

Αξιοποίηση Ηλεκτρονικών Πόρων στην Ελληνική Φιλολογία-Digital Humanities (2): Κωδικοποίηση λογοτεχνικών και άλλων κειμένων- Text Encoding Initiative

Στέλλα Μαρκαντωνάτου

Υπολογιστική Γλωσσολόγος

Διευθύντρια Ερευνών

Ινστιτούτο Επεξεργασίας του Λόγου/ΕΚ ΑΘΗΝΑ

stilianimarkantonatou@gmail.com

Ηλεκτρονικά κείμενα,
επισημείωση και markup
E-Texts, Annotation & Markup

Ηλεκτρονικά κείμενα, σχολιασμός και markup

E-Texts, Annotation & Markup

markup: The symbols used in some types of computer documents that control the way text **appears on a screen or page**, or the process of adding these symbols:

- XML markup has strict rules about what can go where.
- Use the right tags in your mark-up or the page will not display correctly.

- Οι σύγχρονοι κειμενογράφοι (editors) όπως το Word ή τα προγράμματα μορφοποίησης (π.χ. LATEX) στην ουσία «επισημειώνουν» το κείμενο με ειδικούς χαρακτήρες που ερμηνεύονται από τα προγράμματα που δημιουργούν την τελική μορφή του (αυτήν που διαβάζουμε ή τυπώνουμε).
- Τα κείμενα ---χειρόγραφα ή τυπωμένα--- που προέρχονται από ψηφιοποίηση ή από κανονική αντιγραφή προφανώς δεν έχουν τέτοιους χαρακτήρες.

Γιατί

**επισημειώνουμε τα κείμενα που
αποθηκεύουμε ως προς την μορφή τους;**

Διότι

η μορφή των κειμένων φέρει

ΣΗΜΑΣΙΑ

Ηλεκτρονικά κείμενα, σχολιασμός & markup
E-Texts, Annotation & Markup

**ΕΠΙΣΗΜΕΙΩΣΗ ΩΣ ΠΡΟΣ ΤΑ
ΓΛΩΣΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ
ΤΟΥ ΚΕΙΜΕΝΟΥ**

ΜΟΡΦΕΣ ΜΕ ΣΗΜΑΣΙΑ

- Παραθέσεις:** Πόσους και ποιους παραθέτει αυτούσιους στα έργα του ο Παναγής Λεκατσάς;
 - Σημειώσεις στο περιθώριο:** Πόσες και ποιες σημειώσεις βλέπουμε στο περιθώριο στο χειρόγραφο του Γιούγκερμαν του Καραγάτση;
 - Γραφικοί χαρακτήρες (για χειρόγραφα):** Πόσοι μοναχοί τελικά δούλεψαν πάνω στο χειρόγραφο της «Πολιτείας» του Πλάτωνα που φυλάσσεται στο Βατικανό;
 - Διορθώσεις (για δακτυλογραφημένα έντυπα):** Πόσες και ποιες διορθώσεις περιέχει το δακτυλογραφημένο κείμενο του Τρίτου Στεφανιού του Ταχτσή;
- Χρησιμοποιούμε μια ειδική γλώσσα, την XML, και αναπτύσσουμε ένα πρότυπο για τον σκοπό αυτό.**

TEXT ENCODING INITIATIVE

(Πρωτοβουλία για την
κωδικοποίηση του κειμένου)

<http://www.tei-c.org/>

Τί είναι η ΤΕΙ;

- Κοινοπραξία που αναπτύσσει συλλογικά και συντηρεί ένα **πρότυπο (standard) για την αναπαράσταση των κειμένων σε ψηφιακή μορφή**.
- Η βασική της συνεισφορά είναι ένα σύνολο **Οδηγιών (Guidelines)** που καθορίζει πώς πρέπει να κωδικοποιούνται τα κείμενα που μπορούν να διαβάσουν οι μηχανές, κυρίως στον τομέα των Ανθρωπιστικών και Κοινωνικών Επιστημών και της Γλωσσολογίας.
- Από το 1994 οι Οδηγίες εφαρμόζονται ευρέως από βιβλιοθήκες, μουσεία, εκδότες, και ιδιώτες διανοούμενους που διαθέτουν κείμενα σε ηλεκτρονική μορφή για έρευνα, διδασκαλία και διατήρηση.
- Εκτός από τις Οδηγίες, η κοινοπραξία διαθέτει ελεύθερα ένα σύνολο από πόρους, διδακτικό υλικό, σχετιζόμενες δημοσιεύσεις και **λογισμικό**.

