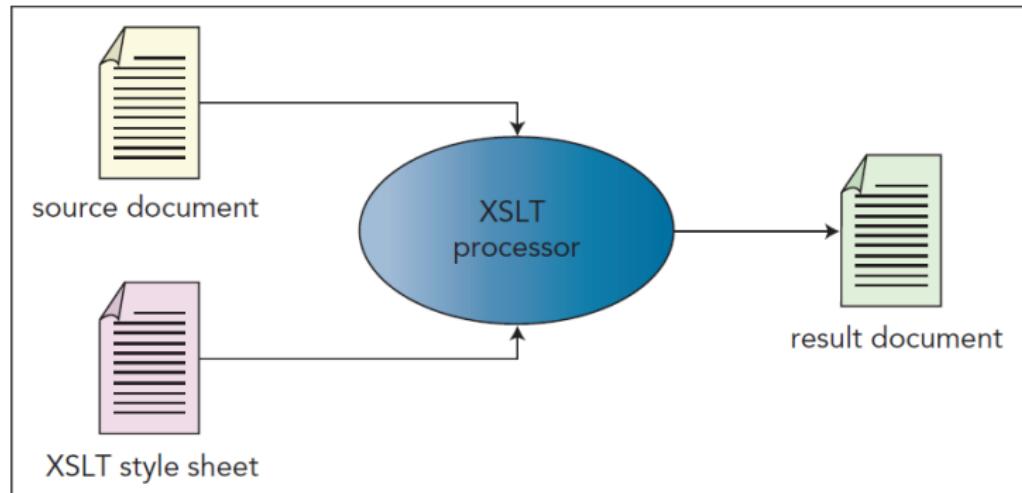
Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni



Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XSL Transformations

- XSLT è un vero e proprio linguaggio di programmazione che usa la sintassi XML
- Usa namespace differenti per distinguere fra istruzioni proprie (precedute da **xsl:**) e output
- Legge e scrive alberi XML (ma è possibile ottenere come output anche del codice HTML o del testo semplice)
- Versione attuale: XSLT 3.0
(<https://www.w3.org/TR/xslt-30/>)

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XSL Capacità di trasformazione

- generazione di testo costante;
- soppressione del contenuto;
- spostamento del testo (es.: scambio ordine di nome e cognome);

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XSL Capacità di trasformazione

- duplicazione del testo (ad es.: tabella di contenuti copiando i titoli);
- ordinamento dei contenuti (ad es.: termini in ordine alfabetico);
- elaborazione di nuove informazioni in base a quelle esistenti (es. statistiche)

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Caratteristiche fondamentali di XSLT

- Basato su regole di trasformazione (*modello pattern-matching*)
- Le regole sono dichiarative (*specificano che cosa deve essere generato quando si incontra un certo modello nel documento*)
- Le regole possono essere disposte in qualsiasi ordine

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Modalità di Trasformazioni XSLT

- **Lato server**, utilizzando script Java, ASP, PHP ecc. , per produrre "al volo" pagine HTML sulla base di documenti XML (es. Cocoon);
- **Lato client**, sui Browser che supportano questa tecnologia;
- tramite un **programma separato** (come ad esempio *oXyGen* oppure *xsltproc*), oppure *saxon*, che permette di applicare uno o più scenari di trasformazione.

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Componenti di Base di un foglio XSLT

- Intestazione XML
- Elemento radice *stylesheet* e namespace
- Eventuali istruzioni di elaborazione
- Serie di template rules

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Intestazione XML

```
<?xml version="1.0" encoding="UTF-8" ?>
```

Elemento radice

```
<xsl:stylesheet version='2.0'  
 xmlns:xsl='http://www.w3.org/1999/XSL/Transform'>
```

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Eventuali istruzioni di elaborazione

```
<xsl:output method="xml" version="1.0"  
indent="yes"/>
```

Serie di template rules

```
<xsl:template match="/" > ...</xsl:template>  
<xsl:template match="title" > ... </xsl:template>
```

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

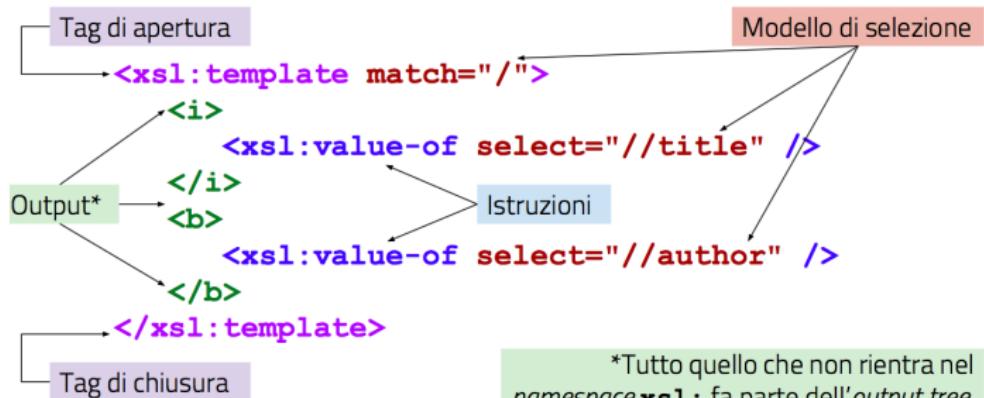
Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni



*Tutto quello che non rientra nel
namespace **xsl:** fa parte dell'*output tree*.

