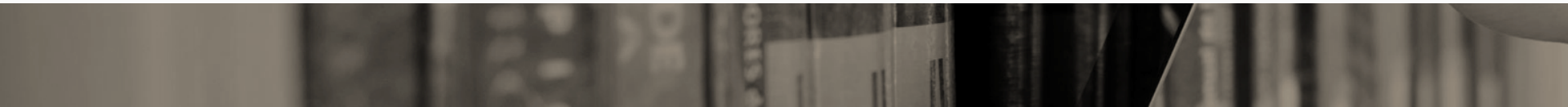
A black and white photograph of a hand holding a tablet computer in a library. The background shows a row of books on a shelf, with some spines visible. The text is overlaid on the bottom left of the image.

Δρόμοι και τρόποι για τη συνάντηση της βιβλιοθήκης
με την ερευνητική/εκπαιδευτική πράξη
στις Ψηφιακές Ανθρωπιστικές Σπουδές

Κατερίνα Τικτοπούλου & Φωτεινή Κοϊδάκη



Δρόμοι και τρόποι για τη συνάντηση της βιβλιοθήκης
με την ερευνητική/εκπαιδευτική πράξη στις Ψηφιακές Ανθρωπιστικές Σπουδές



α) η παραδειγματική περίπτωση ενός project

β) από την ψηφιακή εικόνα στο ψηφιακό περιεχόμενο



βιβλιοθήκη & γνώση

If you want to get laid, go to college. If you want an education, go to the library.

Frank Zappa

A university is just a group of buildings gathered around a library.

Shelby Foote

Bad libraries build collections, good libraries build services, great libraries build communities.

R. David Lankes

για τη νέα Βιβλιοθήκη...

- δεν θα στηρίζεται στα βιβλία αλλά στη γνώση και τη μάθηση // δημιουργία κοινοτήτων αναγνωστών // διευκόλυνση της παραγωγής γνώσης στις κοινότητες
- librarian → databrarian

Ψηφιακές Ανθρωπιστικές Σπουδές

Νέα Βιβλιοθήκη

In capturing how libraries and librarians are contributing to digital humanities at their institutions, it's important to note how much community engagement goes hand in hand with building capacity for DH. Indeed, more than 40% of survey respondents said their libraries are advocating for coordinated, cross-campus support for digital humanities. The interdisciplinarity of projects, which almost three-quarters of survey respondents confirmed their libraries encourage and facilitate, can also generate a sense of community [...]

<https://americanlibrariesmagazine.org/2016/01/04/special-report-digital-humanities-libraries/>

project / Πεζογραφία στα Επτάνησα (19^{ος}-20^{ός})

τομέας ΜΝΕΣ - τμήμα Φιλολογίας

ΣΤΟΧΟΙ

- Διερεύνηση του φαινομένου της πεζογραφίας στον ιδιαίτερο γεωπολιτικό χώρο των Επτανήσων του 19^{ου} και 20^{ού} αιώνα
- Διερεύνηση της θεωρίας και της πράξης των ψηφιακών εκδόσεων κειμένων
- Διαμόρφωση ενός νέου είδους φιλόλογου-ερευνητή που να μπορεί να φέρει σε διάλογο τις νεοελληνικές σπουδές και τις νέες τεχνολογίες (digital scholar)

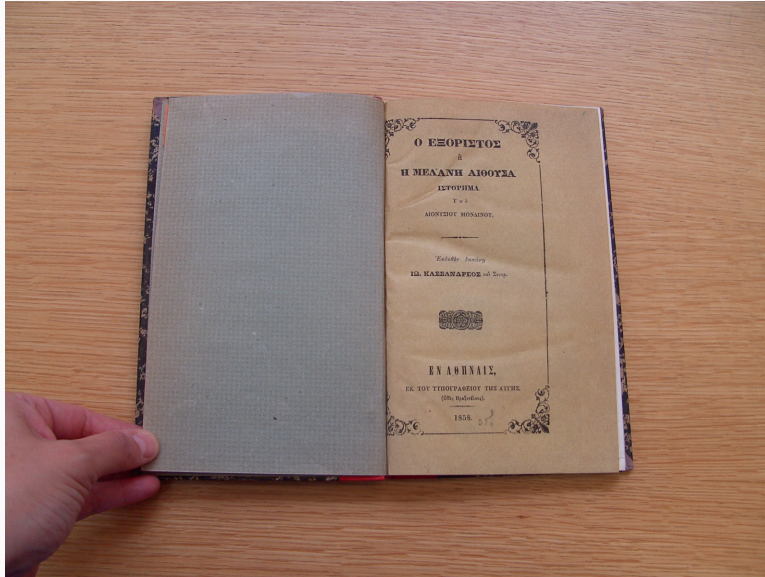
ομάδα ψηφιακών νεοελληνικών σπουδών

digines

- κριτική ενημέρωση σχετικά με τις εξελίξεις στις ψηφιακές νεοελληνικές / ανθρωπιστικές σπουδές
- κατάρτιση και επιμόρφωση νέων ερευνητών στις ψηφιακές τεχνολογίες
- σχεδιασμός και εκπόνηση προγραμμάτων με έμφαση στις ψηφιακές φιλολογικές εκδόσεις και την ψηφιακή διαχείριση αρχειακού υλικού που αφορά τις νεοελληνικές σπουδές
- συνεργασίες / σεμινάρια / συμμετοχή σε συνέδρια / δημοσιεύσεις

