ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

Διαχείριση Εγγράφων SGML σε Βάσεις Γνώσης

Αθηνά Α.Χ. Τραψιώτη

Μεταπτυχιακή Εργασία

Ηράκλειο, Ιανουάριος 1997

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

Διαχείριση Εγγράφων SGML σε Βάσεις Γνώσης

Εργασία που υποβλήθηκε από την
ΑΘΗΝΑ Α.Χ. ΤΡΑΨΙΩΤΗ
ως μερική εκπλήρωση των απαιτήσεων
για την απόκτηση
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Αθηνά Α.Χ. Τραψιώτη
Τμήμα Επιστήμης Υπολογιστών

Εισηγητική Επιτροπή:

Πάνος Κωνσταντόπουλος, Αναπληρωτής Καθηγητής, Επόπτης

Απόστολος Τραγανίτης, Αναπληρωτής Καθηγητής, Μέλος

Κατερίνα Χούστη, Αναπληρώτρια Καθηγήτρια, Μέλος

Δεκτή:

Πάνος Κωνσταντόπουλος
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Ηράκλειο, Ιανουάριος 1997

Διαχείριση Εγγράφων SGML σε Βάσεις Γνώσης.

Αθηνά Α.Χ. Τραψιώτη

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

ΠΕΡΙΛΗΨΗ

Η SGML είναι ένα διεθνές πρότυπο ψηφιακής παράσταστης εγγράφων που επιτρέπει την κωδικοποίηση της δομής, του περιεχομένου και του τρόπου επεξεργασίας των εγγράφων. Η υιοθέτηση του προτύπου SGML στη διαχείριση εγγράφων, επιτρέπει την ανταλλαγή και την επεξεργασία τους από διαφορετικά υπολογιστικά συστήματα. Η SGML τυγχάνει ευρείας διάδοσης και αποδοχής τα τελευταία χρόνια. Κατά συνέπεια, η ανάπτυξη συστημάτων διαχείρισης εγγράφων SGML αποτελεί ένα επίκαιρο και δημοφιλές ερευνητικό και επιχειρηματικό πεδίο.

Αντικείμενο της παρούσας εργασίας είναι η σχεδίαση και υλοποίηση του συστήματος "Κώδιξ", ενός σύστηματος για τη διαχείριση εγγράφων SGML. Το σύστημα Κώδιξ επιτρέπει την αποθήκευση, τον σημασιολογικό ευρετηριασμό, την αναζήτηση, τον σχολιασμό και την καταγραφή του ιστορικού εγγράφων SGML.

Για την αποθήκευση των εγγράφων υιοθετήθηκε ένα υβριδικό σχήμα, το οποίο αποτελείται από το σύστημα αρχείων του λειτουργικού συστήματος και μια βάση γνώσης που παριστάνει τη γνώση σε σημασιολογικό δίκτυο. Στο σύστημα αρχείων του λειτουργικού συστήματος αποθηκεύονται τα έγγραφα SGML, ενώ στη βάση γνώσης παριστάνονται πληροφορίες που αφορούν τη λογική δομή, το νοηματικό περιεχόμενο και τη διαχείρισή τους. Ο ευρετηριασμός γίνεται με την συσχέτιση εγγράφων μεταξύ τους ή με όρους που αφορούν στο σημασιολογικο περιεχόμενό τους. Υποστηρίζεται επίσης ο ευρετηριασμός εγγράφων με όρους που ανήκουν σε θησαυρό.

Το σύστημα Κώδιξ επιτρέπει την αναζήτηση εγγράφων ή τμημάτων τους με βάση τον τύπο εγγράφου που ακολουθούν, τη λογική δομή τους, και το σημασιολογικό τους περιεχόμενο. Η αναζήτηση πληροφορίας στο σύστημα Κώδιξ γίνεται με πλοήγηση και με

τη χρήση προκαθορισμένων ερωτήσεων για την άντληση της πληροφορίας από τη βάση γνώσης του συστήματος. Το αποτέλεσμα των ερωτήσεων αυτών μπορεί να είναι κείμενο ή να εμφανίζεται σε γραφική μορφή.

Λειτουργικός πυρήνας του Κώδικα είναι το Σύστημα Σημασιολογικού Ευρετηριασμού (Semantic Index System - SIS), που αναπτύχθηκε στο Ινστιτούτο Πληροφορικής του Ιδρύματος Τεχνολογίας και Ερευνας.

Η ιδιαιτερότητα του συστήματος Κώδιξ συνίσταται στη εκφραστική παράσταση, το σημασιολογικό ευρετηριασμό και την αποδοτική αναζήτηση των εγγράφων SGML.

Το σύστημα Κώδιξ έχει χρησιμοποιηθεί για τη διαχείριση ιστορικών εγγράφων του Τουρκικού Αρχείου Χανίων και πρόκειται να χρησιμοποιηθεί στα πλαίσια του έργου ”AQUARELLE” του ευρωπαϊκού προγράμματος Τηλεματικής το οποίο αφορά στην πρόσβαση σε πολιτισμικές πληροφορίες.

Επόπτης : Πάνος Κωνσταντόπουλος
Αναπληρωτής Καθηγητής Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

SGML Document Management in Knowledge Bases.

