

# Text Encoding Initiative -TEI

Σ. Αγγελής, Ε. Πάζα,

Λ. Παπαχριστόπουλος, Χ. Παπαθεοδώρου

Μονάδα Ψηφιακής Επιμέλειας

Ινστιτούτο Πληροφοριακών Συστημάτων

Ερευνητικό Κέντρο «Αθηνά»

{s.angelis, e.paza, l.papachristopoulos, c.papatheodorou}@dcu.gr



# Κωδικοποίηση κειμένου (text markup, text encoding)

- Η διαδικασία διάκρισης δομικών ή σημασιολογικών (semantic) χαρακτηριστικών κειμένου με βάση κάποιους κανόνες.
- Οι πρώτοι κανόνες κωδικοποίησης κειμένου δημοσιεύθηκαν το 1994, μετά από 7 χρόνια προσπάθειας εκατοντάδων επιστημόνων από διαφορετικά πεδία.



# Για την ιστορία

- Οι κανόνες αυτοί αποτελούσαν την πρώτη έκδοση του προτύπου που ονομάζεται *Text Encoding Initiative* (TEI).
- Στόχος του προτύπου είναι να δημιουργήσει ένα περιβάλλον για την κωδικοποίηση τεκμηρίων όλων των ειδών του γραπτού λόγου (πεζός λόγος-prose, ποίηση-verse, δράμα-drama), οποιασδήποτε γλώσσας και χρονικής περιόδου.
- Τον Ιανουάριο του 2001 δημιουργήθηκε ο οργανισμός TEI consortium (<http://www.tei-c.org>) ως ένα δίκτυο επιστημόνων που δραστηριοποιούνται κυρίως στο χώρο των ανθρωπιστικών επιστημών, επιχορηγούμενο από γνωστούς και μεγάλους επαγγελματικούς Οργανισμούς.
- 2002: οι κανόνες δημοσιεύθηκαν ως ένα XML DTD με στόχο την παροχή ενός συνόλου στοιχείων και γνωρισμάτων της XML, συνδεδεμένων με μια σύνταξη που θα επιτρέπει την απεικόνιση της λογικής που διέπει τη δομή ενός κειμένου.



# Παράδειγμα

*π. έλεγείας*

Introduction (*προοίμιον*)

As soon as Gyges came to the throne, he too, like others, led an army into the lands of Miletus and Smyrna.

*π. έλεγείας:*

Introduction (*προοίμιον*)

As soon as Gyges came to the throne, he too, like others, led an army into the lands of Miletus and Smyrna.



# Παράδειγμα

```
<title xml:lang="gr">π. ἐλεγείας</title>
```

```
Introduction (<foreign xml:lang="gr">προοίμιον</foreign>)
```

```
<p>As soon as <name type="pers">Gyges</name> came to  
the throne, he too, like others, led an army into the lands  
of <name type="place" reg="Miletus [27.3,37.5]  
(Perseus)"><placeName>Miletus</placeName></name>  
and <name type="place" reg="Smyrna [27.1667,38.4167]">  
<placeName> Smyrna</placeName> </name>.</p>
```



# Βασική δομή

- Τα στοιχεία του ΤΕΙ οργανώνονται σε ένα σύνολο 21 ενοτήτων (modules).
- Οι ενότητες περιλαμβάνουν τους ορισμούς των στοιχείων που ανήκουν σε αυτές.
- Κάθε στοιχείο ανήκει, επίσης, σε μια ή περισσότερες κλάσεις.
- Ανάλογα με τις κλάσεις που ανήκει ένα στοιχείο, το περιεχόμενο και τα γνωρίσματά του μπορεί να είναι διαφορετικά.
- Οι κλάσεις που ανήκει κάθε στοιχείο, το περιεχόμενο και τα γνωρίσματα του περιγράφονται πλήρως στον ορισμό του.
- Ο δισδιάστατος χώρος ενοτήτων – κλάσεων των στοιχείων προσφέρει ιδιαίτερη δυναμική και ευελιξία στο πρότυπο.



# ΤΕΙ - σχήματα

- Ένα ΤΕΙ-σχήμα μπορεί να δομηθεί από ένα συνδυασμό ενότητων.
- Οι ενότητες μπορούν να συνδυάζονται και να σχηματίζουν ΤΕΙ-σχήματα που καλύπτουν συγκεκριμένες ανάγκες κωδικοποίησης.
- Κατά τον ορισμό ενός ΤΕΙ-σχήματος, το πρότυπο παρέχει τη δυνατότητα να προστεθούν στοιχεία σε κάποιες κλάσεις ή να αποκλειστούν κάποια άλλα στοιχεία από συγκεκριμένες ενότητες.
- Κάποιες ενότητες αποτελούν απαραίτητα συστατικά στις περισσότερες περιπτώσεις δόμησης σχημάτων (module tei, χρησιμοποιείται από όλες τις υπόλοιπες ενότητες).
- Παράδειγμα σχήματος: ΤΕΙ lite, tei corpus, tei-ms



# Ενότητες

Ενότητα	Τυπική ονομασία	Περιγραφή
tei	TEI Infrastructure	περιλαμβάνει δηλώσεις και ορισμούς που αφορούν στη βασική υποδομή του προτύπου, όπως δηλώσεις τύπων δεδομένων, κλάσεων γνωρισμάτων κλπ
core	Common Core	περιλαμβάνει στοιχεία που υπάρχουν σε όλα τα TEI τεκμήρια όπως δηλώσεις ονομάτων, ημερομηνιών, διευθύνσεων, σημειώσεις, συντομογραφίες, λέξεις από ξένες γλώσσες κλπ
header	Common Metadata	περιλαμβάνει στοιχεία για τα μεταδεδομένα κωδικοποίησης ενός κειμένου
textstructure	Default Text Structure	περιλαμβάνει στοιχεία που προδιαγράφουν τη βασική δομή ενός TEI τεκμηρίου και κωδικοποιούν τη δομή των κειμένων όπως στοιχεία για παράγραφους, κεφάλαια, κλπ
drama	Performance Texts	περιλαμβάνει στοιχεία που χρησιμοποιούνται για την κωδικοποίηση κειμένων που αφορούν σε παραστάσεις (θεατρο, μουσικόδραμα, κα)
verse	Verse	περιλαμβάνει στοιχεία για την κωδικοποίηση του ποιητικού λόγου
gaiji	Character and Glyph	περιλαμβάνει στοιχεία για την τεκμηρίωση χαρακτήρων και των γραμματοσειρών τους
namesdates	Names, Dates, People, and Places	περιλαμβάνει στοιχεία για την κωδικοποίηση ονομάτων, προσώπων και τοπωνυμίων
figures	Tables, Formulae, Figures Documentation	περιλαμβάνει στοιχεία για πίνακες, σχήματα και μαθηματικούς τύπους που υπάρχουν σε κείμενα
spoken	Transcribed speech	περιλαμβάνει στοιχεία για τη μεταγραφή προφορικού λόγου