project / Πεζογραφία στα Επτάνησα 19ος-20ός

Βιβλιοθήκες



Εντοπισμός κειμένων σε βιβλιοθήκες

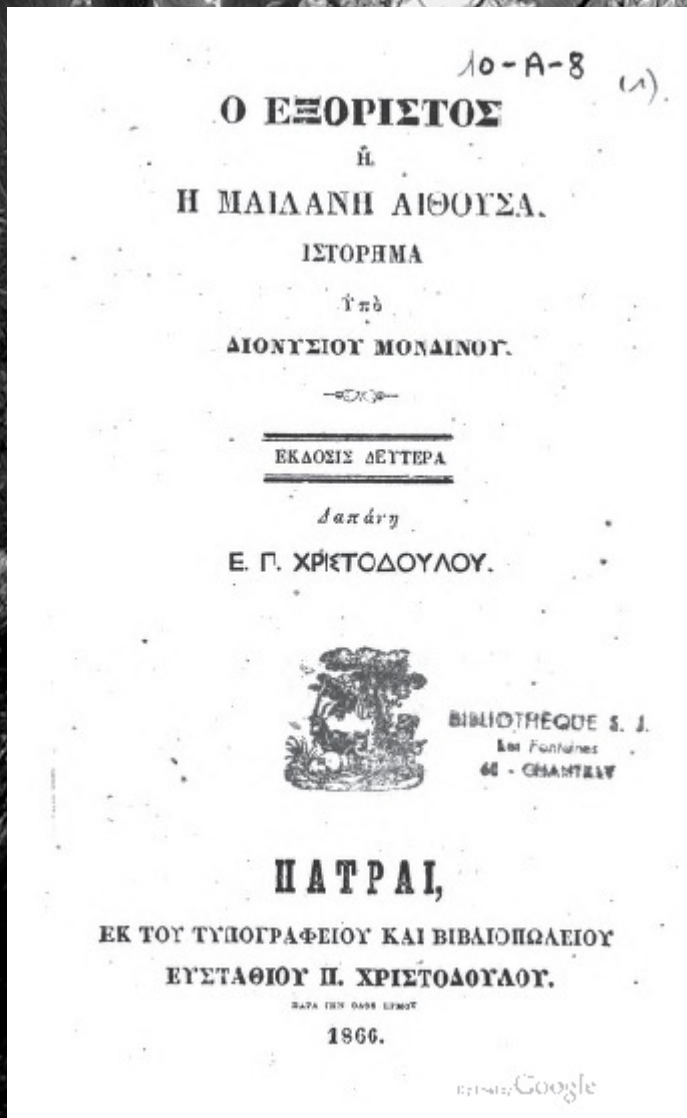
- μη ψηφιοποιημένα
- ψηφιοποιημένα

project

Βιβλιοθήκες

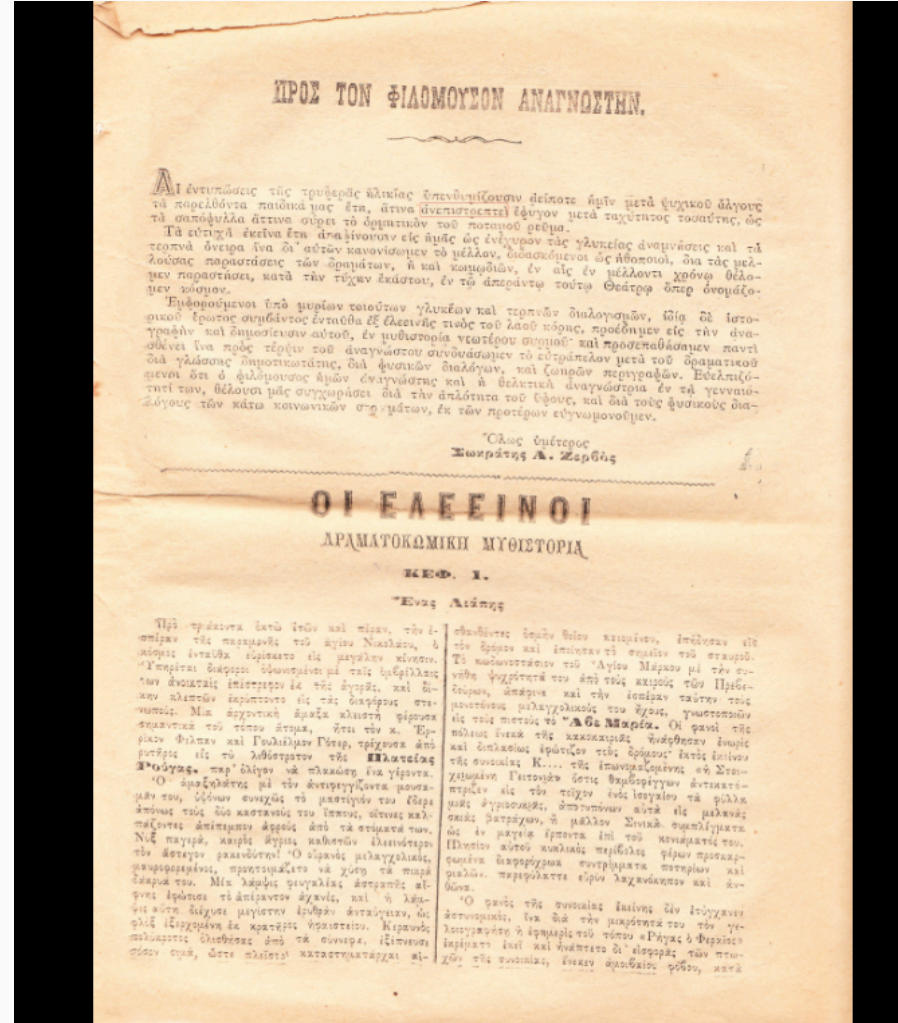
Εντοπισμός κειμένων σε βιβλιοθήκες

- μη ψηφιοποιημένα
- ψηφιοποιημένα = εικόνα



Φωτογράφιση κειμένου

Η έντυπη έκδοση του μυθιστορήματος *Οι Ελεεινοί*
του Σωκράτη Α. Ζερβού (Ζάκυνθος 1897)



2

Πληκτρολόγηση και φιλολογική επεξεργασία του κειμένου

The screenshot shows a Microsoft Word 2010 window titled "ΕΛΕΕΙΝΟΙ 1-18 (10.12.15).docx". The ribbon is set to the "Review" tab, with the "Comments" group active. The document text is in Greek, discussing the myth of the blind men and an elephant. The text includes:

Τα εύρημα έδειξε ότι απεικονίζουν εις ήμισυ ως ένδεχρον τὰς γλυκείας ἀναμνήσεις και τὰ τεργνὰ όνειρα. Ίνα δέ' αὐτῶν κανονίωσμεν τὸ μέλλον, διδασκόμενοι ὡς ήθιστοί, διά τὰς μελλούσας περιστάσεις τῶν δραμάτων, ή και κομωδιῶν, εν αὐς εν μέλλοντι χρόνῳ θέλωμεν παραστήσει, κατά τήν τήχην έκάστου, εν τῷ ἀπεράντιω τούτῳ θεάτρῳ ὅσοι ἀνομάζομεν κόσμον.

Εμφανόμενοι ὅσοι μέρουσι τοιαύτων γλυκείων και τεργνῶν διαλογισμῶν, ήδεις δὲ ιστορικῶ εἶδος συμβάντος ένταθα εἶ, ἐλακτώθης τινός του λαοῦ κόρης, προήβημεν εις τήν ἀναγραφήν και δημοσίωσιν αὐτοῦ ἐν μνηστορίῳ νεωτέρου σμυμοῦ, και προσεπαθήσομεν παντί σθένει, ίνα πρὸς τέρψιν του ἀναγνώστου συνδούσομεν τὸ εὐτράπελον μετά του δραματικοῦ διά γλώσσης δημοτικωτέτης, διά φυσικῶν διαλόγων, και ζωηρῶν περιγραφῶν. Εὐέλπιόμηναι ὅτι ὁ φιλόμουσος ήμῶν ἀναγνώστης και ή θελκτικὴ ἀναγνώστρια, εν τῇ γενναιωτέτῃ των, θέλωσι μάς συγχρηρίει διά τήν ἀπλότητα του ὅρου, και διά τούς φυσικῶς διαλόγους τῶν κάτω κοινωνικῶν στρωμάτων, ἐκ τῶν προτέρων εὐγνωμοσοῦμεν.

"Ὅλος ὀμιλέτος
Σωκράτης Α. Ζερβός

ΟΙ ΕΛΕΕΙΝΟΙ
ΔΡΑΜΑΤΟΚΙΜΙΚΗ ΜΥΘΙΣΤΟΡΙΑ

ΚΕΦ. 1
Ένας Αἰτίπης

Πρὸ τριάκοντα ὀκτώ ἐτῶν και πέραν, τήν ἔσπραν τῆς παραμονῆς του Αγίου Νικολάου, ὁ κόσμος ένταθα εὐρίσκειτο εις μεγάλην κίνησιν. Ὑπάρχει διάφοροι ὀφνωσημένοι με τὰς διηφθάλειας των ἄνοικταί ἐπίστρεφον ἐκ τῆς ἀγοράς, και δίσην κλεπτικῶν ἐκρύπτοντο εις τὰς διαφόρους στενωπούς. Μία ἀρχοντιῶ ἄμαξα κλεστικῆ φέρουσα σημαντικὰ του τύπου ἄτομα, ήτοι τὸν κ. Πέτρον Φέλιππιν και Γουλιέλμον Γ΄στερ, τρέχουσα ἀπὸ μνηστήρ εις τὸ λιθόστρωτον τῆς Πλατείας Ρούγας, παρ' ὀλίγον νὰ πλεακώσῃ ένα γέροντα.

