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

conf. dr. Mihai Maga

Digital Humanities

Digital Humanities for Medieval Philosophical Sources

2nd semester, 2023–2024

HME2415

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

Course outline

0. Presentation

0.1. Course 1: Introduction to Digital Humanities

1. What does Digital Humanities mean?
   - a technique
   - a philosophy
2. History of Digital Humanities
3. Domains of Digital Humanities

0.2. Course 2: Semantic encoding

1. Visual representation vs. semantic representation
2. The semantic paradigm in the digital world
3. Utility of semantic encoding: integration, interfaces, data processing

0.3. Course 3: Critical editing

1. Methods of approach for critical editions
   - Best manuscript method
   - Eclectic method
   - Stemmatic (lachmannian) method
   - Unoriented (material) method
2. Types of critical editions
   - Facsimile type edition
   - Eclectic edition
   - Literary (critical) edition
   - Diplomatic edition
   - Material edition
3. Elements of a critical edition: Introduction, Text, Critical apparatus

0.4. Course 4: Principles of TEI-XML

1. The XML format
   - XML syntax: tag, attribute, text, declaration, comment
2. About TEI[1]
   - The TEI Consortium P5 Guidelines and the document structure
     - http://tei-c.org/

[1] Text Encoding Initiative

XML

```xml
<?xml version="1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
    <text>
        <p xml:id="par01" lang="la">Hic est textus editionis.</p>
    </text>
</TEI>
```

0.5. Course 5: Representation of manuscripts in TEI

1. Indication and description of manuscripts
   - manuscript identifier: place, libray, shelf number
   - description parts: physical, historical, contents, bibliography
2. Instruction set for describing manuscripts

<msDesc>	manuscript description block
<msIdentifier>	identifier: <settlement>, <repository>, <idno>...
<head>	heading
<msContents>	manuscript contents block: <msItem>, <locus>, <title>...
<physDesc>	physical description: <objectDesc>, <supportDesc>, <dimensions>, <layoutDesc>...
<history>	manuscript history: <origin>, <provenance>, <acquisition>...
<additional>	additional information
<msPart>	information about a section of the manuscript

0.6. Course 6: Representation of textual variation in TEI

1. Differences between manuscript copies
   - variants of redaction
   - copyist errors
2. Describing textual variations in TEI:

<app>	apparatus
<lem>	lemma
<rdg>	reading

3. Encoding methods:

LRM	Location-referenced Method
DEPAM	Double End-Point Attachment Method
PSM	Parallel Segmentation Method

4. Statistics and counters

0.7. Course 7: Representation of source apparatus in TEI

1. Ontologies and trees

- tree-structured data
- objects, classes ontologies

2. Describing the elements of a textual reference
   - references to author, title, work, section...: <title>, <author>...
   - bibliographies: <bibl>...
   - other types of references
3. Attaching the references
   - inline
   - using pointers

0.8. Course 8: Visualisation of digital editions

1. Conversion to classic format
   - typesetting rules for editions
   - specialized DTP: LaTeX
2. Interactive interfaces
   - web interfaces: client-side, server-side
   - native applications
3. Inclusion of manuscript images
   - TEI instructions set: <facsimile>, <surface>, <zone>, <graphic>
   - image attachment: inline, pointer
4. Pitfalls of the visual

0.9. Course 9: Integration and digital processing

1. Indices and concordance tables
2. Query languages: XQuery, XPath
3. Search in text: simple, wildcard, stemmatized, lemmatized
4. Lemmatization, normalization, dictionaries
5. Digital corpora
6. Data-mining

0.10. Course 10: Artificial intelligence

1. What is artificial intelligence (AI) and machine learning (ML)?
2. Algorithms and models in ML
3. Training and prediction in ML
4. Classification of approach types in ML

1. Introduction to Digital Humanities

1.1. What does Digital Humanities mean?

Digital humanities [DH] is an area of research and teaching at the intersection of computing and the disciplines of the humanities. Developing from the fields of humanities computing, humanistic computing, and digital humanities praxis, digital humanities embraces a variety of topics, from curating online collections to data mining large cultural data sets.[1]

- a technique
- a philosophy

A list of over 260 definitions of DH can be accessed at:
http://www.artsrn.ualberta.ca/taporwiki/index.php/How_do_you_define_Humanities_Computing_/_Digital_Humanities%3F

See also: M. Terras, J. Nyhan, E. Vanhoutte (eds.), *Defining Digital Humanities: a reader*, Farnham/Burlington:Ashgate 2014

1.1.1. A technique

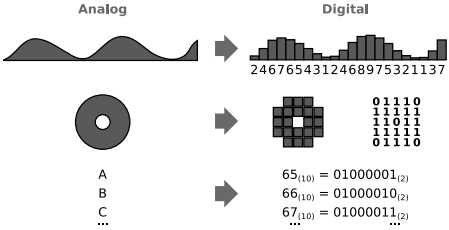
- the humanities domain is so vast that no person or group of persons can have the full control over it
  - even in a very specialised subdomain, due to the vast and multidisciplinary nature of knowledge, nobody can affirm that he knows everything
- exact sciences have progressed enormously in the last 70 years using electronic, cybernetic, digital instruments
  - in fact, the digital progress is dued partially to the desire to push further the boundaries of knowledge

➥ how can this technical advance help the humanities?

- by taking as a model the way in which other sciences use the computer to exploit large volumes of information and to compute solutions to their problems
- by developing computer systems which are appropriate for the specific problems of humanities, which differ highly from other sciences

What does *digital* mean?
digit$_{Engl.}$ = figure, number **digitus**$_{Lat.}$ = finger; measure unit
digital format = discontinuous and discrete numerical representation
digitisation = conversion of analog values into numeric values through sampling or through encoding tables

[1] https://en.wikipedia.org/wiki/Digital_humanities



- the computer processes and stores data only in binary form

1 | 0 +5V_{cc} | 0V_{cc} yes | no true | false

= bits

4

Number and meaning

- Calculate the result of the following operation:

$$430 - 354 = ?$$

- The result is:

$$76$$

- What does **76** mean? **76** of what?
  - Augustine of Hippo (**354** – **430**) lived **76** years
  - "... of the **430** verified persons, **354** were frauding the money..."[2] and **76**?
  - *Afterword*, pp. **354**–**430** (total **76** pages)

The computer does abstract numeric operations, independent of the meaning

but humanities deal primarily with the meaning

In DH we need to attach a meaning to the computational values

5

1.1.2. A philosophy

- Digital Humanities is part of a large movement of thought, which has begun in the 20th century
- The bases of this movement come from:

structuralism the meaning is constructed as relation in a structure

see semiotics, Ferdinand de Saussure, linguistic arbitrariness

[2] Statement by Minister of Labour on 23.10.2011

poststructuralism critique of structuralism; the meaning is constructed only inside the structure; deconstruction reveals the meaning
cognitivism universal grammar (N. Chomsky)

Poststructuralism

...language refers to the position of the listener and the speaker, that is, to the contingency of their story. To seize by inventory all the contexts of language and all possible positions of interlocutors is a senseless task. Every verbal signification lies at the confluence of countless semantic rivers. Experience, like language, no longer seems made of isolated elements lodged somehow in a Euclidean space... [Words] signify from the "world" and from the position of one who is looking.[3]

6

A poststructuralist and deconstructivist philosophy

- We do not have complete access to the original meaning of the medieval philosophy
- We deconstruct the intratextual significations and the historical interpretations to reveal the original relations
- We build our own structure of understanding
  - the differences, the cleavages, the context are important
  - the synchronic and diachronic relations must be determined

➥ the conscience that the author of an edited text is the editor

7

Text, context, author

Text example	**Intratextual analysis**	**Intertextual analysis**	**Deconstruction**
Libertus de Oficiis, *On Life*, III, 4:	Libertus de Oficiis, *On Life*, III, 4:	Libertus de Oficiis, *On Life*, III, 4:	~~Libertus de Oficiis~~,[4] *On Life*, III, 4:
My definition of the soul is the following: the soul is the first actuality of a natural organic body. By this I reject the thesis affirmed by Aristotle in *On life in the deep*, where he states: "in the deep of the sea live creatures which do not have organs."	My definition of the soul is the following: the soul is the first actuality of a natural organic body. By this I reject the thesis affirmed by Aristotle in *On life in the deep*, where he states: "in the deep of the sea live creatures which do not have organs."	~~My~~ definition of the soul is the following: the soul is the first actuality of a natural organic body. By this I reject the thesis ~~affirmed by~~ Aristotle in *On life in the deep*, ~~where he states~~: "in the deep of the sea live creatures which do not have organs."	My definition of the soul is the following: the soul is the first actuality of a natural organic body. By this I reject the thesis affirmed by Aristotle in *On life in the deep*, where he states: "in the deep of the sea live creatures which do not have organs."
	main author quoted author	main author source author Arist. *De anima* II.1 412b	real author 'main' author src. author for r.a. src. author for m.a.

we believe that the meaning is constructed by the author, but "the author" is a contextual concept

in fact, we are constructing the meaning, including the meaning of the "author" concept

[3] E. Lévinas, *Signification and Sense*

[4] This author name and the text are just an invention of this course's author

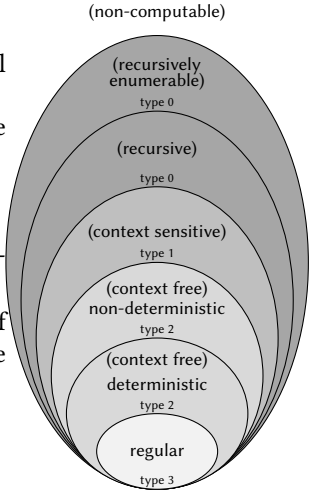
8

Formal grammar

Noam Chomsky – universal grammar, transformational generative grammar

- syntactic knowledge is partially inborn ← language acquisition device ⇒ **universal grammar**
- language consists in
  - **surface structures** (spoken utterances)
  - **deep structures** (relations between words and conceptual meaning)

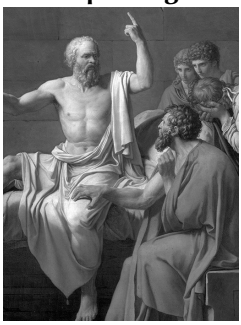
➥ **transformative grammar** consists of a limited set of rules of transformation of deep structures into surface structures



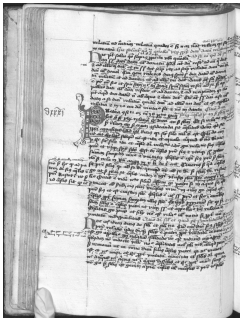
Chomsky's hierarchy

☞ psycholinguistics, computer programming, artificial intelligence

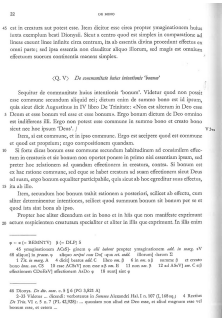
A new paradigm of the text



orality



manuscript



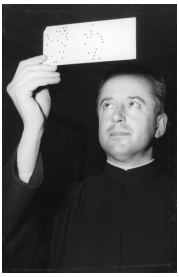
print

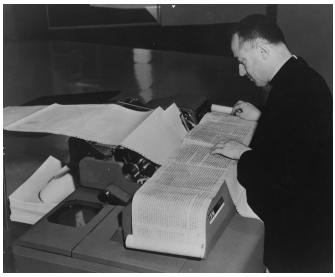
digital

~5300 BC ~1440 1949

1.2. History of Digital Humanities

1949 Roberto Busa, SJ, intends to compose an exhaustive concordance index for Thomas Aquinas' work; he meets Thomas J. Watson, the founder of IBM, and convinces him to sponsor *Index Thomisticus*; for 30 years he enters into a computer (on punched cards) all the Thomistic works and he publishes a printed version in 56 volumes (then available on CD-ROM and now online[5])





'60–'70 several electronic concodrances were produced (texts in Early Middle High German, poems of W. B. Yeats etc.)
–//– projects of statistical evaluation of texts' authorship appeared (Pauline Epistles, Lettres of Junius, *Federalist Papers* etc.)
! computers were highly limited and inaccessible

