Digital Humanities

Digital Humanities for Medieval Philosophical Sources

8. Visualisation of digital editions

conf. dr. Mihai Maga

Babeș–Bolyai University, Cluj–Napoca
Master in Ancient and Medieval Philosophy

2nd semester, 2023–2024
HME2415/08

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

8. Visualisation of digital editions

conf. dr. Mihai Maga   HME2415/08   2 / 14

Because `TEI` is not a visual format, the human user requires a conversion into a familiar format. By origin, the digital documents belong to 2 types:
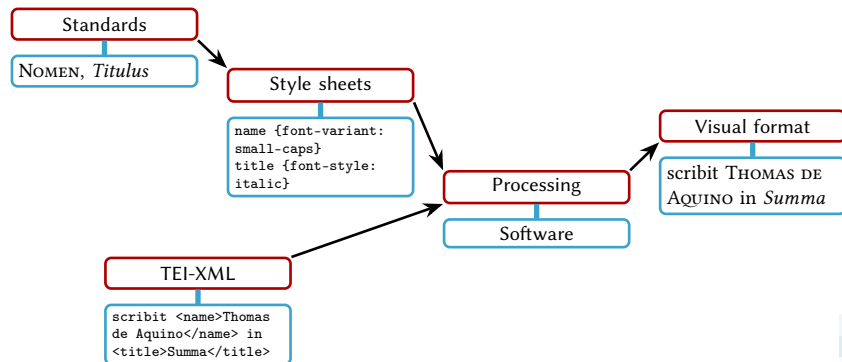
analog-to-digital   editions issued initially in analog format (e.g. on paper), subsequently electronically transcribed and encoded

born-digital   editions elaborated from the beginning in semantic digital format

The born-digital editions require a conversion for the human reader. The formats which are familiar to the reader include the printed formats and those on screen that emulate analog forms.

analog-to-digital

born-digital
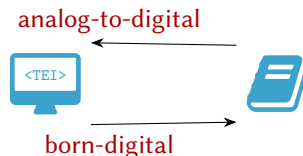
Due to the separation of layers (see *Course 2*), the process of conversion in visual format does not require the modification of the `TEI-XML` source. The conversion is done with the aid of style sheets written based on standards and which specify how each semantic element is rendered visually.

Standards

NOMEN, *Titulus*

Style sheets

```
name {font-variant:
small-caps}
title {font-style:
italic}
```

Processing

Software

Visual format

scribit THOMAS DE AQUINO in *Summa*

TEI-XML

```
scribit <name>Thomas
de Aquino</name> in
<title>Summa</title>
```

1. attaching a `CSS`* style sheet (Cascaded Style Sheet)
   - one specifies the formatting following `CSS` principles for `HTML`
     (see documentation in `http://www.w3schools.com/css/`)
   - one links the `CSS` document in the header of the `TEI` document
     (with the declaration `<?xml-stylesheet type="text/css"`
     `href="stylesheet"?>`; the CSS style can be an external document `"style.css"` or
     an internal element `"#style"`)
   - ❶ does not allow changing the order and composition of elements
2. attaching a `XSLT`* transformation sheet(eXtensible Stylesheet Language
   Transformations)
   - one specifies the transformation following `XSLT` principles for `XML`
     (see documentation in `http://www.w3schools.com/xml/xsl_intro.asp`)
   - one links the `XSLT` document in the header of the `TEI` document
     (with the declaration `<?xml-stylesheet type="text/xsl"`
     `href="transform.xsl"?>`)
   - ❶ the `XSLT` language is more complicated, needs additional support
3. transforming in `LaTeX` or other `DTP` format
   - one transforms with `XSLT` and other scripts to the target application
     format
   - the resulting document is compiled/exported in a common format (e.g.

CSS Style Sheets

- Specifies the style for each `XML` element in the format:

```
tag {parameter: value; parameter: value}
```

- Common `CSS` parameters

```
font-family: "Times New Roman", Arial, serif, sans-serif ...
font-size: 10pt, 12px ...;          font-weight: bold, normal;
font-style: italic, normal;         font-variant: small-caps, normal;
text-align: center, left, right, justify;
color: red, blue, #00FF00 ...;
background-color: red, blue, #00FF00 ...;
display: inline, block, none
```

CSS Example

```
head {font-family: Arial; font-size: 20px; text-align: center}
p {font-family: Garamond; font-size: 12px; text-align: justify}
name {font-variant: small-caps; color: red}
title {font-style: italic; color: green}
quote {font-style: italic; color: blue}
pb, cb, lb {display: none}
```

☞ the format is applied in cascade to the specified element and to its descendants (if those don't have another format). E.g.: `quote` inside `p` has still `font-size: 12px`

- the CSS language specifies the style in the format
  `identifier {parameter: value}`

- the identifier can contain several elements, separated by comma:

  ↪ `quote, title {font-style: italic}` (quotes and titles are in italic)

- the style can be limited to the descendants of a superior element:

  ↪ `bibl title {font-weight: bold}` (only titles in bibliography are bold)

- the identifier can be a tag name (`tag`), an id attribute (`#id`) or a class
  (`.class`):

  ↪ `#codex1 {color: blue} .uncertain {color: gray}` (only the element with id=codex1 is blue, all elements with class=uncertain are gray)

- multiple parameters can be specified, separated by semicolon:

  ↪ `title {font-style: italic; color: blue}` (the title will be italic and blue)

- if the same identifier and the same parameter is present multiple times,
  the last one applies:

  ↪ `p {color:red} p {color:blue}` (the paragraph will be blue)

The display medium determines the restrictions and the possibilities for formatting.

