



Introduction into XML

Georg Vogeler
@gvogeler@hcommons.social



XML is Mark up

- Textmarker
- Type setting annotations in preparation of printed books
- Signs in scholarly editions:
 - | = | , | - - | , < > , []

Die sieben Kriterien der Textualität

1. Kohäsion (sichtbare Merkmale des Zusammenhangs)

□□ / C / (SP) "Four score and (7) years ago our fathers brought forth
/ 9 on this continent a new nation, conceived in ~~in~~ liberty, and
w.f. / S dedicated to the proposition, that all men are created equal. w.f.
a / tr. Now we're engaged in a great Civil War, testing whether
tr. that nation, or any nation (conceived (so) and (dedicated) (so)
can long endure. e.c. // 1 / -
tr.

run on We are met on a great battle field of that war) we have C / O W 7
that/ come to dedicate a portion of this field as a final resting
gave/ place for those who here have given their lives (that this
O S / nation might live/it is altogether proper and fitting) that
x / = } Eg #
= / C
O / e.c.

S lead S / 1 / here, have have consecrated it, far above our power to de-
S / detract tract or add. The world w/ little note, nor long remember,
w.f. / #
C / they / rom

out-are copy e.c. to the great Task remaining before us, -that from these
C / S Dr.
/
here/ stat. / w.f. we now highly resolve that these dead shall not have died
w.f.
tr. / that / 7 / in vain—that this nation, under God, shall have a new
/
x birth of freedom—and Government of the people, by the
not / ✓
people, for the people, shall never perish from the earth.

Syntax

- „pointy brackets“: `< >`
- Bracketing by corresponding `<tag>` and `</tag>` forming an **element**
- All elements have to be nested strictly hierarchical:

- `<phrase><w>Vae</w> <w>victis</w>!</phrase>`
- ~~`<phrase><w>Vae<w></w>victis</w>!</phrase>`~~

- Elements can have attributes by additional text in the starting tag: `<element attribut="Attributinhalt">`



Terminology

- **Tag**

- Code in pointy brackets: `<tag>`

- **Element**

- Corresponding tags („start tag“ `<tag>`, „end tag“ `</tag>`; „opening tag“ `<tag>`, „closing tag“ `</tag>`) and the text inbetween `<tag>content </tag>`
- **Content of the element:** everything between to corresponding tags, i.e. text and other elements

- **Attribute**

- Description of the element in the start tag `<tag attribute="...">`
- Consists of an attribute name `attribute=".."` followed by an equal sign and the attribute content `attribute="content"` in quotation marks.

Syntax

- „pointy brackets“: `< >`
- Bracketing by corresponding `<tag>` and `</tag>` forming an **element**
- All elements have to be nested strictly hierarchical:
 - `<phrase><w>Vae</w> <w>victis</w>!</phrase>`
- Each XML-document has exactly one root element.
- Some character have to replace by „entities“:
 - `< => <`
 - `& => &`

XML Ressourcen

- W3Schools: XML Tutorial (<https://www.w3schools.com/xml/>)
- IDE XML Short References (in German): <https://www.i-d-e.de/publikationen/weitereschriften/#Kleinschriften>

Experiment

"I, being born a woman and distressed", by Edna St. Vincent Millay
I, being born a woman and distressed by all the needs and notions of my kind, am urged by your propinquity to find your person fair, and feel a certain zest to bear your body's weight upon my breast: So subtly is the fume of life designed, to clarify the pulse and cloud the mind, and leave me once again undone, possessed. Think not for this, however, the poor treason of my stout blood against my staggering brain, I shall remember you with love, or season my scorn with pity, — let me make it plain: I find this frenzy insufficient reason for conversation when we meet again. (1923)

(source: <https://www.poetryfoundation.org/poems/148564/i-being-born-a-woman-and-distressed>)



XPath 2.0 XPath ausführen auf 'Aktuelle Datei'



Gliederung

Elementnamen-Filter

- poem I, Being Born a Woman and Distressed
- title I, Being Born a Woman and Distressed
- author Edna St. Vincent Millay
- text
 - octave I, being born a woman and distressed
 - stanza I, being born a woman and distressed
 - sestet Think not for this, however, think