The TEI Consortium is a nonprofit membership organization composed of academic institutions, research projects, and individual scholars from around the world. Members contribute financially to the Consortium and elect representatives to its Council and Board of Directors. In commemoration of the TEI community's 30th anniversary, it was awarded the 2017 Antonio Zampolli Prize from the Alliance of Digital Humanities Organizations.

Το πρότυπο πρέπει να μπορεί να αναπαραστήσει
όσα κειμενικά χαρακτηριστικά είναι χρήσιμα στην
έρευνα:

- απλά και με σαφήνεια
- με οποιοδήποτε λογισμικό
- επιτρέποντας την επεξεργασία κειμένου
- με δυνατότητα επεκτασιμότητας
- σε συμφωνία με τα πρότυπα
- διευκολύνοντας το μοίρασμα των κειμένων

<http://www.tei-c.org/guidelines/customization/Lite/>

Παράδειγμα: ΤΕΙ ΛΙΤΕ, σελ. 3, 4, 5

- Διαφορά OCR από σάρωση
- Δεδομένα
 - Εμφανή
 - Υπόρρητα
- Μεταδεδομένα
 - Όλα τα δεδομένα γίνονται εμφανή

Extensive Markup Language (XML)

Extensible Markup Language (XML)

Η XML είναι μία μεταγλώσσα

Η XML είναι το κυρίως χρησιμοποιούμενο σχήμα κωδικοποίησης

- για την ανταλλαγή πληροφοριών και την επικοινωνία στο World Wide Web
- Για τον ορισμό μεθόδων αποθήκευσης και επεξεργασίας ηλεκτρονικών κειμένων με τρόπο ανεξάρτητο της εκάστοτε χρησιμοποιούμενης μηχανής (πχ. φορητός, έξυπνο κινητό) και του εκάστοτε χρησιμοποιούμενου συστήματος (πχ Windows, Linux).

Ηλεκτρονικά κείμενα, σχολιασμός και markup

E-Texts, Annotation & Markup

markup:

-Παραδοσιακά, δηλώνει τον σχολιασμό και γενικά τα σημάδια πάνω σε ένα κείμενο που καθοδηγούν τον τυπογράφο στην μορφοποίηση.

-Στην ηλεκτρονική εποχή, ο όρος καλύπτει και τους ειδικούς χαρακτήρες που εισάγονται στα ηλεκτρονικά κείμενα για να καθοδηγούν την μορφοποίηση, το τύπωμα και, σε κάποιες περιπτώσεις, την επεξεργασία των κειμένων.

Ένας ακόμη τρόπος **ερμηνείας** ενός κειμένου

Η XML είναι

- Δηλωτική και όχι διαδικασιακή: απλά κατηγοριοποιεί τμήματα ενός κειμένου.
- Διακρίνει τις έννοιες: **Document Type Definition (DTD)** πχ επιστολή, αναφορά, ποίημα , και **συντακτική ορθότητα (validation)**, την οποία ελέγχει ειδικό πρόγραμμα (**parser**)
- Είναι ανεξάρτητη hardware και λογισμικού: όλα τα αρχεία XML είναι σε Unicode.



All Fields & Tags



- My Library
 - academia book
 - kalamata course
 - bib
 - my material
 - My Publications
 - Duplicate Items
 - Unfiled Items
 - Trash

Title	Creator	
> Cookbook for Open Access books	Nordhoff and Press	•
> Anemoskala. Corpus of Modern Greek Poetry – TEI: Text Encoding Initiative		•
> Anemoskala - Akritidou 2015.pdf		•
> Developing Research Tools via Voices from the Field – dh+lib		•
> Greek Manuscripts in Sweden – TEI: Text Encoding Initiative		•
> Introduction to Metadata: Intro to Metadata		•
> Introduction to XML for Text		•
> Perseus Project – TEI: Text Encoding Initiative		•
> Projects Using the TEI – TEI: Text Encoding Initiative		•
> TEI by Example. Module 8: Customising TEI, ODD, Roma		•
> TEI: Text Encoding Initiative		•
> The Development of a Local Thesaurus to Improve Access to the Anthropological Collections of ...		•
> TEI Lite: Encoding for Interchange: an introduction to the TEI — Final revised edition for TEI P5	Burnard and Sperberg-McQueen	•
> tei_lite.doc.pdf		•
> When Metadata Becomes Outreach: Indexing, Describing, and Encoding For DH	Wilson and Alexander	•
> v. A Gentle Introduction to XML - The TEI Guidelines		•
> Version 3.4.0. Last updated on 23rd July 2018, revision 1fa0b54		
> v. A Gentle Introduction to XML - The TEI Guidelines		•

