

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

conf. dr. Mihai MAGA

Digital Humanities

DIGITAL HUMANITIES PENTRU SURSELE FILOSOFICE MEDIEVALE

semestrul II, 2023–2024

HME2415

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

Planul cursului

0. Prezentare	3
0.1. Curs 1: Introducere în Digital Humanities	3
0.2. Curs 2: Codificarea semantică	3
0.3. Curs 3: Editarea critică	3
0.4. Curs 4: Principiile TEI-XML	3
0.5. Curs 5: Reprezentarea manuscriselor în TEI	4
0.6. Curs 6: Reprezentarea variațiilor textuale în TEI	4
0.7. Curs 7: Reprezentarea aparatului de surse în TEI	4
0.8. Curs 8: Vizualizarea edițiilor digitale	5
0.9. Curs 9: Integrarea și procesarea digitală	5
0.10. Curs 10: Inteligența artificială	5
1. Introducere în Digital Humanities	6
1.1. Ce înseamnă DIGITAL HUMANITIES?	6
1.2. Istoria DIGITAL HUMANITIES	9
1.3. Domenii ale DIGITAL HUMANITIES	11
 Temă pentru acasă	12
2. Codificarea semantică	13
2.1. Reprezentare vizuală vs. reprezentare semantică	13
2.2. Paradigma semantică în lumea digitală	15
2.3. Utilitatea codificării semantice	17
 Temă pentru acasă	18
3. Editarea critică	19
3.1. Metode de abordare a edițiilor critice	19
3.2. Tipuri de ediții critice	21
3.3. Elementele unei ediții critice	24
 Temă pentru acasă	25
4. Principiile TEI-XML	27
4.1. Formatul XML	27
4.2. Despre TEI	28
 Temă pentru acasă	31
5. Reprezentarea manuscriselor în TEI	32
5.1. Indicarea și descrierea manuscriselor	32
5.2. Setul de instrucțiuni pentru descrierea manuscriselor	33
 Temă pentru acasă	35

6. Reprezentarea variațiilor textuale în TEI	36
6.1. Diferențe între copiile manuscrise	36
6.2. Descrierea variațiilor textuale în TEI	37
6.3. Metode de codare	38
6.4. Statistici și numărători	38
 Temă pentru acasă	38
7. Reprezentarea aparatului de surse în TEI	39
7.1. Ontologii și arbori	39
7.2. Descrierea elementelor unei referințe textuale	39
7.3. Alte tipuri de referințe	40
7.4. Atașarea referințelor	41
 Temă pentru acasă	41
8. Vizualizarea edițiilor digitale	42
8.1. Conversia în format clasic	42
8.2. Interfețe interactive	45
8.3. Includerea imaginilor manuscriselor	46
8.4. Capcanele vizualului	46
 Temă pentru acasă	47
9. Integrarea și procesarea digitală	48
9.1. Indici și tabele de concordanțe	48
9.2. Limbaje de interogare: XQuery, XPath	48
9.3. Căutarea în text	48
9.4. Lematizare, normalizare, dicționare	49
9.5. Digital corpora	49
9.6. Data-mining	50
 Temă pentru acasă	50
10. Inteligența artificială	51
10.1. Inteligență artificială, Machine Learning	51
10.2. Algoritmi și modele	51
10.3. Antrenament și predicție în ML	52
10.4. Tipuri de abordare în ML	52

0. Prezentare

0.1. Curs 1: Introducere în Digital Humanities

1. Ce înseamnă DIGITAL HUMANITIES?
 - o tehnică
 - o filosofie
2. Istoria DIGITAL HUMANITIES
3. Domenii ale DIGITAL HUMANITIES

1

0.2. Curs 2: Codificarea semantică

1. Reprezentare vizuală vs. reprezentare semantică
2. Paradigma semantică în lumea digitală
3. Utilitatea codificării semantice: integrare, interfețe, procesarea datelor

2

0.3. Curs 3: Editarea critică

1. Metode de abordare a edițiilor critice
 - Metoda celui mai bun manuscris
 - Metoda eclectică
 - Metoda stemmatică (lachmanniană)
 - Metoda neorientată (materială)
2. Tipuri de ediții critice
 - Ediția tip facsimil
 - Ediția eclectică
 - Ediția literară (critică)
 - Ediția diplomatică
 - Ediția materială
3. Elementele unei ediții critice: Introducere, Text, Aparat critic

3

0.4. Curs 4: Principiile TEI-XML

1. Formatul XML
 - sintaxa XML: tag, atribut, text, declarație, comentariu
2. Despre TEI¹
 - Ghidul Consorțiului TEI P5 și structura documentului
 - <http://tei-c.org/>

