


Text Encoding Initiative

επισκόπηση, προβλήματα και εφαρμογές

Λίνα Μπουντούρη
13ο Πανελλήνιο Συνέδριο
Ακαδημαϊκών Βιβλιοθηκών -
Κέρκυρα 2004



Πληροφορική για τις Ανθρωπιστικές Σπουδές

- Προσπάθειες για πρότυπα και εργαλεία
- Χρήστες με αυξημένες και εξειδικευμένες ανάγκες
- Ηλεκτρονικό κείμενο
- Γλώσσες Σήμανσης
 - SGML
 - HTML
 - XML

Text Encoding Initiative

- Πηγές σε μορφή κειμένου
- TEI: κωδικοποίηση και ανταλλαγή κειμένων
- TEI Consortium
- Έκδοση TEI P4
- Περιγραφικό – Διαχειριστικό – Δομικό πρότυπο

TEI DTD (P4)

- Ποικιλία κόμβων
- Core tag sets: υποχρεωτική χρήση
 - TEI.core.dtd (π.χ. παράγραφος)
 - TEI.header.dtd (μεταδεδομένα TEI εγγράφου)
- Base tag sets: για συγκεκριμένους τύπους κειμένων
 - TEI.prose, (πεζός λόγος)
 - TEI.verse, (ποίηση)
 - TEI.drama, (θεατρικά έργα)
 - TEI.spoken, (προφορικά κείμενα)
 - TEI.dictionaries, (έντυπα λεξικά)
 - TEI.terminology, (ορολογία)

TEI DTD (P4)

- Συνδυασμός Base tag sets ετικετών μέσα από:
 - TEI.general
 - TEI.mixed
- Additional tag sets (π.χ. TEI.linking, TEI.names,dates)
- User defined tags: δημιουργία, μετατροπή και επέκταση ετικετών
- TEI Lite DTD

TEI header

- Βιβλιογραφικές και περιγραφικές πληροφορίες του TEI εγγράφου
- Ηλεκτρονική σελίδα τίτλου
- Περιλαμβάνει
 - File Description
 - Encoding Description
 - Profile Description
 - Revision Description

TEI header

- Χρήσιμες πληροφορίες
- Κοινή βάση με βιβλιοθηκονομικά πρότυπα
 - Διαλειτουργικότητα
 - Δημιουργία βιβλιογραφικής εγγραφής
 - Π.χ. Oxford Text Archive
 - Διαχείριση συλλογών
 - Αναζήτηση, επιλογή και πρόσβαση

TEI header

- Αναλυτικότερα
 - File Description:
 - <title>, <author>
 - <sourceDesc>, <editionStmt>
 - <notesStmt>
 - Πληροφορίες πρόσκτησης και πρόσβασης στο περιγραφόμενο υλικό
 - Εξειδικευμένα σημεία πρόσβασης

Εργαλεία

- Δημιουργία TEI DTD και TEI XML Schema
 - Pizza Chef: TEI P4 (DTD)
 - Roma: TEI P4 και P5 (DTD, Relax NG, W3C Schema)
- Δημιουργία TEI εγγράφων
 - SGML και XML επεξεργαστές
 - XMetal, Emacs, <oxygen/>
 - Επεξεργαστές κειμένου (π.χ. UltraEdit)
- SGML και XML βάσεις δεδομένων
- teiPublisher

ΤΕΙ – Πλεονεκτήματα

- Ανεξάρτητο από πλατφόρμες, συστήματα και εφαρμογές
- Αξιοποίηση υπαρχόντων τεχνολογιών και εργαλείων
- eXtensible Stylesheet Language
- Δυνατότητα παραμετροποίησης
- Αυξημένες δυνατότητες αποτύπωσης πληροφοριών

Επικαλυπτόμενες ιεραρχίες

- ΤΕΙ: κωδικοποίηση κειμένου ως ένα σύνολο ιεραρχιών που δεν επικαλύπτονται
- Συνύπαρξη πολλαπλών ιεραρχικών δομών