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Come vengono applicate le regole XSLT

Il processore XSLT

- Legge il documento XML in input e crea l'albero corrispondente
- Inizia a percorrere l'albero leggendo i singoli nodi
- Confronta ogni nodo con le regole presenti nel foglio di stile
- Produce l'output secondo le istruzioni della regola
- Restituisce un albero di output

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

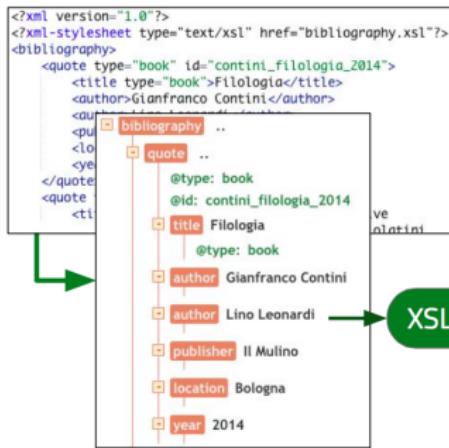
Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

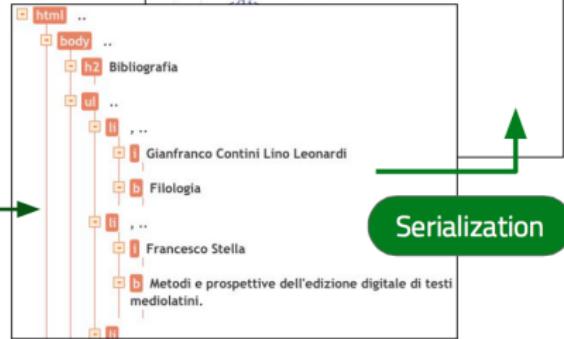
Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni



XSLT



Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio di trasformazione

Costruire un foglio di stile XSLT con alcune regole di
trasformazione e lanciare il comando **java saxon** da terminale.

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Tipi di nodo nell'albero XML

- **Radice del Documento**
- **Elementi** con contenuto del sotto albero
- **Attributi**
- **Testo** compresi gli spazi vuoti
- **Commenti** (contenuto tra <!-- -->)
- **Namespace** con riferimenti e URI
- **Istruzioni di elaborazione** (contenuto tra <? ?>)

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Tipi di nodo nell'albero XML

- Il **documento** stesso costituisce la radice (*l'elemento radice XML non è la radice dell'albero di rappresentazione!*)
- L'intero albero è suddivisibile in sotto-alberi
- I nodi più importanti sono gli elementi e i loro attributi
- Le entità vengono "tradotte" nel testo loro assegnato al momento della dichiarazione
- Lo spazio "vuoto" può essere considerato o no

Fondamenti Extensible Stylesheet Language

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

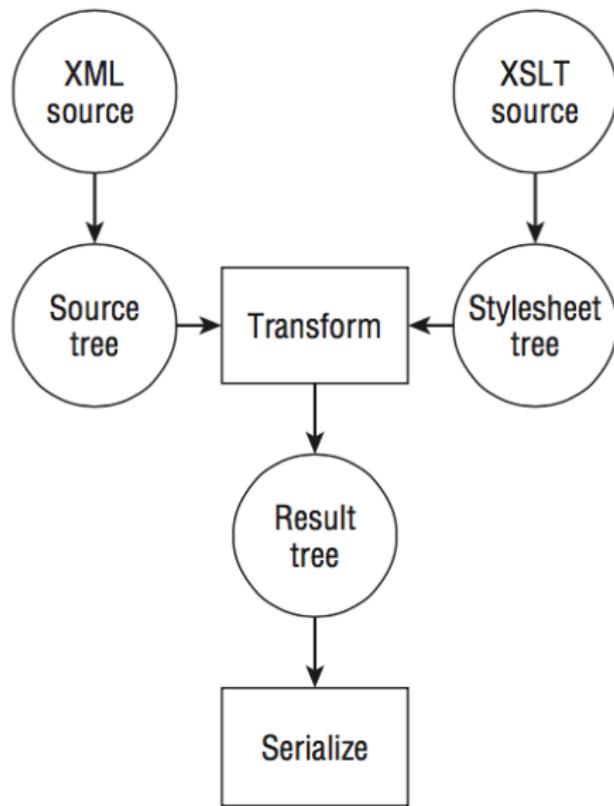
Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni



Progress status

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

1 Introduzione

2 XSL Transformations: Caratteristiche Fondamentali

3 Template (rules/named) e principali istruzioni XSLT

4 XPath: selezione dei nodi ed expression axes

5 Conclusioni

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:template>

Definisce una regola (ovvero un modello) di trasformazione per i nodi di un particolare tipo/contesto.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="template" substitutionGroup="xsl:declaration">  
  <xs:complexType>  
    <xs:complexContent mixed="true">  
      <xs:extension base="xsl:versioned-element-type">  
        <xs:sequence>  
          <xs:element ref="xsl:param" minOccurs="0" maxOccurs="unbounded"/>  
          <xs:group ref="xsl:sequence-constructor-group" minOccurs="0"  
            maxOccurs="unbounded"/>  
        </xs:sequence>  
        <xs:attribute name="match" type="xsl:pattern"/>  
        <xs:attribute name="priority" type="xs:decimal"/>  
        <xs:attribute name="mode" type="xsl:modes"/>  
        <xs:attribute name="name" type="xsl:QName"/>  
        <xs:attribute name="as" type="xsl:sequence-type" default="item()*/>  
      </xs:extension>  
    </xs:complexContent>  
  </xs:complexType>  
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:template>

- **name:** nome del template;
- **match:** pattern che indica l'elemento su cui applicare il modello;
- **priority:** priorità del modello;
- **mode:** modalità di elaborazione, che consente all'elemento di essere elaborato più volte per produrre un risultato diverso ogni volta.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:template>

I template XSLT possono avere due forme:

- **"template rules"** che specificano una regola con pattern-matching (<xsl:apply-templates>)
- **named templates** che specificano regole che possono essere chiamate esplicitamente con <xsl:call-template>

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:value-of>

Restituisce il contenuto del nodo selezionato secondo l'espressione XPath indicata.

(Il contenuto di un elemento è costituito da tutti i caratteri che si trovano fra tag di apertura e tag di chiusura)

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="value-of" substitutionGroup="xsl:instruction">  
  <xs:complexType>  
    <xs:complexContent mixed="true">  
      <xs:extension base="xsl:sequence-constructor">  
        <xs:attribute name="select" type="xsl:expression"/>  
        <xs:attribute name="separator" type="xsl:avt"/>  
        <xs:attribute name="disable-output-escaping" type="xsl:yes-or-no" default=  
          "no"/>  
      </xs:extension>  
    </xs:complexContent>  
  </xs:complexType>  
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:value-of>

- **select:** espressione XPath da valutare nel contesto corrente
- **disable-output-escaping:** default "no"; se "yes", il testo di output non esclude i caratteri XML dal testo

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:value-of>

```
<xsl:template match="fileDesc" >  
  <h1>File Desc</h1>  
  <p>  
    <xsl:value-of select="titleStmt/title"  
      disable-output-escaping="no" />  
  </p>  
</xsl:template>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:apply-templates>

Elaborare in modo ricorsivo i nodi di un documento XML a partire da un punto preciso dell'albero XML.

Elemento <xsl:apply-templates>

Confronta ogni nodo presente nella lista di nodi (*node-set*) selezionati con le template rules del foglio di stile e se viene trovata una regola applicabile (*match*) questa viene applicata.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="apply-templates" substitutionGroup="xsl:instruction">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="xsl:element-only-versioned-element-type">
        <xs:choice minOccurs="0" maxOccurs="unbounded">
          <xs:element ref="xsl:sort"/>
          <xs:element ref="xsl:with-param"/>
        </xs:choice>
        <xs:attribute name="select" type="xsl:expression" default="child::node()"/>
        <xs:attribute name="mode" type="xsl:mode"/>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:apply-templates>

- **select:** espressione XPath
- **mode:** modalità di elaborazione

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile
A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali
Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:apply-templates>

```
<xsl:template match="/">  
  <html><head>  
    <title><xsl:value-of  
      select="TEI/teiHeader/fileDesc/title"/></title>  
  </head><body><div><span>1.</span>  
    <xsl:apply-templates  
      select="TEI/teiHeader/fileDesc" />  
  </div></body></html></xsl:template>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:for-each>

Se sono presenti più nodi con lo stesso nome (e manca un'istruzione ricorsiva precedente) <xsl:value-of> restituisce il valore del primo che incontra.