- in print
  - requires page layout according to typographic norms
    - macro-typography: the layout of text blocks in page
    - micro-typography: at the level of letter parts and letters joining
  - the resolution and graphical quality must be high
  - ❗ limits: the lines and pages must be broken; the space is limited (the number of pages is finite, large empty spaces are to be avoided); the repetition of structure elements is avoided using references; the usage of colors and images is restricted
- on screen
  - requires formatting according to the display application
    - compatibility: conformation with the application standards (e.g. HTML in browser)
    - adaptability: the display devices have various dimentions and forms, the text must be correctly displayed on any hardware (🖥 🎥 💻 📱 …)
  - the quantity of text and images is virtually unlimited; interactivity
  - ❗ limits: some functions and elements are not available on all systems; the rendering quality varies depending on terminal; the file size must be optimized; standards frequently change, sometimes breaking compatibility

Screen vs. print (example)

Accessing the text through a computer allows the user to interact with the semantic, textual and multimedia elements.

- selective display
  - due do the large quantity and diversity of information which TEI may contain, the elements displayed on screen at a certain moment must be limited for readability
  - depending on usage patterns, certain element types may be hidden (and accessible through additional interactions) (e.g. the images of manuscripts attached to a text or its textual variants may be viewed only after pressing a button)
- filtering and extraction
  - the content can be filtered by certain parameters, parts of document can be extracted according to the user's intention
  - there can be semantic filters (selecting contents by their semantic) (e.g. generating a list of titles) or text filters (e.g. free text search) or mixed filters (e.g. text search in author names)
- processing and computation
  - operations executed by the computer in text, according to user's demand (e.g. statistic of citations, comparison with other text, language analysis)

Inclusion of manuscript images

The attribute `facs=""` can be attached to many `TEI` elements and specifies the link to an image which corresponds to the element. The link may be done in 2 modes:

- **file**: an image file is directly specified (jpg, png, tiff, gif...)

  ```
  <locus facs="photo0801.jpg"/>
  ```

- **unique `facsimile` identifier**: a unique `xml:id` identifier is specified, which is declared in a `<facsimile>` element

  ```
  <facsimile xml:id="img2"><graphic url="photo0802.jpg"/></facsimile>
  <p facs="#img2">Omnes homines natura scire desiderant.<p>
  ```

  - the `facsimile` element allows a more precise control of the image by specifying multiple images, surfaces and zones

  ```
  <facsimile xml:id="facs1">
      <surface xml:id="sf1">
          <zone xml:id="zone1" ulx="10" uly="10" lrx="210" lry="297">
              <graphic url="photo01.jpg"/>
          </zone>
      </surface>
      <surface xml:id="sf2"><graphic url="page02a.jpg"/>
                           <graphic url="page02b.jpg"/></surface>
  </facsimile>
  ```

- IIIF (International Image Interoperability Framework) is an increasingly common standard for publishing, transmitting and reuse of manuscript images (and not only) with metadata

🌐 http://iiif.io/

IIIF uses the JSON file format (JavaScript Object Notation)

Sites that use IIIF: British Library, Bibliothèque nationale de France, Bibliotheca Apostolica Vaticana, Virtual Manuscript Library of Switzerland, IRHT (CNRS), Bayerische Staatsbibliothek etc. (over 70 institutions)

- TEI can also be used for publishing facsimiles with metadata

Sites that use TEI for facsimiles: Manuscriptorium, Oxford Bodleian Library etc.

- The mode of displaying the elements induces a pre-interpretation of contents

  ☞ the elements displayed larger, more colored or more upwards look more important

- The selection of elements initially displayed causes de ignorance of those displayable through complex interactions

  ☞ the inexperienced user may believe that the edition is reduced to what is seen on the front page

- The predetermined interactions induce the feeling of exhaustive analysis

  ☞ however complex and numerous, pre-programmed interactions do not exhaust the questions

Fictitious example of display which can mislead:

🚫

THOMAS DE NOVUM EBORACUM, *Quest for Love**

Utrum haec passio quae est amor sui, sit causa omnis peccati. [...] Ad primum ergo dicendum quod amor sui ordinatus est debitus et naturalis, ita scilicet quod velit sibi bonum quod congruit. Sed [...]   Thomas Aquinas, Summa theologica, I^{a}II^{ae}, q. 77, a. 4

Search for source authors: **10** authors found!

Aristoteles, Aristoteles, Aristoteles, Augustinus, Aristoteles, Augustinus, Biblia, Aristoteles, Biblia, Augustinus

* *Application developed by Thomas of New York in 2016. It searches random internet sources.*

Operate at least 5 visible and valid changes in the following style
sheet:

CSS style sheet

```
titleStmt title, titleStmt author, publicationStmt distributor {
  display: block;
  text-align: center;
  font-weight: bold;
}
p {
  display: block;
  text-align: justify;
  text-indent: 40px;
}
title {
  font-style: italic;
}
name, author {
  font-variant: small-caps;
}
```