☞ serial processing on enormous machines owned by a few institutions and which used punched cards and magnetic tapes

1963– *Centre for Literary and Linguistic Computing*, Cambridge
1964 the conference *Literary Data Processing* organized by IBM at Yorktown Heights
1966– the publication of *Computers and the Humanities* journal begins
1965– the COCOA software established a first standard in "humanities computing" (University College London)
'70–'85 consolidation of the methodology DH
–//– more international conferences: Edinburgh (1972), Cardiff (1974), Oxford (1976), Birmingham (1978), Cambridge (1980); more publications appear
–//– centers of "humanities computing" are multiplying, courses on DH are introduced
! computers become more powerful and more accessible
1971– *Project Gutenberg*, Michael Hart (books in public domain)
1972– *Thesaurus Linguae Graecae* (TLG), Univ. of California, Irvine
1976– *Oxford Text Archive* (OTA), evolutions in electronic texts archiving
'85–'90 development, networking, standardisation
–//– textual analysis software in DOS: Word-Cruncher, TACT, MicroOCP
! personal computers, electronic mail, internet and WWW appear
1988–90 Ian Lancashire and Willard McCarty, *Humanities Computing Yearbook*
1987 meeting in Vassar College, Poughkeepsie, for creating a standard encoding scheme for DH ("*Poughkeepsie Principles*") ⟹ TEI (Text Encoding Initiative)
1990 first TEI draft
1994 first complete version of *TEI Guidelines*

[5] http://www.corpusthomisticum.org

'90... the internet becomes a vital part of the academic activity
14
the internet is used for publication, and also for promoting DH
–//– many projects and prototypes for online publication of digital editions appear
- efforts are more concentrated on display, interface, interactivity

–//– online collections and corpora appear
- new ways of scholarly writing are discovered:
  - extraction of documents from a database and their reconnstruction in a new material
  - collaborative editing of documents (manuscript transcriptions, annotation of corpora, digital libraries)
  - possibility of providing multimedia content (images, video, audio)

–//– TEI becomes the de facto standard for DH

2007 release of *TEI Guidelines* P5

1.3. Domains of Digital Humanities
15

- **by discipline**:
  - palaeography and critical edition
  - codicology and bibliotheconomy
  - philology and linguistic analysis
  - general history and applied histories
  - archaeology, art, museography
  - linguistic computing and artificial intelligence
- **by period**:
  - ancient
  - medieval
  - modern
  - contemporary
- **după aplicație/by application**:
  - stylistic analysis and authorship studies
  - assembly and analysis of linguistic corpora
  - scholarly critical editing in digital form
  - textual analysis
  - structuring of thematic collections
  - production of materials in printed and multimedia format
  - speculative computing (A.I.)
- **by digital source**:
  - born-digital
  - converted from analog

DH tends to become an universal instrument

16

Examples of usage for Digital Humanities

- Online library catalogues

http://www.manuscriptorium.com/

- Online text editions

http://scta.lombardpress.org/

- Integrated collections of selected texts

http://www.perseus.tufts.edu/hopper/

- Complete digital corpora

Thesaurus Linguae Graecae CD-ROM

- Large data sets

http://storage.googleapis.com/books/ngrams/books/datasetsv2.html

17

Bibliography

[1] "Part I: History" in *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, John Unsworth. Oxford: Blackwell, 2004. http://www.digitalhumanities.org/companion/

[2] "Digital humanities" in *Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/wiki/Digital_humanities

[3] *Journal of the Text Encoding Initiative* https://jtei.revues.org/

18

Homework

Search on Internet for projects which use elements of Digital Humanities. Compose a list of 3 relevant projects and describe them.

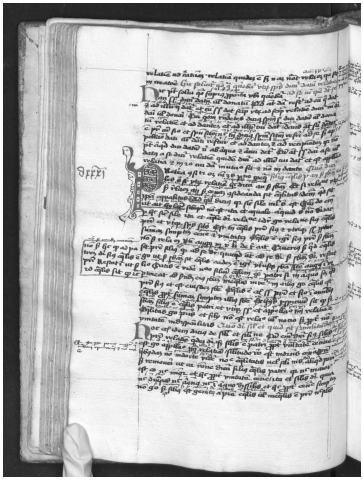
Model for description:

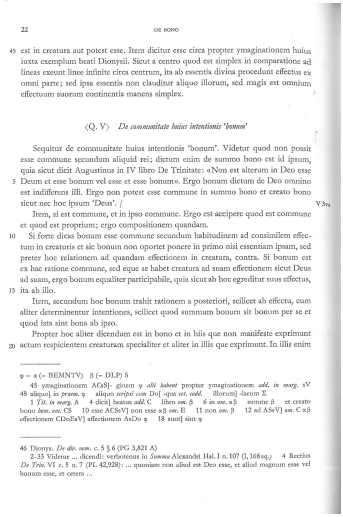
URL: http://...
Title: the title of the project
Coordinator: which institution or person develops the project?
Goal: what does the project intend to do?
Contents: which sorts of contents does the project involve?
Stage: at what stage is the project now?

2. Semantic encoding

2.1. Visual representation vs. semantic representation

The signification of a text is conveyed through:
- the text itself (characters, words)
- the meta-textual elements (format, position etc.)





22 DE BONO

45 est in creatura aut potest esse. Item dicitur esse circa propter ymaginationem huius iuxta exemplum beati Dionysii. Sicut a centro quod est simplex in comparatione ad lineas exeunt linee infinite circa centrum, ita ab essentia divina procedunt effectus ex omni parte; sed ipsa essentia non cluditur aliquo illorum, sed magis est omnium effectuum suorum continentia manens simplex.

⟨Q. V⟩ *De communitate huius intentionis 'bonum'*

Sequitur de communitate huius intentionis 'bonum'. Videtur quod non possit esse commune secundum aliquid rei; dictum enim de summo bono est id ipsum, quia sicut dicit Augustinus in IV libro De Trinitate: «Non est alterum in Deo esse Deum et esse bonum vel esse et esse bonum». Ergo bonum dictum de Deo omnino est indifferens illi. Ergo non potest esse commune in summo bono et creato bono sicut nec hoc ipsum 'Deus'. / V 37a

Item, si est commune, et in ipso commune. Ergo est accipere quod est commune et quod est proprium; ergo compositionem quandam.

Si forte dicas bonum esse commune secundum habitudinem ad consimilem effectum in creaturis et sic bonum non oportet ponere in primo nisi essentiam ipsam, sed preter hoc relationem ad quandam effectionem in creatura, contra. Si bonum est ex hac ratione commune, sed eque se habet creatura ad suam effectionem sicut Deus ad suam, ergo bonum equaliter participabile, quia sicut ab hoc egreditur suus effectus, ita ab illo.

Item, secundum hoc bonum trahit rationem a posteriori, scilicet ab effecto, cum aliter determinentur intentiones, scilicet quod summum bonum sit bonum per se et quod ista sint bona ab ipso.

Propter hoc aliter dicendum est in bono et in his que non manifeste expriment actum respicientem creaturam specialiter et aliter in illis que exprimunt. In illis enim

φ = κ (= BEMNTV) β (= DLP) S

45 ymaginationem AGßδ- ginem φ *alii habent* propter ymaginationem *add. in marg.* sV 48 aliquo] *in praem.* φ aliquo *scripsi cum* Dnj -qua *ut. vulg.* illorum]-larum S

1 *Tit. in marg.* A 4 dicit] beatus *add.* C libro *om.* ß 6 in *ante.* αδ summo β et creato bono *bon. om.* CS 10 esse ACSbV] non esse αβ *om.* E 11 non *om.* β 12 ad AbsV] *om.* C αβ effectionem CDoßaV] affectionem AsDo φ 18 sunt] sint φ

46 Dionys. *De div. nom.* c. 5 § 6 (PG 3,821 A)

2–33 Videtur ... dicendi: verbotenus in *Summa* Alexandri Hal. I n. 107 (I, 168 sq.) 4 Recitus *De Trin.* VI c. 8 n. 7 (PL 42,928): ... quoniam non aliud est Deo esse, et aliud magnum esse vel bonum esse, et ceteris ...

In visual (physical) format, the meta-textual signification is inferred from general or local rules
- sometimes the rules are explicitly indicated
  - by general known conventions

☞ the title is bigger, centered and separated from text

  - by an explicit label

☞ sections entitled *Notes*, *Index*, *Contents*

  - by explicit declaration

☞ sometimes the typographic conventions are described in the *Preamble*

- other times the conventions are unclear and they are inferred from the context

Example[1]

Ifilosofi yindlela yokucinga okanye yokuqiqa ngehlabathi, ngephakade, kwanangasekuhlaleni. Apha kwafilosofi, izimvo kuthiwa zilulwazi *oluvela noluhlalutywa ngamathamb'engqondo*, into ethetha okokuba lulwazi ngezinto [...]

The process of extracting the visual signification, although apparently simple, is a **complex** operation which implies vision, memory, experience, comprehension and synthesis; **errors** may occur in any of these functions.

- Computers, as for now, cannot coherently reproduce this process.
- Sometimes even the human readers may mistake regarding the visual semantics.

Example[2]

{TITLE:Ifilosofi} yindlela yokucinga okanye yokuqiqa {CONCEPT:ngehlabathi}, {CONCEPT:ngephakade}, {CONCEPT:kwanangasekuhlaleni}. Apha kwafilosofi, izimvo kuthiwa zilulwazi {QUOTE:oluvela noluhlalutywa ngamathamb'engqondo}, into ethetha okokuba lulwazi ngezinto [...]

Computer typesetting ⇐≠⇒ **Digital edition**

- uses a common word editor software
- produces a visual format

- uses specialized software
- produces a semantically encoded file

RTF source

```
... sed \b Auctor \b0 dicit |89b| in
\i Opera \i0 quod \i semper sit \i0,
id est\chftn {\footnote id est] inest
W}...
```

HTML source

```
... sed <b>Auctor</b> dicit |89b| in
<i>Opera</i> quod <i>semper sit</i>, id
est<a href="#fn1"><sup>1</sup></a>...
<div><a name="fn1"><sup>1</sup></a> id
est] inest W</div>
```

visual render

... sed **Auctor** dicit |89b| in *Opera* quod *semper sit*, id est[1]...