Athena A.C. Trapsioti

Master of Science Thesis

Department of Computer Science
University of Crete

ABSTRACT

SGML is an international standard for document representation that supports the encoding of the structure, content, and processing of documents. Management of documents using SGML allows their interchange and processing by different computer systems.

SGML has gained widespread acceptance. Consequently, the development of SGML document management systems is a hot research and business field.

This work is concerned with the design and implementation of Codex, an SGML document management system. Codex deals with the storage, semantic indexing, search, annotation and history recording of SGML documents.

Document storage is implemented using a hybrid scheme consisting of the file system of an operating system together with a knowledge base in which knowledge is represented in the form of a semantic network. SGML documents are stored in the file system and information related to the structure, content and processing of documents is represented in the knowledge base.

Documents are indexed with terms related to their content as well as with related documents. Moreover indexing of documents with terms belonging to a thesaurus is supported.

Codex offers document (or document element) search according to their type, logical structure and semantic criteria. Retrieval is performed in two ways: navigation and predefined queries. The result of the queries can be in text and graphical format.

The operational kernel of Codex is the Semantic Index System - SIS, developed of the Institute of Computer Science, Foundation for Research and Technology - Hellas.

The basic characteristics of Codex are expressive representation, semantic indexing and

effective retrieval of SGML documents.

Codex has been employed in the management of historical documents of the Turkish Archive of Chania as well as within the framework of the EC Telematics project ”AQUARELLE” which deals with accessing cultural information.

Supervisor : Panos Constantopoulos
Associate Professor of Computer Science
University of Crete

Ευχαριστίες

Νιώθω την ανάγκη να ευχαριστήσω τον κ. Γιάννη Τζίτζικα για την αμέριστη συμπαράσταση και τις ουσιαστικές παρατηρήσεις του στην πορεία της εργασίας αυτής. Ο καθαρός και εποικοδομητικός τρόπος σκέψης του καθώς και η αγάπη του για την επιστήμη με βοήθησαν πολύ κατά τη διάρκεια εκπόνησης της εργασίας αυτής.

Επίσης νιώθω μεγάλη τιμή και ευχαριστώ ιδιαίτερα τον επόπτη καθηγητή μου κ. Πάνο Κωνσταντόπουλο που μου εμπιστεύθηκε την εργασία αυτή. Τον ευχαριστώ επίσης για τον χρόνο που διέθετε ακούραστα στις συζητήσεις μας και για τις πολύτιμες επιστημονικές συμβουλές του καθ' όλη τη διάρκεια της συνεργασίας μας η οποία αποτελεί για μένα το σημαντικότερο απόκτημα των τελευταίων χρόνων.

Επίσης ευχαριστώ θερμά τους γονείς μου Αργύρη και Χρυσούλα και την αδελφή μου Πετρούλα για το κουράγιο που μου εμπνέει η εμπιστοσύνη και η απέραντη αγάπη με την οποία με περιβάλλουν.

Θα ήθελα να ευχαριστήσω ακόμη την κ. Μαρία Χριστοφοράκη, τον κ. Βασίλη Χριστοφίδη καθώς επίσης και όλα τα μέλη (νυν και πρώην) της Ομάδας Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του Ινστιτούτου Πληροφορικής του Ιδρύματος Τεχνολογίας και Έρευνας για την πολύτιμη συνεργασία και τη συμπαράστασή τους.

Τέλος, θα ήθελα να ευχαριστήσω τα μέλη της εξεταστικής επιτροπής κ. Απόστολο Τραγανίτη και κ. Κατερίνα Χούστη για τα σχόλιά τους στην τελική διαμόρφωση του κειμένου καθώς επίσης και το Ινστιτούτο Πληροφορικής του Ιδρύματος Τεχνολογίας και Ερευνας για την οικονομική και υλικοτεχνική υποστήριξη που μου παρείχε κατά τη διάρκεια των προπτυχιακών και μεταπτυχιακών μου σπουδών.

Περιεχόμενα

Περίληψη **i**

Abstract **iii**

Ευχαριστίες **v**

1 Εισαγωγή **1**

1.1 Το πρότυπο SGML 1

1.2 Το πρόβλημα 1

1.3 Αποτελέσματα της Εργασίας 2

1.4 Υλοποίηση 4

1.5 Οργάνωση της εργασίας 4

2 Σχετικές Εργασίες **7**

2.1 Εισαγωγή 7

2.2 Γενικές Παρατηρήσεις 7

2.2.1 Ερευνητικά Συστήματα 8

2.2.2 Εμπορικά Συστήματα 11

2.3 Σύγκριση συστημάτων 15

3 Το πρότυπο SGML **17**

3.1 Ιστορική Αναδρομή 17

3.2 Σκοπιμότητα και Χρήση 18

3.3 Τα στοιχειώδη εργαλεία της SGML 19

3.4 Το Περιβάλλον μιας Εφαρμογής SGML 20

3.4.1 Δημιουργία εγγράφων SGML 20

3.4.2 Αποθήκευση εγγράφων SGML 21

3.4.3 Εξαγωγή πληροφορίας από έγγραφα SGML 22

3.5 Περιγραφή της SGML 22
3.5.1 Δήλωση SGML 23
3.5.2 Πρόλογος 23
3.5.3 Περιπτώσεις εγγράφων (Document Instance Set) 27
3.5.4 Κωδικοποίηση Εγγράφων SGML. 27
3.5.5 Αρχικοί τύποι στην SGML 28