Info	Notes	Tags	Related
Item Type	Web Page		
Title	v. A Gentle Introduction to XML - The TEI Guidelines		
Author	(last), (first)		
Abstract			
Website Title	Text Encoding Initiative		
Website Type			
Date			
Short Title			
URL	http://www.tei-c.org/release/doc/tei-p5-do...		
Accessed	11/8/2018, 9:08:13 μ.μ.		
Language			
Rights			
Extra			
Date Added	11/8/2018, 9:08:13 μ.μ.		
Modified	12/8/2018, 9:14:17 π.μ.		

lecture_1 my_material

ΒΙΒΛΙΟΓΡΑΦΙΑ



ΠΑΡΑΔΕΙΓΜΑ 1

ΕΡΜΗΝΕΙΑ 1

```
<anthology>
  <poem>
    <author>Γιώργος Σεφέρης</author>
    <date >16 Γενάρη '42</date>
    <place_of_creation> Πρετόρια </place of creation >
    <title>Τριζόνια</title>
    <stanza>
      <line>Το σπίτι γέμισε τριζόνια</line>
      <line>χτυπούν σαν άρρυθμα ρολόγια</line>
      <line>λαχανιασμένα. Και τα χρόνια</line>
    </stanza>
    <stanza>
      <line>που ζούμε σαν αυτά χτυπούν</line>
      <line>καθώς οι δίκαιοι σιωπούν</line>
      <line>σα να μην είχαν τι να πουν.</line>
    </stanza>
  </poem>
  <!-- more poems go here -->
</anthology>
```

DTD-Validation του παραδείγματος 1

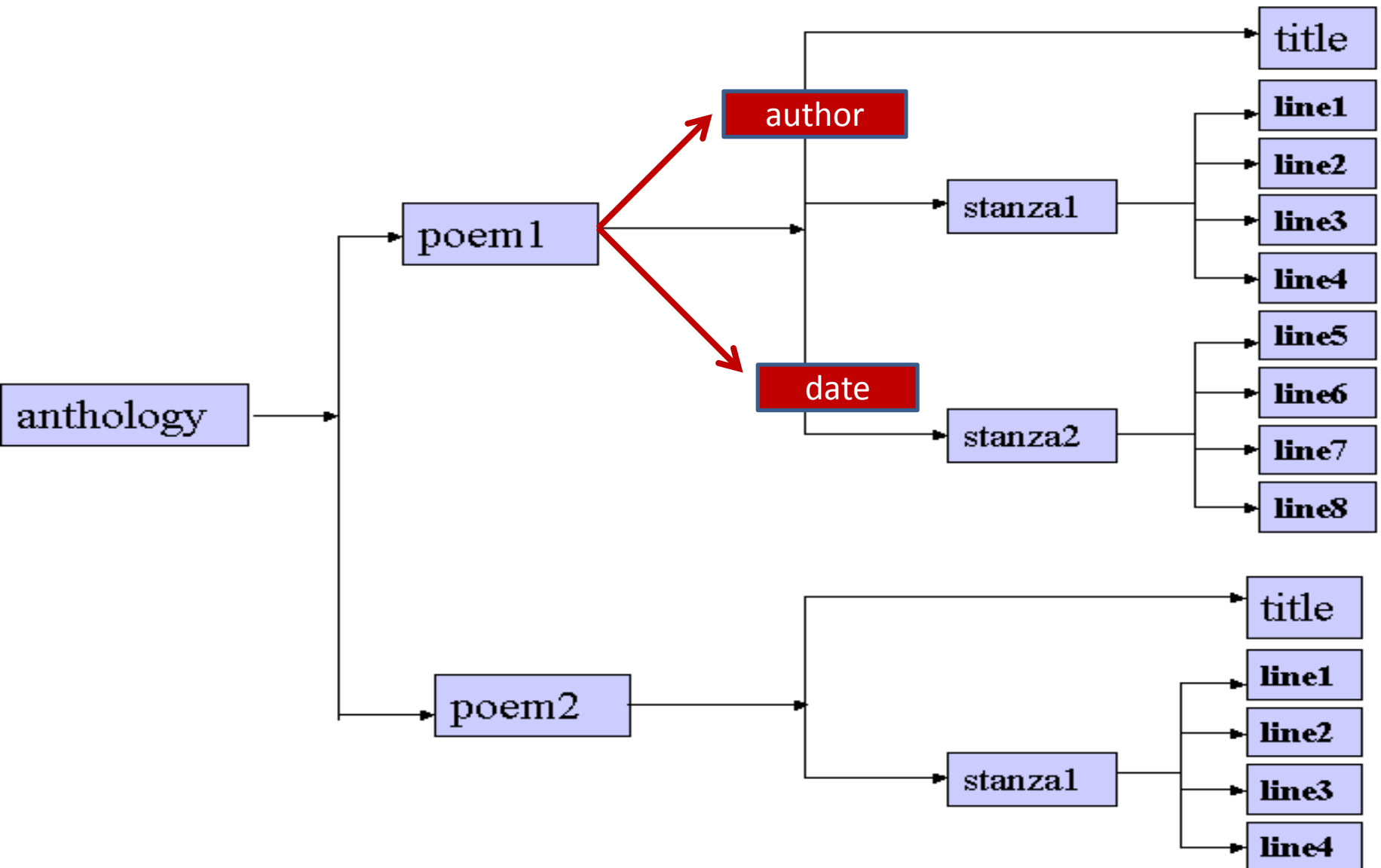
```
anthology_p = element anthology {poem_p+}
  poem_p = element poem {author?, date?,
    place_of_creation?, title_p?, stanza_p+ }
  author = element author {text}
  date = element date {text}
  place_of_creation = element place_of_creation {text}
  stanza_p = element stanza {line_p+}
  title_p = element title {text}
  line_p = element line {text}
start = anthology_p
```