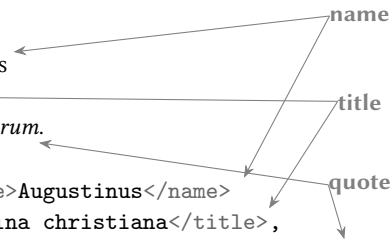
[1] id est] inest W

XML source

```
... sed <name>Auctor</name>
dicit
<cb n="89b" facs="ms1.jpg"/> in
<title>Opera</title>
quod
<quote source="#Liber3cap5">semper
sit</quote>,
<app>
<lem>id est</lem>
<rdg wit="W">inest</rdg>
</app>...
```

From visual to semantic

Ut enim egregius doctor Augustinus
ait in libro *De doctrina christiana*,
omnis doctrina vel rerum est vel signorum.




```
Ut enim egregius doctor <name>Augustinus</name>
ait in libro <title>De doctrina christiana</title>,
<quote>omnis doctrina vel rerum est vel signorum</quote>.
```

[1] Source: `http://xh.wikipedia.org` (article in Xhosa language, South Africa)
[2] Source: `http://xh.wikipedia.org` (article in Xhosa language, South Africa)

XML

```
<p xml:id="par01" lang="la">
    Ut enim egregius doctor
    <name type="person" ref="#Augustinus">Augustinus</name> ait in libro
    <title ref="#Aug-DeDoctrChrist">De doctrina christiana</title>,
    <quote source="#Aug-DeDoctrChrist-Lib1Cap2">omnis doctrina vel rerum
    est vel signorum</quote>.
</p>
```

2.2. The semantic paradigm in the digital world

6

- Every 3–4 days we produce as much information as it was produced 10 years ago during the entire year
- The difference between digital and analogic content is lost, what matters is the **relevance**

Semantic Web, Web 2.0

- emphasizes: user-generated content, usability, interoperability
- is opposed to passive viewing of content
- the digital content is semantically structured so that it can be processed by computers

7

Example: Generating an index

Classic procedure for printed book

1. procurement of the book
2. recording all the occurences of the significant words
3. ordering and verification of records (cards)
4. transcription of the records
5. typesetting the layout

it takes some hours or days

Computer aided procedure, multimedia

1. digitally copying the book
2. marking the significant occurences in the document
3. **dynamic** extraction of marked words
4. –
5. applying a style template

it takes some minutes or hours

- Transition steps:
  visual ⟶ **syntactic** ⟶ **semantic**

8

Proslogion → `<i>Proslogion</i>` → `<title ref="#Ans_Prosl">Proslogion</title>`

- Procedures for extracting the semantic content:
  - manual
    a human user selects and marks the semantic content

example: Wikipedia

- semi-automatic (assisted)
  the human user decides the semantic content and then uses an automatic markup system

example: anti-spam filters for email

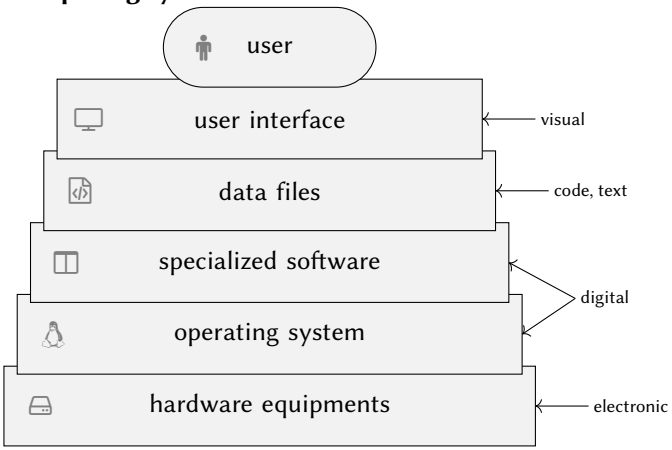
- automatic
  the computer uses statistical data and algorithms to markup the semantic content

example: Google Translate

From semantic representation to visual one can pass easily using automatic formatting based on style sheets.

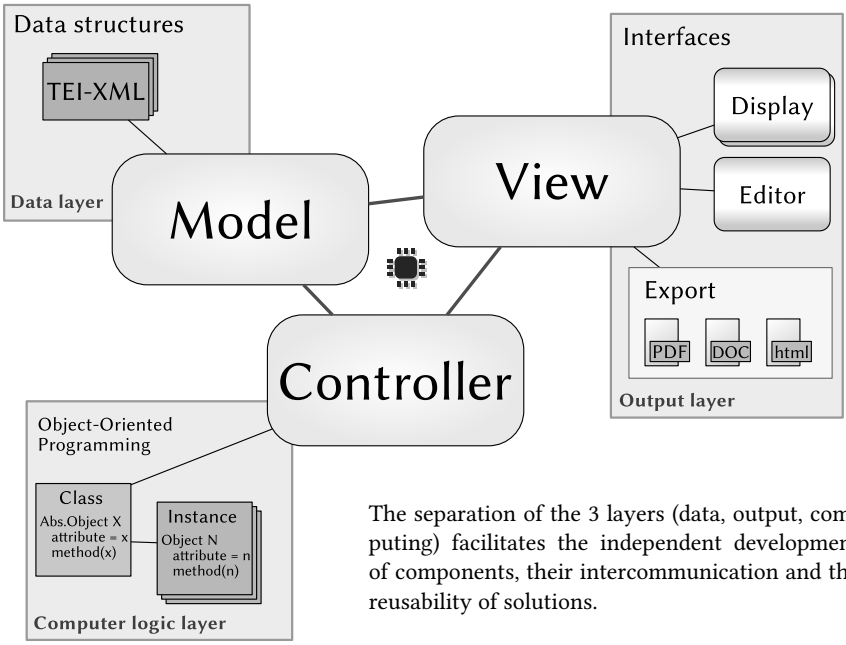
9

Computing systems architecture



10

Reusable design, abstract structures, interfaces



The separation of the 3 layers (data, output, computing) facilitates the independent development of components, their intercommunication and the reusability of solutions.

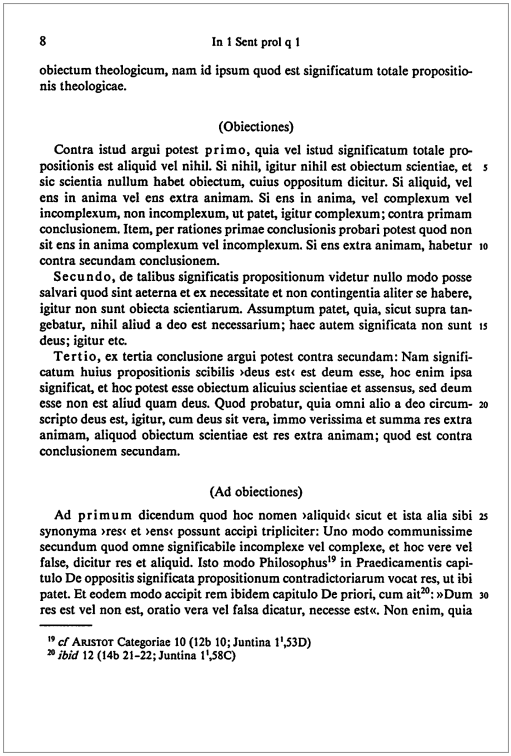
11

2.3. Utility of semantic encoding

1. Computer data processing
   - any document generated according to the standards can be transmitted and processed by any software which implements the standards
   - the computational approach of the large quantity of existing texts relieves the researcher from repetitive and time consuming operations
2. Visualisation according to the user's intention
   - the same document can be displayed in multiple modes, with or without user interaction
3. Enriching and connection of metatextual information
   - any number of details can be attached to the text, the texts can be interconnected, the information can be provided by other sources

12

Homework



8 In 1 Sent prol q 1

obiectum theologicum, nam id ipsum quod est significatum totale propositionis theologicae.

(Obiectiones)

Contra istud argui potest primo, quia vel istud significatum totale propositionis est aliquid vel nihil. Si nihil, igitur nihil est obiectum scientiae, et sic scientia nullum habet obiectum, cuius oppositum dicitur. Si aliquid, vel ens in anima vel ens extra animam. Si ens in anima, vel complexum vel incomplexum, non incomplexum, ut patet, igitur complexum; contra primam conclusionem. Item, per rationes primae conclusionis probari potest quod non sit ens in anima complexum vel incomplexum. Si ens extra animam, habetur contra secundam conclusionem.

Secundo, de talibus significatis propositionum videtur nullo modo posse salvari quod sint aeterna et ex necessitate et non contingentia aliter se habere, igitur non sunt obiecta scientiarum. Assumptum patet, quia, sicut supra tangebatur, nihil aliud a deo est necessarium; haec autem significata non sunt deus; igitur etc.

Tertio, ex tertia conclusione argui potest contra secundam: Nam significatum huius propositionis scibilis ›deus est‹ est deum esse, hoc enim ipsa significat, et hoc potest esse obiectum alicuius scientiae et assensus, sed deum esse non est aliud quam deus. Quod probatur, quia omni alio a deo circumscripto deus est, igitur, cum deus sit vera, immo verissima et summa res extra animam, aliquod obiectum scientiae est res extra animam; quod est contra conclusionem secundam.

(Ad obiectiones)

Ad primum dicendum quod hoc nomen ›aliquid‹ sicut et ista alia sibi synonyma ›res‹ et ›ens‹ possunt accipi tripliciter: Uno modo communissime secundum quod omne significabile incomplexe vel complexe, et hoc vere vel false, dicitur res et aliquid. Isto modo Philosophus[19] in Praedicamentis capitulo De oppositis significata propositionum contradictoriarum vocat res, ut ibi patet. Et eodem modo accipit rem ibidem capitulo De priori, cum ait[20]: »Dum res est vel non est, oratio vera vel falsa dicatur, necesse est«. Non enim, quia

[19] *cf* Aristot Categoriae 10 (12b 10; Juntina 1¹,53D)
[20] *ibid* 12 (14b 21–22; Juntina 1¹,58C)

Identify all the non-textual meaning carrier elements from the adjoined page (signs, formatting, position). Make a legend of these elements.

(G. Ariminensis *Lectura super ... Sententiarum*, Vol. 1, de Gruyter, 1981, p. 8)

3. Critical editing

3.1. Methods of approach for critical editions

- Best manuscript method
- Eclectic method
  - *Lectio difficilior*
- Stemmatic (lachmannian) method
  - *Stemma codicum*
- Unoriented (material) method

3.1.1. Best manuscript method

- the oldest and the simplest method
- a manuscript, considered to be the best, is chosen and transcribed
  - how is it chosen?
    - the oldest
    - the most complete
    - the easiest to read, etc.

  ☞ sometimes the reference manuscript is changed from a section to another
- useful method when among copies one of particular relevance is identified, e.g. an autograph

3.1.2. Eclectic method

- the editor's goal is to obtain a final text that is as close as possible to the author's intention and as intelligible as possible
- the editor can set a scientific or semi-scientific methodology, but some options will be subjective, authorial
  - the editor assumes the task of interpreting the material evidence for the reader
- Eclectic procedures:

  recensio sorting and collating the manuscripts
  examinatio attempt to establish the earliest version of the text
  emendatio correcting the text (sometimes called divinatio)
    - the editor's interventions can lack transparency and can corrupt the text; the intelligibility norms can be subjective
- due to the complex procedures of text restoration, the editorial decisions remain obscure

Lectio difficilior potior

= the more difficult reading is the stronger

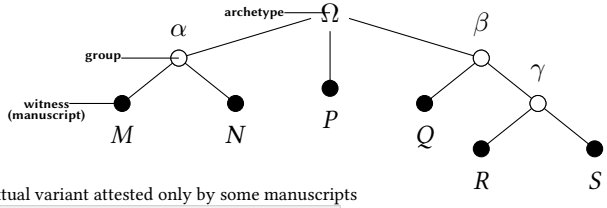
- it may be considered an eclectic method
- used in the 15th–18th c. for editing the *Bible* and other sources, as an objective criterion for selecting textual variants
- **Rule**: there where the manuscripts differ, the most difficult reading is chosen
  - **Assertion**: when copyists don't understand a difficult word or fragment, they replace it with a simpler one, committing a mistake

Ex: Toletanus | Tolemeus | toleramus ; supervincentis | supervenientis | super intendentis

Mt 6,9: πάτερ ἡμῶν, ὁ ἐν τοῖς οὐρανοῖς, ἁγιασθέτω τὸ ὄνομα σου ...
Lc 11,2: πάτερ, ἁγιασθέτω τὸ ὄνομα σου ...

3.1.3. Stemmatic (lachmannian) method

- the reconstruction by philological methods of the derivation relationships of the manuscripts from an archetype
  archetype = virtual model of all the manuscript copies, possibly but not necessary the original text
- by evaluating the common textual differences of the manuscripts (opposed readings, then common errors), a **stemma codicum** is obtained



reading textual variant attested only by some manuscripts
Ex: autem | enim ; ergo | igitur ; sive | seu ; *inv.* ...
error textual variant obviously mistaken, but attested by some manuscripts
Ex: erat | errat ; anima | alia ; sum | suum ; *om.* ...