4 Η γλώσσα SIS-Telos και το σύστημα Σύστημα Σημασιολογικού Ευρετηριασμού 31
4.1 Παράσταση γνώσης στη γλώσσα SIS-Telos 32
4.2 Βασικά χαρακτηριστικά και Πλεονεκτήματα της SIS-Telos 34
4.3 Περιγραφή του Συστήματος Σημασιολογικού Ευρετηριασμού - SIS 34
4.3.1 Γραφικό περιβάλλον του ΣΣΕ 35
4.3.2 Προγραμματική διεπαφή χρήσης του ΣΣΕ 36

5 Ένα μοντέλο παράστασης εγγράφων SGML σε Telos 39
5.1 Παρουσίαση του μοντέλου 39
5.2 Αποψη SGML 40
5.2.1 Πρώτη στάθμη 42
5.3 Αποψη Επεξεργασίας 47
5.4 Εκτεταμένο Παράδειγμα χρήσης του Γενικού Μοντέλου της SGML 51
5.4.1 Περιγραφή του CI DTD. 53
5.4.2 Εισαγωγή του τύπου εγγράφου CI στο ΣΣΕ 53
5.4.3 Ονοματολογία των οντοτήτων της άποψης SGML 56

6 Περιγραφή του συστήματος διαχείρισης Εγγράφων SGML, Κώδιξ 59
6.1 Κύκλος ζωής εγγράφου σε μια εφαρμογή SGML 59
6.1.1 Συλλογή εγγράφων 60
6.1.2 Προ-επεξεργασία 60
6.1.3 Δημιουργία Βασικών Εγγράφων 61
6.1.4 Διαχείριση βασικών Εγγράφων 61
6.1.5 Επεξεργασία 61
6.1.6 Παραγωγή προϊόντων 61
6.2 Δομικά τμήματα της αρχιτεκτονικής του συστήματος Κώδιξ 62
6.2.1 Αρχιτεκτονική client-server 62
6.3 Κύκλος ζωής εγγράφου στο σύστημα Κώδιξ 64
6.3.1 Συλλογή εγγράφων 64

6.3.2 Προ-επεξεργασία 64
6.3.3 Δημιουργία Βασικών Εγγράφων SGML 65
6.3.4 Διαχείριση βασικών Εγγράφων 65
6.3.5 Επεξεργασία 67
6.3.6 Παραγωγή προϊόντων 71
6.3.7 Τεχνικά χαρακτηριστικά του συστήματος Κώδιξ 72
6.3.8 Επεκτάσεις του συστήματος Κώδιξ 73
6.3.9 Χρήση Εξωτερικών Εργαλείων 75
6.4 Παρουσίαση των Προκαθορισμένων Ερωτήσεων 79
6.4.1 Ερωτήσεις πρώτης τάξεως 79
6.4.2 Αναδρομικές Ερωτήσεις 87
6.4.3 Ερωτήσεις πολλαπλών παραμέτρων 90

7 Μια Εφαρμογή: Το Τουρκικό Αρχείο Χανίων **93**
7.1 Περιγραφή του Σημασιολογικού Μοντέλου 93
7.1.1 Περιγραφή της οντολογίας της άποψης SGML 94
7.1.2 Περιγραφή της οντολογίας της άποψης Εννοιών 96
7.1.3 Περιγραφή της οντολογίας της άποψης Επεξεργασίας 96
7.2 Περιγραφή του Τύπου Εγγράφου που υιοθετήθηκε 97
7.2.1 Δομή των Ιεροδικαστικών Εγγράφων 97
7.2.2 Αναλυτική περιγραφή του TAX dtd 98
7.2.3 Ιδιαιτερότητες - Επέκταση του TAX dtd 104

8 Επίλογος **107**
8.1 Συμπεράσματα 107
8.2 Επεκτάσεις 108

Βιβλιογραφία **110**

Γλωσσάρι ελληνικών όρων **117**

Γλωσσάρι αγγλικών όρων **119**

A Τύπος εγγράφου ιεροδικαστικών εγγράφων **121**

B Τύπος εγγράφου Σουλτανικών εγγράφων **125**

C Τύπος εγγράφου Αποψης Επεξεργασίας **129**

D Γενικό μοντέλο άποψης SGML **131**

Κεφάλαιο 1

Εισαγωγή

1.1 Το πρότυπο SGML

Η SGML είναι ένα πρότυπο ψηφιακής παράστασης εγγράφων που αναπτύχθηκε στις αρχές της δεκαετίας του ’80 και επιτρέπει τον εμπλουτισμό του περιεχομένου των εγγράφων με πληροφορία που αφορά στη δομή, το περιεχόμενο και τον τρόπο επεξεργασίας τους.
Η χρησιμότητα της SGML έγκειται στην αποδοτικότερη αναζήτηση και γενικά στη αποδοτική διαχείριση των εγγράφων που ακολουθούν το πρότυπο αυτό με την αξιοποίηση της πληροφορίας που προστίθεται σε αυτά. Επίσης, λόγω του ότι η SGML έχει υιοθετηθεί ως διεθνές πρότυπο, επιτρέπει την ανταλλαγή και επεξεργασία των εγγράφων από διαφορετικά υπολογιστικά συστήματα.