# Ενότητες

Ενότητα	Τυπική ονομασία	Περιγραφή
link	Linking, Segmentation, and Alignment	περιλαμβάνει στοιχεία για παραπομπές, για τη δόμηση του κειμένου σε επιμέρους τμήματα και τη συσχέτιση τμημάτων ή ολόκληρων κειμένων
dictionaries	Print Dictionaries	περιλαμβάνει στοιχεία για έντυπα λεξικά
corpus	Metadata for Language Corpora	περιλαμβάνει στοιχεία μεταδεδομένων και κωδικοποίησης σωμάτων κειμένων (language corpora). Ένα σώμα κειμένου περιλαμβάνει συλλογές κειμένων ή τμημάτων τους με συγκεκριμένο γλωσσολογικό ενδιαφέρον π.χ. κείμενα σε μια διάλεκτο μιας γλώσσας, ή κείμενα που αναπαριστούν όλες τις εκδοχές μιας γλώσσας
msdescription	Manuscript Description	περιλαμβάνει στοιχεία για την κωδικοποίηση χειρογράφων
analysis	Analysis and Interpretation	περιλαμβάνει στοιχεία για τη γλωσσολογική ανάλυση των κειμένων, όπως στοιχεία για τη δήλωση προτάσεων, λέξεων, φράσεων, φωνημάτων κλπ
certainty	Certainty and Uncertainty	περιλαμβάνει στοιχεία για τη δήλωση της αβεβαιότητας με την οποία γίνεται μια κωδικοποίηση ενός κειμένου (πχ μπορεί να μην είναι βέβαιο κάποιο μέρος ενός κειμένου (να έχουν χαθεί λέξεις, προτάσεις κλπ) ή να μην είναι απόλυτα βέβαιη η πνευματική ιδιοκτησία (ο συγγραφέας) ενός κειμένου)
iso-fs	Feature Structures	περιλαμβάνει στοιχεία για την κωδικοποίηση καταλόγων χαρακτηριστικών με τις αντίστοιχες τιμές τους, πχ οι χώρες της Ευρώπης με τις πρωτεύουσές τους
textcrit	Text Criticism	περιλαμβάνει στοιχεία που συνοδεύουν ένα γραπτό κείμενο, κριτικά υπομνήματα, αναλύσεις, πρώιμες μεταφράσεις κλπ
transcr	Transcription of Primary Sources	περιλαμβάνει στοιχεία για την κωδικοποίηση του υποστρώματος στο οποίο υπάρχει ένα κείμενο ή χειρόγραφο, πχ μιας τηλεμοιτυπίας (fax) που περιλαμβάνει ένα κείμενο ή ένα χειρόγραφο
tagdocs	Documenta tion Elements	περιλαμβάνει στοιχεία για την τεκμηρίωση των αλλαγών που γίνονται στη σημασία ή τη συντακτική δομή των στοιχείων του TEI κατά τη δημιουργία TEI-σχημάτων

# Κλάσεις Γνωρισμάτων

- Κλάσεις γνωρισμάτων (attribute classes): στοιχεία που “μοιράζονται” κοινά γνωρίσματα.
  - Τα ονόματα αυτών των κλάσεων έχουν το πρόθεμα “att.”
  - Ιεραρχία κλάσεων: τα γνωρίσματα ενός στοιχείου κληρονομούνται είτε από την ταξινόμησή του σε μια κλάση, είτε από τις κλάσεις-προγόνους της.
  - Η κλάση att.divLike ανήκει στην ενότητα tei και διαθέτει τα γνωρίσματα org, part και sample. Η κλάση att.metrical, ορίζεται μόνο στην ενότητα verse και διαθέτει τα γνωρίσματα met, real και rhyme. Η κλάση αυτή ορίζεται ως υπερκλάση της κλάσης att.divLike. Επομένως μέσα στην ενότητα verse τα μέλη της κλάσης att.divLike κληρονομούν και τα έξι παραπάνω γνωρίσματα.



# Κοινά γνωρίσματα

- `xml:id`, παρέχει ένα μοναδικό κωδικό ταυτοποίησης στο στοιχείο που χρησιμοποιεί το γνώρισμα.
- `η`, παρέχει έναν αριθμό ή μια άλλη ετικέτα αναφοράς (label) σε ένα στοιχείο, η οποία δεν είναι υποχρεωτικά μοναδική μέσα στο TEl τεκμήριο.
- `xml:lang`, υποδηλώνει τη γλώσσα του περιεχομένου του στοιχείου το οποίο συνοδεύει. Η ένδειξη κάθε γλώσσας είναι ένα σύνολο χαρακτήρων που καθορίζεται από το πρότυπο BCP 47.
- `rend` (rendition): υποδηλώνει τον τρόπο παρουσίασης (μορφοποίησης) του περιεχόμενου του στοιχείου μέσα στο κείμενο.



# Κοινά γνωρίσματα

- rendition: έχει περιεχόμενο ένα δείκτη προς κάποιο στοιχείο που περιγράφει την εμφάνιση/παρουσίαση που χρησιμοποιείται στο στοιχείο του οποίου προσαρτάται το γνώρισμα.
- xml:base: παρέχει ένα URI ως βάση για τον πλήρη (απόλυτο) προσδιορισμό αναφορών που γίνονται σε URIs με σχετικό τρόπο.
- xml:space: σηματοδοτεί τον επιθυμητό τρόπο διαχείρισης “λευκών” χώρων (κενά, στηλοθέτες και κενές γραμμές) από εφαρμογές που επεξεργάζονται τα XML τεκμήρια. Οι τιμές του γνωρίσματος είναι οι “default” και “preserve”. Με την πρώτη τιμή ακυρώνεται η χρήση των “λευκών” χώρων, ενώ με την τιμή “preserve”, οι “λευκοί” που υπάρχουν στο αρχικό τεκμήριο διατηρούνται (π.χ. διατηρούνται τα σημεία αλλαγής των στίχων σε ένα ποίημα, δηλαδή τα κενά (λευκοί χώροι) που ακολουθούν το τέλος του κάθε στίχου μέχρι το τέλος της αράδας του κειμένου).