Elemento <xsl:for-each>

è quindi possibile usare l'istruzione <xsl:for-each> e applicare un'istruzione <xsl:value-of> a tutti i nodi identificati dalla regola.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="for-each" substitutionGroup="xsl:instruction">
    <xs:complexType>
        <xs:complexContent mixed="true">
            <xs:extension base="xsl:versioned-element-type">
                <xs:sequence>
                    <xs:element ref="xsl:sort" minOccurs="0" maxOccurs="unbounded"/>
                    <xs:group ref="xsl:sequence-constructor-group"
minOccurs="0" maxOccurs="unbounded"/>
                </xs:sequence>
                <xs:attribute name="select" type="xsl:expression" use="required"/>
            </xs:extension>
        </xs:complexContent>
    </xs:complexType>
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:for-each>

- **select:** espressione XPath

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:for-each>

```
<xsl:template match="/">  
  <html><head>  
    <title><xsl:value-of  
      select="TEI/teiHeader/fileDesc/title"/></title>  
  </head><body><div>  
    <xsl:for-each select="//div">  
      <div><xsl:value-of select=".//p" /></div>  
    </xsl:for-each></div></body></html>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:text>

Permette di inserire una stringa di testo nell'albero di output.

Elemento <xsl:text>

Molto utile se si è deciso di eliminare tutti gli spazi e gli a capo.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="text" substitutionGroup="xsl:instruction">
    <xs:complexType>
        <xs:simpleContent>
            <xs:extension base="xsl:text-element-base-type">
                <xs:attribute name="disable-output-escaping" type="xsl:yes-or-no"
default="no"/>
            </xs:extension>
        </xs:simpleContent>
    </xs:complexType>
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:text>

- **disable-output-escaping**: se "yes", consente di copiare i caratteri di marcatura non identificati nell'albero di output.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:text>

```
<xsl:for-each select="$attr" >  
  <xsl:value-of  
    select="concat(' [',position(),'] ',current())" />  
  <xsl:text>#32;</xsl:text>  
</xsl:for-each>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:if>

Identifica una condizione semplice: la regola viene eseguita soltanto se la condizione viene soddisfatta.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="if" substitutionGroup="xsl:instruction">  
  <xs:complexType>  
    <xs:complexContent mixed="true">  
      <xs:extension base="xsl:sequence-constructor">  
        <xs:attribute name="test" type="xsl:expression" use="required"/>  
      </xs:extension>  
    </xs:complexContent>  
  </xs:complexType>  
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:if>

- **test:** l'espressione di test. Se restituisce true, il contenuto di <xsl:if> viene valutato e inserito nell'albero di output; altrimenti viene ignorato

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:if>

```
<xsl:if test="@n='23'" >...</xsl:if>  
<xsl:if test="title[@level='m']" >...</xsl:if>  
<xsl:if test="count(verse) > 3" >...</xsl:if>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento `<xsl:choose>`

Permette di definire condizioni multiple tra cui scegliere.

Elemento `<xsl:choose>`

Il content model di *choose* prevede uno o più elementi `<xsl:when>` e opzionalmente l'elemento `<xsl:otherwise>`.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="choose" substitutionGroup="xsl:instruction">  
  <xs:complexType>  
    <xs:complexContent>  
      <xs:extension base="xsl:element-only-versioned-element-type">  
        <xs:sequence>  
          <xs:element ref="xsl:when" maxOccurs="unbounded" />  
          <xs:element ref="xsl:otherwise" minOccurs="0" />  
        </xs:sequence>  
      </xs:extension>  
    </xs:complexContent>  
  </xs:complexType>  
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:choose>

```
<xsl:template match="tei:title" >  
<xsl:choose>  
<xsl:when test="@level = 'm' or @level = 'u'" >  
<i><xsl:apply-templates/>. </i> </xsl:when>  
<xsl:when test="@level = 'j'" >  
<i><xsl:apply-templates/></i>  
</xsl:when>  
<xsl:otherwise>  
<i><xsl:apply-templates/></i></xsl:otherwise>  
</xsl:choose></xsl:template>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:sort>

Permette di riorganizzare l'ordine in cui vengono scritti i nodi nell'albero di output.

Elemento <xsl:sort>

Deve comparire all'interno di un'istruzione
<xsl:apply-templates> o <xsl:for-each>.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="sort">
  <xs:complexType>
    <xs:complexContent mixed="true">
      <xs:extension base="xsl:sequence-constructor">
        <xs:attribute name="select" type="xsl:expression"/>
        <xs:attribute name="lang" type="xsl:avt"/>
        <xs:attribute name="data-type" type="xsl:avt" default="text"/>
        <xs:attribute name="order" type="xsl:avt" default="ascending"/>
        <xs:attribute name="case-order" type="xsl:avt"/>
        <xs:attribute name="collation" type="xsl:avt"/>
        <xs:attribute name="stable" type="xsl:yes-or-no"/>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:sort>