Για το λόγο αυτό η SGML χρησιμοποιείται από μεγάλους εκδοτικούς οργανισμούς και εταιρείες των οποίων ένα μεγάλο ποσοστό πληροφορίας που διαχειρίζονται και ανταλλάσουν βρίσκεται σε μορφή ηλεκτρονικών εγγράφων.

1.2 Το πρόβλημα

Η ευρεία διάδοση και αποδοχή της SGML έχει ως συνέπεια την ανάγκη ύπαρξης ειδικών συστημάτων διαχείρισης εγγράφων (Document Management Systems, DMS) τα οποία θα αξιοποιούν τα πλεονεκτήματα του προτύπου αυτού.
Βασική διαχειριστική απαίτηση για τα συστήματα αυτά είναι η δυνατότητα σημασιολογικής αναζήτησης εγγράφων η οποία επιτρέπει την αναζήτηση τους βάσει της δομής και του περιεχομένου τους.

Επίσης, σημαντική είναι η δυνατότητα πλοήγησης στη λογική δομή των εγγράφων. Τέλος, είναι σημαντική η υποστήριξη λειτουργιών συγγραφής εγγράφων (authoring), δημιουργίας νέων εκδόσεων, ροής εργασίας (workflow) καθώς και η εξασφάλιση των κλασικών μηχανισμών ελέγχου που παρέχουν τα συστήματα διαχείρισης βάσεων δεδομένων (ταυτόχρονη πρόσβαση, ανάνηψη κοκ.).
Τα συστήματα διαχείρισης εγγράφων SGML που έχουν υλοποιηθεί και χρησιμοποιούνται για την ικανοποίηση των παραπάνω απαιτήσεων αποτελούν συνήθως ολοκληρωμένες λύσεις που αποτελούνται από συνεργαζόμενα υποσυστήματα (modules) κάθε ένα από τα οποία είναι υπεύθυνο για την εκτέλεση ενός συνόλου λειτουργιών. Είναι χαρακτηριστικό ότι οι επιμέρους λειτουργίες που επιτελούν τα συστήματα αυτά εξαρτώνται από την συγκεκριμένη εφαρμογή την οποία καλούνται να υποστηρίξουν.

1.3 Αποτελέσματα της Εργασίας

Αποτέλεσμα την εργασίας αυτής είναι το σύστημα διαχείρισης εγγράφων SGML Κώδιξ. Το σύστημα αυτό επιτρέπει την αποθήκευση, τον σημασιολογικό ευρετηριασμό, την αναζήτηση, τον σχολιασμό και την καταγραφή του ιστορικού εγγράφων SGML. Το σύστημα είναι ικανό για τη διαχείριση πολλαπλών και διαφορετικών τύπων εγγράφων.

Για τις ανάγκες **αποθήκευσης** των εγγράφων του συστήματος υιοθετήσαμε ένα ιβρυδικό σχήμα το οποίο αποτελείται από το σύστημα αρχείων του λειτουργικού συστήματος και μια βάση γνώσης που παριστάνει τη γνώση σε σημασιολογικό δίκτυο. Στο σύστημα αρχείων του λειτουργικού συστήματος αποθηκεύονται τα πραγματικά έγγραφα με τη μορφή αρχείων SGML ενώ στη βάση γνώσης παριστάνονται πληροφορίες που αφορούν τη λογική δομή, το νοηματικό περιεχόμενο και τη διαχείρισή τους.

Για τη σχεδίαση της βάσης γνώσης του συστήματος Κώδιξ υιοθετήθηκε ένα μοντέλο **τριών απόψεων**.

- Στην πρώτη άποψη (άποψη SGML) παριστάνονται οι βασικές έννοιες που χρησιμοποιεί η SGML για την περιγραφή της δομής των εγγράφων, οι τύποι εγγράφων και η δομή των εγγράφων. Η άποψη αυτή αποτελείται από ένα γενικό μοντέλο που αποτελεί τον αρχικό πληθυσμό της και από οντότητες και σχέσεις που προστίθενται στη συνέχεια και παριστάνουν τη δομή των τύπων εγγράφων (Document Type Definition, DTD) και των εγγράφων. Το γενικό μοντέλο παριστάνει τις βασικές έννοιες της SGML και

τις σχέσεις μεταξύ τους. Το μοντέλο αυτό χρησιμοποιήθηκε (για επαλήθευση της ορθότητάς του) για την παράσταση ενός πολύπλοκου τύπου εγγράφου (CI DTD) που αναπτύχθηκε στα πλαίσια του προγράμματος AQUARELLE[1] και είναι κατάλληλο για παρουσίαση πολιτισμικής πληροφορίας.

- Στη δεύτερη άποψη (άποψη εννοιών) παριστάνονται πληροφορίες που αφορούν το νοηματικό περιεχόμενο των εγγράφων

- Η τρίτη άποψη περιέχει οντότητες και σχέσεις που χρησιμοποιούνται για την παρουσίαση, τον ευρετηριασμό και τον σχολιασμό των εγγράφων.