# Κλάσεις Μοντέλου Περιεχομένου

- Κλάσεις μοντέλου περιεχομένου (model classes): στοιχεία που συνήθως εμφανίζονται μαζί.
  - Τα ονόματα αυτών των κλάσεων έχουν το πρόθεμα model.
  - Δημιουργούν ιεραρχία κλάσεων
  - Αν μια κλάση μοντέλου περιεχομένου περιέχει στο όνομά της τη λέξη like, τότε τα στοιχεία - μέλη της κλάσης έχουν μια πρόσθετη κοινή ιδιότητα που χαρακτηρίζει σημασιολογικά το περιεχόμενο των στοιχείων – μελών της. Για παράδειγμα η κλάση model.bibLike υποδηλώνει ότι τα στοιχεία της περιλαμβάνουν περιεχόμενο βιβλιογραφικής περιγραφής.



# Κλάσεις Μοντέλου Περιεχομένου

- Δομικών υποδιαιρέσεων (divisions): περιλαμβάνουν τις κλάσεις `model.divLike`, `model.div1Like` κ.α., που παρέχουν στοιχεία για την κωδικοποίηση της δομής των κειμένων (μέρη, κεφάλαια, υποκεφάλαια κλπ).
- Παραγράφων (chunks): οι κλάσεις αυτές παρέχουν στοιχεία για την κωδικοποίηση του κειμένου εντός των παραγράφων ή τμημάτων ενός μνημείου του γραπτού λόγου. Τα στοιχεία αυτά εντάσσονται κυρίως στις κλάσεις `model.divPart`, `model.pLike`, `model.entryLike`.
- Φράσεων (phrase-level): τα στοιχεία των κλάσεων αυτών δεν συσχετίζουν διαφορετικά τμήματα και παραγράφους ενός κειμένου, αλλά κωδικοποιούν φράσεις που υπάρχουν μέσα σε αυτές, όπως τίτλοι, εμφατικές φράσεις, παρεμβάσεις του επιμελητή του κειμένου. Τα στοιχεία αυτά δημιουργούν την κλάση `model.phrase.3`.



# Μακροεντολές

- standard content models: κατηγορία μακροεντολών που χρησιμοποιούνται για τη δήλωση μοντέλων περιεχομένου ή μερών τους.
- datatype macros: κατηγορία μακροεντολών που χρησιμοποιούνται για τη δήλωση των τύπων δεδομένων των γνωρισμάτων.
- Μακροεντολή μοντέλου περιεχομένου macro.paraContent (στις δηλώσεις των ακόλουθων στοιχείων του TEI: ab, add, camera, caption, case, cell, colloc, corr, damage, def, del, docEdition, emph, gen, gram, hi, hyph, iType, imprimatur, l, lang, lbl, mod, mood, number, orig, orth, p, per, pos, pron, ref, reg, restore, retrace, rhyme, seg, sic, sound, stress, subc, supplied, surplus, syll, tech, title, titlePart, tns, unclear, usg, writing)
  - (text | model.gLike | model.phrase | model.inter | model.global)\*
  - Η μακροεντολή καθορίζει το μοντέλο περιεχομένου των παραγράφων και άλλων σχετικών στοιχείων, δηλώνοντας ότι τα συγκεκριμένα στοιχεία έχουν ως περιεχόμενο κείμενο (text), είτε ανήκουν στις κλάσεις μοντέλου περιεχομένου model.gLike ή model.phrase ή model.inter ή model.global.



# Παράδειγμα

```
element p {  
  att.global.attributes,  
  att.global.linking.attributes,  
  att.global.analytic.attributes,  
  att.global.facs.attributes,  
  att.global.change.attributes,  
  att.declaring.attributes,  
  macro.paraContent }
```

Η δήλωση `macro.paraContent` σημαίνει ότι το στοιχείο ανήκει στις κλάσεις μοντέλων περιεχομένου που ορίζονται από την μακροεντολή `macro.paraContent`. Δηλαδή το στοιχείο `<r>` περιλαμβάνει ως υποστοιχεία ένα πλήθος μεγαλύτερο από 50 στοιχεία που δηλώνονται στις κλάσεις μοντέλου περιεχομένου `model.gLike`, `model.phrase`, `model.inter`, `model.global`.





# Παράδειγμα

```
data.duration.iso = token { pattern =  
« [ 0-9 . , DHMPRSTWYZ / : + \ - ] + " } }
```

- προσδιορίζει το πεδίο τιμών του γνωρίσματος dur-iso, το οποίο χρησιμοποιείται στα στοιχεία date και time και κωδικοποιεί δεδομένα της μορφής:
- `<time dur-iso="PT0,75H">τρία τέταρτα της ώρας</time>`
- `<date dur-iso="P1,5D">μία ημέρα και μισή</date>`