Παράδειγμα

`<I n="1">Είπες<quotation>«Θα πάγω σ' άλλη γη, θα πάγω σ' άλλη θάλασσα</I>`

`<I n="2">...</I> ...`

`<I n="8">που τόσα χρόνια πέρασα και ρήμαξα και χάλασα.></quotation></I>`

ΕΠΙΚΑΛΥΠΤΟΜΕΝΕΣ ΙΕΡΑΡΧΙΕΣ

- Δυσκολίες αποθήκευσης, αναζήτησης – ανάκτησης και μορφοποίησης
- Λανθασμένη απόδοση κόμβων
- Λύσεις
 - Κενά στοιχεία (Milestones)
 - Τεμαχισμός (Fragmentation)
 - Εικονικά στοιχεία (Virtual elements)
- TEI Overlapping Markup SIG discussion list

Κανόνες περιγραφής – TEI Header

- Έλλειψη καταγραφής δεδομένων
 - Δημιουργία ελλιπών ή λανθασμένων μεταδεδομένων
- Έλλειψη κανόνων περιγραφής
- TEI Consortium: δεν έχει ορίσει πολιτική
- TEI P5

Πληροφορική για τις Ανθρωπιστικές Σπουδές – Ελλάδα

- Ακαδημαϊκή παράδοση και σχετικό υλικό στο χώρο των Ανθρωπιστικών Σπουδών
- Ελληνικά κείμενα κωδικοποιημένα σε ΤΕΙ
 - Perseus Digital Library
(<http://www.perseus.tufts.edu/>)
 - EpiDoc (<http://www.unc.edu/awmc/epidoc/>)

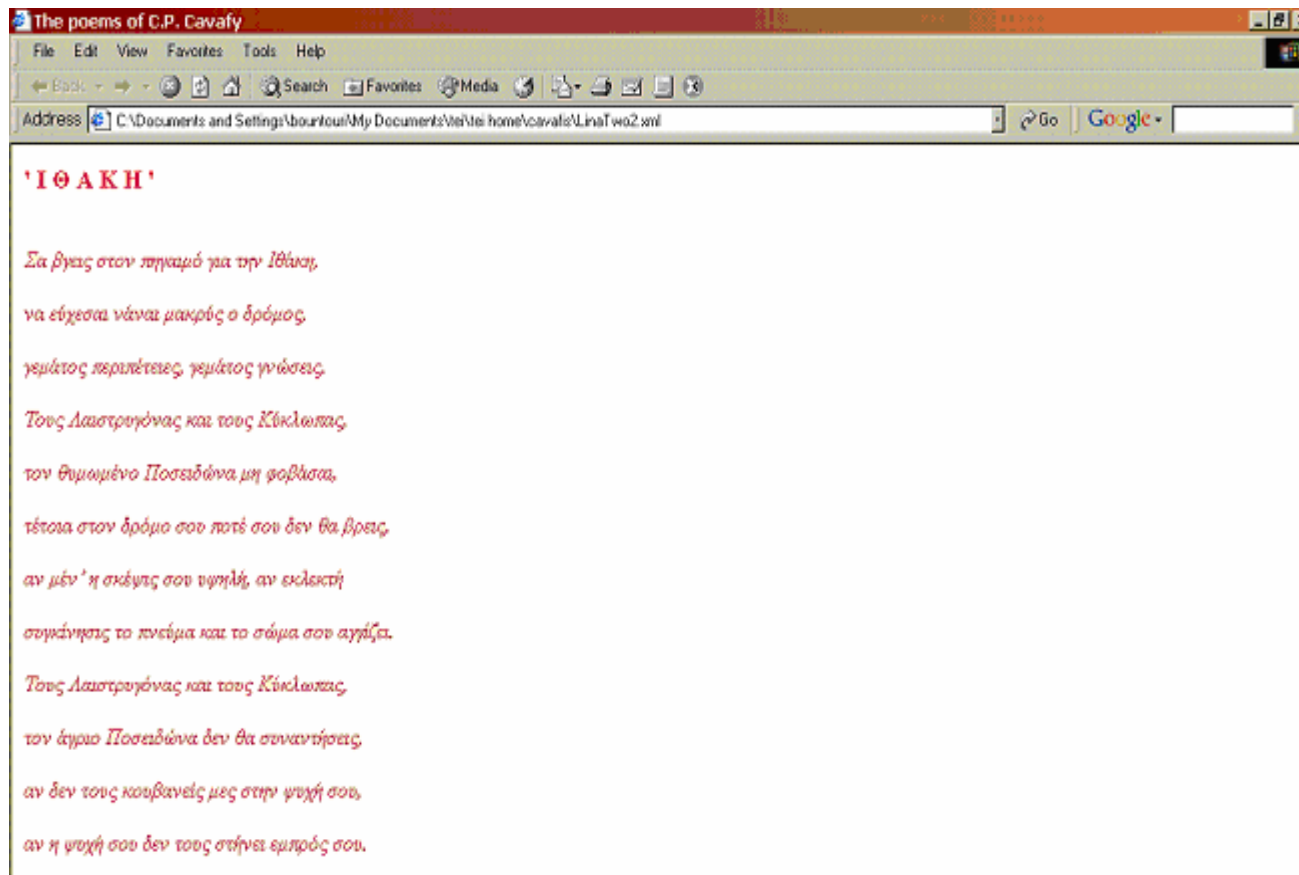
Παράδειγμα

```
<TEI.2>
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Τα ποιήματα του Κ.Π. Καβάφη: ηλεκτρονική
έκδοση</title>
      <author> Κ.Π. Καβάφης</author>
    </titleStmt>
    <publicationStmt>
      <publisher>Εκδόσεις Ίκαρος</publisher>
      <date>1966</date>
    </publicationStmt>
  </fileDesc>
</teiHeader>
```