- **select:** espressione XPath per individuare gli elementi in base ai quali effettuare l'ordinamento.
- **lang:** linguaggio utilizzato per l'ordinamento.
- **data-type:** il tipo degli elementi rispetto ai quali stiamo effettuando l'ordinamento.
- **order:** ordinamento crescente (ascending) o discendente (descending)
- **case-order:** indica se dare precedenza ai caratteri minuscoli (lower-first) o maiuscoli (upper-first)

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:sort>

```
<div><ul><xsl:for-each select="TEI/text/body/div">  
<xsl:sort select="@n" data-type="number"  
order="descending" />  
<li><xsl:value-of select="@n" />  
<xsl:text>|</xsl:text>  
<xsl:value-of select="current()" /></li>  
</xsl:for-each></ul></div>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:variable>

Permette di definire una variabile, ovvero una posizione di memorizzazione denominata in un modo personalizzato, che contiene i risultati di una espressione valutata a runtime.

Elemento <xsl:variable>

L'accesso ad una variabile avviene anteponendo il carattere \$ al nome della variabile (es.: \$unaVariabile).

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="variable">
  <xs:complexType>
    <xs:complexContent mixed="true">
      <xs:extension base="xsl:sequence-constructor">
        <xs:attribute name="name" type="xsl:QName" use="required"/>
        <xs:attribute name="select" type="xsl:expression" use="optional"/>
        <xs:attribute name="as" type="xsl:sequence-type" use="optional"/>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Transformations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:variable>

- **name:** nome della variabile.
- **select:** seleziona il contenuto della variabile, se presente;
altrimenti come contenuto viene usato il contenuto
dell'istruzione

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

La definizione di variabili può assumere tre distinte forme:

Esempio Elemento <xsl:variable>

- creazione di una variabile il cui valore è una stringa vuota
- creazione di una variabile avente valore definito dall'attributo select
- creazione mediante inclusione di contenuto nel corpo dell'elemento

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

La definizione di variabili può assumere tre distinte forme:

Esempio Elemento <xsl:variable>

- `<xsl:variable name="myVar" />`
- `<xsl:variable name="myVar" select="150" />`
- `<xsl:variable name="myVar" >`
`<xs:value-of select="@n"/>`
`</xsl:variable>`

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:param>

E' simile ad una <xsl:variable>, ma il suo valore può essere modificato in base al modo in cui il template viene chiamato o dal foglio di stile stesso.

Elemento <xsl:param>

Può essere inserita come primo figlio di un <xsl:template>

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="param">
  <xs:complexType>
    <xs:complexContent mixed="true">
      <xs:extension base="xsl:sequence-constructor">
        <xs:attribute name="name" type="xsl:QName" use="required"/>
        <xs:attribute name="select" type="xsl:expression"/>
        <xs:attribute name="as" type="xsl:sequence-type"/>
        <xs:attribute name="required" type="xsl:yes-or-no"/>
        <xs:attribute name="tunnel" type="xsl:yes-or-no"/>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:param>

- **name:** nome della variabile.
- **select:** seleziona il contenuto della variabile, se presente;
altrimenti come contenuto viene usato il contenuto
dell'istruzione stessa

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:param>

```
<xsl:template name="body" >  
  <xsl:param name="style" >color:blue</xsl:param>  
  <div><xsl:attribute name="style" >  
    <xsl:value-of select="$style" />  
  </xsl:attribute>  
  <xsl:value-of select=". " /></div>  
</xsl:template>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento `<xsl:call-template>`

Dopo aver assegnato un nome ad un template, è possibile richiamarlo con l'istruzione `<xsl:call-template>`.

Elemento `<xsl:call-template>`

Per invocare un template passando dei parametri è possibile utilizzare l'elemento `<xsl:with-param>` nel corpo dell'elemento `<xsl:call-template>` o `<xsl:apply-templates>` indicando il nome del parametro ed il valore.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="call-template" substitutionGroup="xsl:instruction">  
  <xs:complexType>  
    <xs:complexContent>  
      <xs:extension base="xsl:element-only-versioned-element-type">  
        <xs:sequence>  
          <xs:element ref="xsl:with-param" minOccurs="0" maxOccurs="unbounded" />  
        </xs:sequence>  
        <xs:attribute name="name" type="xsl:QName" use="required"/>  
      </xs:extension>  
    </xs:complexContent>  
  </xs:complexType>  
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Attributi Elemento <xsl:call-template>

- **name:** il nome del template da richiamare. Il foglio di stile deve necessariamente contenere un <xsl:template> con tale nome specificato