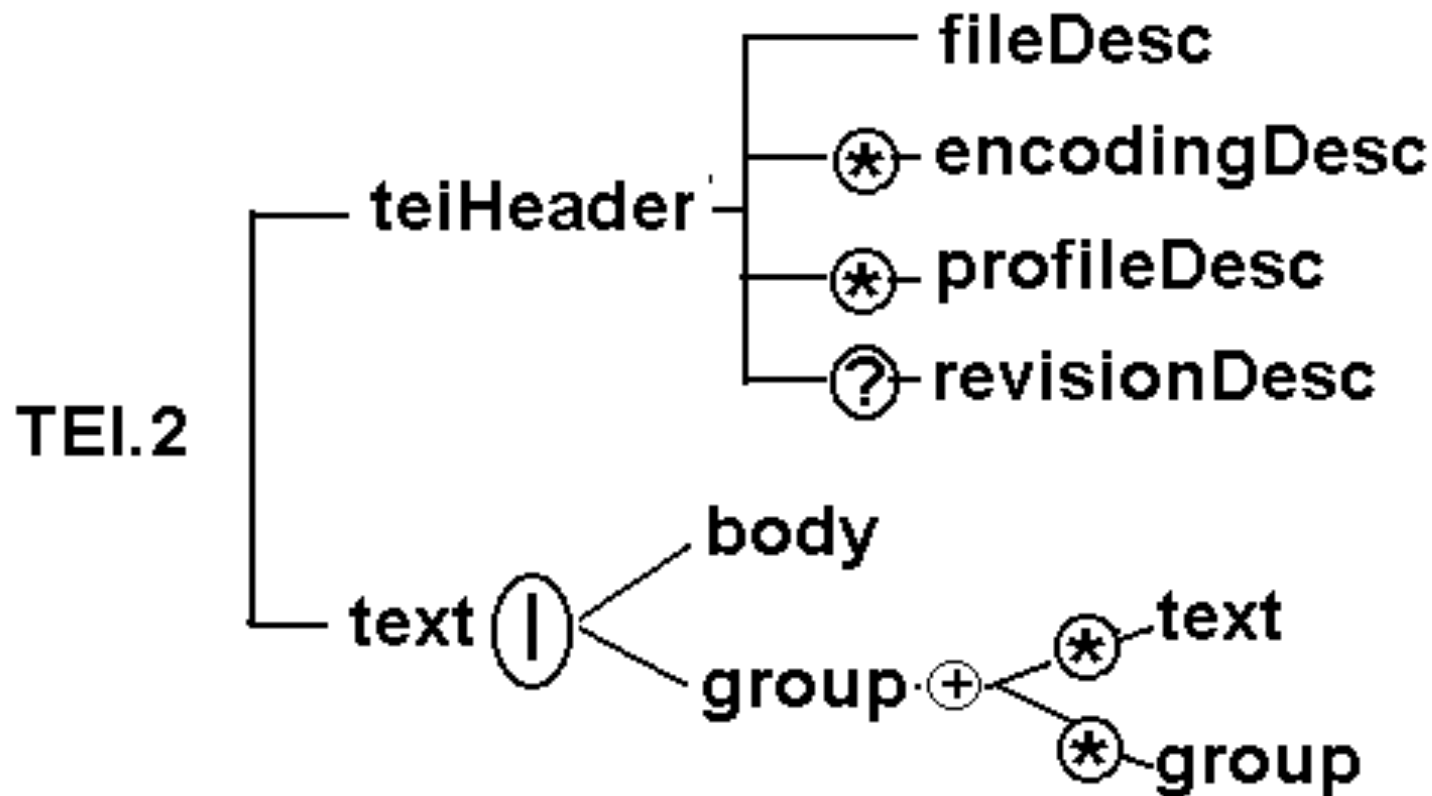


# Μακροεντολές

- 7 standard content models μακροεντολές που ορίζουν τα μοντέλα περιεχομένου των περισσότερων από πεντακόσια στοιχεία του προτύπου.
- 23 datatype macros (μακροεντολές τύπων δεδομένων) γνωρισμάτων προκειμένου να καλύψει όλους τους τύπους δεδομένων των γνωρισμάτων των στοιχείων του.
  - τύποι δεδομένων για: τις γλώσσες στις οποίες είναι γραμμένα τα περιεχόμενα των στοιχείων (`data.language`), κωδικούς ταυτοποίησης (`data.code`), κωδικοποίηση του φύλλου των προσώπων (`data.sex`), τους χώρους ονομάτων της XML και τα URIs που αναφέρονται στα κείμενα (`data.namespace` και `data.pointer` αντίστοιχα).



# Βασική δομή TEI τεκμηρίου



A Group element must contain at least one text or group element

# Βασική δομή

```
<TEI.2>
  <teiHeader> <!-- ... --> </teiHeader>
  <text>
    <front>
      <!-- εξώφυλλο (front matter). -->
    </front>
    <body>
      <!-- κυρίως κείμενο. -->
    </body>
    <back>
      <!-- οπισθόφυλλο (back matter). -->
    </back>
  </text>
</TEI.2>
```



# Πολυμερή τεκμήρια

```
<text>
  <front>
    ...[book front content here]...
  </front>
  <group>
    <text>
      <front> ...[part 1 front content here]...</front>
      <body> ...[part 1 body content here]...</body>
      <back> ...[part 1 back content (if any) here]...</back>
    </text>
    <text> ...[content of part 2]...</text>
    <text> ...[content of part 3]...</text>

    ...etc...
  </group>
  <back>
    ...[book back content(if any) here]...
  </back>
</text>
```



# Δομικά στοιχεία

<!ELEMENT TEI.2

(teiHeader, text) >

<!ELEMENT teiHeader

(fileDesc, encodingDesc\*, profileDesc\*, revisionDesc?) >

<!ELEMENT text

((index | interp | interpGrp | lb | milestone | pb | gap  
| anchor)\*, (front, (index | interp | interpGrp | lb | milestone  
| pb | gap | anchor)\*)?, (body | group), (index | interp  
| interpGrp | lb | milestone | pb | gap | anchor)\*, (back,  
(index | interp | interpGrp | lb | milestone | pb | gap  
| anchor)\*)?) >

<!ELEMENT group

((argument | byline | docAuthor | docDate | epigraph | head  
| opener | salute | signed | index | interp | interpGrp  
| lb | milestone | pb | gap | anchor)\*, (text | group),  
(text | group | index | interp | interpGrp | lb | milestone  
| pb | gap | anchor)\*, ((byline | closer | epigraph | salute  
| signed | trailer), (index | interp | interpGrp | lb |  
milestone | pb | gap | anchor)\*)\* >



```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader type="corpus">
<!-- corpus-level metadata here -->
</teiHeader>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader type="text">
<!-- metadata specific to this text here -->
</teiHeader>
<text>
<!-- ... -->
</text>
</TEI>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader type="text">
<!-- metadata specific to this text here -->
</teiHeader>
<text> <!-- ... -->
</text>
</TEI>
```

# Corpus



# TEI Header

- <fileDesc> υποχρεωτικό, παρέχει μια πλήρη βιβλιογραφική περιγραφή του κωδικοποιημένου κειμένου στοιχεία τίτλου, έκδοσης και προέλευσης του
- <encodingDesc> προαιρετικό, περιγράφει τη σχέση του κωδικοποιημένου κειμένου με την πηγή ή τις πηγές του. Παρέχει πληροφορίες για το εάν και πώς το κείμενο κανονικοποιήθηκε κατά τη μεταγραφή του και τις μεθόδους στις οποίες βασίστηκε η κωδικοποίησή του.
- <profileDesc> προαιρετικό, παρέχει πληροφορίες για τον τρόπο δημιουργίας του κειμένου, τις γλώσσες που χρησιμοποιούνται σε αυτό και τη θεματική του περιγραφή και ταξινόμηση. Ιδιαίτερα χρήσιμο στοιχείο σε περιπτώσεις σωμάτων ή μεγάλων συλλογών κειμένων γιατί συνεισφέρει στην αναζήτηση και ανάκτηση συγκεκριμένων κειμένων.
- <revisionDesc> προαιρετικό, χρησιμοποιείται για την καταγραφή του ιστορικού των αλλαγών έγιναν και γίνονται στο κωδικοποιημένο κείμενο κατά την επεξεργασία του.