The right side of the screen displays a list of comments:

- Comment [1]:** ὀνειρα (-ήνερα εν. πλ. Ο,π.) βλ. και παραστάσι και παραστάσι με ένα -:-
- Comment [2]:** α:9* αη(ν...)
- Comment [3]:** κομωδιῶν (τονομας)
- Comment [4]:** ελακτώθης (τονομας)
- Comment [5]:** -
- Comment [6]:** λέγειτε το κόμμα -.
- Comment [7]:** ὀφθάλμους (τονομας) βλ. ὀφθαλμοί
- Comment [8]:** ἀνοικταί (τονομας)
- Comment [9]:** -.
- Comment [10]:** ὀνειρα (βλ. 1,64)
- Comment [11]:** ρουφῶς (Α) = β
- Comment [12]:** λιθόστρωτον (Α) βλ.

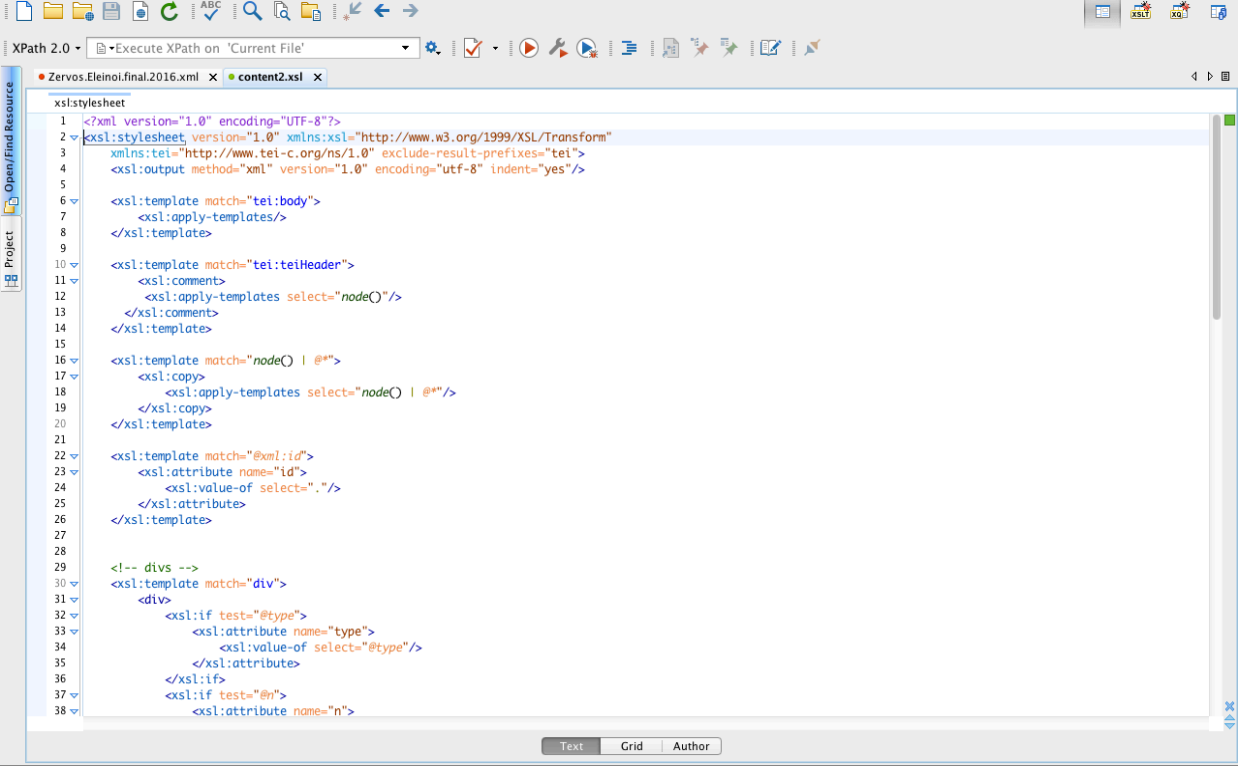
The status bar at the bottom indicates "Print Layout View", "Sec 1", "Pages: 1 of 34", "Words: 0 of 13152", and "100%" zoom.

Μαρκάρισμα/σήμανση του ηλεκτρολογημένου κειμένου με τη γλώσσα μαρκαρίσματος TEI XML

```
XPath 2.0 - Execute XPath on 'Current File'
Zervos.Eleinoi.final.2016.xml x
processing-instruction
262 <pb n="1"/>
263
264 <front>
265
266 <div type="preface">
267 <head> ΠΡΟΣ ΤΟΝ ΦΙΛΟΜΟΥΣΙΟΝ ΑΝΑΓΝΩΣΤΗΝ</head>
268 <p>Αἱ ἐντυπώσεις τῆς τριετέρας ἡλικίας ὑπενηθίζουσιν ἀείποτε ἡμῖν μετὰ ψυχικῶν
269 ἄλλοις τὰ παρελθόντα παιδικὰ μᾶς ἔτη, ἄτυνα ἀνεπιστρέφει ἔμφρον μετὰ ταχύτητας
270 τσαούτης, ὡς τὰ σαπόφωλα ἄτυνα σφύρει τὸ ὀρητικὸν τοῦ ποταμοῦ ρεῖμα.</p>
271 <p>Τὰ εὐτυχῆ ἐκεῖνα ἔτη ἀπαφίνουσιν εἰς ἡμᾶς ὡς ἐνέχυρον τὰς γλυκείας ἀναμνήσεις καὶ
272 τὰ τερπνὰ ὄνειρα, ἵνα δι' αὐτῶν κανονίσωμεν τὸ μέλλον, διδασκόμενοι ὡς ἠθασοῖαι,
273 διὰ τὰς μελλούσας παραστάσεις τῶν δραμάτων, ἢ καὶ κωμωδιῶν, ἐν αἷς ἐν μέλλοντι
274 χρόνῳ θέλωμεν παραστήσει, κατὰ τὴν τύχην ἐκάστου, ἐν τῷ ἀπεράντῳ τούτῳ θεάτρῳ
275 ὄπερ ὀνομάζομεν κόσμον.</p>
276 <p><ref target="#memories">Ἐμφοροῦμενοι ὑπὸ μυρίων τοιοῦτων γλυκῶν καὶ τερπνῶν
277 διαλογισμῶν, ἰδίᾳ δὲ ἱστορικῷ ἔρωτος συμβάντος ἐνταῦθα ἐξ ἐλλεινῆς τινός
278 τοῦ λαοῦ κόρης, προέβημεν εἰς τὴν ἀναγραφὴν καὶ δημοσίευσιν</ref>
279 <ref target="#ionian.novel.19th">αὐτοῦ ἐν μυθιστορίᾳ νεωτέρου συρμοῦ.</ref>
280 <ref target="#entertain">καὶ προεπαθήσωμεν παντὶ σθένει, ἵνα πρὸς τέρψιν τοῦ
281 ἀναγνώστου συνδυάσωμεν τὸ εὐτράπελον μετὰ τοῦ δραματικῶς</ref> διὰ γλώσσης
282 δημοτικωτάτης, διὰ φυσικῶν διαλογῶν, καὶ ζωηρῶν περιγραφῶν. Εὐελπιζόμενοι ὅτι ὁ
283 φιλόμουσος ἡμῶν ἀναγνώστης καὶ ἡ θελεκτικὴ ἀναγνώστρια, ἐν τῇ γενναϊότητι των,
284 <ref target="#ionian.novel.19th">θέλουσι μᾶς συγχωρήσει διὰ τὴν ἀπλότητα τοῦ
285 ὕφους, καὶ διὰ τοὺς φυσικοὺς διαλόγους τῶν κάτω κοινωνικῶν στρωμάτων</ref>,
286 ἐκ τῶν προτέρων εὐγνωμονοῦμεν.</p>
287 <closer>
288 <salute rend="align(right)">Ὁπως ἡμέτερος</salute>
289 <signed rend="align(right)">Σωκράτης Α. Ζερβός</signed>
290 </closer>
291 </div>
292 </front>
293 <body>
294 <ab type="title">
295 <title type="main"> Οἱ ΕΛΛΕΙΝΟΙ </title>
296 </ab>
297 <title type="desc"> ΔΡΑΜΑΤΟΚΙΜΙΚΗ ΜΥΘΙΣΤΟΡΙΑ </title>
298 </ab>
299 <div n="1" type="chapter">
```