¹Text Encoding Initiative

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>
```

5

0.5. Curs 5: Reprezentarea manuscriselor în TEI

1. Indicarea și descrierea manuscriselor
 - identificatorul manuscrisului: loc, bibliotecă, cotă
 - părți ale descrierii: fizică, istorică, de conținut, bibliografie
2. Setul de instrucțiuni pentru descrierea manuscriselor
 - <msDesc> bloc descriere manuscris
 - <msIdentifier> identificator: <settlement>, <repository>, <idno>...
 - <head> rubrică
 - <msContents> bloc conținut manuscris: <msItem>, <locus>, <title>...
 - <physDesc> descriere fizică: <objectDesc>, <supportDesc>, <dimensions>, <layoutDesc>...
 - <history> istorie manuscris: <origin>, <provenance>, <acquisition>...
 - <additional> informații adiționale
 - <msPart> informații despre o secțiune a manuscrisului

6

0.6. Curs 6: Reprezentarea variațiilor textuale în TEI

1. Diferențe între copiile manuscrise
 - variante redacționale
 - erori de copist
2. Descrierea variațiilor textuale în TEI:
 - <app> apparatus
 - <lem> lemma
 - <rdg> reading
3. Metode de codare:
 - LRM Location-referenced Method
 - DEPAM Double End-Point Attachment Method
 - PSM Parallel Segmentation Method
4. Statistici și numărători

7

0.7. Curs 7: Reprezentarea aparatului de surse în TEI

1. Ontologii și arbori

- date structurate arborescent
 - obiecte, clase, ontologii
2. Descrierea elementelor unei referințe textuale
 - trimiteri la autor, titlu, operă, secțiune...: <title>, <author>...
 - bibliografii: <bibl>...
 - alte tipuri de referințe
 3. Atașarea referințelor
 - inline
 - prin pointeri

8

0.8. Curs 8: Vizualizarea edițiilor digitale

1. Conversia în format clasic
 - norme de paginare pentru ediții
 - DTP specializat: $\text{\LaTeX}$
2. Interfețe interactive
 - interfețe web: client-side, server-side
 - aplicații native
3. Includerea imaginilor manuscriselor
 - setul de instrucțiuni TEI: <facsimile>, <surface>, <zone>, <graphic>
 - atașarea imaginilor: inline, pointer
4. Capcanele vizualului

9

0.9. Curs 9: Integrarea și procesarea digitală

1. Indici și tabele de concordanțe
2. Limbaje de interogare: XQuery, XPath
3. Căutarea în text: simplă, wildcard, stematizată, lematizată
4. Lematizare, normalizare, dicționare
5. Digital corpora
6. Data-mining

10

0.10. Curs 10: Inteligența artificială

1. Ce este inteligența artificială (AI) și învățarea automată (Machine Learning, ML)?
2. Algoritmi și modele în ML
3. Antrenament și predicție în ML
4. Clasificarea tipurilor de abordare în ML

1. Introducere în Digital Humanities

1.1. Ce înseamnă DIGITAL HUMANITIES?

Digital humanities [DH] (umanioare digitale) este o arie de cercetare și educație la intersecția dintre tehnica de calcul și disciplinele umanistice. Dezvoltându-se din domeniile umanisticii computerizate, tehnicii de calcul umaniste și practicii digitale în discipline umaniste, digital humanities cuprinde o varietate de subiecte, de la îngrijirea colecțiilor online până la extragerea de cunoștințe din seturi mari de date culturale.¹

- o tehnică
- o filosofie

O listă de peste 260 de definiții ale DH poate fi consultată la:

http://www.artsrn.ualberta.ca/taporwiki/index.php/How_do_you_define_Humanities_Computing/_Digital_Humanities%3F

Vezi și: M. TERRAS, J. NYHAN, E. VANHOUTTE (eds.), *Defining Digital Humanities: a reader*, Farnham/Burlington: Ashgate 2014

1.1.1. O tehnică

- domeniul umanist este atât de vast încât nicio persoană sau grup de persoane nu pot avea în întregime control asupra lui
 - chiar și într-un subdomeniu foarte specializat, datorită naturii vaste și multidisciplinare a cunoștințelor necesare, nimeni nu poate afirma că știe tot
 - științele exacte au progresat enorm în ultimii 70 de ani folosind instrumente electronice, cibernetice, digitale
 - de fapt, progresul digital se datorează în parte dorinței de a împinge mai departe limitele cunoașterii
- ➡ cum poate acest avans tehnic să ajute disciplinele umaniste?
- luând ca model modul în care alte științe folosesc calculatorul pentru a exploata volume mari de informație și pentru a calcula soluții la problemele lor
 - dezvoltând sisteme informatice adecvate problemelor specifice ale disciplinelor umaniste, care diferă mult de celelalte științe

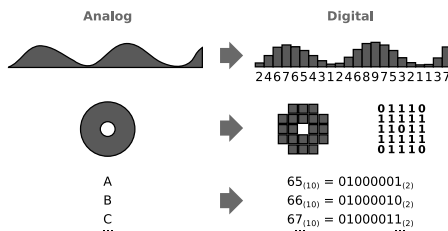
Ce înseamnă *digital*?

digit_{Engl.} = cifră, număr **digitus**_{Lat.} = deget; unitate de măsură

format digital = reprezentare numerică discontinuă și discretă

digitizare = convertirea valorilor analogice în valori numerice prin eșantionare sau prin tabele de codare

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



- calculatorul procesează și stochează datele doar în format binar

1 | 0
 +5V_{cc} | 0V_{cc}
 da | nu
 adevărat | fals
 = biți

Număr și sens

- Calculați rezultatul următoarei operații:

$$430 - 354 = ?$$

- Rezultatul este:

76

- Ce înseamnă 76? 76 de ce?
 - Augustin din Hipona (354 – 430) a trăit 76 ani
 - „... din 430 de persoane verificate, 354 fraudau banii...”² și 76?
 - *Postfață*, pp. 354–430 (total 76 pagini)

Calculatorul face operații numerice abstracte, independente de sens

dar disciplinele umaniste se ocupă în primul rând cu sensul

În DH este nevoie să atașăm un sens valorilor computaționale

1.1.2. O filosofie

- DIGITAL HUMANITIES face parte dintr-o mișcare de gândire amplă, care a început în secolul al XX-lea
- Fundamentele acestei mișcări provin din:
 - structuralism
 - sensul se construiește ca relație într-o structură

v. semiotică, FERDINAND DE SAUSSURE, arbitraritate lingvistică

²Declarație a Ministrului Muncii din 23.10.2011

poststructuralism critica structuralismului; sensul se construiește doar în interiorul structurii; deconstrucția revelează sensul
cognitivism gramatica universală (N. CHOMSKY)

Poststructuralism

De plus, le langage se réfère à la position de celui qui écoute et de celui qui parle, c'est-à-dire à la contingence de leur histoire. Saisir, par inventaire, tous les contextes du langage et des positions où peuvent se trouver les interlocuteurs, est une entreprise insensée. Chaque signification verbale est au confluent de fleuves sémantiques innombrables. Tout comme le langage, l'expérience n'apparaît plus faite d'éléments isolés, logés, en quelque façon, dans un espace euclidien où ils pourraient s'exposer, chacun pour son compte, directement visibles, signifiant à partir de soi... [Les mots] signifient à partir du « monde » et de la position de celui qui regarde.³

O filosofie poststructuralistă și deconstructivistă

- Nu avem acces complet la sensul original al filosofiei medievale
 - Deconstruim semnificațiile intratextuale și interpretările istorice pentru a revela relațiile originale
 - Construim propria noastră structură de înțelegere
 - diferențele, clivajele, contextul sunt importante
 - trebuie determinate relațiile sincronice și diacronice
- ➡ conștiința că autorul unui text editat este editorul

Text, context, autor

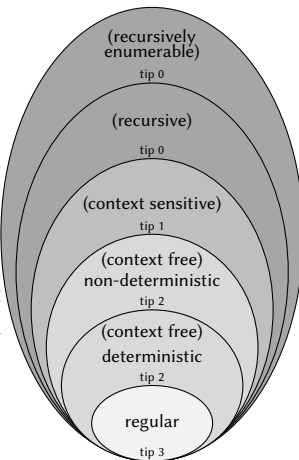
Exemplu de text	Analiză intratextuală	Analiză intertextuală	Deconstrucție
LIBERTUS DE OFICIIS, Despre viață, III, 4:	LIBERTUS DE OFICIIS, Despre viață, III, 4:	LIBERTUS DE OFICIIS, Despre viață, III, 4:	Libertus de Oficiis, ⁴ Despre viață, III, 4:
Definiția mea pentru suflet este următoarea: sufletul este actul prim al corpului natural dotat cu organe. Prin ea resping teza afirmată de Aristotel în Despre viața din adâncuri unde spune: „în adâncul mării trăiesc creaturi care nu au organe.”	Definiția mea pentru suflet este următoarea: sufletul este actul prim al corpului natural dotat cu organe. Prin ea resping teza afirmată de Aristotel în Despre viața din adâncuri unde spune: „în adâncul mării trăiesc creaturi care nu au organe.”	Definiția mea pentru suflet este următoarea: sufletul este actul prim al corpului natural dotat cu organe. Prin ea resping teza afirmată de Aristotel în Despre viața din adâncuri unde spune: „în adâncul mării trăiesc creaturi care nu au organe.”	Definiția mea pentru suflet este următoarea: sufletul este actul prim al corpului natural dotat cu organe. Prin ea resping teza afirmată de Aristotel în Despre viața din adâncuri unde spune: „în adâncul mării trăiesc creaturi care nu au organe.”
	autor principal autor citat	autor principal autor sursă Arist. De anima II.1 412b	autor real autor „principal” autor sursă pt. a.r. autor sursă pt. a.p.
☞ credem că sensul este construit de autor, dar „autorul” este un concept contextual			
☞ de fapt, noi construim sensul, inclusiv sensul conceptului de „autor”			

³E. LÉVINAS, *La signification et le sens*
⁴Acest nume de autor și textul sunt doar o invenție a autorului acestui curs

Gramatica formală

NOAM CHOMSKY – gramatica universală, gramatica transformățională generativă

- cunoștințele sintactice sunt parțial innăscute ← dispozitiv de achiziție lingvistică ⇒ **gramatică universală**