Projekt

SaarbrückenSchool2024.xpr

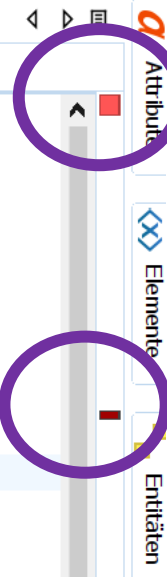
- BayHstAStV13/8.docx
- BayHstAStV1378.xml
- Millay_Sonnet_1923.txt
- Millay_Sonnet_1923.xml
- Petrarca_Sonnet.txt
- Petrarca_Sonnet_1923.txt

Die Hauptdatei-Unterstützung ist deaktiviert.

Aktivieren [Mehr lesen](#)

```
poem text octave stanza
1 <?xml version="1.0" encoding="UTF-8"?>
2 <poem>
3   <title>I, Being Born a Woman and Distressed</title>
4   <author>Edna St. Vincent Millay</author>
5   <text>
6     <octave>
7       <stanza>
8         I, being born a woman and distressed
9         By all the needs and notions of my kind,
10        Am urged by your propinquity to find
11        Your person fair, and feel a certain zest</stanza>
12      </stanza>
13      To bear your body's weight upon my breast:
14      So subtly is the fume of life designed,
15      To clarify the pulse and cloud the mind,
16      And leave me once again undone, possessed.</stanza>
17    </octave>
18    <sestet>
19      <stanza>Think not for this, however, the poor treason
20      Of my stout blood against my staggering brain,
21      I shall remember you with love, or season</stanza>
22      <stanza>My scorn with pity,—let me make it plain:
23      I find this frenzy insufficient reason
24      For conversation when we meet again.</stanza>
25    </sestet>
26  </text>
27 </poem>
```

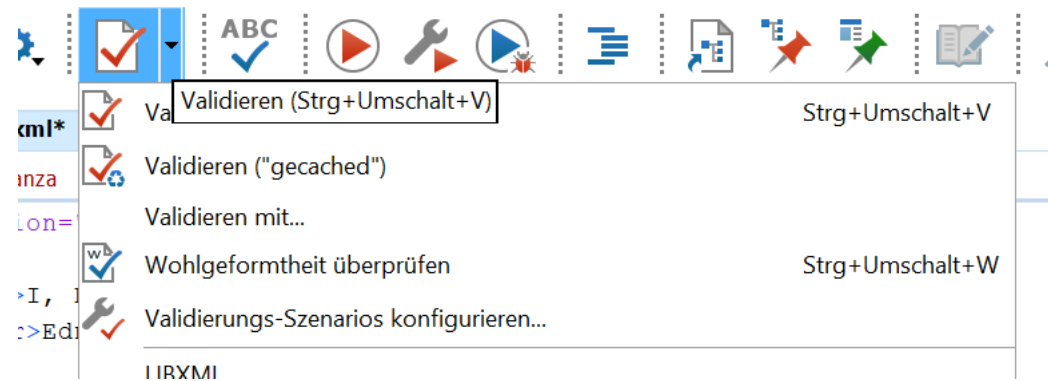
The element type "stanza" must be terminated by the matching end-tag "</stanza>".



