

Digital Humanities

DIGITAL HUMANITIES FOR MEDIEVAL PHILOSOPHICAL SOURCES

2. Semantic encoding

conf. dr. Mihai MAGA

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

2nd semester, 2023–2024
HME2415/02

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

1 Visual representation vs. semantic representation

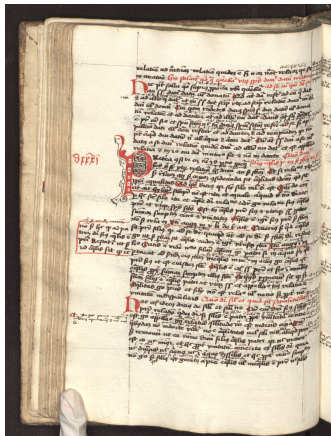
2 The semantic paradigm in the digital world

3 Utility of semantic encoding

 Homework17

The signification of a text is conveyed through:

- the text itself (characters, words)
- the meta-textual elements (format, position etc.)



22

118 ADON

43 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 excurrunt lineae infinite circa centrum, ita ab essentia divina procedunt effectus ex omni parte; sed ipsa essentia non clauditur aliquo illorum, sed magis est omnium effectuum suorum continentia inmens 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 'Deum'. /

V2n

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

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

32 ita ab illo.

Item, secundum hoc bonum trahit rationem a posteriori, scilicet ab effectu, cum alter determinatur intentionem, 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 exprimentur necum respicientem creaturam specialiter et aliter in illis que exprimentur. In illis enim

q - a (- BEBNIV) β (- DLP) S

45 ymaginationem ACaS: gnam q ali habent propter ymaginationem add. in marg. vV
48 aliquo in prae. q aliquo xpi cuse DeJ - qm est. add. libent] dierm Σ
1 Tr. in marg. A 4 alio] beatus add. C ltho m β 6 in soc. a β summa β et creato bono bon. om. CS 10 soc. ACaS] non esse a β m. 11 non om. β 12 ad ASV] om. C a β effectum CDoS[V] effectum AnDo q 18 sum] sint q

46 Dierro. De div. om. c. 5 § 6 (PG 3821 A)

2-35 Videtur ... diceret: veritates in Summa Alexandri Hal. I. n. 107 (l. 148 sq.) 4 Rectius De Tric. VI c. 5 n. 7 (PL 42,928): ... quoniam non aliud est Deo esse, et aliud maximum esse vel bonum esse, et creatura ...

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

Example¹

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

¹Source: <http://xh.wikipedia.org> (article in Xhosa language, South Africa)

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.

➡ We need a clear and explicit representation system.

Example²

{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 [...]

²Source: <http://xh.wikipedia.org> (article in Xhosa language, South Africa)

Computer typesetting $\Leftrightarrow \neq \Rightarrow$

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

Digital edition

- uses specialized software
- produces a semantically encoded file

RTF source	visual render	XML source
<pre>... sed \b Auctor \b0 dicit 89b in \i Opera \i0 quod \i semper sit \i0, id est\chftn {\footnote id est} inest W}...</pre>	<pre>... sed Auctor dicit 89b in <i>Opera</i> quod <i>semper sit</i>, id est¹...</pre> ¹ id est] inest W	<pre>... 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>...</pre>
HTML source		
<pre>... sed Auctor dicit 89b in <i>Opera</i> quod <i>semper sit</i>, id est<sup>1</sup>... <div><sup>1</sup> id est] inest W</div></pre>		

From visual to semantic

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

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>
```

From visual to semantic

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

name

title

quote

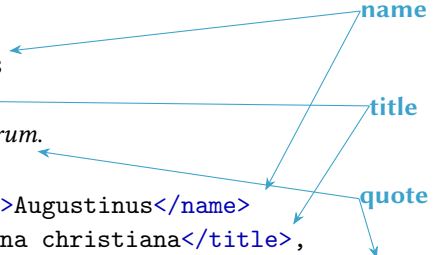
XML

```
<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>
```

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>`.



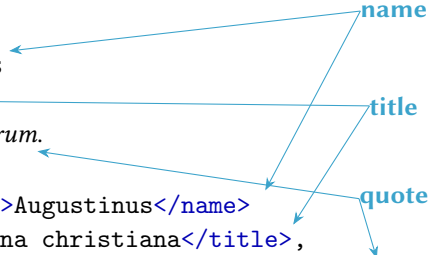
XML