- limbajul constă în
 - **structuri superficiale** (enunțurile vorbite)
 - **structuri profunde** (relațiile dintre cuvinte și înțelesul conceptual)
- ➔ **gramatica transformativă** constă într-un set limitat de reguli de transformare a structurilor profunde în structuri superficiale

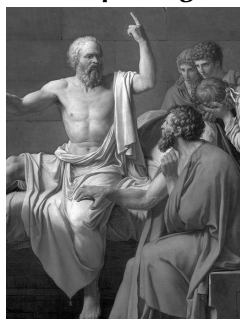


☞ Aplicații: psiholingvistică, programare pe calculator, inteligență artificială

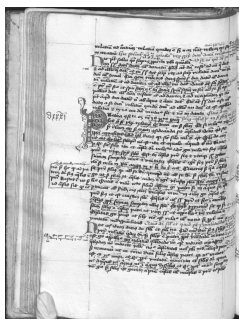
O nouă paradigmă a textului

Ierarhia lui Chomsky

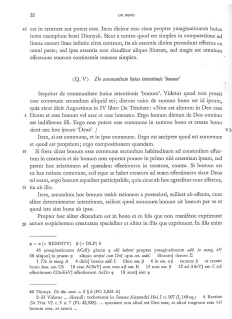
9



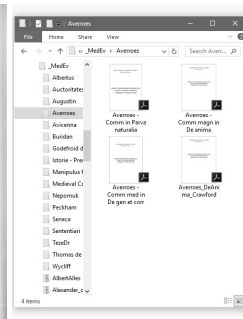
oralitate



manuscris



tipar

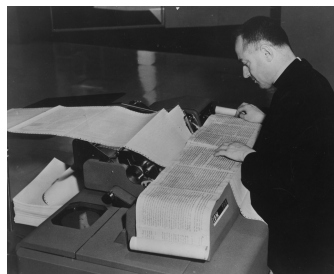


digital

**1.2. Istoria DIGITAL HUMANITIES**

10

1949 ROBERTO BUSA, SJ, intenționează să alcătuiască un index de concordanțe exhaustiv al operei lui TOMA DIN AQUINO; îl întâlnește pe THOMAS J. WATSON, fondatorul IBM, și îl convinge pe acesta să sponsorizeze *Index Thomisticus*; timp de 30 de ani introduce în calculator (pe cartele perforate) toată opera tomistă și publică o versiune tipărită în 56 de volume (disponibilă apoi pe CD-ROM și acum online⁵)



'60-'70 s-au produs mai multe concordanțe electronice (texte în Mittelhochdeutsch, poeme ale lui W. B. Yeats etc.)

-- au apărut proiecte de evaluare statistică a paternității unor texte (Epistolele Pauline, Scrisorile lui Junius, *Federalist Papers* etc.)

! computerele erau extrem de limitate și inaccesibile



prelucrare serială pe mașini de calcul enorme deținute de câteva instituții și care foloseau cartele perforate și benzi magnetice

1963- *Centre for Literary and Linguistic Computing*, Cambridge

1964 conferința *Literary Data Processing* organizată de IBM la Yorktown Heights

1966- începe publicarea revistei *Computers and the Humanities*

1965- programul COCOA stabilește un prim standard în „umanistica computerizată” (University College London)

'70-'85 consolidarea metodologiei DH

-- mai multe conferințe internaționale: Edinburgh (1972), Cardiff (1974), Oxford (1976), Birmingham (1978), Cambridge (1980); apar mai multe publicații

-- se înmulțesc centrele de „umanistică computerizată”, se introduc cursuri în DH

! computerele devin mai performante și mai accesibile

1971- *Project Gutenberg*, Michael Hart (cărți în domeniul public)

1972- *Thesaurus Linguae Graecae* (TLG), Univ. of California, Irvine

1976- *Oxford Text Archive* (OTA), evoluții în arhivarea textelor electronice

'85-'90 dezvoltare, interconectare, standardizare

-- programe de analiză textuală în DOS: Word-Cruncher, TACT, MicroOCP

! apar computerele personale, poșta electronică, internetul și WWW

1988-90 Ian Lancashire și Willard McCarty, *Humanities Computing Yearbook*

1987 întâlnire la Vassar College, Poughkeepsie, pentru crearea unei scheme standard de codare pentru DH („Principiile Poughkeepsie”) ⇒ TEI (Text Encoding Initiative)

1990 prima propunere TEI

1994 prima versiune completă a *Ghidului TEI* (*TEI Guidelines*)

⁵<http://www.corpusthomisticum.org>

'90... internetul devine parte vitală a activității academice



internetul e folosit atât pentru publicare, cât și pentru promovarea DH

--//-- apar multe proiecte și prototipuri pentru publicarea online a edițiilor digitale

- eforturile se concentrează mai mult pe afișare, interfață, interactivitate

--//-- apar colecțiile și corpus-urile online

- se descoperă noi moduri de scriere academică:

- extragerea de documente dintr-o bază de date și reconstruirea lor într-un material nou

- editarea colaborativă a documentelor (transcrieri de manuscrise, adnotarea corpus-urilor, biblioteci digitale)

- posibilitatea furnizării de conținut multimedia (imagini, video, audio)

--//-- TEI devine standardul de facto pentru DH

2007 lansarea *Ghidului TEI P5 (TEI Guidelines P5)*

15

1.3. Domenii ale DIGITAL HUMANITIES

■ după disciplină:

- paleografie și ediție critică
- codicologie și biblioteconomie
- filologie și analiză lingvistică
- istorie generală și istorii aplicate
- arheologie, artă, muzeografie
- calcul lingvistic și inteligență artificială

■ după perioadă:

- antică
- medievală
- modernă
- contemporană

■ după aplicație:

- analiză stilistică și studii de paternitate
- asamblarea și analiza corpus-urilor lingvistice
- editarea critică erudită în format digital
- analiza textuală
- structurarea colecțiilor tematice
- producerea de materiale în format tipărit și multimedia
- calcul speculativ (A.I.)

■ după sursa digitală:

- born-digital (născute digital)
- convertite din analog



DH tinde să devină un instrument universal

Exemple de folosire pentru Digital Humanities

■ Cataloage bibliotecare online

 <http://www.manuscriptorium.com/>

■ Ediții de text online

 <http://scta.lombardpress.org/>

■ Colecții integrate de texte selectate

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

■ Corpus-uri digitale complete

 Thesaurus Linguae Graecae CD-ROM

■ Seturi mari de date

 <http://storage.googleapis.com/books/ngrams/books/datasetv2.html>

17

Bibliografie

- [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

Temă pentru acasă

Căutați pe internet proiecte care folosesc elemente de Digital Humanities. Alcătuiți o listă cu 3 proiecte relevante și descrieți-le.

Model pentru descriere:

URL: <http://...>

Titlu: titlul proiectului

Coordonator: ce instituție sau persoană dezvoltă proiectul?

Scop: ce intenționează proiectul să facă?

Conținut: ce fel de conținuturi implică proiectul?

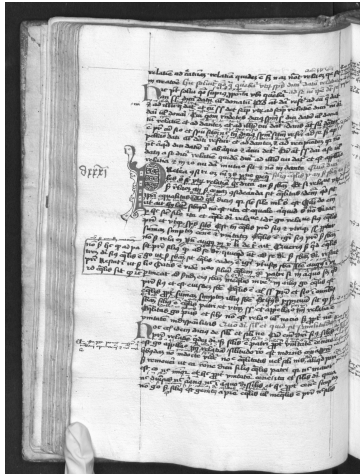
Stadiu: în ce stadiu se află proiectul acum?

2. Codificarea semantică

2.1. Reprezentare vizuală vs. reprezentare semantică

Semnificația unui text este transmisă prin:

- textul în sine (litere, cuvinte)
- elementele meta-textuale (format, poziție etc.)



22 DE ROMO

45 est in creatura aut potest esse. Item dicitur esse circa propter ymaginationem huius
ista exemplum huius Dionysii. Sicut a centro quod est simplex ita compositio ad
lineas excent lineis infinite circa centrum, ita ab essentia divina procedunt effectus ex
omni parte; sed ipsa essentia non clauditur aliquo illorum, sed magis est omnium
effectuum eorum concentrica mensura simplex.

(Q.V.) *De communitate huius intentionis 'bonum'*

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

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

10 Si forte dicitur bonum esse commune secundum habitudinem ad continentem effec-
tum in creatura et ad bonum non oportet potest in primo nisi essentiam ipsum, sed
preter hoc rationem ad quendam effectum in creatura, contra. Si bonum est
ex hac ratione commune, sed equae se habet creatura ad eam effectum sicut Deus
ad eam, ergo bonum equaliter participabilis, quia sicut ab hoc egreditur non effectus,
15 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 ita sit bonum ab ipso.

Propter hoc aliter dicendum est in bono et in illis que non manifeste expriment
20 autem respicientem creaturas specialiter et aliter in illis que expriment in illis enim

q = a (= HENRY) b (= DLP) 8
45 ymaginationem AGS: gloriæ y ali habet propter ymaginationem add. in marg. vY
46 aliquid in fronte y aliquid in fine autem add. HENRY item 3
1 Ita in marg. A 4 dicitur bonum add. C Item ore b 6 in ore. e3 summe b et creato
boni bon. ore. CS 10 dicit AGS non ore e3 ore. E 11 non ore. 12 ad AGS v ore. C e3
effectum CAGS v effectum AGS y 18 sunt sunt y

46 Dionys. De div. nom. c.5 §4 (PG 3,821 A)
2-23 Videtur... ibidem videtur in Summa Alexandri Hal. 1. m. 107 (l. 168a.) 4 Rationis
De Trin. VI c.5 n.7 (PL 42,302b) ... quoniam non alius est Deus esse, et alius rationem esse vel
bonum esse, et contra ...

În format vizual (fizic), semnificația meta-textuală este dedusă din reguli generale sau locale

- uneori regulile sunt indicate explicit

- prin convenții general cunoscute

☞ titlul este mai mare, centrat, separat de text

- printr-o etichetă explicită

☞ secțiuni intitulate *Note, Glosar, Cuprins*

- prin declarare explicită

☞ uneori convențiile tipografice sunt descrise în *Introducere*

- alteori convențiile sunt neclare și sunt deduse din context

Exemplu¹

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

Procesul de extragere a semnificației vizuale, deși aparent simplu, este o operație **complexă** care implică vederea, memoria, experiența, înțelegerea și sinteza; în oricare dintre aceste funcții se pot produce **erori**.

- Computerele nu pot, deocamdată, să reproducă coerent acest proces.
- Uneori și cititorii umani se pot înșela în privința semanticii vizuale.

Exemplu²

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

Formatare pe calculator

- folosește un program de editare text comun
- produce un format vizual

sursă RTF

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

sursă HTML