ΣΧΗΜΑ 1

Γιατί φτιάχνουμε σχήματα;

- Μια σταθερή υπόθεση για την δομή ενός κειμένου μέσα στο ίδιο πρόγραμμα κωδικοποίησης—όσο μινιμαλιστική και να είναι αυτή—
- Δυνατότητα προτυποποίησης για πολλά προγράμματα κωδικοποίησης

XML: ιεραρχική αναπαράσταση της κειμενικής δομής



ΠΑΡΑΔΕΙΓΜΑ 2

```
<anthology>
  <poem>
    <poem author="Γιώργος Σεφέρης" date="16 Γενάρη '42" place_of
      _creation="Πρετόρια"/>
    <title>Τριζόνια</title>
    <stanza>
      <line>Το σπίτι γέμισε τριζόνια</line>
      <line>χτυπούν σαν άρρυθμα ρολόγια</line>
      <line>λαχανιασμένα. Και τα χρόνια.</line>
    </stanza>
    <stanza>
      <line>που ζούμε τα πάντα χτυπούν</line>
      <line>καθώς ο δόκτορας τσιπών</line>
      <line>σε μια μηχανή είχαν τι να πουν.</line>
      <line>Don't let thy life destroy.</line>
    </stanza>
  </poem>
<!-- more poems go here -->
</anthology>
```

Βλ. Ανεμόσκαλα, 663

DTD-Validation του παραδείγματος 2

att.author = attribute author {text}

att. date = attribute date {text}

att. place_of_creation = attribute place_of_creation {text}

anthology_p = element anthology {poem_p+}

poem_p = element poem {att.author?, att.date?,
att.place_of_creation?, title_p?, stanza_p+ }