4

Μετατροπή του εγγράφου XML σε XSL μέσω του λογισμικού οxygen



The screenshot shows the Oxygen XML Editor interface. The main window displays an XSLT stylesheet named 'content2.xsl'. The toolbar at the top includes icons for file operations, execution, and navigation. The 'Execute XPath' menu is open, showing the current file. The stylesheet code is as follows:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
3   xmlns:tei="http://www.tei-c.org/ns/1.0" exclude-result-prefixes="tei">
4   <xsl:output method="xml" version="1.0" encoding="utf-8" indent="yes"/>
5
6   <xsl:template match="tei:body">
7     <xsl:apply-templates/>
8   </xsl:template>
9
10  <xsl:template match="tei:teiHeader">
11    <xsl:comment>
12      <xsl:apply-templates select="node()"/>
13    </xsl:comment>
14  </xsl:template>
15
16  <xsl:template match="node() | @*">
17    <xsl:copy>
18      <xsl:apply-templates select="node() | @*" />
19    </xsl:copy>
20  </xsl:template>
21
22  <xsl:template match="@xml:id">
23    <xsl:attribute name="id">
24      <xsl:value-of select="."/>
25    </xsl:attribute>
26  </xsl:template>
27
28
29 <!-- divs -->
30 <xsl:template match="div">
31   <div>
32     <xsl:if test="@type">
33       <xsl:attribute name="type">
34         <xsl:value-of select="@type"/>
35       </xsl:attribute>
36     </xsl:if>
37     <xsl:if test="@n">
38       <xsl:attribute name="n">
```

The status bar at the bottom shows the file path: /Users/.../omekaFiles/geo-mapping/content2.xsl, the URI: U+003C, the page number: 2 : 1, and a notification for 7 new messages.



Δρόμοι και τρόποι για τη συνάντηση της βιβλιοθήκης
με την ερευνητική/εκπαιδευτική πράξη στις Ψηφιακές Ανθρωπιστικές Σπουδές



α) η παραδειγματική περίπτωση ενός project

β) από την ψηφιακή εικόνα στο ψηφιακό περιεχόμενο



OCR

Optical Character Recognition

Η Οπτική Αναγνώριση Χαρακτήρων αποδεικνύεται δημοφιλής και λειτουργική πρακτική για μεγάλης κλίμακας ψηφιοποιήσεις.

Αδυναμίες στην αναγνώριση:

- ιδιαίτερη μορφοποίηση/εμφάνιση κειμένου
- χαρακτήρες (γλωσσών που δεν περιλαμβάνει)
- ιδιαίτερα μαθηματικά σύμβολα
- εικόνες και σχεδιαγράμματα
- χειρόγραφα

Αποδοτικότητα βάση της

- ποιότητας της εικόνας του κειμένου
- μηχανική εκμάθηση (machine learning algorithm)

Στόχευση & ερωτήματα

- περιεχόμενο επεξεργάσιμο/ πλήρως ερευνήσιμο / ακρίβεια αναζήτησης
- μετατρεψιμότητα σε πολλούς τύπους αρχείων (PDF, TXT, EPUB, TIFF, HTML)
- διαμοιρασμός δεδομένων και μεταδεδομένων / τμηματική επανεπεξεργασία
- προοπτικές διατήρησης
- ...



Κωδικοποίηση κειμένου

SGML/XML κωδικοποίηση

Επεκτάσιμη

Δυνατότητα επέκτασης του λεξιλογίου κωδικοποίησης. Δημιουργία ετικετών επισημείωσης ανάλογα με τις ανάγκες του εκάστοτε κειμένου

Ανεξάρτητη

Μη ιδιόκτητη. Ανεξάρτητη από πλατφόρμες, από λειτουργικό ή λογισμικό. (System & device independent)

Repurposable data

Το περιεχόμενο μπορεί να μεταφερθεί αυτόματα με τη χρήση XSLT σε άλλα format: inDesign, html, e-book, EPUB, HTML, TIFF, TXT, PDF

Interoperable

Διεπικοινωνία και διαλειτουργικότητα μεταξύ διαφορετικών εφαρμογών και πλατφόρμων

SGML/XML κωδικοποίηση

Ευελιξία

Ευελιξία στην
κωδικοποίηση
δεδομένων

Διαμεταφορά

Διαμεταφορά & ανταλλαγή
περιεχομένου

Διαλειτουργικότητα

Μειωμένες πιθανότητες
για απώλεια δεδομένων

Διατήρηση

Διατήρηση του
περιεχομένου παρά τις
τεχνολογικές εξελίξεις



ΤΕΙ στη Βιβλιοθήκη



Text Encoding Initiative

“ These Guidelines apply to texts in any natural language, of any date, in any literary genre or text type, without restriction on form or content. **”**

(Burnard & Brauman, 2013)

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)



Front Matter

[Title](#)

- i. [Releases of the TEI Guidelines](#)
- ii. [Dedication](#)
- iii. [Preface and Acknowledgments](#)
- iv. [About These Guidelines](#)
- v. [A Gentle Introduction to XML](#)
- vi. [Languages and Character Sets](#)

Back Matter

- Appendix A [Model Classes](#)
- Appendix B [Attribute Classes](#)
- Appendix C [Elements](#)
- Appendix D [Attributes](#)
- Appendix E [Datatypes and Other Macros](#)
- iv. Appendix F [Bibliography](#)
- v. Appendix G [Prefatory Notes](#)
- Appendix H [Colophon](#)

Text Body

- iv. 1 [The TEI Infrastructure](#)
- v. 2 [The TEI Header](#)
- vi. 3 [Elements Available in All TEI Documents](#)
- vii. 4 [Default Text Structure](#)
- viii. 5 [Characters, Glyphs, and Writing Modes](#)
- ix. 6 [Verse](#)
- x. 7 [Performance Texts](#)
- xi. 8 [Transcriptions of Speech](#)
- xii. 9 [Dictionaries](#)
- xiii. 10 [Manuscript Description](#)
- xiv. 11 [Representation of Primary Sources](#)
- xv. 12 [Critical Apparatus](#)
- xvi. 13 [Names, Dates, People, and Places](#)
- xvii. 14 [Tables, Formulae, Graphics and Notated Music](#)
- xviii. 15 [Language Corpora](#)
- xix. 16 [Linking, Segmentation, and Alignment](#)
- xx. 17 [Simple Analytic Mechanisms](#)
- xxi. 18 [Feature Structures](#)
- xxii. 19 [Graphs, Networks, and Trees](#)
- xxiii. 20 [Non-hierarchical Structures](#)
- xxiv. 21 [Certainty, Precision, and Responsibility](#)
- xxv. 22 [Documentation Elements](#)
- xxvi. 23 [Using the TEI](#)

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)

TEI sourcecode

- [Getting and Using the TEI Sources.](#)
- [TEI GitHub Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

Επίπεδα Κωδικοποίησης

Αυτοματοποιημένη μεταγραφή & κωδικοποίηση

Το κείμενο παράγεται μέσω OCR και υπάγεται στην εικόνα της σελίδας με βάση κειμενικούς δείκτες πλοήγησης που λαμβάνονται αυτόματα.

1

Το κείμενο παράγεται μέσω OCR, εξαρτάται από την εικόνα της σελίδας και δεν προορίζεται να σταθεί μόνο του ως ένα ηλεκτρονικό κείμενο.

2

Ελάχιστη κωδικοποίηση

Απλή ανάλυση

Το κείμενο παράγεται με τη διόρθωση του OCR ή με πληκτρολόγηση και μπορεί να σταθεί μόνο του ως ηλεκτρονικό κείμενο. Η κωδικοποίηση περιγράφει όχι μόνο τη δομή, αλλά και το περιεχόμενο.

Εξειδικευμένη κωδικοποίηση

3

Το κείμενο δημιουργείται μέσω της μετατροπής από μια ηλεκτρονική πηγή, είτε μέσω OCR ή μέσω πληκτρολόγησης. Κωδικοποιούνται μονάχα τα δομικά χαρακτηριστικά του κειμένου.

4

Βασική ανάλυση περιεχομένου

Το κείμενο παράγεται με τη διόρθωση του OCR ή με πληκτρολόγηση, όπως στο επ. 4. Πρόσθετα, η σήμανση του απαιτεί ειδική παρέμβαση από άτομα που είναι καλοί γνώστες του κειμένου, του είδους του και του περιεχομένου του.

5

Αυτοματοποιημένη μεταγραφή & κωδικοποίηση

Στόχος

Η δημιουργία ενός ηλεκτρονικού κειμένου με σκοπό:

- την αναζήτηση λέξεων
- τη σύνδεση με την εικόνα της σελίδας.

Κύριο πλεονέκτημα:

- επισύναψη δεδομένων με μεταδεδομένα (teiHeader)

Σκεπτικό

Το κείμενο εξαρτάται από την εικόνα της σελίδας

- δεν είναι αυτόνομο
- δεν επαρκεί για κειμενική και γλωσσική ανάλυση

Το αποτέλεσμα ανταποκρίνεται στους στόχους μιας μονάδας διαφύλαξης ή μιας πρωτοβουλίας μαζικής ψηφιοποίησης

Αυτοματοποιημένη μεταγραφή & κωδικοποίηση

workflow

Η κωδικοποίηση γίνεται αυτόματα

Το κείμενο δεν διορθώνεται

Κωδικοποιείται :

- το σημείο έναρξης νέας σελίδας (<rb/>), έτσι ώστε να έχουμε το πλήρες κείμενο ως μια λίστα από σελίδες
- το κείμενο συνολικά ως ένα τμήμα (<div>)
- το περιεχόμενο κάθε σελίδας ως ένα κομμάτι (<ab>)

