

Digital Humanities

DIGITAL HUMANITIES PENTRU SURSELE FILOSOFICE MEDIEVALE

4. Principiile TEI-XML

conf. dr. Mihai MAGA

Universitatea Babeș-Bolyai, Cluj-Napoca
Masteratul de Filosofie Antică și Medievală

semestrul II, 2023–2024
HME2415/04

<https://www.dhcluj.ro/dhm/>

1 Formatul XML

2 Despre TEI

- Programe pentru editarea TEI-XML

Temă pentru acasă¹²

XML (eXtended Markup Language) = text în format digital cu marcaje metatextuale

Doar 5 tipuri de conținut:

1 **tag-uri** (conțin elemente metatextuale)

```
<tag></tag>
```

sintaxă: între semnele < > și închise cu /

2 **attribute** (precizează parametri ai tag-urilor)

```
<tag attr="val"/>
```

sintaxă: nume (fără spații) urmat de = urmat de valoare între ghilimele " "

3 **text** (textul ca atare)

```
text
```

sintaxă: orice caractere (cu excepția <>) și care nu sunt tag-uri

4 **declarații** (comenzi de procesare)

```
<? decl ?>
```

5 **comentarii** (conținut ignorat la procesare)

```
<!-- comm -->
```

XML

```
<text>
```

Acesta este un text XML. Marcheaza ceva **<bold>important</bold>**.

A fost scris la **<oras tara="Romania">Cluj</oras>**.

<!-- aici este un comentariu -->

<tag>Un tag poate contine

<subtag>un subtag care poate si el contine

<subsubtag>sub-subtag-uri**</subsubtag>**

</subtag>

</tag>

Un tag poate avea zero, **<separator/>** unul sau mai multe attribute

<oras tara="RO" judet="CJ" prefix="0264">Cluj-Napoca</oras>

```
</text>
```

- tag-urile au două forme:
 - **pereche**: delimitează o porțiune de text (`<tag>text</tag>`)
 - ❶ trebuie închise în ordinea inversă deschiderii (`<a><b><c>...</c></b></a>`)
 - ❷ doar tag-ul de deschidere poate avea atribute (`<tag attr="val">...</tag>`)
 - **singure**: ca element de sine stătător, fără text (`<tag/>`)
 - ❶ pot avea atribute (`<tag attr="val"/>`)
- conținutul dintre tag-uri poate avea în el tag-uri, creând un arbore
`<root><branch><leaf/></branch><branch><leaf/></branch></root>`
- în XML bine-format, trebuie să existe un singur tag rădăcină care să conțină tot documentul
- spațiile suplimentare și sfârșiturile de linii sunt în mod normal ignorate, dar sunt folosite pentru editarea mai ușoară a codului

```
<document>
```

```
  Format:
```

```
    <b>bold</b> ,
```

```
    <i>italic</i> .
```

```
</document>
```

→ Format: **bold**, *italic*.

nod orice element care face parte din structură

rădăcină element care subordonează toate celelalte elemente ale documentului

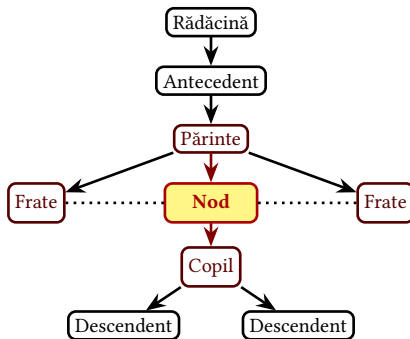
părinte relație între elemente în care elementul vizat are sub-elemente

copii relație în care elementele vizate sunt imediat subordonate elementului părinte

frați relație în care elementele vizate sunt pe același nivel și au un părinte comun

descendenți relație în care elementele vizate sunt inferioare unui element superior

antecedenti relație în care elementele vizate sunt superioare ierarhic și conectate cu un element descendent



- Ghidul Consorțiului Text Encoding Initiative (TEI) stabilește sistemul de adnotare pentru documentele din domeniul umanist
- TEI folosește XML ca format de fișier
 - TEI este un subset de instrucțiuni XML cărora li se atribuie o semantică
- în TEI sunt specificate elementele XML folosite la edițiile digitale
- elementul rădăcină al unui document TEI-XML este `<TEI> </TEI>`
- un document TEI are de obicei două părți obligatorii:
 - un preambul `<teiHeader> </teiHeader>`
 - în preambul sunt descrise proprietățile documentului: titlu, autor, versiune, surse etc.
 - corpul textului `<text> </text>`
 - conține textul cu tag-uri TEI; textul principal este conținut între `<body> </body>`; paragrafele de text sunt cuprinse între `<p> </p>`
- toate specificațiile TEI se găsesc pe site-ul:
<http://www.tei-c.org/>

TEI-XML

```
<?xml version="1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Titulus operae</title>
        <author>Nomen auctoris</author>
      </titleStmt>
      <sourceDesc>
        <p>Textus fictivus</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p xml:id="par01" lang="la">Hic est textus editionis.
      Transcriptus est in linguam programmandi qui
      nomen <ref target="http://www.tei-c.org/">TEI</ref> habet.</p>
    </body>
  </text>
</TEI>
```