- the editorial decisions regarding the textual variants are made then based on the relation of the manuscripts with the model

M. L. West, *Textual Criticism & Editorial Technique*, Stuttgart: B. G. Teubner, 1973

Stemma codicum

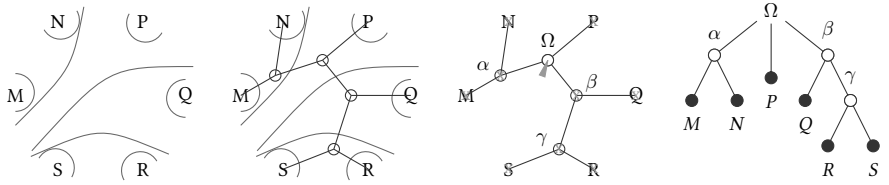
1. all the different **readings** from each manuscript and group of manuscripts is counted

M=15 ; N=11 ; P=21 ; Q=14 ; R=11 ; S=12 ; MN=17 ; RS=11 ; QRS=19

2. the sigils are disposed on a circle and the groups are isolated
3. groups are connected in a common point, obtaining the **non-oriented stemma**
4. common **errors** of the manuscripts and groups are counted in order to decide how to orient the stemma

M=4 ; N=3 ; P=9 ; Q=5 ; R=2 ; S=3 ; MN=5 ; RS=4 ; QRS=3

5. the stemma is rised in the point where there are no common errors, obtaining the **oriented stemma**



3.1.4. Unoriented (material) method

- is focused on the material object (e.g. the manuscript)
- renders exhaustively the properties of the source
- avoids the arbitrary interpretations, conjectures, emendations
  - ensures maximum scientificity and factuality
  - offers instruments for the interpretation of the source text
- can produce so-called **headless edition**

headless edition = critical edition of all the manuscript sources without establishing a unified text

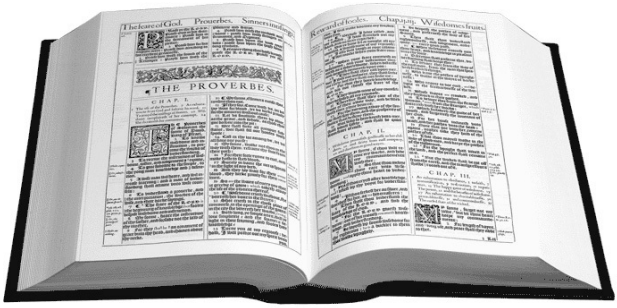
Ms. A	Ms. B	Ms. C
text and description ...	text and description ...	text and description ...

3.2. Types of critical editions

- Facsimile type edition
- Eclectic edition
- Literary (critical) edition
- Diplomatic edition
- Material edition

3.2.1. Facsimile type edition

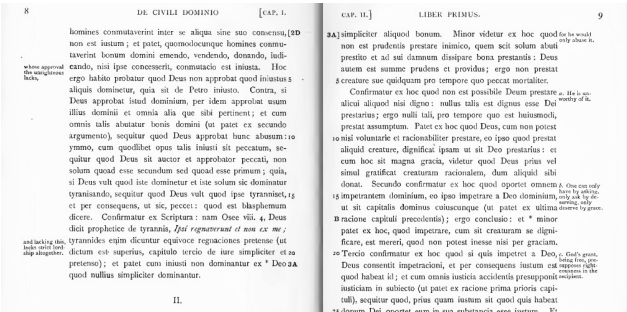
- photographic reproduction of the textual source
- contains eventually an introduction, a minimal apparatus and/or annexes (index, bibliography etc.)



1611 King James First Edition, 400th Anniversary Edition, The Bible Museum (2011)

3.2.2. Eclectic edition

- centered on the text established by the editor, usually through an eclectic method or the best manuscript method
- the critical apparatus is minimal, incomplete or inexistent
- the text fragments reconstituted by the editor are sometimes marked with conventional signs

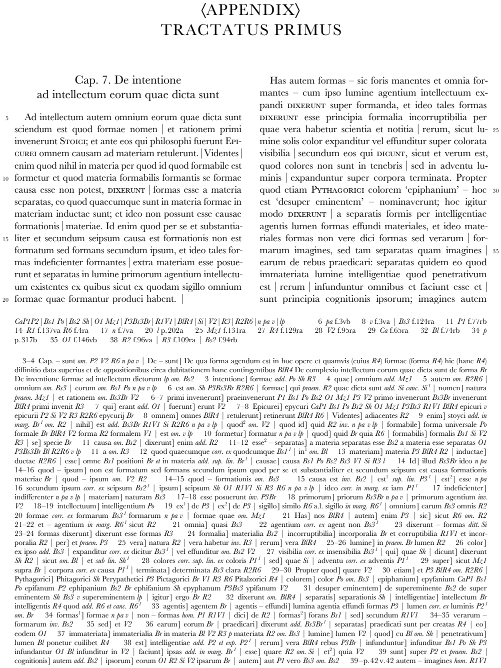


J. Wycliffe, *De civili dominio*, lib. I, ed. Reginald Lane Poole, London, Wyclif Society, 1885

3.2.3. Literary (critical) edition

- centered on the text and its tradition
- contains an extended critical apparatus, mentions the textual variants and editorial decisions

- uses a scientific method to establish the text (e.g. stemmatic m.)



⟨APPENDIX⟩
TRACTATUS PRIMUS

Cap. 7. De intentione
ad intellectum eorum quae dicta sunt

Ad intellectum autem omnium eorum quae dicta sunt sciendum est quod formae nomen | et rationem primi invenerunt Stoici; et ante eos qui philosophi fuerunt Epicurei omnem causam ad materiam retulerunt. | Videntes | enim quod nihil in materia per quod id quod formabile est formetur et quod materia formabilis formantis se formae causa esse non potest, dixerunt | formas esse a materia separatas, eo quod quaecumque sunt in materia formae in materiam inductae sunt; et ideo non possunt esse causae formationis | materiae. Id enim quod per se et substantialiter et secundum seipsum causa est formationis non est formatum sed formans secundum ipsum, et ideo tales formas indeficienter formantes | extra materiam esse posuerunt et separatas in lumine primorum agentium intellectuum existentes ex quibus sicut ex quodam sigillo omnium formae quae formantur produci habent. |

Has autem formas – sic foris manentes et omnia formantes – cum ipso lumine agentium intellectuum expandi dixerunt super formanda, et ideo tales formas dixerunt esse principia formalia incorruptibilia per quae vera habetur scientia et notitia | rerum, sicut lumine solis color expanditur vel effunditur super colorata visibilia | secundum eos qui dicunt, sicut et verum est, quod colores non sunt in tenebris | sed in adventu luminis | expanduntur super corpora terminata. Propter quod etiam Pythagorici colorem 'epiphanium' – hoc est 'desuper eminentem' – nominaverunt; hoc igitur modo dixerunt | a separatis formis per intelligentiae agentis lumen formas effundi materiales, et ideo materiales formas non vere dici formas sed verarum | formarum imagines, sed tam separatas quam imagines | earum de rebus praedicari: separatas quidem eo quod immateriata lumine intelligentiae quod penetrativum est | rerum | infunduntur omnibus et faciunt esse et | sunt principia cognitionis ipsorum; imagines autem

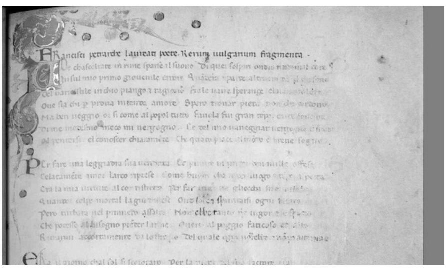
CaP1P2 | *Bs1 Po* | *Bs2 Sh* | *O1 Mz1* | *P3Bs3Br* | *R1V1* | *BlR4* | *Si* | *V2* | *R3* | *R2R6* | *n pa v* | *lp*   6 *pa* f.3vb   8 *v* f.3va | *Bs3* f.124ra   11 *P1* f.77rb 14 *R1* f.137va *R6* f.4ra   17 *n* f.7va   20 *l* p.202a   25 *Mz1* f.131ra   27 *R4* f.129ra   28 *V2* f.95ra   29 *Ca* f.65ra   32 *Bl* f.74rb   34 *p* p.317b   35 *O1* f.146vb   38 *R2* f.96va | *R3* f.109ra | *Bs2* f.94rb

3–4 Cap. – sunt *om. P2 V2 R6 n pa v* | De – sunt] De qua forma agendum est in hoc opere et quamvis (cuius *R4*) formae (forma *R4*) hic (hanc *R4*) diffinitio data superius et de oppositionibus circa dubitationem hanc contingentibus *BlR4* De complexio intellectum eorum quae dicta sunt de forma *Br* De inventione formae ad intellectum dictorum *lp om. Bs2*   3 intentione] formae *add. Po Sh R3*   4 quae] omnium *add. Mz1*   5 autem *om. R2R6* | omnium *om. Bs3* | eorum *om. Bs1 Po n pa v lp*   6 est *om. Sh P3Bs3Br R2R6* | formae] qui *praem. R2* quae dicta sunt *add. Si canc. Si*[1] | nomen] natura *praem. Mz1* | et rationem *om. Bs3Br V2*   6–7 primi invenerunt] praeinvenerunt *P1 Bs1 Po Bs2 O1 Mz1 P3 V2* primo invenerunt *Bs3Br* invenerunt *BlR4* primi invenit *R3*   7 qui] erant *add. O1* | fuerunt] erunt *V2*   7–8 Epicurei] epycuri *CaP1 Bs1 Po Bs2 Sh O1 Mz1 P3Bs3 R1V1 BlR4* epicuri *v* epicurii *P2 Si V2 R3 R2R6* epycurij *Br*   8 omnem] omnes *BlR4* | retulerunt] retinerunt *BlR4 R6* | Videntes] adiacentes *R2*   9 enim] stoyci *add. in marg. Br*[1] *om. R2* | nihil] est *add. Bs3Br R1V1 Si R2R6 n pa v lp* | quod[2] *om. V2* | quod id] quid *R2 inv. n pa v lp* | formabile] forma universale *Po* formale *Br BlR4 V2* forma *R2* formalem *V1* | est *om. v lp*   10 formetur] formatur *n pa v lp* | quod] quid *Br* quia *R6* | formabilis] formalis *Bs1 Si V2 R3* | se] specie *Br*   11 causa *om. Bs2* | dixerunt] enim *add. R2*   11–12 esse[2] – separatas] a materia separatas esse *Bs2* a materia esse separatas *O1 P3Bs3Br Bl R2R6 v lp*   11 a *om. R3*   12 quod quaecumque *corr. ex* quodcumque *Bs1*[1] | in[1] *om. Bl*   13 materiam] materia *P3 BlR4 R2* | inductae] ductae *R2R6* | esse] omne *Bs1* positioni *Br et* in materia *add. sup. lin. Br*[1] | causae] causa *Bs1 Po Bs2 Bs3 V1 Si R3 l*   14 Id] illud *Bs3Br* ideo *n pa*   14–16 quod – ipsum] non est formatum sed formans secundum ipsum quod per se et substantialiter et secundum seipsum est causa formationis materiae *Br* | quod – ipsum *om. V2 R2*   14–15 quod – formationis *om. Bs3*   15 causa est *inv. Bs2* | est[1] *sup. lin. P3*[1] | est[2]] esse *n pa*   16 secundum ipsum *corr. ex* seipsum *Bs2*[1] | ipsum] seipsum *Sh O1 R1V1 Si R3 R6 n pa v lp* | ideo *corr. in marg. ex* iam *P1*[1]   17 indeficienter] indifferenter *n pa v lp* | materiam] naturam *Bs3*   17–18 esse posuerunt *inv. P3Br*   18 primorum] priorum *Bs3Br n pa v* | primorum agentium *inv. V2*   18–19 intellectuum] intelligentium *Po*   19 ex[1]] de *P3* | ex[2]] de *P3* | sigillo] similo *R6* a.l. sigillo *in marg. R6*[1] | omnium] earum *Bs3* omnis *R2*   20 formae *corr. ex* formarum *Bs3*[1] formarum *n pa v* | formae quae *om. Mz1*   21 Has] nos *BlR4* | autem] enim *P3* | sic] sicut *R6 om. R2*   21–22 et – agentium *in marg. R6*[1] sicut *R2*   21 omnia] quasi *Bs3*   22 agentium *corr. ex* agent non *Bs3*[1]   23 dixerunt – formas *ditt. Si*   23–24 formas dixerunt] dixerunt esse formas *R3*   24 formalia] materialia *Bs2* | incorruptibilia] incorporalia *Br* et corruptibilia *R1V1* et incorporalia *R2* | per] et *praem. P3*   25 vera] natura *R2* | vera habetur *inv. R3* | rerum] vera *BlR4*   25–26 lumine] in *praem. Br* lumen *R2*   26 color] ex ipso *add. Bs3* | expanditur *corr. ex* dicitur *Bs3*[1] | vel effunditur *om. Bs2 V2*   27 visibilia *corr. ex* insensibilia *Bs3*[1] | qui] quae *Sh* | dicunt] dixerunt *Sh R2* | sicut *om. Bl* | et *sub lin. Sh*[1]   28 colores *corr. sup. lin. ex* coloris *P1*[1] | sed] quae *Si* | adventu *corr. ex* adventis *P1*[1]   29 super] sicut *Mz1* supra *Br* | corpora *corr. ex* causa *P1*[1] | terminata] determinata *Bs3* clara *R2R6*   29–30 Propter quod] quare *V2*   30 etiam] et *P3 BlR4 om. R2R6* | Pythagorici] Phitagorici *Sh* Perypathetici *P3* Pictagorici *Br V1 R3 R6* Pitalzorici *R4* | colorem] color *Po om. Bs3* | epiphanium] epyfanium *CaP1 Bs1 Po* epifanum *P2* ephipanium *Bs2 Br* ephifanium *Sh* epyphanum *P3Bs3* ypifanum *V2*   31 desuper eminentem] de supereminente *Bs2* de super eminentem *Sh Bs3 v* supereminentem *lp* | igitur] ergo *Br R2*   32 dixerunt *om. BlR4* | separatis] separationis *Sh* | intelligentiae] intellectum *Br* intelligentis *R4* quod *add. R6 et canc. R6*[1]   33 agentis] agentem *Br* | agentis – effundi] lumina agentia effundi formas *P3* | lumen *corr. ex* luminis *P2*[1] *om. Br*   34 formas[1]] formae *n pa v* | non – formas *hom. P1 R1V1* | dici] de *R2* | formas[2]] forans *Bs1* | sed] secundum *R1V1*   34–35 verarum – formarum *inv. Bs2*   35 sed] et *V2*   36 earum] eorum *Br* | praedicari] dixerunt *add. Bs3Br*[1] | separatas] praedicati sunt per creatas *R4* | eo] eodem *O1*   37 immateriata] immaterialia *Br* in materia *Bl V2 R3 p* materiata *R2 om. Bs3* | lumine] lumen *V2* | quod] cu *Bl om. Sh* | penetrativum] lumen *Bl* ponetur cuilibet *R4*   38 est] intelligentiae *add. P2 et exp. P2*[1] | rerum] vera *BlR4* rebus *P3Br* | infunduntur] infunditur *Bs1 Po Sh P3* infundantur *O1 Bl* infunditur in *V2* | faciunt] ipsas *add. in marg. Br*[1] | esse] quare *R2 om. Si* | et[2]] quia *V2*   39 sunt] super *P2* et *praem. Bs2* | cognitionis] autem *add. Bs2* | ipsorum] eorum *O1 R2 Si V2* ipsarum *Br* | autem] aut *P1* vero *Bs3 om. Bs2*   39–p.42 v.42 autem – imagines *hom. R1V1*