```
<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>
```

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>`.



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>
```

- 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

Example: Generating an index

Classic procedure for printed book Computer aided procedure, multi-media

- | | |
|---|--|
| 1 procurement of the book | 1 digitally copying the book |
| 2 recording all the occurrences of the significant words | 2 marking the significant occurrences in the document |
| 3 ordering and verification of records (cards) | 3 dynamic extraction of marked words |
| 4 transcription of the records | 4 – |
| 5 typesetting the layout | 5 applying a style template |
|  it takes some hours or days |  it takes some minutes or hours |

visual $\longrightarrow$ **syntactic** $\longrightarrow$ **semantic**

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

- 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

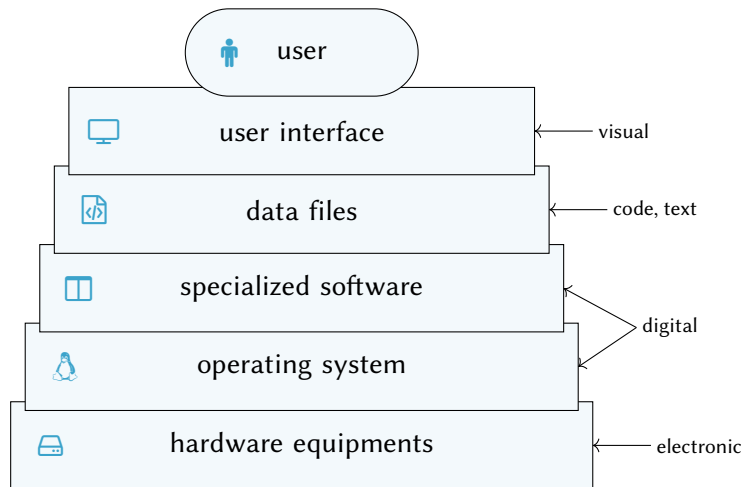
example: anti-spam filters for email

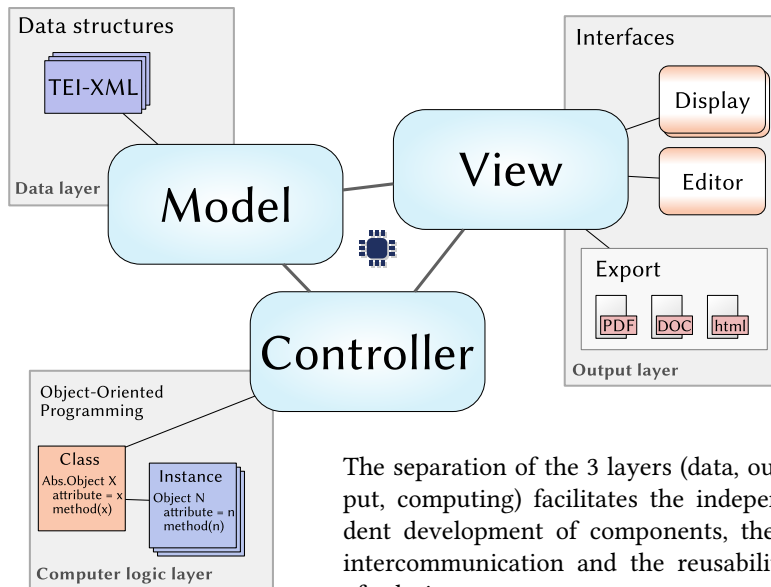
- automatic

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.





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

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



8

In 1 Sent prol q 1

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

(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 incomplexum vel complexum, et hoc vere vel false, dicitur res et aliquid. Isto modo Philosophus¹⁹ in Praedicamentis capitulo De oppositis significata propositionum contradictoriarum vocat res, ut ibi patet. Et eodem modo accipit rem ibidem capitulo De priori, cum ait²⁰: »Dum res est vel non est, oratio vera vel falsa dicatur, necesse est«. Non enim, quia

¹⁹ cf. ARISTOT. Categoriae 10 (12b 10; Juntina 1^a,53D)

²⁰ ibid 12 (14b 21–22; Juntina 1^a,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)