```
... 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>
```

redare vizuală

```
... sed Auctor dicit |89b| in
Opera quod semper sit, id
est1...
```

¹ id est] inest W



Ediție digitală

- folosește programe specializate
- produce un fișier codat semantic

sursă XML

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

De la vizual la 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>.

nume

titlu

citat

¹Sursa: <http://xh.wikipedia.org> (articol în limba Xhosa, Africa de Sud)

²Sursa: <http://xh.wikipedia.org> (articol în limba Xhosa, Africa de Sud)

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. Paradigma semantică în lumea digitală

6

- La fiecare 3–4 zile producem atâta informație câtă a fost produsă acum 10 ani timp de un an întreg
- Diferența între conținutul digital și cel analogic se pierde, ceea ce contează este **relevanța**

Web semantic, Web 2.0

- pune accent pe: conținut generat de utilizator, uzabilitate, interoperabilitate
- se opune vizualizării pasive a conținutului
- conținutul digital este structurat semantic astfel încât să poată fi prelucrat de computere

Exemplu: Generarea unui index

7

Procedura clasică pentru carte tipărită

1. procurarea cărții
 2. fișarea tuturor ocurențelor cuvintelor semnificative
 3. ordonarea fișelor și verificarea lor
 4. transcrierea fișelor
 5. punerea în pagină
- ☹ durează câteva ore sau zile

Procedura computerizată, multimedia

1. copierea electronică a cărții
 2. marcarea ocurențelor semnificative în document
 3. extragerea **dinamică** a cuvintelor marcate
 4. –
 5. aplicarea unui șablon de stil
- ☺ durează câteva minute sau ore

- Etapele tranziției:
vizual → **sintactic** → **semantic**

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

- Proceduri de extragere a conținutului semantic:

- manuală
un utilizator uman selectează și marchează conținutul semantic

☞ exemplu: Wikipedia

- semiautomată (asistată)

utilizatorul uman decide conținutul semantic și folosește apoi un sistem automat de marcare

☞ exemplu: filtrele anti-spam pentru email

- automată

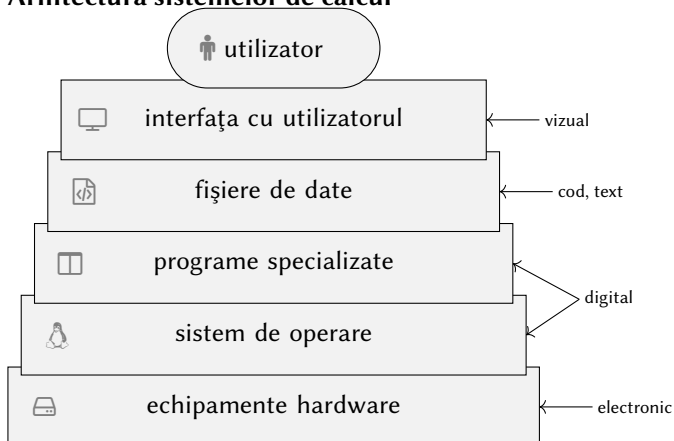
computerul folosește date statistice și algoritmi pentru a marca conținutul semantic

☞ exemplu: Google Translate

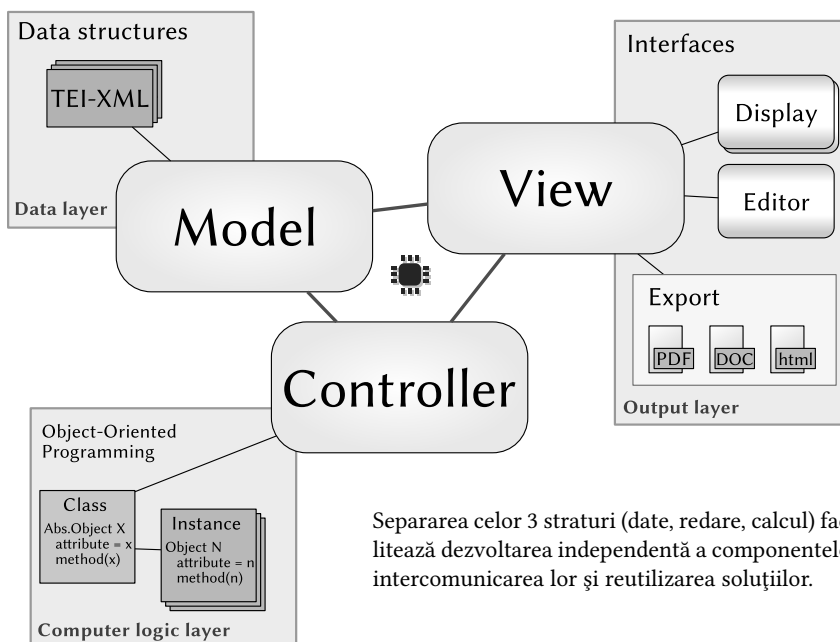
☼ De la reprezentarea semantică la cea vizuală se poate trece ușor prin formătări automate pe baza unor foi de stil.

9

Arhitectura sistemelor de calcul



10

Design reutilizabil, structuri abstracte, interfețe**2.3. Utilitatea codificării semantice**

1. Prelucrarea computerizată a datelor
 - orice document generat conform standardelor poate fi transmis și prelucrat de orice program care implementează standardele
 - abordarea computațională a cantității mari de texte existente scutește cercetătorul de operații repetitive și consumatoare de timp
2. Vizualizare conform intenției utilizatorului
 - același document poate fi afișat în multiple moduri, cu sau fără interacțiunea utilizatorului
3. Îmbogățirea și conectarea informațiilor metatextuale
 - oricâte detalii pot fi atașate textului, textele pot fi interconectate, informațiile pot fi furnizate de alte surse

Temă pentru acasă

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, aliquid 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 dicitur, 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)

Identificați toate elementele non-textuale purtătoare de sens din pagina alăturată (semne, formataři, poziție). Întocmiți o legendă a acestor elemente.

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

3. Editarea critică

3.1. Metode de abordare a edițiilor critice

- Metoda celui mai bun manuscris
- Metoda eclectică
 - *Lectio difficilior*
- Metoda stemmatică (lachmanniană)
 - *Stemma codicum*
- Metoda neorientată (materială)

1

3.1.1. Metoda celui mai bun manuscris

- cea mai veche și mai simplă metodă
- este ales un manuscris considerat cel mai bun și este transcris
 - cum este ales?
 - cel mai vechi
 - cel mai complet
 - cel mai ușor de citit, etc.



uneori manuscrisul de referință este schimbat de la o secțiune la alta

- metodă utilă când între copii se identifică una de o relevanță aparte, ex. un autograf

2

3.1.2. Metoda eclectică

- scopul editorului este de a obține un text final cât mai apropiat de intenția autorului și cât mai inteligibil
- editorul își poate stabili o metodologie științifică sau semi-științifică, dar unele opțiuni vor fi subiective, autoriale
 - editorul își asumă sarcina de a interpreta probele materiale pentru cititor
- Proceduri eclecticice:
 - recensio sortarea și colaționarea manuscriselor
 - examinatio încercarea de a stabili cea mai timpurie versiune a textului
 - emendatio corectarea textului (uneori numită divinatio)
 - intervențiile editorului pot fi netransparente și pot corupe textul; normele de inteligibilitate pot fi subiective
- datorită procedurilor complexe de reconstituire a textului, deciziile editoriale rămân obscure

Lectio difficilior potior

= lecțiunea mai dificilă este mai puternică

- poate fi considerată o metodă eclectică
- folosită în sec. XV–XVIII pentru editarea *Bibliei* și a altor surse, drept criteriu obiectiv în selectarea variantelor textuale
- **Regula:** acolo unde manuscrisele diferă, se alege lecțiunea cea mai dificilă
 - **Aserțiune:** când copiiștii nu înțeleg un cuvânt sau fragment dificil, îl înlocuiesc cu unul mai simplu, comițând o greșeală

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

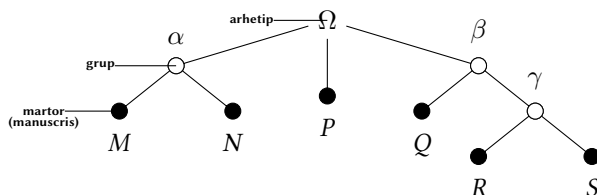
Mt 6,9: πάτερ ἡμῶν, ὁ ἐν τοῖς οὐρανοῖς, ἁγιασθήτω τὸ ὄνομα σου ...

Lc 11,2: πάτερ, ἁγιασθήτω τὸ ὄνομα σου ...

3.1.3. Metoda stemmatică (lachmanniană)

5

- reconstrucția prin metode filologice a relațiilor de derivare a manuscriselor dintr-un arhetip
 - arhetip = model virtual al tuturor copiilor manuscrise, posibil dar nu obligatoriu textul original
- prin evaluarea diferențelor textuale comune ale manuscriselor (lecțiuni opuse, apoi erori comune), se obține o **stemma codicum**



lecțiune variantă textuală atestată doar de unele manuscrise

Ex: autem | enim ; ergo | igitur ; sive | seu ; inv ...

eroare variantă textuală evident greșită, dar atestată de unele manuscrise

Ex: erat | errat ; anima | alia ; sum | suum ; om ...

- deciziile editoriale în privința variantelor textuale se iau apoi în funcție de relația manuscriselor cu modelul

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

Stemma codicum

6

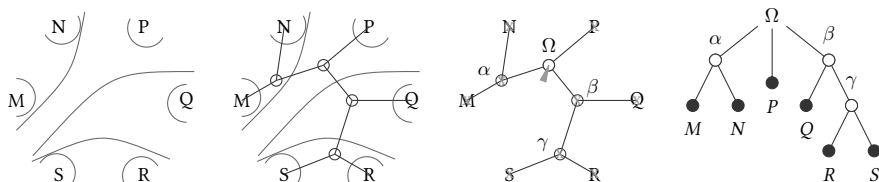
1. se numără **lecțiunile** diferite din fiecare manuscris și grup de manuscrise

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

2. se dispun în cerc siglele și se izolează grupurile
3. se conectează într-un punct comun grupurile, obținând **stemma neorientată**
4. se numără **erorile** comune ale manuscriselor și grupurilor pentru a decide cum se orientează stemma

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

5. se ridică stemma în punctul în care nu există greșeli comune, obținând **stemma orientată**



3.1.4. Metoda neorientată (materială)

7

- este centrată pe obiectul material (ex. manuscrisul)
- redă exhaustiv proprietățile sursei
- evită interpretările arbitrare, conjecturile, emendările
 - asigură maximă științificitate și factualitate
 - oferă instrumente pentru interpretarea textului sursă
- poate produce așa-numita **headless edition**

headless edition = ediție critică a tuturor surselor manuscrise fără stabilirea unui text unificat

Ms. A	Ms. B	Ms. C
text și descriere	text și descriere	text și descriere
...	...	...

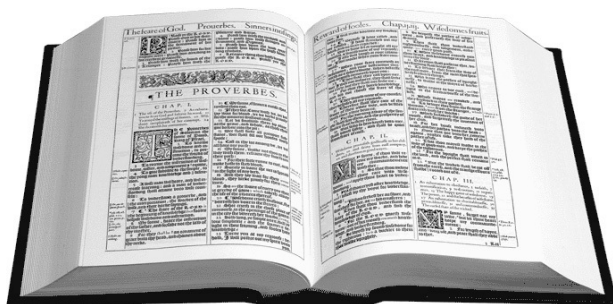
3.2. Tipuri de ediții critice

8

- Ediția tip facsimil
- Ediția eclectică
- Ediția literară (critică)
- Ediția diplomatică
- Ediția materială

3.2.1. Ediția tip facsimil

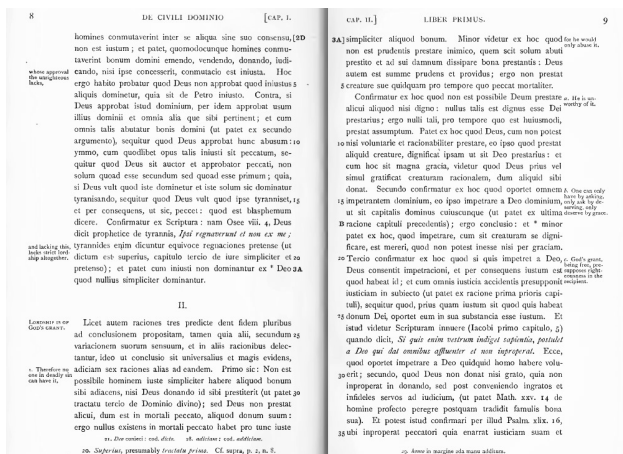
- reproducere fotografică a sursei textuale
- conține eventual o introducere, un aparat minimal și/sau anexe (index, bibliografie etc.)

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

3.2.2. Ediția eclectică

- centrată pe textul stabilit de editor, de obicei printr-o metodă eclectică sau a celui mai bun manuscris
- aparatul critic este minimal, incomplet sau inexistent
- fragmentele de text reconstituite de editor sunt uneori marcate cu semne convenționale

10

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

3.2.3. Ediția literară (critică)

- centrată pe text și pe tradiția acestuia
- conține un aparat critic extins, menționează variantele textuale și deciziile editoriale

11

3.2.5. Ediția materială

- centrată pe forma materială conservată a textului
- descrie și redă exhaustiv toate detaliile sursei
- uneori nu conține textul unificat, editat, ci doar variante (= ediție headless)
- preferabil în format digital, datorită cantității mari de informație

Feuillet(s)	Genre	Incipit	Remarques
6	Séquence	Alle celeste nec non et perhenne luia	1.4 ajout du mot -Christi
12	Offertoire	Ad te Domine	V. Respicie, 1.8 ajout du mot -mei
15	Introit	Gaudete in Domino	grattage du texte du psaume
18	Offertoire	Confortamini	1.11 grattage [—enim deus]
23	Hymne	Benedictus es	1.5 grattage et ajustement du texte sous la mélodie
23	Hymne	Benedictus es	1.3.6 & 7 ajout -es, main 1 musique (cf. encre)
25	Offertoire	Exsulta satis	V. Loquetur, 1.4 grattage et ajustement du texte sous la mélodie
48	Offertoire	Reges Tharsis	V. Suscipiant, 1.6 rajout postérieur de la reprise —et adorabunt
51	Introit	Omnis terra	1.5 ajout postérieur -terra
53	Offertoire	Jubilate Deo universa terra	V. Locutum est 1.3 grattage —mea (mauvaise place de la syllabe)
53	Offertoire	Jubilate Deo universa terra	V. Locutum est ajout postérieur de <est> (oubli)
73	Introit	Invocabit	1.8 ajout de -me
74	Graduel	Angelis suis	1.5 ajout postérieur de -ne unquam
76	Trait	Qui habitat	V. Quoniam ang., 1.2 grattage du texte
94	Trait	De necessitatibus	1.4 grattage et réécriture du texte, main postérieure
98	Offertoire	Miserere mihi	V. Tibi soli, 1.6 ajout du mot —soli, main 3
104	Offertoire	Domine in auxilium	rubrique Co grattée (erreur de rubrique)
107	Offertoire	Illumina oculos	V. Respicie, 1.2 grattage de la syllabe —di de —exaudi (placement)
108	Trait	Ad te levavi	1.8 grattage du mot —oculos
117	Graduel	Oculi omnium	1.3 grattage du m de —manum

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

3.3. Elementele unei ediții critice

- Introducere
- Text
- Aparat critic



se pot adăuga alte elemente auxiliare: index, bibliografie, diagrame, ilustrații etc.

3.3.1. Introducere

- O ediție critică conține în introducere:
 1. descrierea surselor
 - descrierea fiecărui manuscris
 - relația dintre manuscrise (stemma codicum)
 2. date
 3. atribuire
 4. prezentare doctrinară
 - conținutul și poziția textului în context istoric
 - sursele explicite și implicite
 5. descrierea procedurii de editare
 - metoda de editare folosită
 - abrevieri și aparat critic

6. bibliografie secundară

- ⇒ ordinea și lungimea părților introducerii pot varia; uneori când unele elemente descriptive au fost deja publicate, o parte poate fi înlocuită cu referința bibliografică

3.3.2. Text

- Textul editat urmează anumite standarde lingvistice și grafice la care editorul aderă:
 - ortografie latină clasică, latină medievală sau ortografie necorectată
 - punctuație punctuație modernă, clasică sau punctuația originală a sursei
 - secționare textul este divizat prin titluri în secțiuni; se pot păstra secțiunile originale sau se poate face o nouă secționare logică
 - formatare se pot folosi forme diferite de caractere (încălate, majuscule reduse etc.) sau tag-uri în format digital pentru a indica elemente specifice: titluri, nume, citate etc.

3.3.3. Aparat critic

- conține toate detaliile paleografice, istorice, filologice atașate textului de către editor
- vizual reprezentat prin: semne în text, note de subsol, note marginale, note finale
- rândurile textului sunt numerotate, aceste numere sunt folosite ca referințe
- Tipuri de aparat:
 - paleografic (filologic) descrie diferențele textuale între manuscrise
 - surse indică sursele primare pentru citate și aluzii
 - surse biblice uneori aparat distinct pentru sursele biblice
 - tradiție indică manuscrisele pe baza cărora este editat textul
 - comparativ când există mai multe redactări ale textului, o redactare secundară poate avea propriul aparat

Utilitatea codificării semantice digitale

- orice tip de ediție poate fi codată în TEI
- aceeași ediție poate fi codată în mai multe moduri
- o ediție poate fi trecută ușor de la un standard la altul
- aceeași ediție poate fi vizualizată prin diverse interfețe
- afișarea aparatului critic este flexibilă și modificabilă
- ediția critică digitală este convertibilă în format clasic imprimat

✍ Temă pentru acasă

Alegeți din bibliotecă sau online o ediție de text medieval. Cercetați textul și introducerea încercând să răspundeți la următoarele întrebări:

1. Ce tip de ediție este?

☐ facsimil ☐ eclectică ☐ critică ☐ diplomatică ☐ materială

Observații: _____

2. Ce metodă de editare a fost folosită?

☐ cel mai bun manuscris ☐ eclectică ☐ stematică ☐ materială

Observații: _____

3. Ce elemente de ediție critică sunt prezente:

■ în introducere: _____

■ în text: _____

■ în aparatul critic: _____

4. Principiile TEI-XML

4.1. Formatul XML

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

Doar 5 tipuri de conținut:

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

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

sintaxă: între semnele < > și închise cu /

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

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

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

3. **text** (textul ca atare)