<titleStmt>

<titleStmt> (title statement) groups information about the title of a work and those responsible for its content. 2.2.1 The Title Statement 2.2 The File Description	
Module	header — The TEI Header
Attributes	att.global (@xml:id, @n, @xml:lang, @xml:base, @xml:space) (att.global rendition (@rend, @style, @rendition)) (att.global linking (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) (att.global analytic (@ana)) (att.global facs (@facs)) (att.global change (@change)) (att.global responsibility (@cert, @resp)) (att.global source (@source))
Contained by	header: biblFull fileDesc
May contain	core: author editor meeting respStmt title header: funder principal sponsor
Example	<pre><titleStmt> <title>Cappgrave's Life of St. John Norbert: a machine-readable transcription</title> <respStmt> <resp>compiled by</resp> <name>J. Lucas</name> </respStmt> </titleStmt></pre> Show all
Content model	<pre><content> <sequence> <elementRef key="title" minOccurs="1" maxOccurs="unbounded"/> <classRef key="model_respLike" minOccurs="0" maxOccurs="unbounded"/> </sequence> </content></pre>
Schema Declaration	<pre><element titleStmt { att.global.attributes, att.global.rendition.attributes, att.global.linking.attributes, att.global.analytic.attributes, att.global.facs.attributes, att.global.change.attributes, att.global.responsibility.attributes, att.global.source.attributes, (title, model_respLike) }</pre>



<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-titleStmt.html>

■ Editoare avansate

i oferă validare TEI, autocompletare, documentație

- **oXygen XML Editor** (cu plată) <https://www.oxygenxml.com/>
- **VS Code** (open source) <https://code.visualstudio.com/>
- **Atom** (open source) <https://atom.io/>
- **XML Copy Editor** (open source) <http://xml-copy-editor.sourceforge.net/>
- **jEdit** (open source) <http://www.jedit.org/>

■ Editoare simple

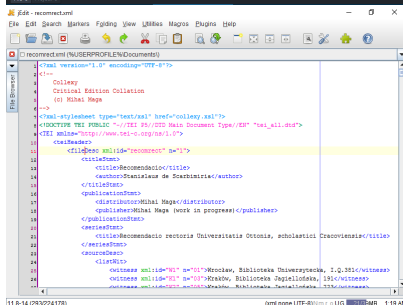
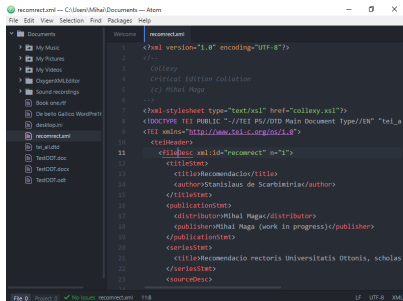
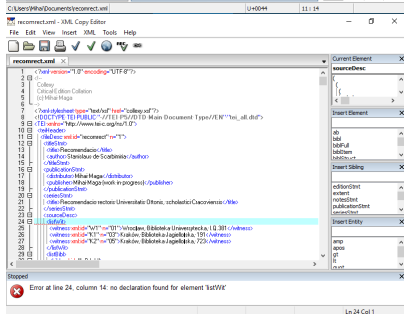
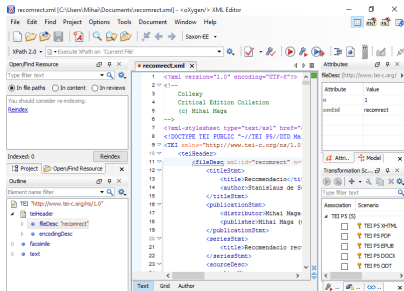
i oferă colorizarea sintaxei, uneori validare XML

- **Notepad++** (open source) <https://notepad-plus-plus.org/>
- **Notepad2** (gratuit) <http://www.flos-freeware.ch/notepad2.html>
- **Eclipse** (open source) <https://eclipse.org/>

■ Editoare web

i funcționează în browser, oferă funcționalități diverse

- **CodeMirror** (open source) <https://codemirror.net/>
- **eXide** (open source) <http://exist-db.org/exist/apps/eXide/index.html>
- **DH Editor** (pentru acest curs) <https://dhcluj.ro/dhm/lab/editor/>



Folosind următoarele tag-uri, codați textul de mai jos:

`<text></text>` elementul rădăcină al documentului

`<p></p>` paragraf

`<title></title>` titlu de operă

`<name></name>` nume de autor

`<quote></quote>` text citat

Quia intellectus habet duas operationes: scilicet unam qua format quiditates, in qua non est falsum, ut dicit ARISTOTELES in III *De anima*; aliam qua componit et dividit; et in hac etiam non est falsum, ut patet per AUGUSTINUM in libro *De vera religione*, qui dicit sic: “nec quisquam intelligit falsa”. Ergo falsitas non est in intellectu.

Praeterea, AUGUSTINUS in libro *LXXXIII quaestionum*, quaestio 32: “omnis qui fallitur, id in quo fallitur, non intelligit”. Ergo in intellectu non potest esse falsitas.

Item ALGAZEL dicit: “aut intelligimus aliquid sicut est, aut non intelligimus”. Sed quicumque intelligit rem sicut est, vere intelligit. Ergo intellectus semper est verus; ergo non est in eo falsitas.

THOMAS DE AQUINO, *Quaestiones disputatae de veritate*, Q. 1, art. 12