Albertus Magnus, *Liber sex principiorum*, p. xli, Institutum Alberti Magni Coloniense

3.2.4. Diplomatic edition

- conserves as much as possible the specific form of the text: orthography, layout
- it does not normalize the words, sometimes it does not expand the abbreviations
- contains an exhaustive philological apparatus
- relevant for legal and diplomatic documents



Petrarchive: An edition of Petrarch's songbook, *Rerum vulgarium fragmenta*. Edited by H. Wayne Storey and John A. Walsh

Petrarch's *Rerum vulgarium fragmenta*: Charta 1 recto

text view diplomatic transcription

charta 1 recto

Francisci petrarche laureati poete. Rerum vulgarium fragmenta.

1 Uoi chascoltate in rime sparse il suono   Di quei sospiri ondio nudriual core
In sul mio primo giouenile errore   Quādera ī parte altruom da ql chio sono
Del uario stile in chio piango & ragiono   Fra le uane sperançe el uan dolore
Oue sia chi p proua intenda amore   Spero trouar pieta non che perdono.
Ma ben ueggio or si come al popol tutto   Fauola fui gran tēpo. onde souente
di me medesmo meco mi uergogno.   Et del mio uaneggiar uergogna el frutto
el pentersi. el conoscer chiaramēte   Che quāto piace al mōdo e breue sogno

2 Per fare una leggiadra sua uendetta   Et punire in un di ben mille offese
Celatamēte amor larco riprese   Come huom cha nocer luogo & tēpo aspetta
Era la mia uirtute al cor ristretta   Per far [iui &] ne gliocchi sue difese

Petrarca, *Rerum vulgarium fragmenta*, 1r, http://dcl.slis.indiana.edu/petrarchive/content/c001r.xml#rvf001

3.2.5. Material edition

- centered on the conserved material form of the text
- exhaustively describes and renders all the details of the source
- sometimes it does not contain the unified, edited text, but only variants (= headless edition)
- preferable in digital format, due to the large quantity of information

Feuillet(s)	Genre	Incipit	Remarques
6	Séquence	Alle celeste nec non et perhenne luia	l.4 ajout du mot *-Christi*
12	Offertoire	Ad te Domine	V. Respice, l.8 ajout du mot *-mei*
15	Introit	Gaudete in Domino	grattage du texte du psaume
18	Offertoire	Confortamini	l.11 grattage [*—enim deus*]
23	Hymne	Benedictus es	l.5 grattage et ajustement du texte sous la mélodie
23	Hymne	Benedictus es	l.3,6 & 7 ajout *-es*, main 1 musique (cf. encre)
25	Offertoire	Exsulta satis	V. Loquetur, l.4 grattage et ajustement du texte sous la mélodie
48	Offertoire	Reges Tharsis	V. Suscipiant, l.6 rajout postérieur de la reprise *—et adorabunt*
51	Introit	Omnis terra	l.5 ajout postérieur *-terra*
53	Offertoire	Jubilate Deo universa terra	V. Locutum est l.3 grattage *—mea* (mauvaise place de la syllabe)
53	Offertoire	Jubilate Deo universa terra	V. Locutum est ajout postérieur de *<est >*(oubli)
73	Introit	Invocabit	l.8 ajout de *-me*
74	Graduel	Angelis suis	l. 5 ajout postérieur de *-ne unquam*
76	Trait	Qui habitat	V.Quoniam ang., l.2 grattage du texte
94	Trait	De necessitatibus	l.4 grattage et réécriture du texte, main postérieure
98	Offertoire	Miserere mihi	V. Tibi soli, l.6 ajout du mot *—soli*, main 3
104	Offertoire	Domine in auxilium	rubrique Co grattée (erreur de rubrique)
107	Offertoire	Illumina oculos	V.Respice, l.2 grattage de la syllabe *—di* de *—exaudi* (placement)
108	Trait	Ad te levavi	l.8 grattage du mot *—oculos*
117	Graduel	Oculi omnium	l.3 grattage du m de *—magum*

Le graduel de Bellelay, Liste des interventions, http://el.enc.sorbonne.fr/bellelay/correctionstexte.php

14

3.3. Elements of a critical edition

- Introduction
- Text
- Critical apparatus

☞ other auxiliary elements may be added: index, bibliography, diagrams, illustrations etc.

15

3.3.1. Introduction

- A critical edition contains in introduction:
  1. sources description
     - description of each manuscript
     - relation between manuscripts (stemma codicum)
  2. datation
  3. attribution
  4. doctrinal presentation
     - the contents and the position of the text in historical context
     - explicit and implicit sources
  5. description of the editing procedure
     - the used method of editing
     - abbreviations and critical apparatus

6. secondary bibliography

☞ the order and the length of introduction parts may vary; sometimes when some descriptive elements have been already published, a part may be replaced with a bibliographic reference

16

3.3.2. Text

- The edited text follows certain linguistic and graphical standards to which the author adheres:

orthography Classical Latin, Medieval Latin or uncorrected orthography
punctuation modern punctuation, classical punctuation or original source punctuation
sectioning the text is divided through titles in sections; the original sections may be kept, or a new logical sectioning can be made
formatting different character shapes (italics, small caps etc.) or tags in digital format may be used to indicate specific elements: titles, names, quotes etc.

17

3.3.3. Critical apparatus

- contains all the palaeographic, historical, philological details attached to the text by the editor
- visually represented by: signs in text, footnotes, marginal notes, endnotes
- text lines are numbered, these numbers are used as references
- Types of apparatus:

palaeographic (philological) describes the textual differences between manuscripts
sources indicates the primary sources for quotes and allusions
biblical sources sometimes a distinct apparatus for the biblical sources
tradition indicates the manuscripts based on which the text is edited
comparative when there are more recensions of the text, a secondary recension may have its own apparatus

18

The utility of semantic digital encoding

- any type of edition can be encoded in `TEI`
- the same edition can be encoded in multiple modes
- an edition can be easily passed from a standard to another
- the same edition can be visualized through various interfaces
- the display of the critical apparatus is flexible and editable
- the digital critical edition is convertible to the classical printed format

19

Homework

Choose from the library or online an edition of a medieval text. Investigate the text and the introduction trying to answer the following questions:

1. What type of edition is it?
   ☐ facsimile ☐ eclectic ☐ critical ☐ diplomatic ☐ material
   Remarks: _____
2. Which edition method was used?
   ☐ best manuscript ☐ eclectic ☐ stemmatic ☐ material
   Remarks: _____
3. Which elements of critical edition are present:
   - in the introduction: _____
   - in the text: _____
   - in the critical apparatus: _____

4. Principles of TEI-XML

4.1. The XML format

XML (eXtended Markup Language) = digital format text with metatextual markup
Only 5 types of content:

1. **tags** (contain metatextual elements)
   `<tag></tag>`
   syntax: between the signs < > and closed with /
2. **attributes** (specify parameters for tags)
   `<tag attr="val"/>`
   syntax: name (without spaces) followed by = followed by the value between quotes " "
3. **text** (text as is)
   `text`
   syntax: any characters (except <>) and which are not tags
4. **declarations** (processing commands)
   `<? decl ?>`
5. **comments** (content which is ignored on processing)
   `<!-- comm -->`

XML Example

XML

```
<text>
    This is XML text. Mark something <bold>important</bold>.
    It was written at <city country="Romania">Cluj</city>.
    <!-- here is a comment -->
    <tag>A tag may contain
        <subtag>a subtag which can also contain
            <subsubtag>sub-subtags</subsubtag>
        </subtag>
    </tag>
    A tag may have zero, <separator/> one or more attributes
        <city country="RO" department="CJ" prefix="0264">Cluj-Napoca</city>
</text>
```

Specifications

- the tags have two forms:
  - **pair**: delimit a portion of text (`<tag>text</tag>`)
    - must always be closed in reverse order of the opening (`<a><b><c>...</c></b></a>`)
    - only the opening tag may have attributes (`<tag attr="val">...</tag>`)

- **single**: as standalone element, without text (<tag/>)
  - may have attributes (<tag attr="val"/>)
- the contents between tags may have tags inside, creating a tree
  <root><branch><leaf/></branch><branch><leaf/></branch></root>
- in a well-formed XML, there must be a single root tag which contains the whole document
- extra spaces and line ends are usually ignored, but they are used for easier code editing

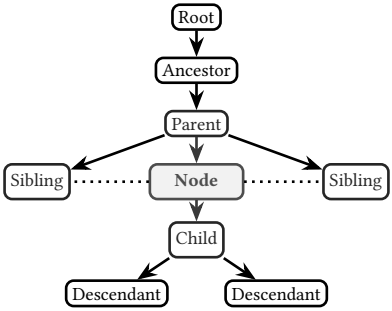
```
<document>
    Format:
        <b>bold</b>,
        <i>italic</i>.
</document>
```

→ Format: **bold**, *italic*.