```
text
```

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

4. **declarații** (comenzi de procesare)

```
<? decl ?>
```

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

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

Exemplu de XML

XML

```
<text>
Acesta este un text XML. Marcheaza ceva <bold>important</bold>.
A fost scris la <oras tara="Romania">Cluj</oras>.
<!-- aici este un comentariu -->
<tag>Un tag poate contine
  <subtag>un subtag care poate si el contine
    <subsubtag>sub-subtag-uri</subsubtag>
  </subtag>
</tag>
Un tag poate avea zero, <separator/> unul sau mai multe attribute
  <oras tara="RO" judet="CJ" prefix="0264">Cluj-Napoca</oras>
</text>
```

Specificații

- tag-urile au două forme:

- **pereche**: delimitează o porțiune de text (<tag>text</tag>)

- trebuie închise în ordinea inversă deschiderii (<a><c>...</c>)
- doar tag-ul de deschidere poate avea attribute (<tag attr="val">...</tag>)

- **singure**: ca element de sine stătător, fără text (<tag/>)

- pot avea atribute (`<tag attr="val"/>`)
- conținutul dintre tag-uri poate avea în el tag-uri, creând un arbore

```
<root><branch><leaf/></branch><branch><leaf/></branch></root>
```
- în XML bine-format, trebuie să existe un singur tag rădăcină care să conțină tot documentul
- spațiile suplimentare și sfârșiturile de linii sunt în mod normal ignorate, dar sunt folosite pentru editarea mai ușoară a codului

```
<document>
```

```
  Format:
```

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

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

```
</document>
```

→

Format: **bold**, *italic*.

Concepte de bază în structuri tip arbore

4

nod orice element care face parte din structură

rădăcină element care subordonează toate celelalte elemente ale documentului

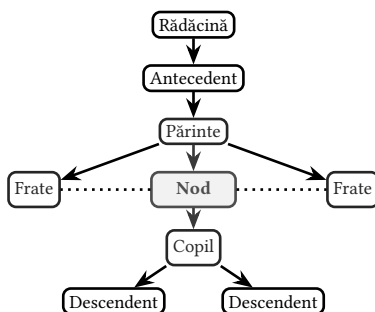
părinte relație între elemente în care elementul vizat are sub-elemente

copii relație în care elementele vizate sunt imediat subordonate elementului părinte

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

descendenți relație în care elementele vizate sunt inferioare unui element superior

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



4.2. Despre TEI

5

- Ghidul Consorțiului Text Encoding Initiative (TEI) stabilește sistemul de adnotare pentru documentele din domeniul umanist
- TEI folosește XML ca format de fișier
 - TEI este un subset de instrucțiuni XML cărora li se atribuie o semantică
- în TEI sunt specificate elementele XML folosite la edițiile digitale
- elementul rădăcină al unui document TEI-XML este `<TEI>` `</TEI>`
- un document TEI are de obicei două părți obligatorii:
 - un preambul `<teiHeader>` `</teiHeader>`
 - în preambul sunt descrise proprietățile documentului: titlu, autor, versiune, surse etc.

- corpul textului <text> </text>
 - conține textul cu tag-uri TEI; textul principal este conținut între <body> </body>; paragrafele de text sunt cuprinse între <p> </p>
- toate specificațiile TEI se găsesc pe site-ul: <http://www.tei-c.org/>

6

Exemplu de structură de bază TEI

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

Exemplu de documentație din Ghidul TEI

<titleStm> (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 analysis (@ana)) (att.global lang (@lang)) (att.global change (@change)) (att.global responsibility (@cert, @resp)) (att.global source (@source))
Contained by	header: pBibliography
May contain	core: author editor meeting respStm title header: funder principal sponsor
Example	<pre> <titleStm> <title>Gregory's Life of St. John Nubert: a machine-readable transcription/</title> <respStm> <resp>compiled by</resp> <name>J. Lucas</name> </respStm> </titleStm> </pre> Show all
Content model	<pre> <content> <sequence> <elementRef key="title" minOccurs="1" maxOccurs="unbounded"/> <classRef key="model_respStm" minOccurs="0" maxOccurs="unbounded"/> </sequence> </content> </pre>
Schema Declaration	<pre> <element titleStm { att.global.attributes, att.global.rendition.attributes, att.global.linkage.attributes, att.global.analysis.attributes, att.global.lang.attributes, att.global.change.attributes, att.global.responsibility.attributes, att.global.source.attributes } title, model_respStm ? </pre> XML syntax

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

4.2.1. Programe pentru editarea TEI-XML

■ Editoare avansate

ofere validare TEI, autocompletare, documentație

- | | | |
|----------------------------|---------------|---|
| ■ oXygen XML Editor | (cu plată) | https://www.oxygenxml.com/ |
| ■ VS Code | (open source) | https://code.visualstudio.com/ |
| ■ Atom | (open source) | https://atom.io/ |
| ■ XML Copy Editor | (open source) | http://xml-copy-editor.sourceforge.net/ |
| ■ iEdit | (open source) | http://www.iedit.org/ |

- Editoare simple

 oferă colorizarea sintaxei, uneori validare XML

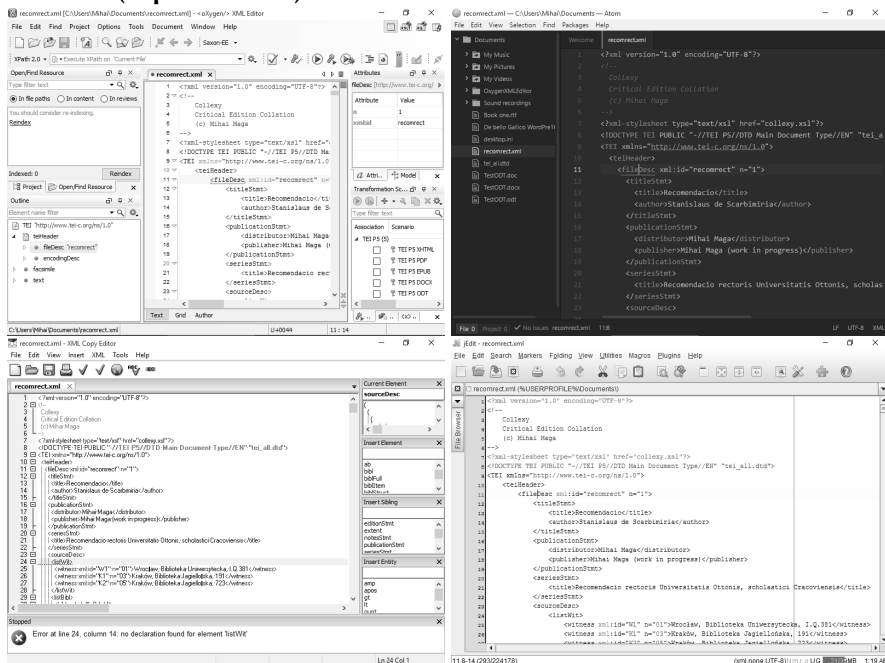
- **Notepad++** (open source) <https://notepad-plus-plus.org/>
- **Notepad2** (gratuit) <http://www.flos-freeware.ch/notepad2.html>
- **Eclipse** (open source) <https://eclipse.org/>

■ Editoare web

 funcționează în browser, oferă funcționalități diverse

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

Editoare (copii de ecran)



Temă pentru acasă

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

<text></text> elementul rădăcină al documentului

<p></p> paragraf

<title></title> titlu de operă

<name></name> nume de autor

<quote></quote> text citat

Quia intellectus habet duas operationes: scilicet unam qua format quidditates, 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. Reprezentarea manuscriselor în TEI

5.1. Indicarea și descrierea manuscriselor

- indicarea unui manuscris se face în general prin locație, bibliotecă, cotă

London, British Library, Arundel Ms. 415

- descrierea manuscrisului conține informații precum:
 - codicologice (material, dimensiuni, legătură, decorație)
 - istorice (dată, proveniență, posesie)
 - conținut (lista textelor conținute, cu detalii: titlu, pagini, incipit/explicit, referințe)

415

Membraneus, 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.

Indicarea și descrierea manuscriselor în TEI

- în funcție de scop, descrierea manuscriselor poate fi:
 - în cazul ediției critice care are în corp textul, manuscrisele sunt descrise în preambul, în interiorul elementului `<sourceDesc>`
 - când întregul document este o listă de manuscrise (un catalog), manuscrisele sunt descrise în corpul textului, în interiorul elementului `<body>`
 - dacă manuscrisul este menționat doar ca referință în text (de ex. într-o prefață), descrierea poate să apară în interiorul altor elemente TEI
- descrierea și indicarea (referențierea) manuscriselor sunt lucruri diferite:
 - un manuscris este descris o singură dată și primește un identificator
 - referințele interne (ex. aparatul critic) indică manuscrisul cu ajutorul identificatorului

5.2. Setul de instrucțiuni pentru descrierea manuscriselor

<msDesc> element părinte care conține descrierea
 xml:id="□" atribut care precizează identificatorul unic
 <msIdentifier> element părinte care conține datele de identificare
 <settlement> localitatea
 <repository> biblioteca
 <idno> cota
 <msContents> element părinte pentru conținutul manuscrisului
 <summary> conține prezentarea generală a ms.
 <date notBefore="□" notAfter="□"> datarea manuscrisului
 <origPlace> originea manuscrisului
 <...> alte elemente
 <msItem> descrie o parte individuală din ms.

4

Instrucțiuni pentru descrierea unei părți de manuscris

<msItem> poate apărea de mai multe ori, poate conține:
 <locus> sau <locusGrp> indicarea paginilor
 <author> nume autor
 <title> titlul părții
 <bibl> sau <listBibl> bibliografie
 <incipit> și <explicit> începutul și sfârșitul textului

5

Alte elemente pentru descrierea manuscriselor

<head> rubrica manuscrisului
 <physDesc> descriere fizică, prin <objectDesc> <supportDesc>
 <dimensions> <extent> <layoutDesc> <handDesc> <decoDesc>
 <bindingDesc>
 <history> date istorice, prin <origin> <provenance> <acquisition>
 <additional> informații adiționale, ex. <listBibl>
 <msPart> informații despre o parte distinctă a unui codex, similar cu <msDesc>

Exemplu de descriere de manuscris

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 Boccaccio</author>
  <title>De claris mulieribus</title>
  <bibl>G. Boccaccio, 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>