```
<TEI xml:id="someid">
<teiHeader xml:lang="en"> <!-- header goes here -->
</teiHeader>
<text xml:lang="en">
<body>
<div1>
<ab>
<pb n="113" facs="00000001.tif"/>
<!-- uncorrected OCR for first page image begins here -->
POINT VIII.
BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S
CONVICTION SHOULD BE VACATED; ALTERNATIVELY,
DISCOVERY AND A HEARING SHOULD BE ORDERED.
The nature and extent of surveillance of Hiss, his
family and associates was not known at the time of trial by
the defense. Even now, with the release of some of the govern-
ment documents concerning FBI investigative techniques regarding
Hiss, the full extent of surveillance -- wiretapping, mail open-
ings, mail covers, physical surveillance, and other intrusive
techniques -- is still not clear. Nevertheless, it is apparent
that information gathered through the exploitation of unlawful
wiretaps and other illegal surveillance was used at trial and
consequently the conviction must be reversed. Alternatively,
further discovery and a hearing is essential to a fair deter-
mination regarding these issues.
FBI surveillance of Hiss began in earnest in 1941 with
the institution of a mail cover on his incoming correspondence
at his home in connection with an FBI investigation of possible
Hatch Act violations. CN Ex. 98A. Another mail cover was placed

-113 -

<!-- uncorrected OCR for first page image ends here -->
<pb n="114" facs="00000002.tif"/>
<!-- uncorrected OCR for second page image begins here -->
on the Hiss mail in 1945, and at the same time the FBI obtained
toll call records from the Hiss residence Telephone for the
years 1943 and 1944 as well. CN Ex. 99. In September, 1945,
the FBI intercepted telegrams to Hiss as well. CN Ex. 100.
In late November, 1945, FBI surveillance of the Hiss
residence in Washington, D.C., escalated. For the third time,
a mail cover was instituted beginning on November 28, 1945,
which was continued at least until 1946. CN Ex. 101 at p. 70;
CN Ex. 102. Continuous physical surveillance of Hiss was begun
as well. CN Ex. 101 at p. 72. Although this twenty-four-hour
surveillance was discontinued on December 14, 1945, physical
surveillance was conducted frequently at various times until
September, 1947. CN Ex. 102; CN Ex. 103.
The most intrusive invasion of petitioner's rights
68/ Also before 1947, a letter from Priscilla Hiss addressed
to her son, Timothy Hobson, was intercepted and its contents
read. CN Ex. 100A at p. 167. In approximately March, 1947,
a letter from a Michael Greenberg addressed to petitioner re-
garding an application for employment with the United Nations
was also intercepted, in a manner not revealed by the docu-
ments. CN Ex. 100B
```

Αυτοματοποιημένη μεταγραφή & κωδικοποίηση

Χαρακτηριστικά projects :

- άμεση διαθεσιμότητα μεγάλου όγκου υλικού
- ψηφιακή εικόνα
- καμία χειροκίνητη παρέμβαση
- προορίζονται για μεγάλες κοινότητες χρηστών
- στόχος η ανάγνωση και η ευρετηρίαση
- δυνατότητα επεκτασιμότητας

Ελάχιστη κωδικοποίηση

Στόχος

Η δημιουργία ενός ηλεκτρονικού κειμένου με σκοπό:

- την αναζήτηση λέξεων
- τη σύνδεση με την εικόνα της σελίδας

Η αναγνώριση της **απλής** δομικής ιεραρχίας με σκοπό:

- τη βελτίωση της πλοήγησης (δυνατότητα δημιουργίας (TOCs)
- την αναζήτηση σε συγκεκριμένα τμήματα

Σκεπτικό

Το κείμενο εξαρτάται από την εικόνα της σελίδας

- δεν είναι αυτόνομο
- δεν επαρκεί για κειμενική και γλωσσική ανάλυση

Αποφυγή του κόστους ανάλυσης των κειμενικών στοιχείων ή/και της λεπτομερούς μεταγραφής του κειμένου

workflow

Η κωδικοποίηση γίνεται αυτόματα

Το κείμενο δεν διορθώνεται

Κωδικοποιείται :

- το σημείο έναρξης νέας σελίδας, όπως και στο επίπεδο 1 (<rb/>)
- το κάθε τμήμα του κειμένου ξεχωριστά (πολλά <div>)
- οι επικεφαλίδες των τμημάτων (<head>)