4

Basic concepts in tree type structures

node	any element which is part of the structure
root	element which subordinates all the other elements of the document
parent	relation between elements in which the target element has sub-elements
children	relation in which the target elements are immediately subordinated to the parent element
siblings	relation in which the target elements are on the same level and have a common parent
descendants	relation in which the target elements are inferior to a superior element
ancestors	relation in which the target elements are hierarchically superiors and connected to a descendant element



5

4.2. About TEI

- The Text Encoding Initiative Consortium Guidelines (TEI) establish the annotation system for the documents from the humanities
- TEI uses XML as file format
  - TEI is a subset of XML instructions to which a semantic is assigned
- in TEI are specified the XML elements used for digital editions
- the root element for a TEI-XML document is <TEI> </TEI>
- a TEI document usually has two mandatory parts:
  - a preamble <teiHeader> </teiHeader>

- in the preamble the document properties are described: title, author, version, sources etc.
- the body of the text `<text> </text>`
  - contains the text with TEI tags; the main text is contained between `<body> </body>`; the text paragraphs are comprised within `<p> </p>`
- all the TEI specifications are on the website: `http://www.tei-c.org/`

6

Example of basic TEI structure

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>
```

7

Example of documentation from TEI Guidelines

<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	`<titleStmt>`
	`<title>Lagrave's Life of St. John Norbert: a machine-readable transcription</title>`
	`<respStmt>`
	`<resp>compiled by</resp>`
	`<name>J.J. Lucas</name>`
	`</respStmt>`
	`</titleStmt>`
	Show all

Content model	`<content>`
	`<sequence>`
	`<elementRef key="title" minOccurs="1" maxOccurs="unbounded"/>`
	`<classRef key="model.respLike" minOccurs="0" maxOccurs="unbounded"/>`
	`</sequence>`
	`</content>`

Schema Declaration	XML syntax 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*) }

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

8

4.2.1. Software for TEI-XML editing

- Advanced editors

they offer TEI validation, autocomplete, documentation

  - **oXygen XML Editor** (paid) 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/
- Simple editors

they offer syntax colorization, sometimes XML validation

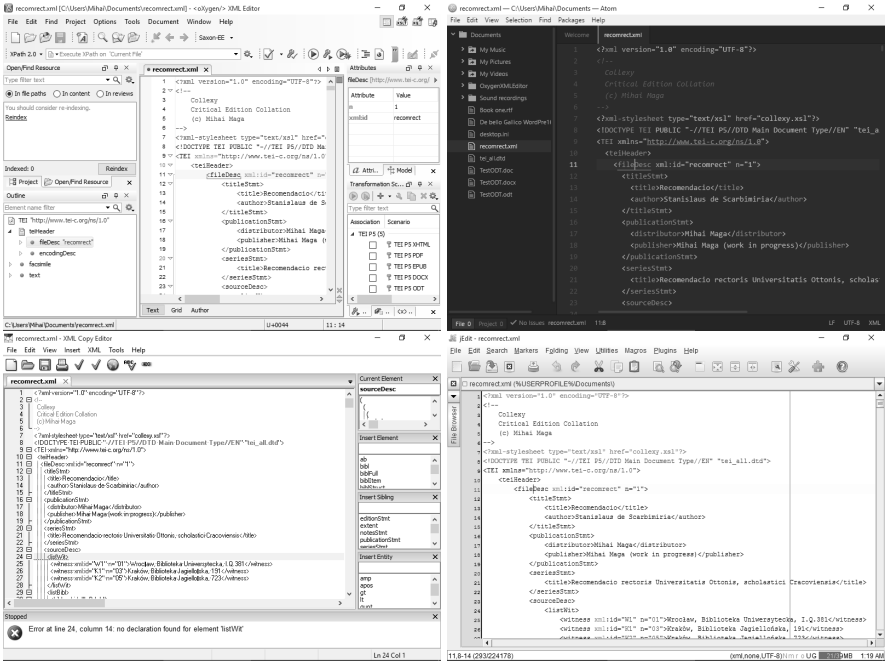
  - **Notepad++** (open source) https://notepad-plus-plus.org/
  - **Notepad2** (free) http://www.flos-freeware.ch/notepad2.html
  - **Eclipse** (open source) https://eclipse.org/
- Web editors

work in browser, they offer various functionalities

  - **CodeMirror** (open source) https://codemirror.net/
  - **eXide** (open source) http://exist-db.org/exist/apps/eXide/index.html
  - **DH Editor** (for this course) https://dhcluj.ro/dhm/lab/editor/

9

Editors (screenshots)



Homework

Using the following tags, encode the text below:

<text></text>	root element for the document
	paragraph
<title></title>	title of a work
<name></name>	name of an author
<quote></quote>	quoted text

> 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

5. Manuscript representation in TEI

5.1. Indication and description of manuscripts

- the manuscript indication is generally done through location, library, shelf no.

London, British Library, Arundel Ms. 415

- the manuscript description contains information like:
  - codicological (material, dimensions, binding, decoration)
  - historical (date, provenance, possession)
  - contents (list of contained texts, with details: title, pages, incipit/explicit, references)

415
Membranaceus, in 4to., ff. 49, sec. fortasse XIII. exeuntis; cum codice superiori compactus.
1. Versus de septem sacramentis. fol. 30.
2. Quædam de ædium partibus. fol. 31. b
   Incip. "Primo, proaulum, id est, præporticus."
3. ...

J. Forshall, *Catalogue of Manuscripts in the British Museum*, 1834, vol. I, p. 118.

Indication and description of manuscripts in TEI

- depending on the purpose, the manuscript description may be:
  - in the case of the critical edition which has the text in the body, the manuscripts are described in the preamble, inside the `<sourceDesc>` element
  - when the entire document is a list of manuscripts (a catalogue), the manuscripts are described in the body of the text, inside the `<body>` element
  - if the manuscript is mentioned only as a reference (e.g. in a preface), the description may appear inside other `TEI` elements
- the description and the indication (referencing) of the manuscripts are different things:
  - a manuscript is described only once and receives an identifier
  - the internal references (e.g. the critical apparatus) indicate the manuscript by means of the identifier

5.2. Instruction set for describing manuscripts

`<msDesc>` parent element which contains the description
`xml:id="␣"` attribute which specifies the unique identifier
`<msIdentifier>` parent element which contains the identification data
`<settlement>` locality
`<repository>` library
`<idno>` shelf no.
`<msContents>` parent element for the contents of the manuscript
`<summary>` contains the overview of the ms.
`<date notBefore="␣" notAfter="␣">` manuscript datation
`<origPlace>` manuscript origin
`<...>` other elements
`<msItem>` describes an individual part from the ms.

Instructions for describing a part of a manuscript

`<msItem>` may appear multiple times, may contain:
`<locus>` or <locusGrp> page indication
`<author>` author name
`<title>` part title
`<bibl>` or <listBibl> bibliography
`<incipit>` and `<explicit>` beginning and end of the text

Other elements for describing the manuscripts

`<head>` the manuscript rubric
`<physDesc>` physical description, with `<objectDesc>` `<supportDesc>` `<dimensions>` `<extent>` `<layoutDesc>` `<handDesc>` `<decoDesc>` `<bindingDesc>`
`<history>` historical data, with `<origin>` `<provenance>` `<acquisition>`
`<additional>` additional information, e.g. `<listBibl>`
`<msPart>` information on a distinct part of a codex, same as `<msDesc>`

Example of a manuscript description

TEI-XML

```
<msDesc xml:id="MsCj9">
   <msIdentifier>
      <settlement>Cluj-Napoca</settlement>
      <repository>Biblioteca Academiei</repository>
      <idno>Cod. lat. 9</idno>
   </msIdentifier>
   <msContents>
      <summary>
```

TEI-XML (cont)

```
        <date notBefore="1451" notAfter="1500">sec. XV, a 2a jum.</date>
        <origPlace>Europa centrala?</origPlace>
      </summary>
      <msItem n="1">
        <locus from="1r" to="64v">ff. 1r-64v</locus>
        <author>Giovanni Bocaccio</author>
        <title>De claris mulieribus</title>
        <bibl>G. Boccacio, De Mulieribus Claris (ed. Virginia Brown),
        Harvard:Harvard University Press 2001</bibl>
      </msItem>
      <msItem n="2">
        <locus from="65r" to="117v">ff. 65r-117v</locus>
        <title>Commentarius super Aristotelis libros</title>
      </msItem>
    </msContents>
  </msDesc>
```

7

Example of catalogue source and of ms. indication

CLUJ-NAPOCA
Biblioteca Academiei

Cod. lat. 9
a. Giovanni Bocaccio, *De claris mulieribus* (1r–64v) — b. *Commentarius super Aristotelis libros* (65r–117v).
Orig.: Europa centrală?; XV^2.

Adrian Papahagi, Adinel-Ciprian Dincă, în colab. cu Andreea Mârza, *Manuscrisele medievale occidentale din România: Census*, Iaşi:Polirom, 2019

TEI-XML

```
<p>
    ... Boccacio in the Cluj manuscript <ptr target="#MsCj9"/>
</p>
```

8

5.2.1. Manuscripts description for the critical apparatus

- for the critical apparatus in TEI, the witnesses are described by the elements `<listWit> <witness>`

`<listWit>` parent element for witnesses list; appears in `<sourceDesc>` or in `<body>`, `<back>`, `<abstract>`...

`<witness>` describes or indicates a witness

`xml:id="␣"` attribute which specifies the unique identifier

- the witness may be summarily described between the tags `<witness></witness>`

☞ `<witness xml:id="Cj9">Cluj, Bibl.Acad., Cod.lat.9</witness>`

- can be indicated through a link to a `msDesc` element

```
<witness xml:id="Cj9" sameAs="#MsCj9"/>
```

- or can be described completely with `msDesc`

```
<witness xml:id="Cj9"><msDesc>[...]</msDesc></witness>
```

9

5.2.2. Details on manuscripts included in text

Some details noticed in manuscripts are indicated inside the transcribed text, in the place where they are noticed.

`<pb/>` page break
`<cb/>` column break
`<lb/>` line break
`<locus/>` place that needs details