Η υιοθέτηση των παραπάνω απόψεων στη σχεδίαση της βάσης γνώσης επιτρέπει την αναζήτηση των εγγράφων με χρήση πολλαπλών και διαφορετικών κριτηρίων.

Η λειτουργία της αναζήτησης εγγράφων ή τμημάτων τους στο σύστημα Κώδιξ γίνεται με δύο τρόπους α) πλοήγηση στη βάση γνώσης του συστήματος και β) χρήση προκαθορισμένων ερωτήσεων. Η αναζήτηση μπορεί να γίνει με βάση τον τύπο εγγράφου που ακολουθούν, τη λογική δομή τους και το σημασιολογικό περιεχόμενό τους. Το αποτέλεσμα των ερωτήσεων μπορεί να είναι κείμενο ή να εμφανίζεται σε γραφική μορφή.

Η λειτουργία του ευρετηριασμού επιτρέπει συσχετισμό εγγράφων με όρους που αναφέρονται στη σημασιολογία τους ή και με άλλα έγγραφα. Επίσης, υποστηρίζεται ο ευρετηριασμός των εγγράφων με όρους που ανήκουν σε θησαυρό.

Το σύστημα Κώδιξ υποστηρίζει τον σχολιασμό σε οποιοδήποτε τμήμα των διαχειριζόμενων εγγράφων.

Η ιδαιτερότητα του συστήματος Κώδιξ συνίσταται στην εκφραστική παράσταση, το σημασιολογικό ευρετηριασμό και την αποδοτική αναζήτηση των εγγράφων SGML.

Το σύστημα Κώδιξ χρησιμοποιήθηκε στα πλαίσια του προγράμματος AQUARELLE το οποίο αφορά στη δημιουργία ενός κατανεμημένου συστήματος πρόσβασης σε πολιτισμική πληροφορία. Επίσης ο Κώδιξ χρησιμοποιήθηκε στη διαχείριση ιστορικών εγγράφων του Τουρκικού Αρχείου Χανίων. Η έλλειψη κάποιου πρότυπου τύπου εγγράφου ειδικού για τη συγκεκριμένη εφαρμογή οδήγησε στο σχεδιασμό ενός νέου τύπου εγγράφου (TAX DTD) ο οποίος είναι σύμφωνος με την περιγραφή της δομής των εγγράφων αυτών (όπως αυτός περιγράφεται στη βιβλιογραφία).

[1] Ο τύπος εγγράφου CI αναπτύχθηκε από την εταιρεία EUROCLID

1.4 Υλοποίηση

Λειτουργικός πυρήνας του Κώδικα είναι το Σύστημα Σημασιολογικού Ευρετηριασμού, ΣΣΕ που αναπτύχθηκε στο Ινστιτούτο Πληροφορικής. Το σύστημα αυτό προσφέρει ένα προσαρμόσιμο εργαλείο επαφής χρήσης, ένα εργαλείο εισαγωγής δεδομένων, δυνατότητα κατασκευής προκαθορισμένων ερωτήσεων και μια προγραμματική διεπαφή χρήσης για σύνδεση με άλλες εφαρμογές. Η επιλογή του ΣΣΕ για την υλοποιηση του λειτουργικού πυρήνα του συστήματος έγινε διότι αυτό είναι κατάλληλο για παράσταση (και αποδοτική αναζήτηση) μεγάλου όγκου πληροφορίας συσχετιζόμενης με πολλαπλούς τρόπους. Επίσης, η χρήση του ΣΣΕ [27, 3] επιτρέπει την επέκταση και προσαρμογή της βάσης γνώσης σε νέες εφαρμογές δίχως να γίνεται αναδιάρθρωση της πληροφορίας που υπάρχει ήδη στο σύστημα.

1.5 Οργάνωση της εργασίας

Στο δεύτερο κεφάλαιο γίνεται κριτική ανασκόπιση άλλων συστημάτων που έχουν αναπτυχθεί με σκοπό τη διαχείριση εγγράφων SGML.

Στο τρίτο κεφάλαιο γίνεται μια ιστορική αναδρομή στους τρόπους κωδικοποίησης εγγράφων. Επίσης, αναφέρονται η σκοπιμότητα και η χρήση του προτύπου SGML καθώς και τα στοιχειώδη εργαλεία μιας εφαρμογής SGML. Τέλος, περιγράφονται αναλυτικά οι βασικές έννοιες της SGML.

Στο τέταρτο κεφάλαιο υπάρχει σύντομη περιγραφή του Συστήματος Σημασιολογικού Ευρετηριασμού και της γλώσσας παράστασης γνώσης SIS-Telos που αποτελούν τον λειτουργικό πυρήνα του συστήματος Κώδιξ.

Στο πέμπτο κεφάλαιο παρουσιάζεται η οντολογία της άποψης SGML και της άποψης Επεξεργασίας. Γίνεται μια αναλυτική περιγραφή των οντοτήτων που αποτελούν το Γενικό μοντέλο της άποψης SGML που χρησιμοποιείται ως βάση για την παράσταση εγγράφων SGML. Επίσης, παρουσιάζεται ο τύπος εγγράφου CI (CI DTD) που ακολουθείται[2] από έγγραφα που περιέχουν πολιτισμική πληροφορία. Επίσης παρουσιάζεται ο τρόπος ενσωμάτωσής του στο ΣΣΕ.