stanza_p = element stanza {line_p+}

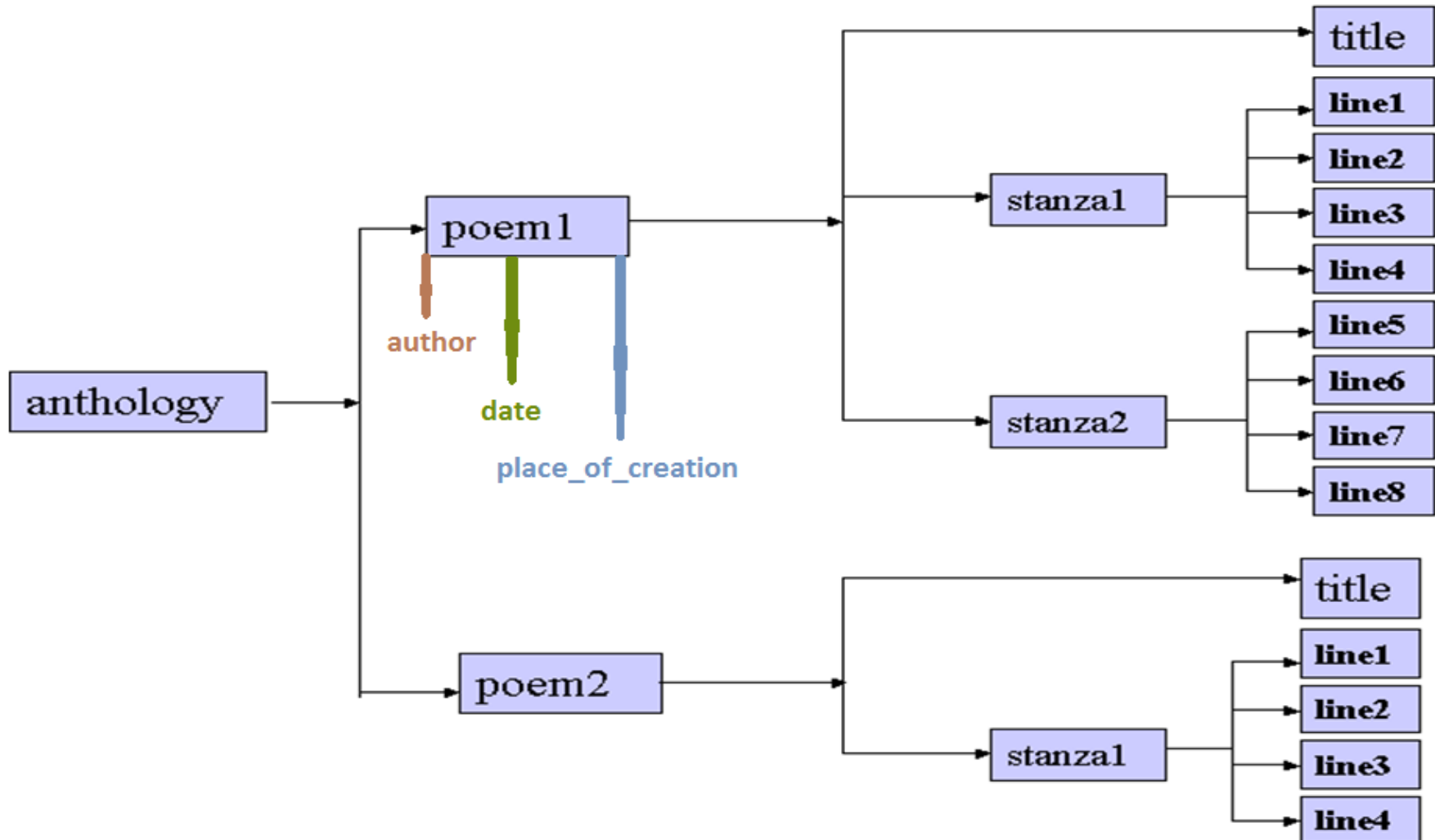
title_p = element title {text}

line_p = element line {text}

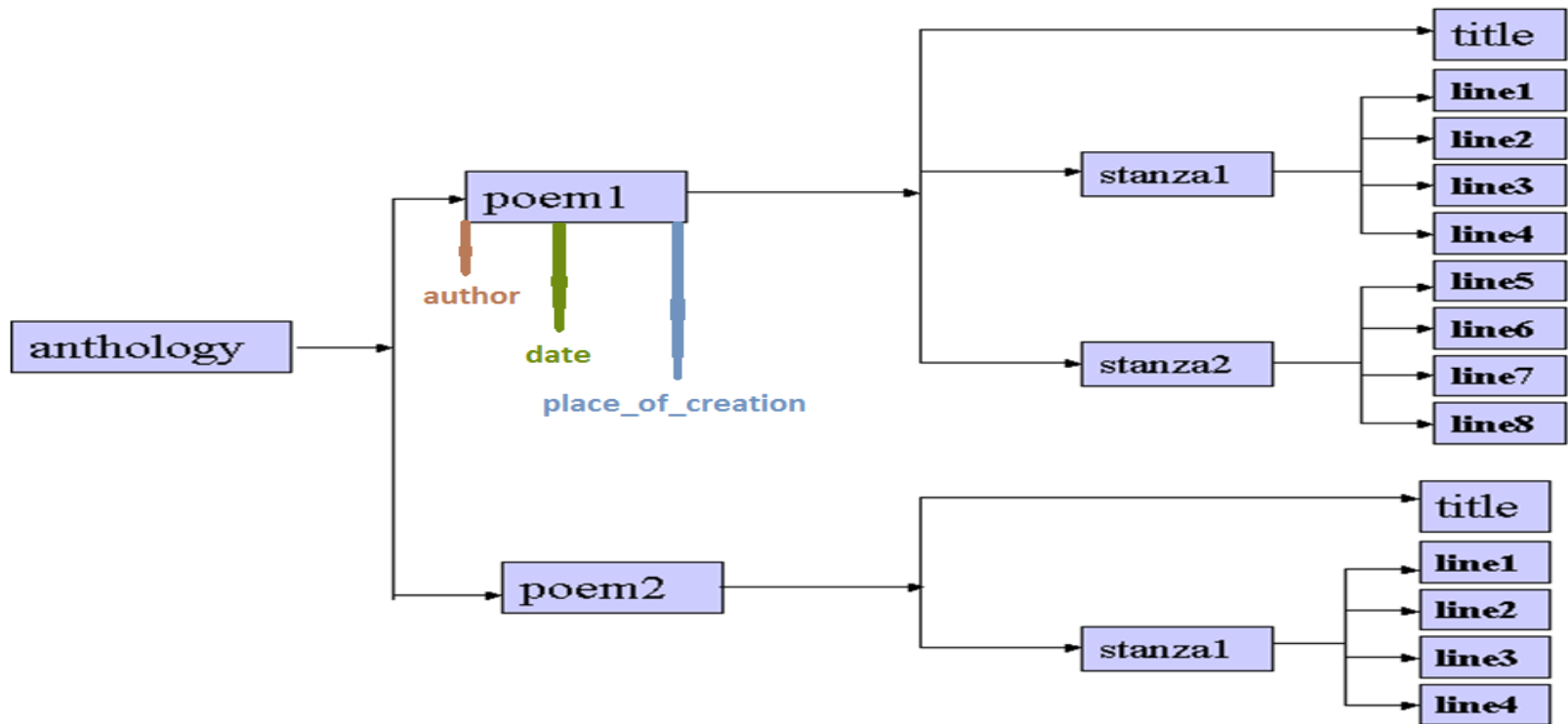
start = anthology_p

ΣΧΗΜΑ 2

XML: ιεραρχική αναπαράσταση της κειμενικής δομής + χαρακτηριστικά (attributes)



XML: ιεραρχική αναπαράσταση της κειμενικής δομής + χαρακτηριστικά (attributes)



Ποια είναι η σχέση των οντοτήτων που ταξινομούνται με την πιο πάνω DTD;

Αγιάνναπα, Β', Γ. Σεφέρης

- http://www.greek-language.gr/Resources/literature/tools/concordance/browse.html?cnd_id=1&text_id=3176

Ο δαίμων της πορνείας, Γ. Σεφέρης

- http://www.greek-language.gr/Resources/literature/tools/concordance/browse.html?cnd_id=1&text_id=3178

- **Μικρό λεξικό φιλολογικών και αρχαιογνωστικών όρων**
- http://www.greek-language.gr/digitalResources/ancient_greek/history/grammatologia/page_093.html

Ανεμόσκαλα

- <http://www.greek-language.gr/Resources/literature/tools/concordance/index.html>

To extend our example further, we might decide to **add a linguistic analysis to each of the poems**, using a set of **PoS** elements such as <aux>, <adj>, etc., derived from some pre-existing XML vocabulary for linguistic analysis.

```
<anthology>
```

```
<!-- anthology markup elements here -->
```

```
<line>
```

```
<gram:itj>O</gram:itj>
```

```
<gram:nom>Rose</gram:nom>
```

```
<gram:pron>thou</gram:pron>
```

```
<gram:aux>art</gram:aux>
```

```
<gram:adj>sick</gram:adj>
```

```
</line>
```

```
</anthology>
```