```

<TEI xml:id="someid">
<teiHeader xml:lang="en"> <!-- header goes here -->
</teiHeader>
<text xml:lang="en">
<body>
<div1>
<pb n="113" facs="00000001.tif"/>
<!-- content of head element was transcribed from page image -->
<head>POINT VIII. BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S
CONVICTION SHOULD BE VACATED; ALTERNATIVELY, DISCOVERY AND A HEARING
SHOULD BE ORDERED.</head>
<ab> <!-- uncorrected OCR for first page image begins here -->
POINT VIII.
BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S
CONVICTION SHOULD BE VACATED; ALTERNATIVELY,
DISCOVERY AND A HEARING SHOULD BE ORDERED.
The nature and extent of surveillance of Hiss, his
family and associates was not known at the time of trial by
the defense. Even now, with the release of some of the govern-
ment documents concerning FBI investigative techniques regarding
Hiss, the full extent of surveillance -- wiretapping, mail open-
ings, mail covers, physical surveillance, and other intrusive
techniques -- is still not 'clear. Nevertheless, it is apparent
that information gathered through the exploitation of unlawful
wiretaps and other illegal surveillance was used at trial and
consequently the conviction must be reversed. Alternatively,
further discovery and a hearing is essential to a fair deter-
mination regarding these issues.
FBI surveillance of Hiss began in earnest in 1941 with
the institution of a mail cover on his incoming correspondence
at his home in connection with an FBI investigation of possible
Hatch Act violations. CN Ex. 98A. Another mail cover was placed

-113 -

<!-- uncorrected OCR for first page image ends here -->
<pb n="114" facs="00000002.tif"/>
<!-- uncorrected OCR for second page image begins here -->
on the Hiss mail in 1945, and at the same time the FBI obtained
toll call records from the Hiss residence Telephone for the
years 1943 and 1944 as well. CN Ex. 99. In September, 1945,
the FBI intercepted telegrams to Hiss as well. CN Ex. 100.
In late November, 1945, FBI surveillance of the Hiss
residence in Washington, D.C., escalated. For the third time,
a mail cover was instituted beginning on November 28, 1945,
which was continued at least until 1946. CN Ex. 101 at p. 70;
CN Ex. 102. Continuous physical surveillance of Hiss was begun
as well. CN Ex. 101 at p. 72. Although this twenty-four-hour
surveillance was discontinued on December 14, 1945, physical
surveillance was conducted frequently at various times until
September, 1947. CN Ex. 102; CN Ex. 103.
The most intrusive invasion of petitioner's rights
68/ Also before 1947, a letter from Priscilla Hiss addressed
to her son, Timothy Hobson, was intercepted and its contents
read. CN Ex. 100A at p. 157. In approximately March, 1947,
a letter from a Michael Greenberg addressed to petitioner re-
garding an application for employment with the United Nations
was also intercepted, in a manner not revealed by the docu-
ments. CN Ex. 100B

-114 -

```

Ελάχιστη κωδικοποίηση

Χαρακτηριστικά projects :

- άμεση διαθεσιμότητα μεγάλου όγκου υλικού
- επιθυμητή η ψηφιακή εικόνα
- δυνατότητες αναζήτησης και προβολής με βάση την κύρια δομή του κειμένου
- έλεγχος κάθε κειμένου για τη διασφάλιση της σωστής κωδικοποίησης των τμημάτων
- δυνατότητα επεκτασιμότητας

Απλή ανάλυση

Στόχος

Η δημιουργία ενός αυτόνομου ηλεκτρονικού κειμένου, όπου:

- αποτυπώνεται αναλυτικά η δομή και η εμφάνιση
- δεν αναλύεται το περιεχόμενο

Σκεπτικό

Αυτόνομο ηλεκτρονικό κείμενο

Απαιτείται ανθρώπινη παρέμβαση, χωρίς να ενδείκνυται εξειδίκευση.

Κωδικοποιούνται χαρακτηριστικά του κειμένου με βάση τη δομή και την τυπογραφική εμφάνισή του

- περισσότερες επιλογές στην αναζήτηση
- περισσότερες επιλογές στην προβολή

Απλή ανάλυση

workflow

Το κείμενο παράγεται μέσω της μεταγραφής από μια ηλεκτρονική πηγή (doc, txt, htm, xml).

Δεν διορθώνεται.

Κωδικοποιούνται:

- τα παρακειμενικά τμήματα (<front>, <back>)
- τα παρένθετα κείμενα (<floatingText>)
- οι τίτλοι, υπότιτλοι, πλαγιότιτλοι (<head>)
- όλες οι παράγραφοι, οι λυρικές στροφές και οι στίχοι (<p>, <lg>, <l>)
- το περιεχόμενο των υποσημειώσεων (<notes>)
- στήλες, λίστες και πίνακες (<cb>, <list>, <table>)
- οι τυπογραφικές μορφοποιήσεις (“italics”, “bold”, “left”, “right”, “upercase”)
- ενδοκειμενικές διασυνδέσεις (<ref>, <ptr>)