Στο έκτο κεφάλαιο περιγράφονται τα στάδια του κύκλου ζωής εγγράφων σε ένα σύστημα διαχείρισης εγγράφων SGML. Επίσης παρουσιάζονται η αρχιτεκτονική και οι

[2] Ο τύπος εγγράφου CI DTD χρησιμοποιείται στα πλαίσια του ESPRIT προγράμματος Aquarelle

λειτουργίες (που αντιστοιχούν σε κάθε στάδιο του κύκλου ζωής των εγγράφων) του συστήματος Κώδιξ. Τέλος, δίνεται το σύνολο των προκαθορισμένων ερωτήσεων του Κώδικα για την ανάκτηση πληροφορίας σχετικής με τα διαχειριζόμενα έγγραφα.

Στο έβδομο κεφάλαιο δίνεται μια αναλυτική περιγραφή της οντολογίας της βάσης γνώσης του συστήματος Κώδιξ για την διαχείριση των ιεροδικαστικών εγγράφων του ΤΑΧ. Επίσης, περιγράφεται αναλυτικά ο τύπος εγγράφου που σχεδιάσαμε για την περιγραφή της δομής των εγγράφων αυτών.

Στο όγδοο κεφάλαιο παρουσιάζονται τα συμπεράσματα από την παρούσα εργασία και προτείνονται κάποιες προτάσεις για σχετικές μελλοντικές εργασίες.

Στο παράρτημα Α παρουσιάζεται ο τύπος εγγράφου ΤΑΧ . Στο παράρτημα Β παρουσιάζεται ο τύπος εγγράφου που σχεδιάσαμε για τις οντότητες που συνιστούν την άποψη επεξεργασίας . Στο παράρτημα Γ παρουσιάζεται ο κατάλογος με τις οντότητες της άποψης SGML του σημασιολογικού μοντέλου του ΣΣΕ για την εφαρμογή του ΤΑΧ. Στο παράρτημα Δ παρουσιάζεται ο κατάλογος με τις οντότητες που συνιστούν το Γενικό μοντέλο της άποψης SGML της βάσης γνώσης του συστήματος.

Κεφάλαιο 2

Σχετικές Εργασίες

2.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται οι εργασίες που λήφθηκαν υπόψη κατά τη σχεδίαση του συστήματος Κώδιξ. Επικεντρώνουμε το ενδιαφέρον μας στον τρόπο αποθήκευσης, ευρετηριασμού και σχολιασμού εγγράφων SGML των συστημάτων αυτών. Για την εκπόνηση της εργασίας αυτής μελετήθηκαν τόσο συστήματα που βρίσκονται σε ερευνητικό στάδιο όσο και εμπορικά συστήματα.

2.2 Γενικές Παρατηρήσεις

Μια γενική παρατήρηση που προκύπτει από τη μελέτη συστημάτων διαχείρισης εγγράφων SGML είναι πως στα ερευνητικά κυρίως συστήματα δίνεται ιδιαίτερο βάρος στην αποδοτική παράσταση της πολλαπλά συσχετιζόμενης πληροφορίας. Για το λόγο αυτό επιλέγεται η οντοκεντρική παράσταση της πληροφορίας [21]. Επίσης η έρευνα επικεντρώνεται στη μελέτη μεθόδων ανάκτησης πληροφορίας βάσει κυρίως σημασιολογικών κριτηρίων. Ακόμη, δίνεται ιδιαίτερη έμφαση στην αξιολόγηση των συστημάτων από διαφορετικές κατηγορίες χρηστών που το γνωστικό τους πεδίο είναι λιγότερο ή περισσότερο σχετικό με το περιεχόμενο των εγγράφων που διαχειρίζεται το σύστημα. Τέλος τα συστήματα αυτά διερευνούν τρόπους υποστήριξης χρηστών με ιδιαίτερες απαιτήσεις για την διαχείριση των εγγράφων όπως ειδικοί μελετητές, ερευνητές κ.ο.κ. (δυνατότητα σχολιασμού, δυναμικού ευρετηριασμού κτλ.).

Από την άλλη πλευρά, τα εμπορικά συστήματα επιμένουν ιδιαίτερα στη χρήση κλασσικών

σχεσιακών συστημάτων διαχείρισης βάσεων δεδομένων (RDBMS). Η επιλογή αυτή προσεγγίζει καλύτερα ένα μεγάλο ποσοστό χρηστών που είναι περισσότερο εξοικειωμένο με το σχεσιακό μοντέλο παράστασης πληροφορίας. Επίσης τα συστήματα αυτά παρέχουν πρότυπη γλώσσα διατύπωσης ερωτήσεων (SQL like) και δυνατότητα διατύπωσης ερωτήσεων με χρήση εντύπων (QBE).