```

Exemplu de sursă din catalog și de indicare a ms.

7

CLUJ-NAPOCA
Biblioteca Academiei

Cod. lat. 9

a. GIOVANNI BOCCACCIO, *De claris mulieribus* (1r–64v) – b. *Commentarius super Aristotelis libros* (65r–117v).

Orig.: Europa centrală?; XV².

ADRIAN PAPAHAGI, ADINEL-CIPRIAN DINĂ, în colab. cu ANDREEA MĂRZA, *Manuscrisele medievale occidentale din România: Census, Iași:Polirom, 2019*

TEI-XML

```

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

```

5.2.1. Descrierea manuscriselor pentru aparatul critic

8

- pentru aparatul critic în TEI, martorii textului sunt descriși prin elementele `<listWit>` `<witness>`

`<listWit>` element părinte pentru lista de martori; apare în `<sourceDesc>` sau în `<body>`, `<back>`, `<abstract>`...

`<witness>` descrie sau indică un martor

`xml:id="□"` atribut care precizează identificatorul unic

- martorul poate fi descris sumar între tag-urile `<witness>``</witness>`



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

- poate fi indicat printr-o legătură la un element msDesc

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

- sau poate fi descris complet cu msDesc

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

5.2.2. Detalii despre manuscrise incluse în text

Anumite detalii observate în manuscrise sunt indicate în interiorul textului transcris, în locul unde sunt observate.

`<pb/>` sfârșit de pagină

`<cb/>` sfârșit de coloană

`<lb/>` sfârșit de rând

`<locus/>` loc care necesită detalii

```
<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"
fac="img12_002.jpg"/> utilitatem propter se ipsos diliguntur, et maxime aliorum qui est
per oculos.</p>
```

📖 Temă pentru acasă

Codați în TEI următoarele manuscrise:

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. GUILIELMUS 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.

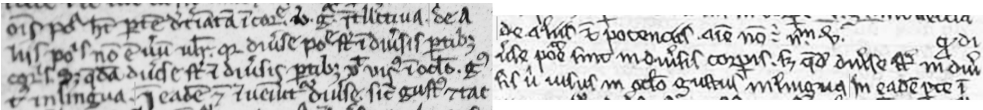
6. Reprezentarea variațiilor textuale în TEI

6.1. Diferențe între copiile manuscrise

- Textul unei opere diferă la nivel de copii manuscrise
 - uneori același text supraviețuiește în mai multe variante redacționale
 - transcriind texte, copiiștii fac modificări și greșeli
- Ediția critică trebuie să redea variațiile textului pentru că
 - textul editat este reconstituit pe baza variantelor
 - variațiile textuale sunt importante pentru tradiție



Exemplu de diferențe:

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

Segmentarea și alinierea variantelor

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

Aparatul critic clasic (imprimat)

- 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. Descrierea variațiilor textuale în TEI

4

Elemente TEI pentru codarea aparatului critic

<app> element părinte pentru intrările din aparatul critic

<rdg> conține o singură lecțiune în variația textuală

<lem> conține lema (sau textul de bază, textul acceptat)

<witDetail> dă informații suplimentare despre manuscris/-e

Elementele <rdg>, <lem> și <witDetail> trebuie să aibă atributul `wit="□"` care specifică identificatorul manuscrisului/-elor ce atestă varianta precedat de #. Identificatorii multipli se separă cu spațiu (`wit="#A #B #C"`). Identificatorul este specificat în lista de martori (atributul `xml:id` din <witness>). Pentru a marca omisiunea, poate fi folosit un tag singur (<rdg wit="□"/>)

■ Alte elemente pentru aparatul critic:

<supplied> text adăugat de editor

<...>

<unclear> nu poate fi transcris clar

dub.

<damage> zonă deteriorată a ms.

†

 text marcat ca șters în ms.

del., exp., eras.

<gap> porțiune ce nu a fost transcrisă pentru motive diverse

Exemplu TEI

5

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. Metode de codare

6

■ metoda referențierii locației (**location-referenced**)

aparatul este stocat separat (sau în text) și este legat de text prin atributul loc

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

■ metoda prin dublu punct de atașare (**double-end-point**)

aparatul este legat de text prin ancore și prin atributele from și to

```
<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>
```

■ metoda segmentării paralele (**parallel-segmentation**)

textul este divizat în segmente, fiecare cu aparatul său

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

6.4. Statistici și numărători

7

Având aparatul critic codat semantic, este ușor să se extragă automat variațiile textuale ale fiecărui manuscris.

- variațiile pot fi prelucrate statistic pentru a aprecia global tipul și numărul lor
- lista variațiilor folosește pentru a alcătui *stemma codicum*
- textul unui anumit martor poate fi reconstituit automat

 Temă pentru acasă

8

Codați în TEI diferențele dintre următoarele variante textuale:

Text acceptat xml:id="A"	Variantă textuală 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. Reprezentarea aparatului de surse în TEI

7.1. Ontologii și arbori

arbore structură abstractă de date care simulează o arborescență ierarhică (vezi în *Curs 3: rădăcină, părinte, descendenți*)

- pot fi concepute ca un arbore: un text (capitole, subcapitole...), o bibliografie (autor, operă...), un corpus de texte etc.

ontologie (în știința informației) definirea tipurilor, proprietăților și relațiilor care epuizează entitățile unui domeniu

- definirea tuturor claselor, atributelor și relațiilor obiectelor dintr-un corpus de texte reprezintă o ontologie sau un model

clasă = tipologie căreia îi aparține obiectul

atribut = tip de proprietate al obiectului

relație = raport cu alte tipuri de obiecte

```
<rdg cause="homeoteleuton" wit="#MsAA01"/>
clasă atribut relație
```

1

7.2. Descrierea elementelor unei referințe textuale

Sursele și referințele textuale se descriu ca elemente bibliografice.

<listBibl> element părinte al unei liste bibliografice

(poate să apară în antet în <sourceDesc> sau după corpul textului în <back>)

<bibl> element părinte al unei indicații bibliografice

xml:id="□" identificator unic

<author> autorul sursei (simplu sau structurat cu <name>...)

<title> titlul sursei (simplu sau structurat)

<publisher> editura

<editor> numele editorului volumului

<pubPlace> locul publicării

<date> data publicării

<biblScope> scopul referinței (ex. pagini, capitol)

<...> alte proprietăți (vezi Ghidul TEI)

2

Exemplu de bibliografie în 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 Minnesota 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. Alte tipuri de referințe

4

prosopografie listă de persoane, opțional cu date biografice

- folosește elementele <listPerson>, <person> etc.

geografie listă de locații geografice

- folosește elementele <listPlace>, <place>, <geoDecl>, <geo> etc.

cronologie listă de evenimente istorice

- folosește elementele <listEvent>, <event> etc.

Referințele pot apărea în text sau, cel mai adesea, sunt redactate separat și referențiate pe baza identificatorului unic.

Exemplu de referințe în TEI

5

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>
      cepirit <place ref="#Rom">Romam</place></ref>.
    </p>
  </body>
</text>
<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>
  </person>

```

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>

```

7.4. Atașarea referințelor

6

- Referințele se atașează de obicei cu ajutorul identificatorului unic (`xml:id="□"`).
- Apelul la identificatorul unic este un identificator local (prefixat cu #) sau un URL extern (parțial sau complet).
- Identificatorul unic poate conține litere, cifre (dar nu în prima poziție), câteva caractere speciale (. - _).

```

<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"/>

```

 Temă pentru acasă

7

Identificați sursele din următorul text și codați-l în TEI cu o listă bibliografică și trimiteri:

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. Vizualizarea edițiilor digitale

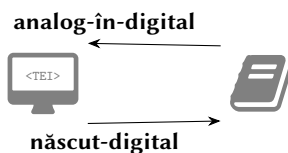
8.1. Conversia în format clasic

Pentru că TEI nu este un format vizual, pentru utilizatorul uman este nevoie de o conversie într-un format familiar. După origine, documentele digitale sunt de 2 tipuri:

analog-în-digital ediții apărute inițial în format analogic (ex. pe hârtie), transcrise electronic și codate ulterior

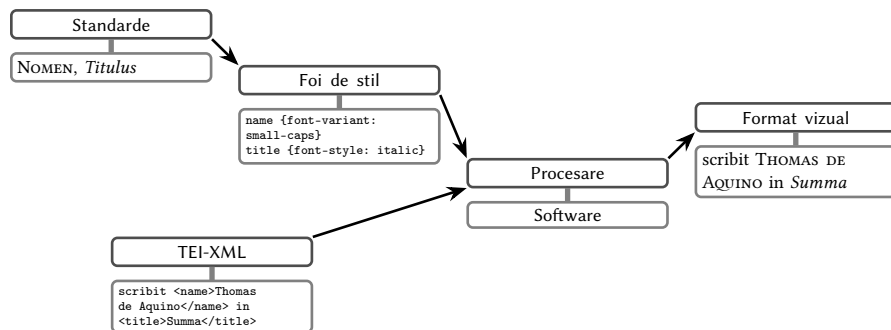
născut-digital ediții elaborate de la început în format digital semantic

Edițiile **născute-digital** au nevoie de conversie pentru cititorul uman. Formatele familiare cititorului includ formatele tipărite și pe cele de ecran care emulează forme analogice.