```
<TEI xml:id="someid">
<teiHeader xml:lang="en"> <!-- header goes here -->
</teiHeader>
<text xml:lang="en">
<body>
<div1>
<pb n="113" facs="00000001.tif"/>
<head>POINT VIII. BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S CONVIC
SHOULD BE VACATED; ALTERNATIVELY, DISCOVERY AND A HEARING SHOULD B
ORDERED.</head>
<p>The nature and extent of surveillance of Hiss, his family and associ
not known at the time of trial by the defense. Even now, with the
some of the govern ment documents concerning FBI investigative tec
regarding Hiss, the full extent of surveillance -- wiretapping, ma
ings, mail covers, physical surveillance, and other intrusive tech
still not 'clear. Nevertheless, it is apparent that information ga
through the exploitation of unlawful wiretaps and other illegal su
was used at trial and consequently the conviction must be reversed
Alternatively, further discovery and a hearing is essential to a f
mination regarding these issues.</p>
<p>FBI surveillance of Hiss began in earnest in 1941 with the instituti
mail cover on his incoming correspondence at his home in connectio
FBI investigation of possible Hatch Act violations. CN Ex. 98A. An
cover was placed <pb n="114" facs="00000002.tif"/> on the Hiss mai
and at the same time the FBI obtained toll call records from the H
residence Telephone for the years 1943 and 1944 as well. CN Ex. 99
September, 1945, the FBI intercepted telegrams to Hiss as well. CN
100.</p>
<p>In late November, 1945, FBI surveillance of the Hiss residence in Wa
D.C., escalated. For the third time, a mail cover was instituted b
November 28, 1945, which was continued at least until 1946. CN Ex.
70; CN Ex. 102. Continuous physical surveillance of Hiss was begun
Ex. 101 at p. 72. Although this twenty-four-hour surveillance was
on December 14, 1945, physical surveillance was conducted frequent
various times until September, 1947. <note place="bottom" anchored
Timothy Hobson, was intercepted and its contents read. CN Ex. 10
167. In approximately March, 1947, a letter from a Michael Green
addressed to petitioner re garding an application for employment
United Nations was also intercepted, in a manner not revealed by
ments. CN Ex. 100B</note> CN Ex. 102; CN Ex. 103.</p>
<p>The most intrusive invasion of petitioner's rights <pb n="115" facs=
from Washington, D.C. to New York City on September 13, 1947. A "t
surveillance," -- a wiretap -- was placed on the Hiss telephone at
residence on F Street-in Washington, D.C. The logs of this surveill
constitute twenty-nine volumes of FBI serials and are roughly 2,50
length, in which an enormous amount of information concerning the
sonal lives, relationships with friends and associates, and habits
recorded.</p>
<p>The wiretap was installed following FBI Director Hoover's applicat
Attorney General for authorization, <note place="bottom" anchored=
information on Hiss. CN Ex. 104. Additional information was furr
letter dated November 30, 1945. CN Ex. 105.</note> although no w
authorization appears in the documents released to Hiss. The purpo
application was to gather information regarding Hiss' alleged con
Soviet espionage agents and communists in government service, gene
allegations which had been made by Elizabeth Bentley and Chambers.
<p>As one would expect, the interception of every telephone/p>
</div1>
</body>
</text>
</TEI>
```


Χαρακτηριστικά projects

- για κοινότητες χρηστών που στοχεύουν στην ανάγνωση και την αναζήτηση
- περισσότερες δυνατότητες προβολής, διαμοιρασμού και ευρετηρίασης
- ποιοτικός έλεγχος κάθε κειμένου για τη διασφάλιση των αποφάσεων κωδικοποίησης
- μικρότερος όγκος (kb) και χαμηλότερες απαιτήσεις γραφικών
- δημιουργία χωρίς εξειδίκευση επάνω στο περιεχόμενο των κειμένων
- δυνατότητα επεκτασιμότητας

Βασική ανάλυση περιχομένου

Στόχος

Η δημιουργία ενός αυτόνομου ηλεκτρονικού κειμένου, όπου:

- αποτυπώνεται αναλυτικά η δομή και η εμφάνιση
- προσδιορίζονται οι λειτουργίες των δομικών στοιχείων
- περιγράφεται η φύση του περιχομένου.

Σκεπτικό

Η αναλυτικότερη απόδοση του περιχομένου επιτρέπει:

- ευελιξία στην προβολή και το διαμοιρασμό
- εξελιγμένες δυνατότητες ευρετηρίασης
- συνδυασμό ευρύτερου πεδίου χρήσεων και αναγνωστών

Αποτελεσματική λειτουργία κειμένου με εφαρμογές e-reading και audio-book.

Βασική ανάλυση περιεχομένου

workflow

- το κείμενο παράγεται μετά από διόρθωση της πηγής.

Κωδικοποιούνται:

- όλα τα δομικά στοιχεία (βλ. Επ. 3)
- όλα τα τυπογραφικά χαρακτηριστικά (βλ. Επ. 3)
- πρόλογοι, επίλογοι, υστερόγραφα, χαιρετισμοί, υπογραφές
- ονόματα ηρώων και ομιλητών
- διάλογοι, αποφθέγματα, διακειμενικά φαινόμενα
- κύρια ονόματα, ανθρωπωνύμια, τοπωνύμια
- αλλόγλωσσο λεξιλόγιο, ιδιώματα, τεχνική ορολογία
- emphatic words or phrases
- συγγραφικές και εκδοτικές παρεμβάσεις
- ενδοκειμενικές διασυνδέσεις λέξεων και τμημάτων
- κ.π.ά.