Παρ' ολ' αυτά, και οι δύο προσεγγίσεις συμφωνούν στην υποστήριξη **πολλαπλών τρόπων αναζήτησης** (διατύπωση ερωτήσεων, πλοήγηση κτλ.) **βάσει διαφορετικών κριτηρίων** (λογική δομή, σημασιολογία, λεξικογραφική ανάλυση των εγγράφων). Ακόμη, η χρήση συστημάτων διαχείρισης βάσεων δεδομένων (οντοκεντρικών ή σχεσιακών) και από τις δύο προσεγγίσεις έχει σαν πλεονέκτημα την χρήση των ειδικών μηχανισμών ελέγχου που παρέχουν τα συστήματα αυτά (έλεγχος ταυτόχρονης πρόσβασης, ανάληψης κ.ο.κ.). Τελος, θεωρείται πολύ σημαντική η **ύπαρξη προγραμματικής διεπαφής** (API) για την δυνατότητα σύνδεσης των συστημάτων αυτών με άλλες συνεργαζόμενες εφαρμογές.

2.2.1 Ερευνητικά Συστήματα

Σύστημα Διαχείρισης Δομημένων εγγράφων σε Ψηφιακές Βιβλιοθήκες.

Το σύστημα που υλοποιήθηκε από το ινστιτούτο IPSI-GMD [35, 16, 17] αποβλέπει σε μια εφαρμογή βάσεων δεδομένων για δομημένα έγγραφα. Η αποθήκευση των εγγράφων αυτών έγινε στην οντοκεντρική βάση δεδομένων Vodak [52]. Το σύστημα είναι ικανό για αποθήκευση διαφόρων τύπων εγγράφων. Για το σχεδιασμό της βάσης δεδομένων κάθε τμήμα του εγγράφου παριστάνεται από ένα αντικείμενο της βάσης. Επίσης κάθε τύπος τμήματος του τύπου εγγράφου παραστάθηκε με μια κλάση αντικειμένων στη βάση δεδομένων. Για αποδοτικότερη εισαγωγή και αναζήτηση εγγράφων και τμημάτων τους υιοθετήθηκε μια υβριδική αναπαράσταση των δομημένων εγγράφων στη βάση. Οι τύποι τμημάτων των εγγράφων που δεν μετέχουν συνήθως σε ερωτήσεις και έχουν τελική πληροφορία θεωρούνται ότι είναι περιπτώσεις της εικονικής κλάσης αντικειμένων FLAT TYPE (παράδειγμα τέτοιων τμημάτων είναι ο τύπος τμήματος paragraph). Για ανάκτηση πληροφορίας με βάση λέξεις και εκφράσεις που περιέχονται στα έγγραφα χρησιμοποιήθηκε το σύστημα λεξικογραφικής αναζήτησης (full text retrieval system) INQUERY. Επίσης, για αποδοτικότερη αναζήτηση πληροφορίας στο σύστημα υλοποιήθηκαν δύο τύποι ευρετηριασμού. Δημιουργήθηκε ένα ευρετήριο (direct index) για τον ευρετηριασμό των πιο πληροφοριακών τμημάτων των εγγράφων (μόνο για τελικούς τύπους τμημάτων). Επίσης χρησιμοποιήθηκε ένα ευρετήριο με πληροφορία δομικού χαρακτηρα για ταχύτερη εξυπηρέτηση ερωτήσεων δομικού χαρακτήρα (πχ. αναζήτηση όλων των υποκεφαλαίων

που βρίσκονται μέσα σε ένα κεφάλαιο).

Διαχείριση Δομημένων Εγγράφων με χρήση της O_2

Στα πλαίσια της εργασίας αυτής[1] [50] χρησιμοποιήθηκε μια οντοκεντρική βάση δεδομένων για τη διαχείριση εγγράφων έτσι ώστε να γίνεται χρήση των μηχανισμών ανάνηψης, ταυτόχρονης πρόσβασης, διατύπωση ερωτήσεων κ.ο.κ. που παρέχονται από τα συστήματα διαχείρισης βάσεων δεδομένων. Επιλέχτηκε η O_2 [41] διότι παρέχει μία εύκολα επεκτάσιμη γλώσσα ερωτήσεων και διότι παρέχει πλούσιο σύστημα τύπων.
Για την συντακτική ανάλυση των εγγράφων SGML χρησιμοποιήθηκε ο συντακτικός αναλυτής [26]. Για την παράσταση των τύπων τμημάτων των εγγράφων δημιουργήθηκαν κλάσεις αντικειμένων στη ΒΔ. Κάθε τέτοια κλάση αποτελείται από ένα *τύπο, περιορισμούς ακεραιότητας, τυπική συμπεριφορά (default behavior).* Για τις ανάγκες αποθήκευσης και ανάκτησης των εγγράφων έγιναν επεκτάσεις τόσο στο μοντέλο δεδομένων όσο και στη γλώσσα ερωτήσεων [49].

- Στο μοντέλο δεδομένων σχεδιάστηκαν ειδικοί τύποι για να παραστήσουν τύπους τμημάτων που εχουν εναλλακτική δομή καθώς επίσης και τύπους τμημάτων των οποίων η δομή αποτελείται από υποτμήματα (που μπορεί να εμφανίζονται ακολουθιακά ή με οποιαδήποτε σειρά).

- Με την επέκταση της γλώσσας ερωτήσεων μπορούν να γίνουν ερωτήσεις σε έγγραφα όπως σε οποιοδήποτε άλλο τύπο δεδομένων.