Oxygen



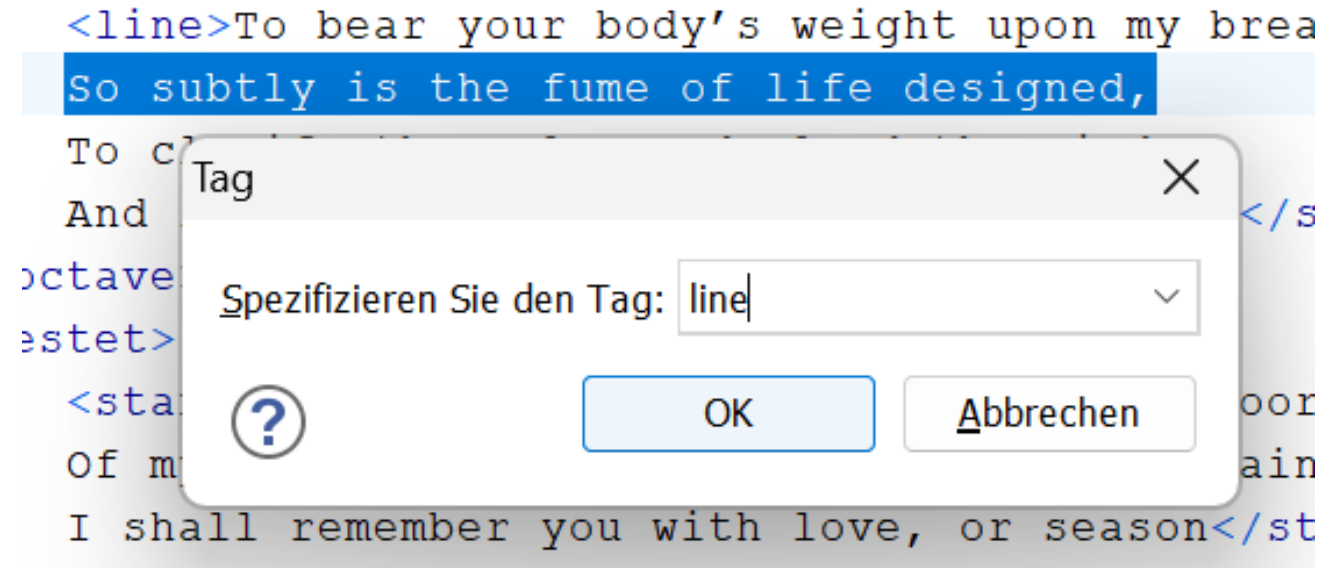
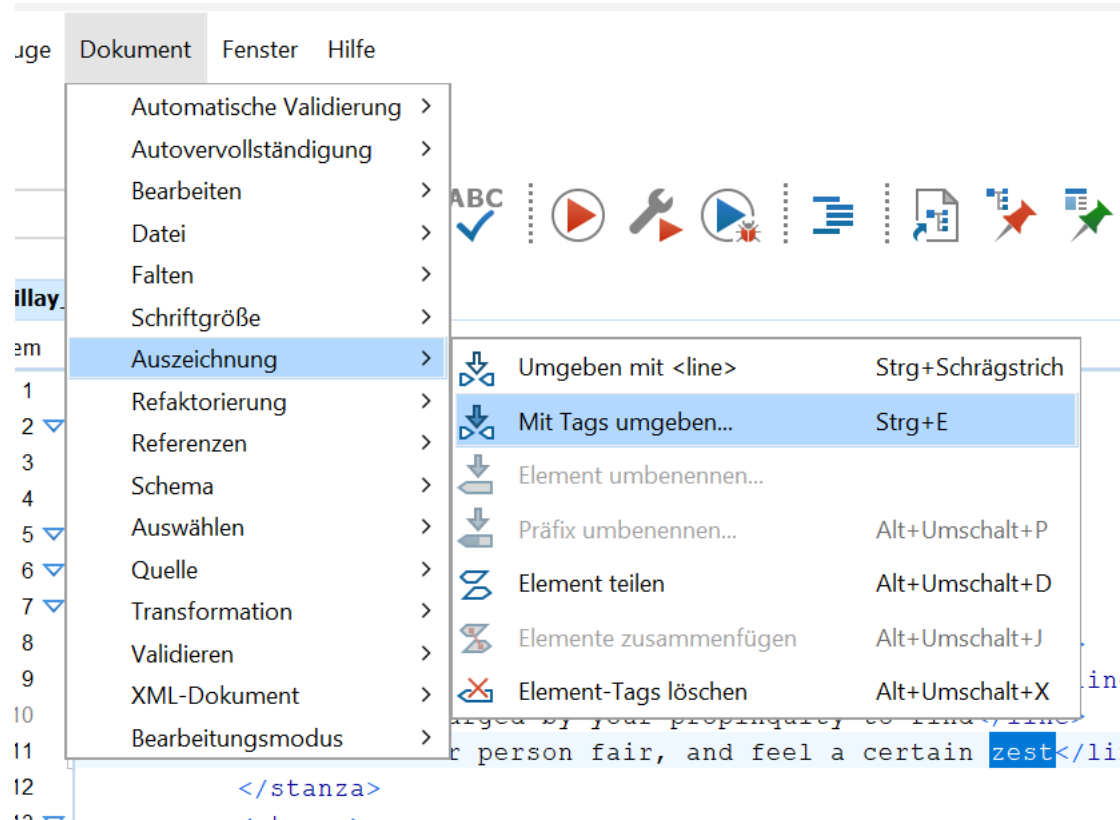
- Validation:
 - Wellformed (correct XML syntax) and valid (correct XML names and structure)