```
<pb n="113" facs="./pageImages/AH4_0113.jpg"/>
<head>POINT VIII.</head>
<head>BECAUSE OF UNLAWFUL SURVEILLANCE, PETITIONER'S
<lb/>CONVICTION SHOULD BE VACATED; ALTERNATIVELY,
<lb/>DISCOVERY AND A HEARING SHOULD BE ORDERED.</head>
<p>The nature and extent of surveillance of Hiss, his
<lb/>family and associates was not known at the time of trial by
<lb/>the defense. Even now, with the release of some of the govern-
<lb break="no"/>ment documents concerning FBI investigative techniques regard-
<lb/>Hiss, the full extent of surveillance -- wiretapping, mail open-
<lb break="no"/>ings, mail covers, physical surveillance, and other intrusive
<lb/>techniques -- is still not clear. Nevertheless, it is apparent
<lb/>that information gathered through the exploitation of unlawful
<lb/>wiretaps and other illegal surveillance was used at trial and
<lb/>consequently the conviction must be reversed. Alternatively,
<lb/>further discovery and a hearing is essential to a fair deter-
<lb break="no"/>mination regarding these issues.</p>
<p>FBI surveillance of Hiss began in earnest in 1941 with
<lb/>the institution of a mail cover on his incoming correspondence
<lb/>at his home in connection with an FBI investigation of possible
<lb/>Hatch Act violations. CN Ex. 98A. Another mail cover was placed
<pb n="114" facs="./pageImages/AH_0114.jpg"/>
on the Hiss mail in 1945, and at the same time the FBI obtained
<lb/>toll call records from the Hiss residence Telephone for the
<lb/>years 1943 and 1944 as well. CN Ex. 99. In September, 1945,
<lb/>the FBI intercepted telegrams to Hiss as well. CN Ex. 100.</p>
<p>In late November, 1945, FBI surveillance of the Hiss
<lb/>residence in Washington, D.C., escalated. For the third time,
<lb/>a mail cover was instituted beginning on November 28, 1945,
<lb/>which was continued at least until 1946. CN Ex. 101 at p. 70;
<lb/>CN Ex. 102. Continuous physical surveillance of Hiss was begun
<lb/>as well. CN Ex. 101 at p. 72. Although this twenty-four-hour
<lb/>surveillance was discontinued on December 14, 1945, physical
<lb/>surveillance was conducted frequently at various times until
<lb/>September, 1947.<note place="bottom" anchored="true" n="68">Also
before 1947, a letter from Priscilla Hiss addressed
<lb/>to her son, Timothy Hobson, was intercepted and its contents
<lb/>read. CN Ex. 100A at p. 167. In approximately March, 1947,
<lb/>a letter from a Michael Greenberg addressed to petitioner re-
<lb break="no"/>garding an application for employment with the United Nation
<lb/>was also intercepted, in a manner not revealed by the docu-
<lb break="no"/>ments. CN Ex. 100B</note> CN Ex. 102; CN Ex. 103.</p>
<p>The most intrusive invasion of petitioner's rights
<pb n="115" facs="./pageImages/AH_0115.jpg"/>
<lb/>occurred from December 13, 1945 until the Hisses moved from
<lb/>Washington, D.C. to New York City on September 13, 1947. A
<lb/>so-called technical surveillance</so-called>, -- a wiretap -- was placed on
<lb/>telephone at their residence on P Street in Washington, D.C.
<lb/>The logs of this surveillance constitute twenty-nine volumes
<lb/>of FBI serials and are roughly 2,500 pages in length, in which
<lb/>an enormous amount of information concerning the Hisses' per-
<lb break="no"/>sonal lives, relationships with friends and associates, and
<lb/>habits is recorded.</p>
<p>The wiretap was installed following FBI Director Hoover's
<lb/>application to the Attorney General for authorization,
<note place="bottom" anchored="true" n="69">Hoover's
initial request was answered by a note reques-
<lb break="no"/>ting information on Hiss. CN Ex. 104<sup>sic</sup>. Additional info
<lb/>was furnished by letter dated November 30, 1945. CN Ex. 105<sup>sic</sup>.</note>
<lb/>Although no written authorization appears in the documents released to
<lb/>Hiss. The purpose of the application was to gather information
<lb/>regarding Hiss' alleged contacts with Soviet espionage agents and
<lb/>communists in government service, general allegations which had
<lb/>been made by Elizabeth Bentley and Chambers.</p>
```

Βασική ανάλυση περιεχομένου

Χαρακτηριστικά projects

- Εξελιγμένες δυνατότητες αναζήτησης και ανάκτησης δεδομένων
- Στόχος η ανάλυση δεδομένων
- Ποιοτικός έλεγχος κάθε κειμένου για τη διασφάλιση των αποφάσεων κωδικοποίησης
- Μικρότερος όγκος (kb) και χαμηλότερες απαιτήσεις γραφικών
- Δυνατότητα επεκτασιμότητας

Εξειδικευμένη κωδικοποίηση

Στόχος

Ηλεκτρονικό κείμενο αναλυτικής κωδικοποίησης

- κατάλληλο για ερευνητικούς σκοπούς

Συμμετοχή εξειδικευμένου ακαδημαϊκού προσωπικού.

Σκεπτικό

Τα projects που έχουν ως βάση το ψηφιακό προϊόν της βιβλιοθήκης να μπορούν να προχωρήσουν σε υψηλού επιπέδου κωδικοποίηση στο πλαίσιο:

- των εγχειρημάτων ψηφιοποίησης τεκμηρίων
- της επιστημονικής επιμέλειας κειμένων
- ακαδημαϊκής υποστήριξης
- ή άλλης έρευνας

Εξειδικευμένη κωδικοποίηση

Δυνατότητες

Η κωδικοποίηση μπορεί να περιγράφει

- σημασιολογικά
- γλωσσολογικά
- μετρικά
- και άλλα χαρακτηριστικά, πέρα από τα βασικά δομικά στοιχεία.

Η κωδικοποίηση επιτρέπει

- τις εκδοτικές, κριτικές και αναλυτικές προσθήκες
- την περιγραφή χειρογράφων
- την προσθήκη μετάφρασης
- τη δημιουργία φιλολογικών υπομνημάτων
- κ.ά

The Chymistry of
Isaac Newton

Home | Browse Manuscripts | Online Tools | Educational Resources | Project Inform

Normalized Transcription | Diplomatic Transcription

Keynes MS. 12, King's College Library, Cambridge University

16 Propositions 1

1. $\odot \mathcal{D} \sigma \lambda \eta \psi$ magnesia *<illeg.>* omnes species artis. Consil: conjug. p 434

7. Lapis componitur ex 4 elementis et 5^{ta} essentiâ quæ omnia sunt ejusdem radicis Consil conjug p 435, 437

8. Item ex omnibus speciebus 7 imperfectis. Nov Lum Cap 8 Myster occultæ naturæ p

2. Omnes species sunt ex una radice Consil conjug p 435

3. Idq̄ in hoc ordine $h. \lambda. \mathcal{D}. \sigma. \odot 1$

4. Sola est Magnesia quæ *<illeg.>* illegible; reason: cancelled and illegible 8 letters

5. Aurum tamen ei nondum præparat

6. Ejusq̄ Et auri vices Chalybs supplet

9. Nihil tingitur pro^m projectione quod non ingreditur compositionem tingentis.

10. Totum opus in unicâ re consistit

Educ salem e metallis: qui est unicus clavis nam omnis generatio fit ex spermate

Metalla ignem non passa debent sumi. Arca arcanorū. Cælū terræ

Εξειδικευμένη κωδικοποίηση

Ενδεικτικά projects

Ανεμόσκαλα, Κέντρο Ελληνικής γλώσσας

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

The Willa Carter Archive, University of Nebraska-Lincoln

<http://cather.unl.edu/>

Pietro Mellini's inventory in verse, Getty Research Institute

<http://www.getty.edu/research/mellini/>

The Bodleian First Folio, University of Oxford e-Research Center

<http://firstfolio.bodleian.ox.ac.uk/>

Ψηφιακές Ανθρωπιστικές Σπουδές

Νέα Βιβλιοθήκη

- Συμβολή στη Συλλογή δεδομένων
linked data [project : «Επτανησιακά»]
- Ανάλυση δεδομένων [project : μεταδεδομένα header]
- Επίπεδο κωδικοποίησης τουλάχιστον 2 ή και 3
παράδειγμα *Εφημερίς των Ιονίων Νήσων* / Βρετανικές εφημερίδες
- Συνεργασία βιβλιοθηκονόμων και ερευνητών (νέων / stage)

Επιλογή Βιβλιογραφίας

- *Best Practices for TEI in Libraries* (version 3.0, published October 2011)
- Digitization Workflows: Scanning, OCR, and Audio Transcription, *Digital Humanities at Berkeley*, Jun. 2015.
- L. D Unger, *Digital Conversion OCR & Text Encoding*, University of Michigan Library, 2014.
- C. Strange, D. McNamara, J. Wodak, I. Wood, "Mining for the Meanings of a Murder: The Impact of OCR Quality on the Use of Digitized Historical Newspapers", *Digital Humanities Quarterly* 8:1 (2014).
- J. McGlone, "Preserving and Publishing Digital Content Using XML Workflows", in A.P. Brown (Ed.), *The Library Publishing Toolkit*, . Geneseo, NY: IDS Project Press, 2013, (pp. 97-108).
- K. Stapelfeldt, D. Moses, "Islandora and TEI: Current and Emerging Applications/Approaches", *Journal of the Text Encoding Initiative* 5 (Jun. 2013).

ABOUT XML-TEI

- C.M. Sperberg-McQueen and Lou Burnard, *TEI P5, Guidelines for Electronic Text Encoding and Interchange*, TEI Consortium, Mar. 2016.
- James Cummings, "TEI, What else?" (ηλεκτρονική παρουσίαση), Prezi, Jan. 2015.
- Α. Τικτοπούλου (συμμετοχή: Φ. Κοϊδάκη), «Κωδικοποιήσεις κειμένου», στο Ξ. Δημητρούλια & Α. Τικτοπούλου, *Ψηφιακές λογοτεχνικές σπουδές*, Ελληνικά Ακαδημαϊκά Συγγράμματα και Βοηθήματα: Κάλλιπος, Θεσσαλονίκη, 2015.

Ψηφιακές Ανθρωπιστικές Σπουδές & Νέα Βιβλιοθήκη