# fileDesc

- **<titleStmt>** (title statement): ομαδοποιεί πληροφορίες για τον τίτλο του έργου του οποίου αποτελεί έκφανση το κωδικοποιημένο ψηφιακό κείμενο και τους υπευθύνους του περιεχομένου του. Το στοιχείο είναι υποχρεωτικό και πρέπει να υπάρχει μόνο μία φορά στο στοιχείο **<fileDesc>**.
  - **<title>** περιέχει τον τίτλο του κωδικοποιημένου κειμένου καθώς επίσης κάθε υπότιτλο ή εναλλακτικό τίτλο.
  - Όταν ένα κείμενο αποτελεί ηλεκτρονική έκδοση ενός υπάρχοντος έργου, τότε οι βέλτιστες πρακτικές του ΤΕΙ επιβάλλουν ο τίτλος της ηλεκτρονικής έκδοσης να προκύπτει από τον τίτλο της αρχικής έκδοσης, αλλά να διαφέρει πλήρως από αυτή. Για παράδειγμα η προσθήκη της φράσης «ψηφιακή έκδοση» στο περιεχόμενο του στοιχείου **<title>** διακρίνει το συγκεκριμένο κείμενο από την αρχική του έκδοση και εξασφαλίζει ότι οι βιβλιογραφικές αναφορές (citations) θα αφορούν στην αντίστοιχη έκφανση στην οποία γίνονται.



# fileDesc

- Το στοιχείο <title> δεν πρέπει να περιέχει ονόματα αρχείων υπολογιστή ή άλλες συμβολοσειρές (κωδικοί) που αναφέρονται στο υπολογιστικό σύστημα που φιλοξενεί την ψηφιακή εκδοχή ενός κειμένου.
- Οι τιμές του γνωρίσματος type είναι:
  - main, που δηλώνει τον κύριο τίτλο
  - sub, που δηλώνει υπότιτλο (subtitle) ή τίτλο μέρους του κειμένου (title of part)
  - alt, που δηλώνει εναλλακτικό τίτλο, π.χ. μεταφρασμένο τίτλο σε άλλη γλώσσα
  - short , δηλώνει μια συντετμημένη μορφή ενός τίτλου
  - desc, που είναι μια παράφραση ενός έργου που λειτουργεί ως τίτλος
- Οι τιμές του γνωρίσματος level είναι:
  - a, που αντιστοιχεί στη λέξη analytic και εφαρμόζεται για τίτλους άρθρων, ποιημάτων που αποτελούν μέρη ενός μεγαλύτερου έργου
  - m, που αντιστοιχεί στη λέξη monographic και εφαρμόζεται σε τίτλους βιβλίων, συλλογών, έργων που εκδίδονται ως αυτόνομες εργασίες, είτε ενός είτε περισσότερων τόμων
  - j, που αντιστοιχεί στη λέξη journal και εφαρμόζεται σε τίτλους περιοδικών
  - s, που αντιστοιχεί στη λέξη series και εφαρμόζεται σε τίτλους σειρών
  - u, που αντιστοιχεί στη λέξη unpublished και εφαρμόζεται σε τίτλους υλικού που δεν έχει εκδοθεί, όπως διατριβές.



# titleStmt

- **<author>** περιέχει το καθιερωμένο όνομα του συγγραφέα ενός κειμένου είτε αυτός είναι πρόσωπο, είτε είναι συλλογικό όργανο.
- **<editor>** περιέχει δευτερεύουσα δήλωση υπευθυνότητας για ένα κείμενο, όπως το όνομα του επιμελητή, μεταφραστή κλπ.
- **<sponsor>** περιέχει το όνομα του φορέα – χορηγού ενός κωδικοποιημένου ψηφιακού κειμένου.
- **<funder>** περιέχει το όνομα του χρηματοδότη (πρόσωπο, φορέας, οργανισμός) του έργου που είχε σαν αποτέλεσμα την παραγωγή και κωδικοποίηση ενός ψηφιακού κειμένου.
- **<principal>** περιέχει το όνομα του κυρίου ερευνητή – υπεύθυνου για τη δημιουργία και κωδικοποίηση του κειμένου σε μηχαναγνώσιμη ψηφιακή μορφή.



# titleStmt

- **<respStmt>** (statement of responsibility) πρόκειται για ένα σύνθετο στοιχείο που χρησιμοποιείται για τη δήλωση της πνευματικής υπευθυνότητας του περιεχομένου ενός κωδικοποιημένου κειμένου ή σώματος κειμένων ή μιας έκδοσης, ή μιας σειράς ή μιας καταγραφής (recording), για την οποία τα προηγούμενα στοιχεία δεν επαρκούν ή δεν εφαρμόζονται. Οποιαδήποτε πρόσωπα ή οργανισμοί έχουν συντελέσει στην παραγωγή, προώθηση, διάχυση και διανομή ενός κειμένου και δεν ανήκουν στις παραπάνω περιπτώσεις (author, editor, sponsor, funder και principal).
  - **<resp>** το οποίο περιέχει μια φράση που περιγράφει το είδος της υπευθυνότητας ενός προσώπου ή οργανισμού στην παραγωγή ή τη διανομή ενός κωδικοποιημένου κειμένου
  - **<name>** το όνομα του προσώπου ή του οργανισμού.



```
<titleStmt>
```

```
  <title level="m" type="main">Ιλιάς: ψηφιακή έκδοση</title>
```

```
  <title type="alt">Heliade: a digital edition</title>
```

```
  <author>Όμηρος</author>
```

```
  <editor>Καζαντζάκης, Ν.</editor>
```

```
  <editor>Κακριδής, Ι.Θ.</editor>
```

```
  <editor>Monro, D. B.</editor>
```

```
  <editor>Allen. T.W.</editor>
```

```
  <principal>Παππάς, Θ.</principal>
```

```
  <respStmt>
```

```
    <name>Γεργατσούλης, Ε.</name>
```

```
    <resp>εισαγωγή δεδομένων και τελικές διορθώσεις</resp>
```

```
  </respStmt>
```

```
  <respStmt>
```

```
    <name>Κόκκωνας, Ι.</name>
```

```
    <resp>κωδικοποίηση με το πρότυπο TEI</resp>
```

```
  </respStmt>
```

```
</titleStmt>
```