```
<p>Omnes homines natura <lb ed="#M"/> scire desiderant. Signum autem <cb ed="#M"
n="12vb"/> est sensuum dilectio; <pb ed="#N" n="42"> praeter enim et <locus target="#M"
facs="img12_002.jpg"/> utilitatem propter se ipsos diliguntur, et maxime aliorum qui est
per oculos.</p>
```

10

Homework

Encode in `TEI` the following manuscripts:

ALBA IULIA
Biblioteca Batthyaneum

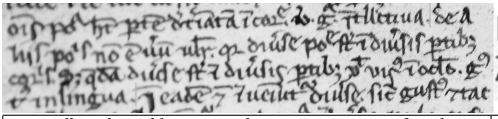
MS I.56
a. *Quaestiones (in Sententias?)* (1r–73r) — b. *Sermones quadragesimales* (74r–165v) — c. *Tabulae astronomicae cum canone* (166r–205v) — d. *Tabulae Alphonsinae* (206v–237v) — e. Ioannes Duns Scottus, *Quaestiones in Aristotelis Analytica Priora* (238r–254r) — f. Guilelmus de Hentisbury?, *Sophismata* (254r–301v).
Orig.: Cracovia?; 1383.
MS I.64
a. Bernardus Claraevallensis, *Sermones in purificatione Beatae Mariae Virginis* (1r–66v) — b. Lotharius Signinus/Innocentius papa III, *De miseria humanae conditionis* (68r–84v) — c. S. Augustinus, *Soliloquia* (84v–111v) — d. Alanus ab Insulis, *Anticlaudianus* (111v–115r) — e. Pseudo-Aristotelis, *Secretum secretorum* (115v–143r) — f. Bernardus Claraevallesis, *Meditationes* (143r–152v) — g. Gregorius papa I?, *De conflictu vitiorum et virtutum* (153r–159r) — h. *'Auctoritates diverse secundum ordinem alphabeti'* (159r–198r) — i. *Excerpta* (198v–199v).
Orig.: Cehia/Moravia/Silezia?; s. XIV/XV.

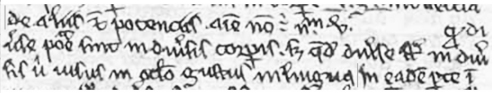
6. Representation of textual variation in TEI

6.1. Differences between manuscript copies

- The text of a work differs at the level of manuscript copies
  - sometimes the same text survives in multiple variants of redaction
  - transcribing the texts, the copyists make changes and mistakes
- The critical edition must render the text variations because
  - the edited text is reconstructed based on variants
  - the textual variations are important for the tradition

Example of differences:





A = Alba Iulia, Biblioteca Batthyaneum, Ms. I.143, f. 30rb	**P** = Praha, Národní knihovna České republiky, IV.D.13, f. 23rb
de a-| liis · potentiis · non est verum universaliter, quia diversæ potentiæ sunt in diversis partibus | corporis. Sed quædam diversæ sunt in diversis partibus, ut visus in oculo, gus-| tus in lingua;	de aliis etiam potentiis animæ non est nullum v. [*lac.*], quod di-| versæ potentiæ sunt in diversis · corporis. Sed quædam diversæ sunt in diver-| sis · , ubi visus in oculo, gustus in lingua;

Segmentation and alignment of variants

A	de	aliis	·	potentiis	·	non	est	verum	universaliter
P	de	aliis	etiam	potentiis	animae	non	est	nullum	v. + *lac.*

A	quia	diversae	potentiae	sunt	in	diversis	partibus	corporis
P	quod	diversae	potentiae	sunt	in	diversis	·	corporis

A	Sed	quaedam	diversae	sunt	in	diversis	partibus	ut	visus
P	Sed	quaedam	diversae	sunt	in	diversis	·	ubi	visus

A	in	oculo	gustus	in	lingua
P	in	oculo	gustus	in	lingua

Classical (printed) critical apparatus

1 de aliis etiam potentiis animae non est verum universaliter, quia diversae potentiae sunt in diversis partibus corporis. Sed quaedam
5 diversae sunt in diversis partibus, ut visus in oculo, gustus in lingua;

1 etiam] *om. A* || animae] *om. A* || **2** verum universaliter] nullum v. *P*; *lac. add. P* || quia] quod *P* || **4** partibus] *om. P* || **5** partibus] *om. P* || **6** ut] ubi *P*

6.2. Describing textual variations in TEI

TEI elements for critical apparatus encoding

`<app>`	parent element for entries in the critical apparatus
`<rdg>`	contains a single reading within a textual variation
`<lem>`	contains the lemma (or base text, accepted text)
`<witDetail>`	gives further details for the manuscript/-s

The elements `<rdg>`, `<lem>` and `<witDetail>` must have the attribute `wit="␣"` which specifies the identifier for the manuscript/-s attesting the variant preceded by `#`. Multiple identifiers are separated by space (`wit="#A #B #C"`).The identifier is specified in the witness list (the attribute `xml:id` from `<witness>`). To mark omission, a single tag may be used (`<rdg wit="␣"/>`)

- Other elements for the critical apparatus:

`<supplied>`	text supplied by the editor	⟨…⟩
`<unclear>`	cannot be clearly transcribed	*dub.*
`<damage>`	damaged ms. zone	†
`<del>`	text marked as deleted in ms.	*del.*, *exp.*, *eras.*
`<gap>`	portion which was not transcribed for various reasons	

TEI Example

TEI-XML

```
<p>
   de aliis
   <app>
      <lem wit="#P">etiam</lem>
      <rdg wit="#A"/>
   </app>
   potentiis
   <app>
      <lem wit="#P">animae</lem>
      <rdg wit="#A"/>
   </app>
   non est
   <app>
      <lem wit="#A">verum universaliter</lem>
      <rdg wit="#P">nullum <unclear>v.</unclear></rdg>
      <witDetail wit="#P" type="lacuna">lac. add.</witDetail>
   </app>,
   <app>
      <lem wit="#A">quia</lem>
      <rdg wit="#P">quod</rdg>
   </app>
   diversae potentiae sunt in diversis
   <app><lem wit="#A">partibus</lem><rdg wit="#P"/></app> corporis.
</p>
```

TEI-XML (cont)

6.3. Encoding methods

- **location-referenced** method

  the apparatus is stored separately (or within the text) and is linked to the text with the `loc` attribute

```
<p n="parI">de aliis etiam potentiis ...</p>
...
<app loc="parI 3"><lem wit="#P">etiam</lem><rdg wit="#A"/></app>
```

- **double-end-point**-attached method

  the apparatus is linked to the text through anchors and through the `from` and `to` attributes

```
<p>... non est <anchor xml:id="i1"/>verum universaliter<anchor xml:id="i2"/>, quia
...</p>
...
<app from="#i1" to="#i2"><lem wit="#A">verum universaliter</lem><rdg wit="#P">nullum
v.</rdg></app>
```

- **parallel-segmentation** method

  the text is divided into segments, each with its apparatus

```
<p>... potentiis <app><lem wit="#P">animae</lem><rdg wit="#A"/></app> non est ...</p>
```

6.4. Statistics and counters

Having the critical apparatus semantically encoded, it is to extract automatically the textual variations of each manuscript.

- the variations can be processed statistically for global appreciation of their type and number
- the list of variations is used to make up the *stemma codicum*
- the text of a witness can be automatically reconstructed

Homework

Encode in `TEI` the differences between the following textual variants:

Accepted text `xml:id="A"`	Textual variant `xml:id="V"`
Omnis doctrina est de rebus vel de signis. Veteris ac novae legis continentiam diligenti indagine etiam atque etiam considerantibus nobis, praevia Dei gratia innotuit sacrae paginae tractatum circa res vel signa praecipue versari. Ut enim egregius doctor Augustinus ait in libro De doctrina Christiana, “omnis doctrina vel rerum est vel signorum”. Sed res etiam per signa discuntur.	Quamvis doctrina est de rebus vel de signis vel etc. Veteris ac novae legis continentiam diligenti indagine etiam considerantibus nobis praevia Dei gratia innotuit circa res vel signa praecipue versari etc. Ut enim auctor eximius Augustinus ait in libro suo De doctrina christiana, “omnis doctrina vel rerum” etc. Sed res etiam per signa dicetur.

7. Representation of source apparatus in TEI

7.1. Ontologies and trees

tree abstract data structure that simulates a hierarchical arborescence (see in *Course 3*: **root**, **parent**, **children**)

- can be conceived as a tree: a text (chapters, subchapters...), a bibliography (author, work...), a text corpus etc.

ontology (in information science) definition of types, properties and relations which exhaust the entities of a domain

- the definition of all classes, attributes and relations of the objects from a text corpus represents an ontology or a model

class = typology to which the object belongs
attribute = type of propriety of the object
relation = rapport with other object types

```
<rdg cause="homeoteleuton" wit="#MsAA01"/>
class attribute                relation
```

7.2. Describing the elements of a textual reference

The sources and textual references are described as bibliographic elements.

`<listBibl>` parent element for a bibliographic list
(may appear in header in `<sourceDesc>` or after the text body in `<back>`)
`<bibl>` parent element for a bibliographic indication
`xml:id="␣"` unique identifier
`<author>` author of the source (simple or structured with **<name>**...)
`<title>` title of the source (simple or structured)
`<publisher>` publishing house
`<editor>` name of the volume editor
`<pubPlace>` publishing place
`<date>` publishing date
`<biblScope>` reference scope (e.g. pages, chapter)
`<...>` other properties (see `TEI Guidelines`)

Example of bibliography in TEI

TEI-XML

```
<listBibl>
   <bibl xml:id="bibl001">
      <author>Aristoteles</author>
      <title>Metaphysica</title>
```

TEI-XML (cont)

```
   </bibl>
   <bibl xml:id="bibl002">
      <author>
         <name>Stephen Ramsay</name>,
         <name>Geoffrey Rockwell</name>
      </author>
      <title level="a">Developing Things: Notes toward an Epistemology of
      Building in the Digital Humanities</title>
      <editor>
         <name>Matthew K. Gold</name>
      </editor>
      <title level="m">Debates in the Digital Humanities</title>
      <publisher>University of Minesotta Press</publisher>
      <pubPlace>Minneapolis,
         <country>USA</country>
      </pubPlace>
      <date when="2012">2012</date>
      <biblScope unit="pp" from="75" to="84">75-84</biblScope>
   </bibl>
</listBibl>
```

7.3. Other types of references

prosopography person list, optionally with biographic data
- uses the elements `<listPerson>`, `<person>` etc.

geography geographic locations list
- uses the elements `<listPlace>`, `<place>`, `<geoDecl>`, `<geo>` etc.

chronology historical events list
- uses the elements `<listEvent>`, `<event>` etc.

The references may appear in the text or, more often, they are separately edited and referenced based on the unique identifier.

Example of references in TEI

TEI-XML

```
<TEI>
  <teiHeader><!-- ... --></teiHeader>
  <text>
    <body>
      <p>
        <name ref="#Aug">Augustinus</name> scribit <title ref="#DeCiv">De civitate</title>
        postquam <ref target="#a410" type="event"><name ref="#Alar">Alaricus</name>
        ceperit <place ref="#Rom">Romam</place></ref>.
      </p>
    </body>
    <back>
      <listPerson>
        <person xml:id="Aug">
          <persName>Augustinus</persName> <addName>Hipponensis</addName>
        </person>
        <person xml:id="Alar">
          <persName>Alaricus</persName> <genName>I</genName> <roleName>rex Visigothorum</roleName>
```

TEI-XML (cont)

```
        </person>
      </listPerson>
      <listPlace>
        <place xml:id="Rom"><placeName>Roma</placeName> <geo>41.90 12.45</geo></place>
      </listPlace>
      <listEvent>
        <event xml:id="a410" when="410"><label>Anno 410 Roma ab Visigothis capta est</label></event>
      </listEvent>
      <listBibl>
        <bibl xml:id="DeCiv">
          <author ref="#Aug">Augustinus</author> <title>De civitate Dei</title>
        </bibl>
      </listBibl>
    </back>
  </text>
</TEI>
```

6

7.4. Attaching the references

- The references are attached with the aid of the unique identifier (`xml:id="␣"`).
- The call to the unique identifier is a local identifier (prefixed with `#`) or an external `URL` (partial or full).
- The unique identifier may contain letters, digits (but not in the first position), some special characters (`.-_`).

```
<span xml:id="id_01"></span>      <ptr target="#id_01"/>

                          <ptr target="sources.xml#id_01"/>

                    <ptr target="http://example.com/sources.xml#id_01"/>
```

7

Homework

Identify the sources from the following text and encode it in `TEI` with a bibliographical list and links:

> 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

8. Visualisation of digital editions

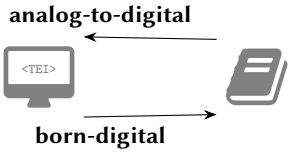
8.1. Conversion to classic format

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

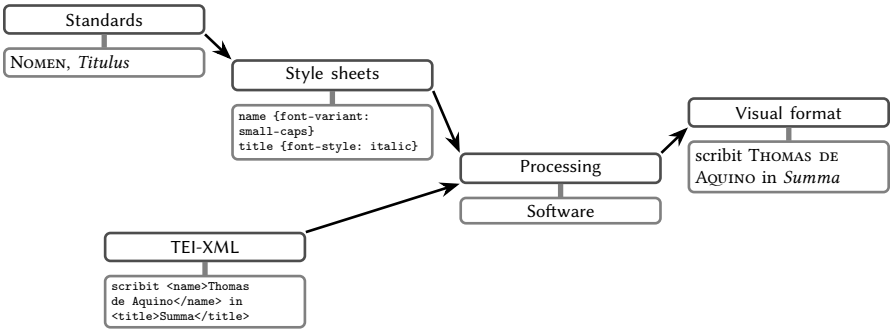
born-digital editions elaborated from the beginning in semantic digital format

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.



Conversion to classic format (2)

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.



Conversion methods

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. PDF)
   - ❶ requires a specialized application and/or advanced knowledge in editing/programming

4

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

5

CSS syntax

- 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)

6

Print vs. screen

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

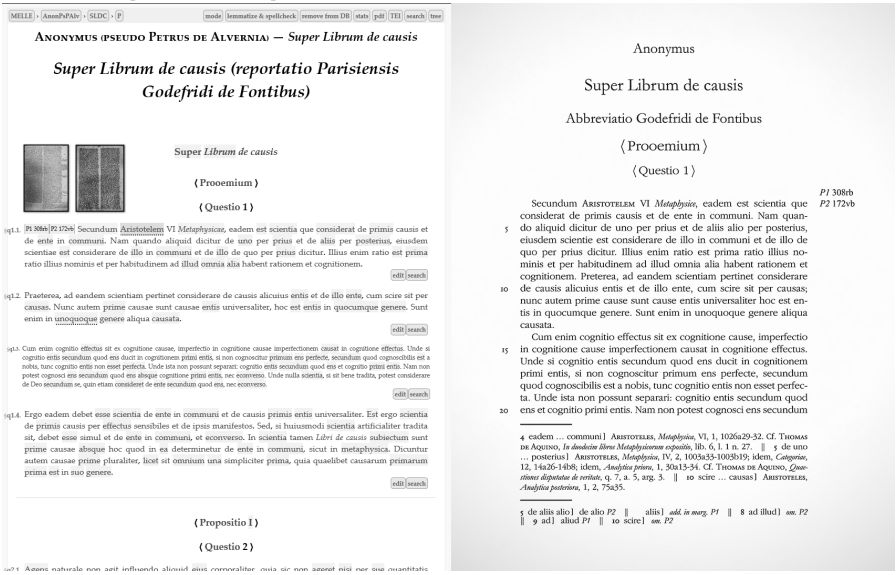
❗ limits: the lines and pages must be broken; the space is limited (the number of pages is finite, large empty spaces are to be avoided); the repetition of structure elements is avoided using references; the usage of colors and images is restricted

- **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

7

Screen vs. print (example)



8.2. Interactive interfaces

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)

8.3. 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>
```

10

Facsimile publication standards

- `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.