Oxygen: Annotating existing text



- Select text range + Ctrl-E
(menu: document > XML-refactoring > surround with tags)



Whitespace: line breaks and indentation

- XML does merges all whitespace (spaces between characters, words, lines, indentions etc.) into one single whitespace:
- „Oh my love is endless“
=> „Oh my love is endless“
- „Oh
my
love is endless“
=> „Oh my love is endless“
- „Oh
 my
 love is endless“
=> „Oh my love is endless“

Oxygen: „Re-format“



- Use your whitespace to make encoding more readable:

```
~EAC~  
<octave><stanza><line>I, being born a woman and distressed</line><line>By all the needs and notions of  
<line>Am urged by your propinquity to find</line><line>Your person fair, and feel a certain zest</line>  
<stanza><line>To bear your body's weight upon my breast:</line><line>So subtly is the fume of life desi  
</octave>
```

Ctrl-
Shift+P



```
~EAC~  
<octave>  
  <stanza>  
    <line>I, being born a woman and distressed</line>  
    <line>By all the needs and notions of my kind,</line>  
    <line>Am urged by your propinquity to find</line>  
    <line>Your person fair, and feel a certain zest</line>  
  </stanza>  
  <stanza>  
    <line>To bear your body's weight upon my breast:</line>  
    <line>So subtly is the fume of life designed,</line>  
    <line>To clarify the pulse and cloud the mind,</line>  
    <line>And leave me once again undone, possessed.</line>  
  </stanza>  
</octave>
```

Making Oxygen even more comfortable



- *Ctrl-shift-y* = soft line breaks (Document > edit > toggle line break)
- Offers Frameworks for specific use cases:
 - File -> New -> XML document, TEI all, ...
- Oxygen is „schema aware“:
 - Suggest fitting tags: closing tag, fitting tags by a pre-defined schema
 - Learns typical structures from a document opened
 - Let you assign an external schema to your document and validates
- => Just press „<“
-



Text Encoding Initiative

<https://tei-c.org>



TEI: the community

- Since 1986 by humanists
- Steered by (policital) board and a (technical) council elected by institutional and individual members (<https://members.tei-c.org/>)
- The SIGs („Special Interest Groups“)
- The journal (<https://jte.revues.org/>)
- The annual conference: <https://members.tei-c.org/Events/meetings> (this year in Buenos Aires)
- The Mailing-List: <https://listserv.brown.edu/archives/cgi-bin/wa?SUBED1=tei-l&A=1>



TEI: The guidelines

- <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- The specs: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- The Oxygen-Framework
- The tools:
 - Schema creation: Roma (<https://roma.tei-c.org>)
 - Document conversion: TEI-garage (<https://teigarage.tei-c.org>)



TEI resources

- <https://teibyexample.org>
- <https://teigarage.tei-c.org>
- <https://teipublisher.com>, Edition Visualisation Toolkit (Roberto Roselli del Turco, <http://evt.labcd.unipi.it/>)
- Example: <https://maria-magdalena.ub.uni-freiburg.de/teidoc/>