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:call-template>

```
<body>
  <xsl:call-template name="body" >
    <xsl:with-param name="style" >
      color:red </xsl:with-param>
    </xsl:call-template></body>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento `<xsl:element>`

Per creare elementi è possibile utilizzare l'istruzione
`<xsl:element>`.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="element" substitutionGroup="xsl:instruction">  
  <xs:complexType mixed="true">  
    <xs:complexContent>  
      <xs:extension base="xsl:sequence-constructor">  
        <xs:attribute name="name" type="xsl:avt" use="required"/>  
        <xs:attribute name="namespace" type="xsl:avt"/>  
        <xs:attribute name="inherit-namespaces" type="xsl:yes-or-no" default="yes"/>  
        <xs:attribute name="use-attribute-sets" type="xsl:QNames" default="" />  
        <xs:attribute name="type" type="xsl:QName"/>  
        <xs:attribute name="validation" type="xsl:validation-type"/>  
      </xs:extension>  
    </xs:complexContent>  
  </xs:complexType>  
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xsl:template match="/">
    <html>
        <body>
            Corpo della pagina
        </body>
    </html>
</xsl:template>

<xsl:template match="/">
    <xsl:element name="html">
        <xsl:element name="body">
            <xsl:text>
                Corpo della pagina
            </xsl:text>
        </xsl:element>
    </xsl:element>
</xsl:template>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:attribute>

E' possibile creare attributi utilizzando l'elemento `<xsl:attribute>` ed indicando negli attributi *name* e *namespace* (opzionale) il nome e il namespace di appartenenza dell'attributo.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="attribute" substitutionGroup="xsl:instruction">
  <xs:complexType>
    <xs:complexContent mixed="true">
      <xs:extension base="xsl:sequence-constructor">
        <xs:attribute name="name" type="xsl:avt" use="required"/>
        <xs:attribute name="namespace" type="xsl:avt"/>
        <xs:attribute name="select" type="xsl:expression"/>
        <xs:attribute name="separator" type="xsl:avt"/>
        <xs:attribute name="type" type="xsl:QName"/>
        <xs:attribute name="validation" type="xsl:validation-type"/>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio Elemento <xsl:attribute>

```
<xsl:element name="div" >  
  <xsl:attribute name="id" >  
    <xsl:value-of select="@id"/>  
  </xsl:attribute>  
</xsl:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Altri elementi di creazione

- **commenti:** mediante `<xsl:comment>` specificando fra i tag di apertura e chiusura il testo del commento
- **processing instruction:** si utilizza `<xsl:processing-instruction>` specificando mediante l'attributo name il nome ed inserendone il contenuto tra i tag di apertura e chiusura
- **testo:** si usa `<xsl:text>` specificando nel corpo il contenuto della sezione CDATA.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Gestione spazi bianchi

La gestione degli spazi bianchi nel documento di origine è specificato dalle regole di scarto attraverso le istruzioni **xsl:preserve-space** e **xsl:strip-space**.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:preserve-space>

Elenca gli elementi dell'albero di origine in cui devono essere conservati gli spazi bianchi originali.

Esempio <xsl:preserve-space>

```
<xsl:preserve-space elements="p head"/>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="preserve-space" substitutionGroup="xsl:declaration">  
    <xs:complexType>  
        <xs:complexContent>  
            <xs:extension base="xsl:element-only-versioned-element-type">  
                <xs:attribute name="elements" type="xsl:nametests" use="required"/>  
            </xs:extension>  
        </xs:complexContent>  
    </xs:complexType>  
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Elemento <xsl:strip-space>

Elenca gli elementi dell'albero di origine in cui devono essere scartati gli spazi bianchi originali

Esempio <xsl:strip-space>