11

8.4. Pitfalls of the visual

- 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.*

Homework

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; }
```

9. Integration and digital processing

9.1. Indices and concordance tables

The extraction of significant elements (e.g. titles, names, quotes, terms) can be done automatically and immediately to the extant that:

- elements are marked with tags and/or attributes
- a procedure of query, extraction and display of elements is used

9.2. Query languages: XQuery, XPath

- `XQuery` (XML Query) is a functional language for querying collections of data in `XML` format
- `XPath` (XML Path Language) is subset of `XQuery`, defined as a query language for selecting nodes from `XML`

XPath Example (1)

If the following XML code is given:

```
<text>
   <body>
     <p>...</p>
   </body>
</text>
```

the expression to select all <p> nodes from <body> is:

```
/tei:text/tei:body/tei:p
```

XPath Example (2)

To extract all the elements with the attribute `class="sic"` from the next XML:

```
<p><w class="sic">unus</w> <w>duo</w> <w class="non">tres</w>
<w class="sic">quattuor</w></p>
```

the XPath expression can be:

```
//tei:w[@class='sic']
```

9.3. Search in text

Text search can be one of many types:

- **simple search**: the given character string is searched

sum ⇒ ego **sum**, **sum**ma, as**sum**ptio, sen**sum**

- **wildcard search**: the words which match the given character string and any characters in place of wildcard are searched

sum* ⇒ ego **sum**, **sum**ma, ~~**assumptio**~~, ~~**sensum**~~

- **structured search**: the given character string is searched in some given structure elements

e.g. XPath: `//tei:title[contains(., 'Sum')]`
⇒ `<title>Summa</title>`, ~~`<place>Sumer</place>`~~

9.4. Lemmatization, normalization, dictionaries

The search by character string is not always satisfying, since the terms may appear in different forms because of the grammar inflexion and of the changing orthographic rules.

res, rei ⇒ **res**; **rei**; **rem**; **re**; **rerum**; **rebus**; ~~**rex**~~; ~~**ires**~~; ~~**reddo**~~; ~~**resto**~~; ~~**recitares**~~

Disambiguation techniques (in order of complexity):

- **normalization**: bringing the text to a standard form
- **stemmatisation**: extraction of the word stem (root)

re-s; **ir**-es; **rest**-o; **recit**-ares ...

- **lemmatisation**: attaching the standard form of the word to each occurence

rerum ~ res,rei; **ires** ~ eo,ire; **recitares** ~ recito,-are ...

TEI Example

```
<p><w lemma="omnis,-is,-e">Omnes</w> <w lemma="homo,-inis">homines</w>
<w lemma="naturaliter">naturaliter</w> <w lemma="scio,-re">scire</w> <w
lemma="desidero,-are">desiderant</w>.</p>
```

- **dictionary**: linking the words to an internal or external dictionary
- **analysis**: attaching the whole grammar analysis to a word

9.5. Digital corpora

- A **text corpus** is a large structured collection of texts selected by a certain criterion (subject, period, author, intellectual milieu etc.)
- A **digital corpus** is a big dataset (text and metadata) which comprises a text corpus with annotations and which covers (at least at concept level) a certain principal criterion.
  - due to the intention of completeness, the corpus is tree-like built as an ontology (see *Course 7*)
  - a digital corpus is organized by scientific and functional principles which must ensure the reliability of its exploitation, including:
    - access to standardized versions of every text
    - simple and structured searches in the whole collection
    - exhaustive statistical analysis
    - automatic content analysis (e.g. linguistic analysis, data validation)
    - hypotheses testing, ensuring a high degree of confidence
    - transparency and predictability of the structures

The design of a corpus is a complex and important task, because it must cover all the possible manifestations of the contents, often without complete access to the contents.

9.6. Data-mining

- **Data mining** is the process of computer analyzing big datasets to identify meaningful patterns which help understanding and interpretation of contents.
  - the process is extremely complex, involving database systems, statistics, semantics, artificial intelligence and intellectual interpretation
  - usage scenarios:
    - validation of work hypotheses

    proving or disproving the existence of a pattern (e.g. influence of a certain factor)
    - discovery of significant models

    discovery of previously unknown regularities (e.g. a major but ignored factor)
    - analysis of pattern predictability

    using verified patterns, the content affiliation can be proved (e.g. authorship)

Homework

Write a `XPath` expression which extracts from the following `TEI` fragment all the occurrences of the verb *sum, esse, fui.*

Aristoteles, *Perihermeneias*

```
<p><w lemma="amplus">Amplius</w>, <w lemma="si">si</w> <w lemma="sum">est</w> <w
lemma="albus">album</w> <w lemma="nunc">nunc</w>, <w lemma="verbum">verum</w> <w
lemma="sum">erat</w> <w lemma="dico">dicere</w> <w lemma="primus">primo</w> <w
lemma="quoniam">quoniam</w> <w lemma="sum">erit</w> <w lemma="albus">album</w>, <w
lemma="quare">quare</w> <w lemma="semper">semper</w> <w lemma="verbum">verum</w> <w
lemma="sum">fuit</w> <w lemma="dico">dicere</w> <w lemma="quilibet">quodlibet</w>
<w lemma="ille">illud</w> <w lemma="is">eorum</w> <w lemma="qui">quae</w> <w
lemma="factus">facta</w> <w lemma="sum">sunt</w> <w lemma="quoniam">quoniam</w>
<w lemma="sum">erit</w>; <w lemma="quod">quod</w> <w lemma="si">si</w> <w
lemma="semper">semper</w> <w lemma="verus">verum</w> <w lemma="sum">est</w> <w
lemma="dico">dicere</w> <w lemma="quoniam">quoniam</w> <w lemma="sum">est</w>
<w lemma="vel">vel</w> <w lemma="sum">erit</w>, <w lemma="non">non</w> <w
lemma="possum">potest</w> <w lemma="hic">hoc</w> <w lemma="non">non</w>
<w lemma="sum">esse</w> <w lemma="nec">nec</w> <w lemma="non">non</w> <w
lemma="futurus">futurum</w> <w lemma="sum">esse</w>.</p>
```

10. Artificial Intelligence

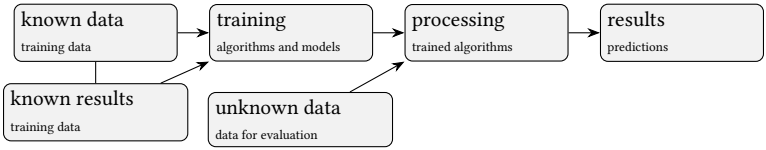
10.1. Artificial Intelligence, Machine Learning

Artificial Intelligence (AI**)** = computer systems which are capable to solve problems by mimicking "cognitive" functions of the human mind.

Machine Learning (ML**)** = part of AI which studies the application of **algorithms** and **statistical models** without precise instructions, but based on **patterns** and **inferences**, through which the system makes **decisions** and emits **predictions**.

Computer process
Instruction-based programming (classical programming):

ML-based programming:



ML: Blackbox model:

10.2. Algorithms and models

- **Model** = a description of a system of data based on statistical assumptions.
- **Algorithm** = a finite sequence of instructions that performs a given task.
- In computer processing, a **data set** is defined by a **model** and the corresponding **data content**. Data is processed by **algorithms** considering the **model**.
- In ML, the **algorithms** are generated by the computer system through inferences and statistical analyses of **existing data** (= **training**), then they can be applied to **new data** which correspond to the same **model** (= **prediction**)

10.3. Training and prediction in ML

The training

- Required components:
  - *Example:* Image recognition
  - **the data model**
    - it must correctly and completely describe the data
    - *Ex:* Data structure description in an image file format
  - **the training data set**
    - it must be statistically representative and comprise many entries
    - *Ex:* Tens of thousands of quality images and as varied as possible
  - **the valid results for the training data**
    - they must be correct and cover any detail expected in the results
    - *Ex:* Labels for each image
  - **the adequate software and hardware system**
    - the quality must be verified, system requirements are very high
    - *Ex:* Software that supports image processing, computing power (cloud?)

The prediction

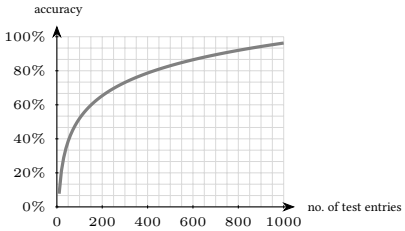
- it will return a result set corresponding to new data, together with an indication of the accuracy of predictions

Ex: The most probable labels for the new images, and the accuracy percentage for each association

!!! The result is not cathegoric, neither necessary correct, but it is the state which the system reaches by aplying the trained algorithms from the model

Accuracy of results

- The accuracy grows logarithmic as more training data is entered:



- It is impossible to reach an accuracy of 100% (that is, absolute certitude), but most often an accuracy of 90%–99% is considered acceptable, depending on the goal.
- A large part of the effort is involved in preparing the data.

10.4. Types of approach in ML

Depending on the types of data and on the types of expected results, there are different approaches of machine learning and of data models.

- The artificial intelligence is extremely specialized and requires the correct choice of the model and learning types in order to produce satisfying results.
- AI is an extremely dynamic domain today, and the approaches frequently change, therefore an exhaustive presentation is not possible.

8

Types of machine learning

- **Supervised learning**
  - training on a known model
- **Unsupervised learning**
  - detecting the structure of an unknown data set
- **Semi-supervised learning**
  - training with a known data set, then improvement with unknown data
- **Reinforcement learning**
  - based on reward
- **Anomaly detection**
  - identification of rare items
- **Association rules**
  - identification of rules in an unknown data set

9

Types of models

- **Artificial neural networks**
  - model based on a collection of processing nodes ("artificial neurons") which are interconnected ("synapses")
- **Decision trees**
  - a predictive model which starts from observing an element to generate conclusions regarding the element value depending on its position in the tree
- **Support vector machines**
  - supervised learning methods which predict if an element falls in one of 2 possible categories
- **Bayesian networks**
  - graphic model which represents a set of variables and their associated features
- **Regression analysis**
  - statistical methods of estimating the relationship between the input variables and their associated features
- **Genetic algorithms**
  - search algorithms which simulate the natural selection to generate new genotypes
- **Foundation models (base models)**
  - very large AI models, trained on a vast quantity of data and adaptable to a multitude of tasks, e.g. chat

10

End of the course