P5: Guidelines for Electronic Text Encoding and Interchange

To extend our example further, we might decide to **add a semantic linguistic analysis to each of the poems...**

```
<anthology>
```

```
<!-- anthology markup elements here -->
```

```
<line>
```

```
<gram:adj>μνησιπήμων</gram:adj>
```

```
<gram:adj pyli.relation="http://www.greek-  
language.gr/digitalResources/ancient_greek/library/search.html?lq  
=word:609">
```

```
</line>
```

```
</anthology>
```

P5: Guidelines for Electronic Text Encoding and Interchange



Μνημοσύνη

Ψηφιακή Βιβλιοθήκη της Αρχαίας Ελληνικής Γραμματείας

Μνημοσύνης δ' ἑξαῦτις ἐράσσατο καλλικόμοιο,
ἔξ ἧς οἱ Μοῦσαι χρυσάμπυκες ἐξεγένοντο
έννεα, τῆσιν ἄδον θαλῖαι καὶ τέρψις αἰοιδῆς.
— Ησίοδος, Θεογονία 915-7

Αρχική Συγγραφείς Γένη / Είδη Συμφραστικός πίνακας

Φίλτρα αναζήτησης

Καθαρισμός Εφαρμογή

- Συγγραφείς
- Γένη / Είδη

Αναζήτηση ΑΕ λέξεων

word:609 Βρες

Επιλογές αναζήτησης

Ακριβής αναζήτηση Ο τονισμός είναι σημαντικός

Λέξη: "μνησιπήμων"

Βρέθηκε 1 εμφάνιση

⏪ ⏩ ⏴ ⏵

ΑΙΣΧ Ἀγ 180 στάζει δ' ἄνθ' ὕπνου πρὸ καρδίας | **μνησιπήμων** πόνος· καὶ παρ' ἅ- | κοντας ἦλθε

- A
- B
- Γ
- Δ
- E
- Z
- H
- Θ
- I

Ένα πρόσφατο ηλεκτρονικό μήνυμα

Gina Hurley <gina.hurley@yale.edu> / 4:52 PM (50 minutes ago) /to TEI-MS-SIG

Dear colleagues,

At this year's International Medieval Congress at Kalamazoo, the Beinecke Rare Book & Manuscript Library and the Digital Editing group will be hosting a special roundtable on the topic of "Digital Editing and Pedagogy" (the full CFP may be found below). I write to invite you to propose papers, and to ask that you circulate our call widely. We would love to hear about how you've brought digital editing into your undergraduate and graduate classrooms. Please feel free to contact me directly with any questions!

Best wishes, Gina M. Hurley

1. The Pedagogy of Digital Editing (Roundtable)

Organizer: Gina M. Hurley

This roundtable asks how digital editing of medieval manuscripts can be brought into graduate and undergraduate classrooms and what pedagogical advantages (and disadvantages) it has. While the editing process introduces students to digital tools, it can also instill more traditional skills, as the tasks involved in creating digital markup—tagging poetic forms, historical figures, or iconographic and paleographic features—require students to engage closely with the textual and material features of a primary source.

We welcome proposals that discuss pedagogical approaches to digital editing of medieval materials. First and foremost, we are interested in the different shapes that this training can take—as a unit, a final project, a workshop, or even as an entire class. Other questions might include: How can the creation of collaborative digital editions encourage multimodal literacies in the classroom? How can this training be added into the structure of existing classes and majors? What are the advantages and disadvantages of including digital humanities training in a traditional liberal arts curriculum?

We are keen that this roundtable should include papers from institutions both with and without extensive manuscript holdings.

Please send 250-word proposals to digitalmanuscriptrolls@gmail.com by September 15, 2018.

Για το επόμενο μάθημα:

A companion to digital humanities

http://digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-4-2&toc.id=0&brand=9781405103213_brand

Για το επόμενο μάθημα:

Βασιλική Αφεντουλίδου και Αναστασία Χριστοφίδου. 2017. Σώμα εποπτείας νεολογισμών της Νέας Ελληνικής: Σχεδιασμός και κειμενική ταξινόμηση. 14^ο Τεύχος του Δελτίου Επιστημονικής Ορολογίας και Νεολογισμών της Ακαδημίας Αθηνών, 129-180