Conversia în format clasic (2)

Datorită separării straturilor (vezi *Curs 2*), procesul de conversie în format vizual nu necesită modificarea sursei TEI-XML. Conversia se face cu ajutorul unor **foi de stil** elaborate în baza unor **standarde** și care specifică cum este redat vizual fiecare element semantic.



Metode de conversie

1. atașarea unei foi de stil CSS* (Cascaded Style Sheet)
 - se specifică formatarea după principiile CSS pentru HTML(vezi documentația pe <http://www.w3schools.com/css/>)

- se leagă documentul CSS în antetul documentului TEI
(cu declarația `<?xml-stylesheet type="text/css" href="stylesheet"?>`; stilul CSS poate fi un document extern "style.css" sau un element intern "#style")
- nu permite schimbarea ordinii și compoziției elementelor
- 2. atașarea unei foi de transformare XSLT* (eXtensible Stylesheet Language Transformations)
 - se specifică transformarea după principiile XSLT pentru XML
(vezi documentația pe http://www.w3schools.com/xml/xsl_intro.asp)
 - se leagă documentul XSLT în antetul documentului TEI
(cu declarația `<?xml-stylesheet type="text/xsl" href="transform.xsl"?>`)
 - limbajul XSLT este mai complicat, necesită suport suplimentar
- 3. transformarea în LaTeX sau alt format DTP
 - se transformă cu XSLT și alte scripturi în formatul aplicației destinație
 - documentul rezultat se compilează/exportă într-un format comun (ex. PDF)
 - necesită o aplicație specializată și/sau cunoștințe avansate de editare/programare

Foi de stil CSS

4

- Specifică stilul pentru fiecare element XML în formatul:

```
tag {parametru: valoare; parametru: valoare}
```

- Parametri CSS comuni

```
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
```

Exemplu CSS

```
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}
```



formatul se aplică în cascadă elementului specificat și descendenților lui (dacă aceștia nu au alt format). Ex.: quote din p are tot font-size: 12pt

Sintaxa CSS

5

- limbajul CSS specifică stilul în formatul `identificator {parametru: valoare}`
- identificatorul poate conține mai multe elemente diferite, separate cu virgulă:



```
quote, title {font-style: italic} (citatele și titlurile sunt în italic)
```

- stilul poate fi limitat la descendenții unui element superior:



```
bibl title {font-weight: bold} (doar titlurile din bibliografie sunt bold)
```

- identificatorul poate fi un nume de tag (tag), un atribut id (#id) sau o clasă (.class):



```
#codex1 {color: blue} .uncertain {color: gray} (doar elementul cu id="codex1" este albastru, toate elementele cu class="uncertain" sunt gri)
```

- mai mulți parametri pot fi specificați, separați cu punct și virgulă:



```
title {font-style: italic; color: blue} (titlul va fi italic și albastru)
```

- dacă același identificator și același parametru este prezent de mai multe ori, se aplică ultimul:



```
p {color:red} p {color:blue} (paragraful va fi albastru)
```

Imprimat vs. ecran

Mediul de afișare determină restricțiile și posibilitățile formătărilor.

■ imprimat

- necesită așezare în pagină după norme tipografice
 - macro-tipografie: așezarea blocurilor de text în pagină
 - micro-tipografie: la nivelul părților de literă și alăturării literelor
- rezoluția și calitatea grafică trebuie să fie ridicate
- ❗ limite: rândurile și paginile trebuie întrerupte; spațiul este limitat (numărul de pagini este finit, spațiile libere mari trebuie evitate); se evită repetarea elementelor structurale prin trimiteri; folosirea culorilor și imaginilor este restricționată

■ pe ecran

- necesită formatarea conform aplicației de afișare
 - compatibilitate: conformarea cu standardele aplicațiilor (ex. HTML în browser)
 - adaptabilitate: dispozitivele de afișare au dimensiuni și forme diferite, textul trebuie să se afișeze corect pe orice echipament (   ...)
- cantitatea de text și imagini este virtual nelimitată; interactivitate
- ❗ limite: unele funcții și elemente nu sunt disponibile pe toate sistemele; calitatea redării variază în funcție de terminal; mărimea fișierelor trebuie optimizată; standardele se schimbă frecvent, uneori spărgând compatibilitatea

MILE // JassonFAC // SLDC // [index](#) // [Semantics & spelling](#) // [convert from TD](#) // [stats](#) // [T1](#) // [search](#) // [info](#)

ANONYMUS (PSEUDO PETRUS DE ALVERNIA) — *Super Librum de causis*

Super Librum de causis (reportatio Parisiensis Godefridi de Fontibus)

Super Librum de causis

(Prooemium)

(Question 1)

[q1]. [PL3086 \(PL3254\)](#) Secundum [Aristotelem](#) VI [Metaphysicae](#), eadem est scientia que considerat de prima causa et de ente in communi. Nam quia aliquis dicitur de uno per primum et de aliis per posterius, eundem scientiam non considerat de illo in communi et de illis de quo per primum dicitur. Illius enim ratio est primum ratio illius nominis et per habitudinem ad illud omnia alia habent rationem et cognitionem.

[edit](#) [search](#)

[q12]. Praeterea, ad eandem scientiam pertinet considerare de causa aliquos entis et de illo ente, cum scire sit per causam. Nunc autem primum primum causae entis universalis, hoc est ens in quocumque genere. Sunt enim in quocumque genere aliqua causae.

[edit](#) [search](#)

[q13]. Cum enim cognoscit ens et cognoscit causas, imperfecto in cognoscere causas imperfectioris causae in cognoscere effectum, unde si cognoscit ens cognoscit quo ens dicitur in cognoscere primum ens, non cognoscit primum ens per se, secundum quo cognoscitur est a toto. Unde si ens non potest agnoscere cognoscit ens secundum quo ens et cognoscit primum ens, non potest cognoscere ens secundum quo ens aliquid cognoscit primum ens, nec e converso. Unde nulla scientia, si et bene tradita, potest considerari de uno secundum se, quia etiam considerat de ente secundum quo ens, nec e converso.

[edit](#) [search](#)

[q14]. Ergo, eadem debet esse scientia de ente in communi et de causis primis entis universalis. Est ergo scientia de primis causis per effectus sensibiles et de primis manifestos. Sed, si huiusmodi scientias artificialiter tradidit aut, debet esse simul et de ente in communi, et e converso. In scientia tamen *Libri de causis* subiectionem primum causae aliquid hoc quod in ea determinatur de ente in communi, scilicet in metaphysica. Dicuntur autem causae primae pluraliter, licet sit omnium una simpliciter prima, quia quaelibet causarum primarum prima est in suo genere.

[edit](#) [search](#)

(Propositio 1)

(Question 2)

[q2]. Aliquis naturalis non autem influendo aliquid aliis corripit, quia sic non invenit nisi ens causa cognoscit

Secundum ARISTOTELUM *P1 Metaphisic*, eadem est scientia que
considerat de primis causis et de eadem in comuni. Nam quando
aliquid dicitur de uno per prius et de alio alio post, ostendit
esse eandem scientiam, quia non est aliud quod sit de uno
per prius dicitur. Illius enim ratio est prima ratio illius no-
minis et per habitudinem ad illud omnia alia habent rationem
et cognoscuntur. Unde et eadem scientiam pertinet cognoscere
de causis aliquis entis et de illo ente, cum scire sit causae;
nunc autem prime causae sunt causae universales hoc genus
entis in quacunque genere. Sunt enim in unoquoque genere alia
causata.

Cum enim cognitio effectus sit ex cognitione causae, imperfectio
in cognitione causae imperfectionem causat in cognitione effectus.
Unde si cognitio entis secundum quod ens dicitur in cognitionem
primam non est cognoscens effectum, quia non est cognitum
quod cognoscibilis est a nobis, tunc cognitio entis non est perfecta.
Unde ista non possunt separari: cognitio entis secundum quod
ens et cognitio primi entis. Nam non potest cognosci ens secundum

4 eadem ARISTOTELIS, *Metaphisic*, VI, 1, 1026a29-32. THOMAS
DE AQUINO, in *Secunda littera Metaphisicorum operum*, lib. 6, 1. 1. 27 § de uno
posteriori. ARISTOTELIS, *Metaphisic*, VI, 1, 1026a29-32. THOMAS DE AQUINO, *Quaestiones*
disputatae de veritate, q. 1, 2. 7, arg. 5. §. 1. ARISTOTELIS, *Metaphisic*, VI, 1, 1026a29-32.
ARISTOTELIS, *Metaphisic*, I, 2. 7. 78a35.

5 de alio alio § de alio P2 alii alii in marg. P1 et in illud) uno P2
§ et alii P2

3

9

8.3. Includerea imaginilor manuscriselor

Atributul `facss=""` se poate atașa multor elemente TEI și specifică legătura cu o imagine care corespunde elementului. Legătura poate fi făcută în 2 moduri:

- **fișier**: se specifică direct un fișier imagine (jpg, png, tiff, gif...)

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

- **identificator unic facsimile**: se specifică un identificator unic `xml:id` declarat într-un element `<facsimile>`

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

- elementul `facsimile` permite controlul mai precis al imaginii prin specificarea de multiple imagini, suprafețe și zone

```
<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>
```

Standarde de publicare a facsimilelor

- **IIIF** (International Image Interoperability Framework) este un standard tot mai răspândit pentru publicarea, transmiterea și re folosirea imaginilor manuscriselor (și nu numai) cu metadata

 <http://iiif.io/>

IIIF folosește formatul de fișier JSON (JavaScript Object Notation)

Site-uri care folosesc IIIF: British Library, Bibliothèque nationale de France, Bibliotheca Apostolica Vaticana, Virtual Manuscript Library of Switzerland, IRHT (CNRS), Bayerische Staatsbibliothek etc. (peste 70 instituții)

- TEI poate fi de asemenea folosit pentru publicarea facsimilelor cu metadata

Site-uri care folosesc TEI pentru facsimile: Manuscriptorium, Oxford Bodleian Library etc.