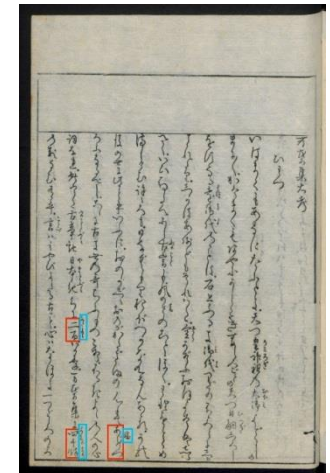
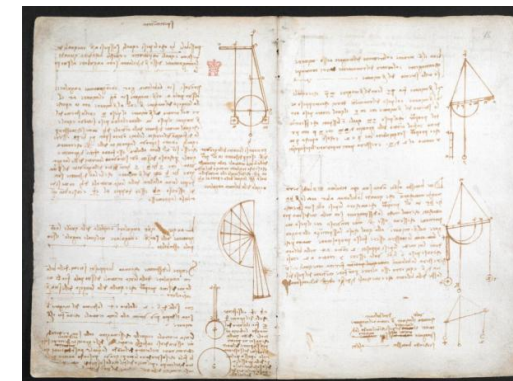
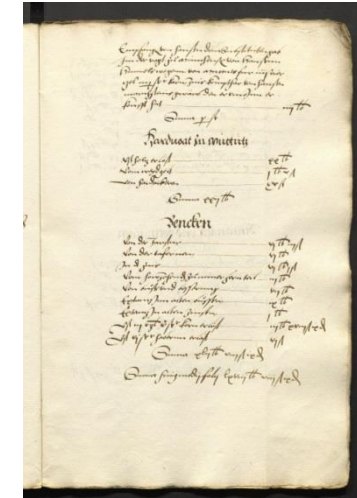
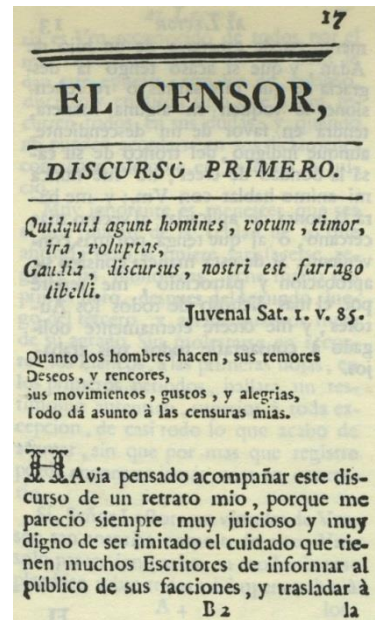


TEI Guidelines: Structure

- Core elements for
 - default text structure
 - Metadata (teiHeader)
- Modules for specific use cases:
 - Poetry, Theater and performanc, Lexica, ...
 - Primary Source: transcription, manuscript descriptions
 - Critical edition
 - Names, dates places

Structure

- ⊕ 1 [The TEI Infrastructure](#)
- ⊕ 2 [The TEI Header](#)
- ⊕ 3 [Elements Available in All TEI Documents](#)
- ⊕ 4 [Default Text Structure](#)
- ⊕ 5 [Characters, Glyphs, and Writing Modes](#)
- ⊕ 6 [Verse](#)
- ⊕ 7 [Performance Texts](#)
- ⊕ 8 [Transcriptions of Speech](#)
- ⊕ 9 [Dictionaries](#)
- ⊕ 10 [Manuscript Description](#)
- ⊕ 11 [Representation of Primary Sources](#)
- ⊕ 12 [Critical Apparatus](#)
- ⊕ 13 [Names, Dates, People, and Places](#)
- ⊕ 14 [Tables, Formulæ, Graphics and Notated Music](#)
- ⊕ 15 [Language Corpora](#)
- ⊕ 16 [Linking, Segmentation, and Alignment](#)
- ⊕ 17 [Simple Analytic Mechanisms](#)
- ⊕ 18 [Feature Structures](#)
- ⊕ 19 [Graphs, Networks, and Trees](#)
- ⊕ 20 [Non-hierarchical Structures](#)
- ⊕ 21 [Certainty, Precision, and Responsibility](#)
- ⊕ 22 [Documentation Elements](#)
- ⊕ 23 [Using the TEI](#)



hijde ride A review journal for digital editions and resources published by the DCE

Home About Issues Tags Data Reviewers search [input] Go

Home / Issues / Issue 1: Carolingian Scholarship and Martianus Capella: The Oldest Commentary Tradition

Carolingian Scholarship and Martianus Capella: The Oldest Commentary Tradition

Carolingian Scholarship and Martianus Capella: The Oldest Commentary Tradition, Markku Tuomi et al. (ed.), 2009. <http://martianus.huygens.knaw.nl> (last Accessed: 13.04.2014). Reviewed by Franz Fischer (CCER, University of Cologne), Franz Fischer (af@uni-koeln.de)

Abstract

This paper reviews a semi-diplomatic edition of the oldest gloss commentary on Martianus Capella's *De nuptiis Philologiae et Mercurii*, a popular fifth-century encyclopaedic allegory of the seven liberal arts. Based on and including facsimiles of a ninth-century manuscript from Leiden University Library (Crescensius Latinus Folio 48), the edition was produced collaboratively by various specialists. While it succeeds in providing full access to an important textual witness to a rich commentary tradition – the complex nature of which defies any attempt at capture in a print edition – the integration of a deeper critical analysis of the textual transmission and an augmentation of the material presented is still desired.