```
<xsl:strip-space elements = "*" />
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

```
<xs:element name="strip-space" substitutionGroup="xsl:declaration">  
  <xs:complexType>  
    <xs:complexContent>  
      <xs:extension base="xsl:element-only-versioned-element-type">  
        <xs:attribute name="elements" type="xsl:nametests" use="required"/>  
      </xs:extension>  
    </xs:complexContent>  
  </xs:complexType>  
</xs:element>
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esercizio

Modificare opportunamente il file template.xsl aggiungendo variabili, parametri e call template.

Progress status

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

1 Introduzione

2 XSL Transformations: Caratteristiche Fondamentali

3 Template (rules/named) e principali istruzioni XSLT

4 XPath: selezione dei nodi ed expression axes

5 Conclusioni

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XPath

XPath è un *expression language* fondamentale per realizzare fogli di stile XSLT.

XPath

- Selezionare nodi in un documento XML
- Fare match nell'albero source per selezionare il corretto template
- Manipolare dati attraverso funzioni

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XPath

XPath offre una sintassi estesa (piuttosto verbosa) e una sintassi abbreviata.

XPath

Le espressioni XPath permettono di selezionare con grande precisione elementi, attributi, ecc.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempi XPath selezione con match

Selezionare *documento XML* oppure tutti i nodi <quote>

- `<xsl:template match="/" >`
- `<xsl:template match="quote" >`

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempi XPath selezione con match

Per selezionare i nodi <title> "figli", "nipoti" o comunque
discendenti di <quote>

- `match="quote/title"` (*discendente diretto - figlio*)
- `match="quote/*/title"` (*discendente di secondo livello
- nipote*)
- `match="quote//title"` (*discendente a qualsiasi livello*)

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempi XPath selezione con match

Per selezionare i nodi `<quote>` con attributo `@type` (e valore `book`)

- `match="quote[@type]"`
- `match="quote[@type='book']"`

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempi XPath selezione con match

Per selezionare i nodi usando il carattere "jolly"

- `match="*"`
- `match="*[@type]"`
- `match="*[@type='book']"`

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempi XPath selezione con match

Per selezionare più di un elemento (operatore OR) oppure per
selezionare un nodo con uno specifico ID

- `match="title | author"`
- `match="id('stella-2007')"`

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XPath selezione con select

Il valore di questo attributo è un'espressione conforme al linguaggio XPath

XPath selezione con select

L'attributo select può ricorrere a una sintassi più complessa rispetto a match

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

L'attributo select è usato con le istruzioni XSLT

- `xsl:apply-templates`
- `xsl:value-of`
- `xsl:copy-of`
- `xsl:for-each`
- `xsl:sort`
- `xsl:variable`
- `xsl:param`

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XPath

Le espressioni XPath permettono di "navigare" l'albero del documento XML usando assi di navigazione (*expression axes*).

XPath

La selezione può essere assoluta o relativa al **nodo corrente** e si compone di tre parti: (**Asse, Test, Predicato**)

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

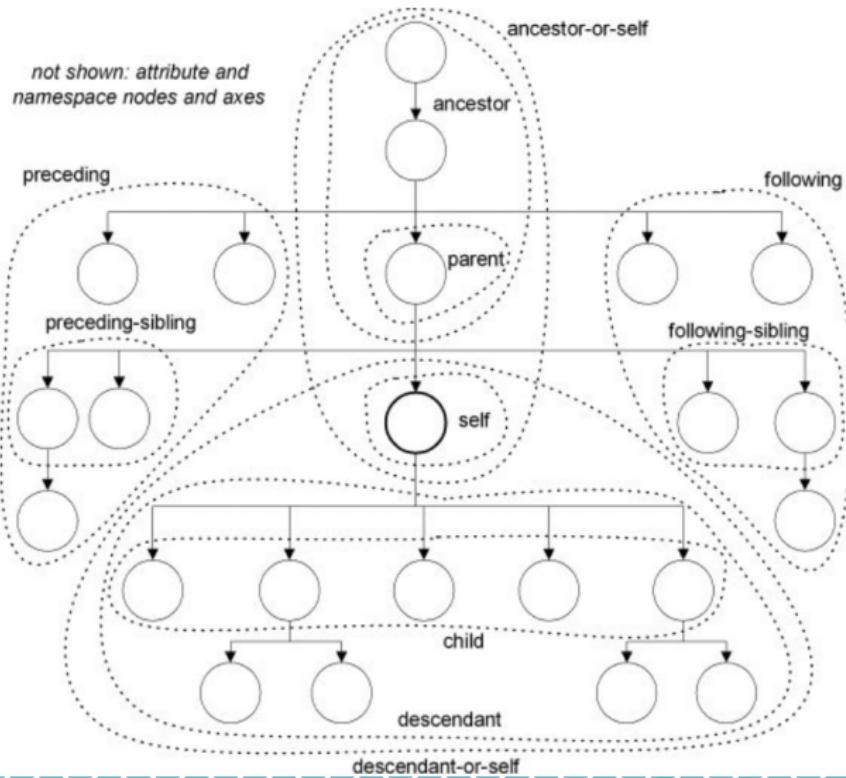
Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni



Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

ancestor	L'antenato del nodo di contesto, ovvero il genitore del nodo di contesto, il genitore del genitore del nodo di contesto, il genitore del genitore del genitore del nodo di contesto e così via fino al nodo radice.
ancestor-or-self	L'antenato del nodo di contesto e il nodo di contesto stesso.
attribute	L'attributo del nodo di contesto.
child	I figli diretti del nodo di contesto.
descendant	I discendenti del nodo di contesto, ovvero i figli del nodo di contesto, i figli dei figli del nodo di contesto e così via.
descendant-or-self	Il nodo di contesto e i suoi discendenti.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

following	Tutti i nodi che iniziano dopo la fine del nodo di contesto, esclusi gli attributi e i nodi dei <i>namespace</i> .
following-sibling	Tutti i nodi che iniziano dopo la fine del nodo di contesto e hanno lo stesso genitore del nodo di contesto.
namespace	Il <i>namespace</i> del nodo di contesto.
parent	Il genitore unico del nodo di contesto.
preceding	Tutti i nodi che finiscono prima dell'inizio del nodo di contesto, esclusi gli attributi e i nodi dei <i>namespace</i> .
preceding-sibling	Tutti i nodi che finiscono prima dell'inizio del nodo di contesto e hanno lo stesso genitore del nodo di contesto.
self	Il nodo di contesto

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

/	Seleziona l'elemento radice (root element).
/quote/author	Seleziona il nodo <author> figlio di <quote>.
//author	Seleziona tutti i nodi <author> a qualsiasi livello.
/quote//author	Seleziona tutti i nodi <author> a qualsiasi livello, ma solo se figli di <quote>.
. e ..	Indicano rispettivamente il nodo corrente e il nodo genitore.
/quote/*	Seleziona qualsiasi nodo figlio di <quote>.
quote/@id	Seleziona l'attributo id dell'elemento <quote>.
quote/@*	Seleziona tutti gli attributi dell'elemento <quote>.
/quote[@type='book']	Seleziona l'elemento con attributo e valore specificati.

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Sintassi abbreviata	Sintassi estesa
.	self::node()
..	parent::node()
name	child::name
@name	attribute::name
//	/descendant-or-self::node()

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

=, !=	Uguaglianza, Disuguaglianza
>; e <;	Maggiore di (>) e minore di (<)
>= e <=	Maggiore o uguale (>=) e minore o uguale (<=)
and, or, not()	AND logico, OR logico, NOT logico
position()	Posizione del nodo in una serie di nodi (numero)
last()	Ultimo nodo di una serie di nodi
	Unione tra condizioni o insieme di nodi
+, -, *	Addizione, sottrazione, moltiplicazione
div e mod	Divisione e resto della divisione (modulo)

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio sintassi estesa - sintassi abbreviata

- `<xsl:value-of select="child::author"/>`
- `<xsl:value-of select="author"/>`

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Esempio sintassi estesa - sintassi abbreviata

- `<xsl:value-of select="parent::quote"/>`
- `<xsl:value-of select="ancestor::quote"/>`
- `<xsl:value-of select=".." />`

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
espressioni
axes

Conclusioni

Esempio predicati

- `//div[@type='chapter']`
- `//div[@type != 'chapter']`
- `//div[@n > 2]`
- `//div[1]`
- `//div[last()]`
- `//div[position() = last() - 1]`
- `//div[position() mod 2 = 0]`

Progress status

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

1 Introduzione

2 XSL Transformations: Caratteristiche Fondamentali

3 Template (rules/named) e principali istruzioni XSLT

4 XPath: selezione dei nodi ed expression axes

5 Conclusioni

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XSLT: come procedere per un buon risultato

- Tracciare una mappa della conversione da XML a (X)HTML
- Lasciare il <teiHeader> per ultimo
- Definire lo scheletro HTML sul primo **template rule** e controllare che funzioni

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XSLT: come procedere per un buon risultato

- Divisione strutturale: uso di intestazioni (`<h1>`, `<h2>`, ecc.) e paragrafi, a capo (`<p>`, `
`) in HTML
- Formattazione degli elementi: usare gli elementi già presenti in HTML (``, `<i>`, ecc.) oppure definire degli ``.
- Scrivere una regola alla volta, testarla e solo se funziona passare alla regola successiva

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

XSLT: come procedere per XML-TEI

Tutti i nomi di elementi TEI devono essere preceduti dal prefisso **tei:!**

XSLT: come procedere per XML-TEI