8.4. Capcanele vizualului

- Modul de afișare al elementelor induce o pre-interpretare a conținutului

☞ elementele afișate mai mare, mai colorat sau mai sus par mai importante

- Selecția elementelor afișate inițial cauzează ignorarea celor afișabile prin interacțiuni complexe

☞ utilizatorul neexperimentat poate crede că ediția se reduce la ce se vede pe prima pagină

- Interacțiunile predeterminate induc sentimentul de analiză exhaustivă

☞ oricât de complexe și numeroase, interacțiunile pre-programate nu epuizează întrebările

Exemplu fictiv de afișare care poate duce în eroare:

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 theologiae, 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.

Temă pentru acasă

12

Operați cel puțin 5 modificări vizibile și valide în următoarea foaie de stil:

Foaie de stil CSS

```
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. Integrarea și procesarea digitală

9.1. Indici și tabele de concordanțe

Extragerea unor elemente semnificative (ex. titluri, nume, citate, termeni) se poate face automat și pe loc în măsura în care:

- elementele sunt marcate cu tag-uri și/sau atribute
- se folosește un procedeu de interogare, extragere și afișare a elementelor

9.2. Limbaje de interogare: XQuery, XPath

- XQuery (XML Query) este un limbaj funcțional de interogare pentru colecțiile de date în format XML
- XPath (XML Path Language) este un subset al XQuery, definit ca limbaj de interogare pentru selectarea nodurilor din XML

Exemplu XPath (1)

Dacă este dat următorul cod XML:

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

expresia pentru a selecta toate nodurile <p> din <body> este:

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

Exemplu XPath (2)

Pentru a extrage toate elementele cu atributul class="sic" din următorul XML:

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

expresia XPath poate fi:

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

9.3. Căutarea în text

Căutarea în text poate fi de mai multe tipuri:

- **căutare simplă**: se caută șirul de caractere dat

➞ sum ⇒ ego sum, summa, assumptio, sensum

- **căutare cu metacaractere**: se caută cuvintele care se potrivesc șirului de caractere dat și cu orice caractere în locul metacaracterelor

➞ sum* ⇒ ego sum, summa, assumptio, sensum

- **căutare structurată:** se caută șirul de caractere dat în anumite elemente de structură



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

9.4. Lematizare, normalizare, dicționare

4

Căutarea după șiruri de caractere nu este întotdeauna satisfăcătoare, întrucât termenii pot apărea în forme diferite din cauza flexiunii gramaticale și a normelor ortografice schimbătoare.



res, rei ⇒ res; rei; rem; re; rerum; rebus; rex; ires; redde; resto; recitares

Tehnici de dezambiguizare (în ordinea complexității):

- **normalizare:** aducerea textului la o formă standard
- **stematizare:** extragerea stemei (rădăcinii) cuvântului



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

- **lematizare:** atașarea formei standard a cuvântului la fiecare ocurență



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

Exemplu TEI

```
<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>
```

- **dicționar:** conectarea cuvintelor la un dicționar intern sau extern
- **analiză:** atașarea întregii analize gramaticale la un cuvânt

9.5. Digital corpora

5

- Un **corpus de texte** este o mare colecție structurată de texte selectate după un anumit criteriu (subiect, perioadă, autor, mediu intelectual etc.)
- Un **corpus digital** este un set mare de date (text și metadata) care cuprinde un corpus de texte cu adnotări și care acoperă (cel puțin la nivel de concept) un anumit criteriu principal.
 - datorită intenției de completitudine, corpus-ul este construit arborescent ca o ontologie (vezi *Curs* 7)
 - un corpus digital este organizat după principii științifice și funcționale care trebuie să asigure fiabilitatea exploatarei lui, incluzând:
 - acces la versiunile standardizate ale fiecărui text
 - căutări simple și structurate în întreaga colecție
 - analize statistice exhaustive
 - analize automate de conținut (ex. analize lingvistice, validarea datelor)
 - testarea ipotezelor, cu asigurarea unui grad mare de încredere
 - transparența și predictibilitatea structurilor



Proiectarea unui corpus este o sarcină complexă și importantă, deoarece trebuie să acopere toate manifestările posibile ale conținutului, deseori fără a avea acces complet la conținut.

6

9.6. Data-mining

■ **Data mining** (mineritul datelor) este procesul de analiză computerizată a seturilor mari de date pentru identificarea de modele semnificative care ajută înțelegerea și interpretarea conținuturilor.

■ procesul este extrem de complex, implicând sisteme de baze de date, statistică, semantică, inteligență artificială și interpretare intelectuală

■ scenarii de utilizare:

■ validarea unor ipoteze de lucru



dovedirea sau respingerea existenței unui model (ex. influența unui anumit factor)

■ descoperirea de modele semnificative



descoperirea de regularități necunoscute anterior (ex. un factor major, dar ignorat)

■ analiza predictibilității modelelor



folosind modele verificate, se poate proba apartenența conținutului (ex. paternitatea)

7

Temă pentru acasă

Scrieți o expresie XPath care să extragă din următorul fragment TEI toate ocurențele verbului *sum*, *esse*, *fui*.

ARISTOTELES, *Perihermeneias*

```
<p><w lemma="amplius">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. Inteligența artificială

10.1. Inteligență artificială, Machine Learning

Inteligență artificială (AI) = sisteme informatice capabile să rezolve probleme mimând funcții “cognitive” ale minții umane.

1

Machine Learning (ML) = partea din AI care studiază aplicarea de **algoritmi** și **modele statistice** fără instrucțiuni precise, ci bazându-se pe **structuri repetitive** și **inferențe**, prin care sistemul ia **decizii** și stabilește **predicții**.

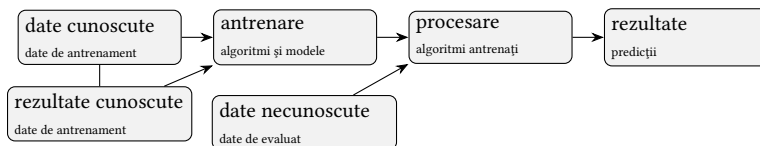
2

Procesul informatic

Programarea bazată pe instrucțiuni (programarea clasică):



Programarea bazată pe ML:



ML: Modelul cutiei negre:



3

10.2. Algoritmi și modele

- **Model** = o descriere a unui sistem de date pe baza unor asumții statistice.
- **Algoritm** = o secvență finită de instrucțiuni care execută o sarcină dată.
- În procesarea informatică, un **set de date** este definit printr-un **model** și **conținutul de date** corespunzător. Datele sunt prelucrate de **algoritmi** ținând cont de **model**.
- În ML, **algoritmii** sunt generați de sistemul informatic prin inferențe și analize statistice ale **datelor existente** (= **antrenament**), apoi pot fi aplicați unor **date noi** care corespund aceluiași **model** (= **predicție**).

10.3. Antrenament și predicție în ML

Antrenamentul

■ Componente necesare:

☞ *Exemplu: Recunoașterea imaginilor*

■ modelul de date

- trebuie să descrie corect și complet datele

☞ *Ex: Descrierea structurii datelor într-un format de fișier de imagine*

■ setul de date de antrenament

- trebuie să fie statistic reprezentativ și să cuprindă multe înțări

☞ *Ex: Zeci de mii de imagini de calitate și cât mai variate*

■ rezultatele valide pentru datele de antrenament

- trebuie să fie corecte și să acopere orice detaliu așteptat în rezultate

☞ *Ex: Etichete pentru fiecare imagine*

■ sistemul software și echipamentul IT adecvat

- calitatea trebuie să fie verificată, cerințele de sistem sunt foarte înalte

☞ *Ex: Program care suportă procesarea imaginilor, putere de calcul (cloud?)*

Predicția

- va returna un set de rezultate corespunzând datelor noi, împreună cu o indicare a acurateței predicțiilor

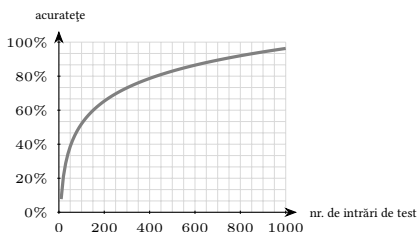
☞ *Ex: Etichetele cele mai probabile pentru noile imagini, și procentajul de acuratețe pentru fiecare asociere*

!!! Rezultatul nu este unul categoric, nici necesar corect, ci este starea la care ajunge sistemul aplicând algoritmi antrenati din model

5

Acuratețea rezultatelor

- Acuratețea crește logaritmically cu cât sunt introduse mai multe date de antrenament:



6

- Este imposibil de atins acuratețea de 100% (adică certitudinea absolută), dar deseori o acuratețe de 90–99% este considerată acceptabilă, în funcție de scop.
- O mare parte din efort în ML se depune în pregătirea datelor.

10.4. Tipuri de abordare în ML

În funcție de tipurile de date și de tipurile de rezultate așteptate, există abordări diferite ale învățării automate și modelelor de date.

7

- Inteligența artificială este extensiv de specializată și necesită alegerea corectă a tipurilor de modele și de învățare pentru a produce rezultate satisfăcătoare.
- AI este un domeniu extrem de dinamic astăzi, iar abordările se schimbă frecvent, de aceea o prezentare exhaustivă nu este posibilă.

Tipuri de învățare automată

- **Învățarea supravegheată**
 - antrenarea pe baza unui model cunoscut
- **Învățarea fără supraveghere**
 - detectarea structurii unui set de date necunoscut
- **Învățarea supravegheată parțial**
 - antrenare cu un set cunoscut, apoi perfecționare cu un set necunoscut
- **Învățarea prin armare**
 - pe bază de recompensă
- **Detectarea anomaliei**
 - identificarea intrărilor rare
- **Reguli de asociere**
 - identificarea regulilor într-o structură de date necunoscută

Tipuri de modele

- **Rețele neuronale artificiale**
 - model bazat pe o colecție de noduri de procesare ("neuroni artificiali") interconectate ("sinapse")
- **Arbori decizionali**
 - un model predictiv care pornește de la observarea unui element pentru a genera concluzii privind valoarea elementului în funcție de poziția lui în arbore
- **Mașini cu vectori de suport**
 - metode de învățare supravegheată care prezic dacă un element cade într-una din 2 categorii posibile
- **Rețele bayesiene**
 - model grafic ce reprezintă un set de variabile aleatoare și caracteristicile asociate lor
- **Analiza regresiei**
 - metode statistice de estimare a relaționării între variabilele de intrare și caracteristicile asociate lor
- **Algoritmi genetici**
 - algoritmi de căutare care simulează selecția naturală pentru a genera noi genotipuri
- **Modele de bază (Foundation models)**
 - modele AI foarte mari, antrenate cu o cantitate vastă de date și adaptabile la o multitudine de sarcini, ex. chat

Sfârșitul cursului