Einleitung

Die Webseite *Carolingian Scholarship and Martianus Capella* bietet eine digitale semi-diplomatische Edition der aus dem 9. Jahrhundert stammenden Handschrift *Leiden, Universiteitsbibliotheek, Crescensius Latinus Folio 48*. Die Handschrift gilt als einer der wichtigsten Textzeugen für die im Zuge der karolingischen Bildungsreform entstandenen umfangreichen Kommentaren der *Hexamer des Philologie et Mercurii*. Das

Social

Facebook Twitter YouTube

ToC

1. Einleitung
2. Inhalte und Methoden
3. Fazit
4. Anmerkungen
5. Bibliographie

Meta

Last updated: June 20, 2014 at 12:30 pm

PDF

XSL of this review, including formal evaluation data

Fachsheet of this project

Citation suggestion



TEI core elements

- Basic structure of the TEI XML document in
 - `<teiHeader>` (Metadata of the file)
 - `<text>` (the very text)
 - (`<front>`), `<body>` (the very-very text), (`<back>`)
- Text contains: sections `<div>`, paragraphs `<p>`, headings `<head>`, (foot-/end-/marginal)notes `<note>`, lists `<list>/<item>`, bibliographic references `<bibl>`, important parts `<emph>`, table of content `<divGen/>`, other language `<foreign>`, quotation `<q> | <quote> | <cit>`, images `<figure>`, tables `<table>/<row>/<cell>` etc.

TEI: Structure

Alternatives: text can be a:

- an **image** (<facsimile>)
- a linguistic structure (<text>)
- a documentary transcription (<sourceDoc>)



TEI: Structure

Alternatives: text can be a:

- an image (<facsimile>)
- a **linguistic structure (<text>)**
- a documentary transcription (<sourceDoc>)

Voi ch'ascoltate in rime sparse il suono
di quei sospiri ond'io nudriva 'l core
in sul mio primo giovenile errore
quand'era in parte altr'uom da quel ch'i' sono,

del vario stile in ch'io piango et ragiono
fra le vane speranze e 'l van dolore,
ove sia chi per prova intenda amore,
spero trovar pietà, nonché perdono.

Ma ben veggio or sì come al popol tutto
favola fui gran tempo, onde sovente
di me medesmo meco mi vergogno;

et del mio vaneggiar vergogna è 'l frutto,
e 'l pentersi, e 'l conoscer chiaramente
che quanto piace al mondo è breve sogno.