# publicationStmt

- Ομαδοποιεί πληροφορίες για την έκδοση – διανομή ενός κωδικοποιημένου έργου του γραπτού λόγου. Το στοιχείο είναι υποχρεωτικό και πρέπει να υπάρχει μόνο μία φορά στο στοιχείο <fileDesc> .
- Υποστοιχεία:
  - **<publisher>** περιέχει το όνομα του οργανισμού που έχει αναλάβει την υπευθυνότητα για την έκδοση ή τη διανομή του κειμένου, δηλαδή έχει τη δικαιοδοσία της δημοσιοποίησης ενός κειμένου.
  - **<distributor>** περιέχει το όνομα του προσώπου ή του φορέα που έχει αναλάβει την υπευθυνότητα της παραγωγής αντιγράφων ενός κειμένου.
  - **<authority>** στην περίπτωση που ένα κείμενο δεν έχει τυπικά εκδοθεί, αλλά κάποιος φορέας ή πρόσωπο έχει τη δικαιοδοσία για την κυκλοφορία του (release authority), τότε αυτός ο φορέας ή το πρόσωπο δηλώνεται στο περιεχόμενο του στοιχείου αυτού.



# publicationStmt

- **<pubPlace>** περιέχει τον τόπο έκδοσης, διανομής ή κυκλοφορίας του κωδικοποιημένου κειμένου.
- **<address>** περιέχει την ταχυδρομική διεύθυνση του προσώπου ή οργανισμού που δηλώνεται στο περιεχόμενο του στοιχείου **<publisher>**, ή **<distributor>** ή **<authority>**.
  - Το στοιχείο περιλαμβάνει το προαιρετικό υποστοιχείο **<addressline>** το οποίο είναι επαναλαμβανόμενο και χρησιμοποιείται όταν μιας διεύθυνση παρατίθεται σε περισσότερες της μιας γραμμής κειμένου. Το στοιχείο περιέχει το μέρος της διεύθυνσης που βρίσκεται σε μια γραμμή.
- **<idno>** παρέχει ένα κωδικό ταυτοποίησης κάποιου αντικειμένου (πχ βιβλιογραφικού τεκμηρίου, προσώπου, οργανισμού). Το γνώρισμα **type** προσδιορίζει την κατηγορία του κωδικού ταυτοποίησης (π.χ. ISBN, ή Αριθμός Μητρώου Κοινωνικής Ασφάλισης).
- **<date>** περιέχει μια ημερομηνία που αφορά σε έκδοση, παραγωγή αντιτύπων ή κυκλοφορία του κειμένου.



# publicationStmt

- **<availability>** περιέχει πληροφορίες για τη διαθεσιμότητα ενός κειμένου, π.χ. περιορισμούς ή άδειες χρήσης, περιγραφή της πνευματικής ιδιοκτησίας για το κείμενο κλπ.
  - Οι πληροφορίες αυτές κωδικοποιούνται από τις τιμές του γνωρίσματος `status` που εφαρμόζεται προαιρετικά στο στοιχείο και μπορούν να είναι `free`, όταν το κείμενο είναι διαθέσιμο χωρίς περιορισμούς, `restricted`, όταν υπάρχουν περιορισμοί στη διαθεσιμότητά του και `unknown` όταν η κατάσταση διαθεσιμότητας δεν είναι γνωστή.
  - Όταν το κείμενο καλύπτεται από μια συγκεκριμένη άδεια, τότε το υποστοιχείο `<licence>` περιέχει είτε το πλήρες κείμενο της άδειας, είτε μια συνοπτική δήλωσή της.
  - Το γνώρισμα `target` μπορεί να εφαρμοσθεί στο συγκεκριμένο στοιχείο για να δηλώσει μια δικτυακή διεύθυνση στην οποία αναφέρεται το είδος ή το κείμενο της άδειας.





```
<publicationStmt>
  <publisher> Πανεπιστημιακές Εκδόσεις Κρήτης</publisher>
  <pubPlace>Ηράκλειο</pubPlace>
  <address>
    <addressline>Ίδρυμα Τεχνολογίας & Έρευνας Τ.Θ. 1385</addressline>
    <addressline>Τ.Κ. 71110, Βούτες Ηρακλείου</addressline>
    <addressline>Κρήτη</addressline>
  </address>
  <date>2005</date>
  <idno type="ISBN">0-19-254705-4</idno>
  <availability status="restricted">
    <licence target="http://creativecommons.org/licenses/by-sa/3.0/">Διανέμεται με την άδεια Creative Commons Attribution-ShareAlike 3.0 </licence>
  </availability>
</publicationStmt>
```



# sourceDesc

- `<sourceDesc>` (source description): περιγράφει το κείμενο-πηγή από την οποία παρήχθη ένα κωδικοποιημένο κείμενο. Ουσιαστικά περιλαμβάνει είτε μια βιβλιογραφική αναφορά σε ένα ψηφιοποιημένο τεκμήριο με τη μορφή κειμένου, είτε μια φράση της μορφής "πρωτότυπο ψηφιακό κείμενο (born digital)" στην περίπτωση που το κωδικοποιημένο κείμενο δεν προϋπήρχε σε άλλη μορφή. Το στοιχείο είναι υποχρεωτικό και μπορεί να επαναλαμβάνεται περρισσότερες από μια φορά μέσα στο στοιχείο `<fileDesc>`.