```
<xsl:stylesheet version="1.0"  
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"  
    xmlns:tei="http://www.tei-c.org/ns/1.0"  
    xmlns="http://www.w3.org/1999/xhtml" >
```

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione

XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Fogli di stile TEI

Gli sviluppatori della TEI mettono a disposizione dei fogli di stile per documenti TEI P4 e posteriori

<http://www.tei-c.org/Tools/Stylesheets/>

XSLT per XML-TEI

<http://sourceforge.net/projects/tei/>

<http://tei.oucs.ox.ac.uk/teideb/binary>

<http://www.tei-c.org/oxgarage/>

<http://wiki.tei-c.org/index.php/Stylesheets>

git hub TEI/TEIC

Visualizzare ed Elaborare documenti XML

Codifica TEI -
Visualizza-
zione ed
Elaborazione:
Fogli di Stile

A.M. Del
Grosso

Introduzione
XSL Trans-
formations:
Caratteristiche
Fondamentali

Template
(rules/named)
e principali
istruzioni
XSLT

XPath:
selezione dei
nodi ed
expression
axes

Conclusioni

Riferimenti

- Michael Kay, XSLT 2.0 and XPath 2.0 Programmer's Reference
- James Clark, XSL Transformations (XSLT) Version 1.0, W3C Recommendation 16 November 1999, <http://www.w3.org/TR/xslt>
- E.R. Harold, XSL Transformations (XSLT), capitolo 14 del libro XML Bible, disponibile in rete:
<http://metalab.unc.edu/xml/books/bible/updates/14.html>