Παράδειγμα

```
<text>
  <body>
    <lg type="free">
      <l>Σα βγεις στον πηγαιμό για την Ιθάκη,</l>
      <l> να εύχεσαι νάναι μακρύς ο δρόμος,</l>
      <l> Τους Λαιστρυγόνας και τους Κύκλωπας,</l>
      <l>τον θυμωμένο Ποσειδώνα μη φοβάσαι,</l>
      <l>τέτοια στον δρόμο σου ποτέ σου δεν βρεις,</l>
      <l>αν μέν' η σκέψις σου υψηλή, αν εκλεκτή</l> ...
    </lg>
  </body>
</text>
</TEI.2>
```


Παράδειγμα



The screenshot shows a web browser window titled "The poems of C.P. Cavafy". The address bar contains the file path "C:\Documents and Settings\bourtout\My Documents\ve\tei home\cavafy\LinaI wo2.xml". The browser's menu bar includes "File", "Edit", "View", "Favorites", "Tools", and "Help". The main content area displays the title "'ΙΘΑΚΗ'" in red, followed by the text of the poem in red italics.

'ΙΘΑΚΗ'

*Σα βρεις στον πηγαρό για την Ιθάκη,
να είχεςαι νάνα μακρός ο δρόμος,
γεμάτος περιπέτειες, γεμάτος γνώσεις,
Τους Λαοστραγόνες και τους Κίκλωπας,
τον θυμαμένο Ποσειδώνα μη φοβήσαι,
τέτοια στον δρόμο σου ποτέ σου δεν θα βρεις,
αν μόν' η σκέψης σου υψηλή, αν εκλεκτή
σφαιρήσης το πνεύμα και το σώμα σου αγάζει.
Τους Λαοστραγόνες και τους Κίκλωπας,
τον άγριο Ποσειδώνα δεν θα συναντήσεις,
αν δεν τους κοιθάνεις μες στην ψυχή σου,
αν η ψυχή σου δεν τους στήνει εμπρός σου.*

Συμπεράσματα

- ΤΕΙ: πρόσβαση και αναζήτηση
- Ξεπέρασε αρχικό στόχο (ανταλλαγή κειμένων)
- Μέσο για μόνιμη αποθήκευση κειμένων
- Ανοικτά ερευνητικά προβλήματα