# sourceDesc

- **source description:** περιγράφει το κείμενο-πηγή από την οποία παρήχθη ένα κωδικοποιημένο κείμενο. Ουσιαστικά περιλαμβάνει είτε μια βιβλιογραφική αναφορά σε ένα ψηφιοποιημένο τεκμήριο με τη μορφή κειμένου, είτε μια φράση της μορφής "πρωτότυπο ψηφιακό κείμενο (born digital)" στην περίπτωση που το κωδικοποιημένο κείμενο δεν προϋπήρχε σε άλλη μορφή. Το στοιχείο είναι υποχρεωτικό και μπορεί να επαναλαμβάνεται περισσότερες από μια φορά μέσα στο στοιχείο <fileDesc>.
- Τα υποστοιχεία του στοιχείου <sourceDesc> μπορούν να ανήκουν στις ακόλουθες κλάσεις:
  - **model.biblLike:** ομαδοποιεί στοιχεία που περιέχουν μια βιβλιογραφική περιγραφή,
  - **model.sourceDescPart:** ομαδοποιεί στοιχεία που χρησιμοποιούνται ως υποστοιχεία του <sourceDesc> και μπορούν να επαναλαμβάνονται
  - **model.listLike** που ομαδοποιεί στοιχεία που περιγράφουν καταλόγους ή λίστες.



```
<sourceDesc><p>Πρωτότυπο ψηφιακό κείμενο.</p></sourceDesc>
```

```
<sourceDesc>
```

```
  <bibl>"Οι Έμποροι των Εθνών" του Α. Παπαδιαμάντη, 1883</bibl>
```

```
</sourceDesc>
```

```
<sourceDesc>
```

```
  <biblStruct xml:lang="gr">
```

```
    <monogr>
```

```
      <author>Παπαδιαμάντης, Αλέξανδρος</author>
```

```
      <title type="main">Οι Έμποροι των Εθνών</title>
```

```
      <imprint>
```

```
        <pubPlace>Αθήνα</pubPlace>
```

```
        <publisher>Εστία</publisher>
```

```
        <date when="2002">2002</date>
```

```
      </imprint>
```

```
    </monogr>
```

```
  </biblStruct>
```

```
</sourceDesc>
```



```
<sourceDesc>
```

```
<!-- το στοιχείο scriptStmt δηλώνει ότι η πηγή είναι καταγραφή -->
```

```
<scriptStmt xml:id="ERT12">
```

```
<bibl>
```

```
<author>NET</author>
```

```
<title>Ειδήσεις NET</title>
```

```
<date when="1991-06-12">12 Ιουνίου 2009</date>
```

```
</bibl>
```

```
</scriptStmt>
```

```
<recordingStmt> <!-- τα χαρακτηριστικά της εγγραφής -->
```

```
<recording type="audio" dur="P30M" xml:id="rec-3001">
```

```
<equipment>
```

```
<p>Πολλαπλές μικροφωνικές συσκευές με μίξη σε  
στερεοφωνική ψηφιακή ταινία ήχου, συχνότητα δειγματοληψίας 44.1 KHz</p>
```

```
</equipment>
```

```
<date when="1991-06-12">12 Ιουνίου 2009</date>
```

```
</recording>
```

```
<recording type="audio" dur="P15M" xml:id="rec-3002">
```

```
<date when="1991-06-12">12 Ιουνίου 2009</date>
```

```
</recording> <recordingStmt> </sourceDesc>
```



# editionStmt

edition statement: ομαδοποιεί πληροφορίες που σχετίζονται με την επιμέλεια έκδοσης ενός κωδικοποιημένου κειμένου. Είναι προαιρετικό, αν και συνιστάται η ύπαρξή του από τις βέλτιστες πρακτικές του TEI και περιλαμβάνει τα ακόλουθα υπο-στοιχεία:

- **<edition>** περιέχει τις ιδιαίτερες πληροφορίες που αφορούν σε μια έκδοση ενός κειμένου (πχ 2η έκδοση επηυξημένη). Μια έκδοση περιλαμβάνει το σύνολο των πανομοιότυπων αντιτύπων ενός έργου που δημιουργούνται από ένα κύριο αντίγραφο (master copy). Αλλαγή έκδοσης συνιστά μια αναπαραγωγή από ένα τροποποιημένο κύριο αντίγραφο και όχι απαραίτητα από ένα άλλο φορέα διανομής του έργου. Στα ψηφιακά τεκμήρια η έννοια του κύριου αντιγράφου δεν υφίσταται για αυτό χρησιμοποιείται συνήθως ο ξένος όρος *version* για να περιγράψει μια διαφορετική απόδοση ενός ψηφιακού πόρου.
- Το περιεχόμενο του στοιχείου συνήθως περιλαμβάνει αριθμούς και ημερομηνίες συνοδευόμενες από τις λέξεις έκδοση, *edition*, *version*, αναθεωρημένη, *revised*, *update*, νέα κλπ. Στην περίπτωση που η περιγραφή της έκδοσης περιλαμβάνει ημερομηνία, τότε πρέπει να γίνεται χρήση του στοιχείου **<date>**, το οποίο θα περιέχει τα στοιχεία της ημερομηνίας, όπως αναγράφονται στο κείμενο.



# editionStmt

- **<respStmt>** όπως προαναφέρθηκε πρόκειται για ένα σύνθετο στοιχείο που αποτελείται από τα υποστοιχεία **<resp>** και **<name>**. Όταν χρησιμοποιείται ως υπο-στοιχείο του στοιχείου **<editionStmt>**, τότε περιλαμβάνει πληροφορίες για την υπευθυνότητα ενός προσώπου ή συλλογικού οργάνου στην έκδοση ενός κειμένου.

```
<editionStmt>
```

```
  <edition>αναθεωρημένη έκδοση<date>Ιούλιος 2012</date></edition>
```

```
  <respStmt>
```

```
    <resp>Εισαγωγή από</resp>
```

```
    <name>Καραβαγγέλη, Ζ.</name>
```

```
  </respStmt>
```

```
</editionStmt>
```



# extent

- περιγράφει το μέγεθος του κειμένου, όπως έχει καταχωρηθεί ή αποθηκευθεί σε κάποιο συμβατικό ή ψηφιακό μέσο, έτσι όπως μπορεί να μετρηθεί σε αντίστοιχες μονάδες μέτρησης.
- Το στοιχείο είναι προαιρετικό και μπορεί να επαναλαμβάνεται περισσότερες από μία φορές στο στοιχείο <teiHeader>

<extent>4532 bytes</extent> ή  
<extent>3200 προτάσεις</extent>





# seriesStmt

- series statement: ομαδοποιεί πληροφορίες που αφορούν στην περιγραφή της σειράς στην οποία ένα κείμενο ανήκει. Μια σειρά ορίζεται ως εξής:
  - ένα σύνολο από τεκμήρια που σχετίζονται μεταξύ τους από το γεγονός ότι κάθε τεκμήριο εκτός από τον τίτλο του, φέρει τον τίτλο ενός συλλογικού έργου δημιουργώντας μια ολότητα. Τα επιμέρους τεκμήρια μπορεί να φέρουν αρίθμηση, χωρίς αυτό να είναι απαραίτητο
  - κάθε ένα από δύο ή περισσότερα τεκμήρια (υπό μορφή, τόμων, άρθρων, διαλέξεων κλπ) με παρόμοια θεματική που εκδίδονται με μια σειρά
  - μια ακολουθία αριθμημένων τόμων στο πλαίσιο ενός περιοδικού ή ενός έργου.