TEI: Structure

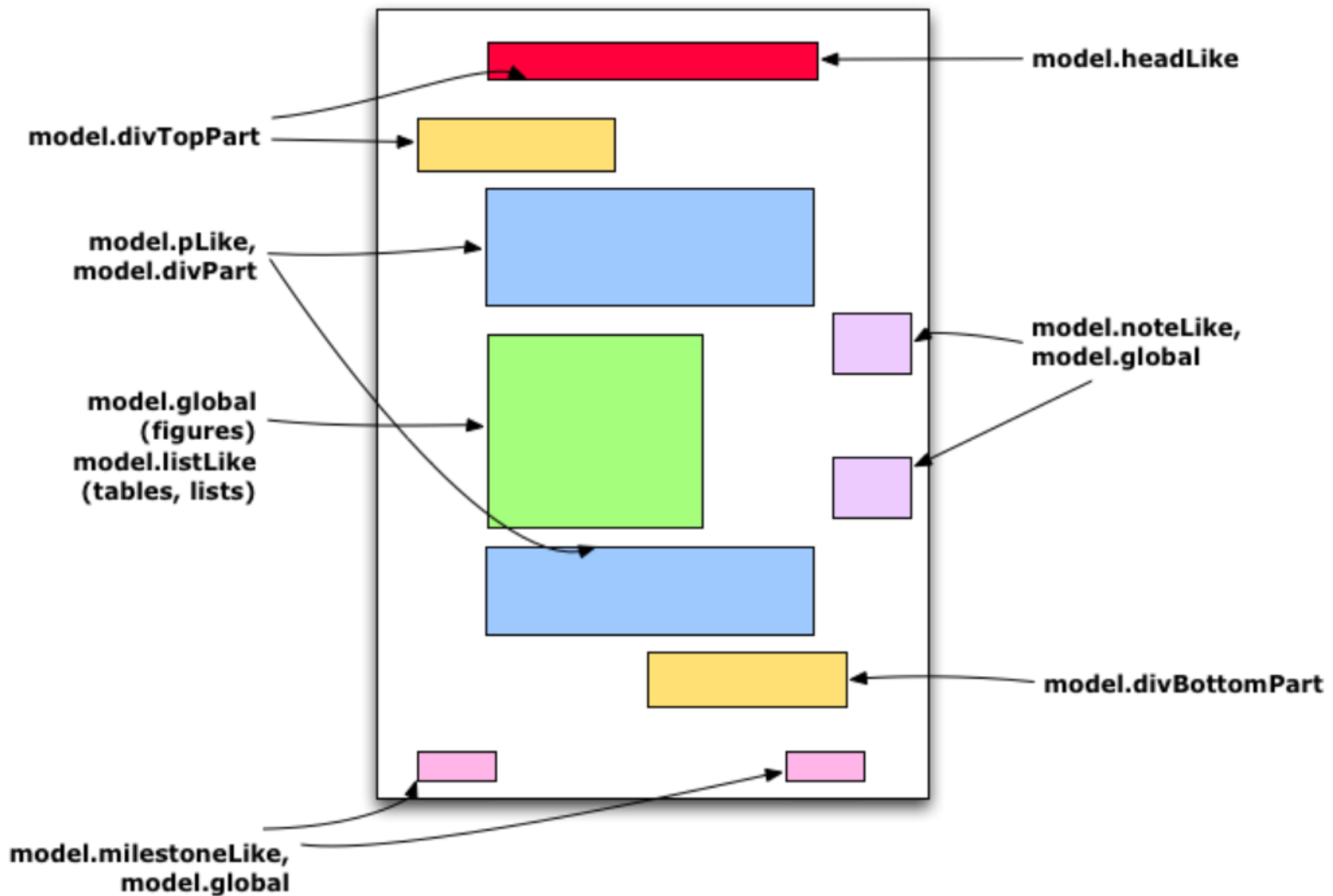
Alternatives: text can be a:

- an image (<facsimile>)
- a linguistic structure (<text>)
- a **documentary transcription** (<sourceDoc>)



VOI CHASOLTATE IN RI
 ME SPARSE EL SVONO
 Di quei sospiri ond'io nudriva 'l core
 In sul mio primo giovenile errore
 Quand'era in parte altr'uom da quel ch'i' son
 Del vario stile in ch'io piango et ragiono
 Fra le vane speranze e 'l van dolore,
 Ove sia chi per prova intenda amore,
 Spero trovar pietà, nonché perdono.
 Ma ben veggio or sì come al popol tutto
 Favola fui gran tempo, onde sovente
 Di me medesimo meco mi vergogno;
 E del mio vaneggiar vergogna è 'l frutto,
 E 'l pentersi, e 'l conoscer chiaramente
 Che quanto piace al mondo è breve sogno.

TEI basic textual units





TEI by teigarage

1. Convert the edition of the account for Vicedominus Peter von Eck 1337 (BayHStAStV1378.docx) with <https://teigarage.tei-c.org/> into TEI P5 XML
2. Open the file in Oxygen
3. Study the XML encoding, read in particular the descriptions in the guidelines on the elements with the names „note“, „p“, „div“, „hi“
4. Discuss with your colleagues how much the annotations cover of the editorial interest
5. try to invent elements that have names that are more precise from the standpoint of an editor.

This work is licensed under a [Creative Commons Namensnennung 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



All works of other author cited here are their intellectual property and are used for academic purpose only.