# seriesStmt

- Το στοιχείο <seriesStmt> είναι προαιρετικό και μπορεί να περιλαμβάνει ως περιεχόμενο είτε ελεύθερο κείμενο είτε τα ακόλουθα υπο-στοιχεία:
  - <title> που περιλαμβάνει τον τίτλο ενός τεκμηρίου.
  - <idno> παρέχει ένα κωδικό ταυτοποίησης κάποιου αντικειμένου (πχ βιβλιογραφικού τεκμηρίου, προσώπου, οργανισμού).
  - <respStmt> όπως προαναφέρθηκε κατά την περιγραφή του στοιχείου <titleStmt> πρόκειται για ένα σύνθετο στοιχείο που αποτελείται από τα υποστοιχεία <resp> και <name>. Όταν χρησιμοποιείται ως υπο-στοιχείο του στοιχείου <seriesStmt>, τότε περιλαμβάνει πληροφορίες για την υπευθυνότητα ενός προσώπου ή συλλογικού οργάνου στην έκδοση μιας σειράς.



# seriesStmt

```
<seriesStmt>
```

```
  <title level="s">Κωδικοποιημένα Κείμενα Αρχαίας  
  Ελληνικής Γραμματείας</title>
```

```
  <respStmt>
```

```
    <resp>επιμέλεια </resp>
```

```
    <name>Παππάς, Θεόδωρος</name>
```

```
  </respStmt>
```

```
  <biblScope type="vol">1</biblScope>
```

```
  <idno type="ISSN">978 345 6789</idno>
```

```
</seriesStmt>
```



# notesStmt

- notes statement, προαιρετικό υποστοιχείο του <fileDesc> και περιέχει τις σημειώσεις που παρέχουν επιπλέον πληροφορίες για ένα κείμενο, πέρα από αυτές που έχουν καταγραφεί κατά τη βιβλιογραφική του περιγραφή.
- Περιλαμβάνει ένα ή περισσότερα υποστοιχεία <note> τα οποία μπορεί να φέρουν σημειώσεις της μορφής:
  - ημερομηνίες που σχετίζονται με την κατάσταση του περιεχομένου ενός κειμένου σε ηλεκτρονική μορφή
  - ονόματα φυσικών προσώπων και συλλογικών οργάνων που συνδέονται με την φάση της παραγωγής ή της διαχείρισης μιας ηλεκτρονικής έκδοσης, αν αυτά δεν περιλαμβάνονται σε δηλώσεις υπευθυνότητας για την έκδοση ενός ηλεκτρονικού τεκμηρίου
  - δηλώσεις για τη διαθεσιμότητα ενός ηλεκτρονικού τεκμηρίου σε άλλη μορφή ή τη διαθεσιμότητα τεκμηρίωσης του
  - δηλώσεις για την ύπαρξη του έργου που αντιστοιχεί το τεκμήριο ή της περίληψής του και σε άλλες γλώσσες εκτός από τη γλώσσα του τεκμηρίου
  - δηλώσεις για άλλους κωδικούς ταυτοποίησης του τεκμηρίου εκτός από τον κωδικό που περιέχει το στοιχείο <idno> και κατάλογοι σχετικών δημοσιευμάτων.



# notesStmt

<notesStmt>

<note>Ιστορικός σχολιασμός από τον Ε. Πρόντζα</note>

<note>Η οπτική αναγνώριση χαρακτήρων έγινε από το Εργαστήριο Ψηφιακών Βιβλιοθηκών του Ιονίου Πανεπιστημίου.</note>

</notesStmt>



# profileDesc

- Έχει στόχο την παροχή πληροφοριών του πλαισίου στο οποίο κωδικοποιήθηκε ένα κείμενο ή ένα σώμα κειμένων.
- Παρέχει μη βιβλιογραφικές πληροφορίες για το κωδικοποιημένο κείμενο, όπως για την ή τις γλώσσες που χρησιμοποιήθηκαν κατά την κωδικοποίηση, τους παράγοντες που συντέλεσαν και τις συνθήκες υπό τις οποίες κωδικοποιήθηκε το κείμενο ή το σώμα.



# profileDesc

- Το στοιχείο περιλαμβάνει τα ακόλουθα υποστοιχεία τα οποία απαρτίζουν την κλάση μοντέλου περιεχομένου `model.profileDesc` class.
  - **<creation>** είναι προαιρετικό υποστοιχείο και περιλαμβάνει πληροφορίες για τη δημιουργία ενός κειμένου, όπως πληροφορίες για την ημερομηνία και τον τόπο δημιουργίας του. Οι πληροφορίες αυτές είναι σημαντικές σε περιπτώσεις που υπάρχουν γλωσσικές παραλλαγές στα κείμενα ενός σώματος κειμένου, αλλά είναι δύσκολο να προκύψουν με μεγάλο βαθμό βεβαιότητας από τη βιβλιογραφική περιγραφή που παρέχεται από το στοιχείο `<fileDesc>`.
  - **<langUsage>** περιγράφει τις χρησιμοποιούμενες γλώσσες και διαλέκτους του κειμένου
  - **<textClass>** περιγράφει θεματικά το τεκμήριο

# profileDesc

```
<profileDesc>
  <creation>
    <date when="1988-07-10">10 Ιουλίου 1988</date>
  </creation>
  <textClass>
    <classCode scheme="http://www.udcc.org">001.9</classCode>
    <keywords scheme="http://authorities.loc.gov">
      <list>
        <item>End of the world</item>
        <item>History - philosophy</item>
      </list>
    </keywords>
  </textClass>
  <langUsage>
    <language ident="fr-CA" usage="60">Québécois</language>
    <language ident="en-GB" usage="20">British English</language>
  </langUsage>
</profileDesc>
```