Digital Library Initiative

Η εργασία αυτή [**?**] φιλοδοξεί στην ανάπτυξη της απαραίτητης υποδομής υποστήριξης μιας ψηφιακής βιβλιοθήκης [2]. Η αρχιτεκτονική του συστήματος αποτελείται από ένα σύνολο κατανεμημένων φυσικών πηγών εισαγωγής πληροφορίας (σε μορφή εγγράφων) και ένα σύνολο (όχι κατ ανάγκη το ίδιο) πελατών που χρησιμοποιούν το σύστημα για άντληση πληροφορίας.
Τα εισαγόμενα έγγραφα είναι επιστημονικού περιεχομένου και μετά την εισαγωγή τους επεξεργάζονται κατάλληλα ώστε να ακολουθούν έναν συγκεκριμένο τύπο εγγράφου (ISO 12083 Article DTD).
Για τον ευρετηριασμό των εγγράφων (και την αποδοτικότερη αναζήτησή τους) χρησιμοποιείται

[1] Η εργασία αυτή εντέσσεται στα πλαίσια του Esprit προγράμματος TecPad
[2] Ως ψηφιακή βιβλιοθήκη αναφέρεται ένα σύνολο κατανεμημένων συστημάτων αποθήκευσης (repositories)

ένας θησαυρός όρων (Inspec thesaurus) και ένας κατάλογος συνεμφανιζόμενων όρων[3]. που δημιουργείται αυτόματα με χρήση ειδικών αλγόριθμων.
Στόχοι της εργασίας αυτής είναι η έρευνα τεχνικών ανάκτησης εγγράφων με χρήση σημασιολογικής πληροφορίας και η αυτόματη ανάλυση του περιεχομένου των εγγράφων. Επίσης η δημιουργία μιας ομοσπονδίας σημασιολογικού δικτύου (semantic federation) για κατανεμημένα συστήματα αποθήκευσης. Τέλος, ιδιαίτερο βάρος δίνεται στην αξιολόγηση της αποτελεσματικότητας του συστήματος.

EDS

Η εργασία αυτή [54] απέβλεπε στη δημιουργία ενός συστήματος ηλεκτρονικού λεξικού με τη χρήση του προτύπου SGML και μιας οντοκεντρικής βάσης δεδομένων. Επίσης χρησιμοποιήθηκε η C++ για τον προγραμματισμό των λειτουργιών του συστήματος. Συγκεκριμένα, στο σύστημα αυτό χρησιμοποιήθηκε η οντοκεντρική βάση δεδομένων ObjectStore και το λεξικό Pequen o Larousse Spanish κωδικοποιημένο σύμφωνα με το πρότυπο SGML.
Η αρχιτεκτονική του συστήματος χωρίζεται σε δύο τμήματα, το **DBP** που είναι υπεύθυνο για τη δημιουργία της βάσης δεδομένων και το **DQP** που είναι υπεύθυνο για τη διατύπωση ερωτήσεων και την ανάκτηση πληροφορίας από το σύστημα.
Κάθε καταχώρηση του λεξικού παριστάνεται στη βάση δεδομένων με μια ιεραρχική δομή (δέντρο). Οι κόμβοι της ιεραρχικής δομής αποτελούν τα στοιχεία καταχώρησης (entry elements). Κάθε τέτοιος κόμβος αντιστοιχεί στους τύπους τμημάτων των εγγράφων SGML. Τα στοιχεία καταχώρησης του λεξικού αποτελούν περιπτώσεις των αντίστοιχων κόμβων της ιεραρχικής δομής στη ΒΔ.

Επίσης στη ΒΔ παραστάθηκαν λίστες κατηγοριών για κάθε λέξη/σύμβολο του λεξικού (π.χ. μέρη του λόγου) για γρηγορότερη και ευκολότερη αναζήτηση. Οι λίστες αυτές συνδέονται με την ιεραρχική δομή που περιγράφηκε παραπάνω.
Η διατύπωση ερωτήσεων γίνεται με τη χρήση μιας γλώσσας ερωτήσεων τύπου SQL-like. Για την αποθήκευση της πληροφορίας που διαχείρίζεται το σύστημα χρησιμοποιήθηκε η ObjectStore.

[3]Οι συνεμφανιζόμενοι όροι ενός όρου είναι αυτοί που εμφανίζονται μαζί (δίπλα, στην ίδια πρόταση, στην ίδια παράγραφο κτλ.) με αυτόν στα περιεχόμενα των εγγράφων. Ο κατάλογος αυτός χρησιμοποιείται εναλλακτικά και συμπληρωματικά με το θησαυρό όρων και υποδηλώνει σχέση μεταξύ των συνεμφανιζόμενων όρων όχι κατ ανάγκη σημασιολογική π.χ. συσχέτιση του ονόματος ενός συγγραφέα με όρους που αναφέρονται σε μια ερευνητική περιοχή

MULTOS

Το σύστημα MULTOS αναπτύχθηκε στα πλαίσια του ESPRIT προγράμματος MULTimedia Office Server [**?**] . Σκοπός του συστήματος είναι η αποθήκευση και ανάκτηση εγγράφων πολύμορφης πληροφορίας (multimedia documents).
Για την παράσταση των εγγράφων σχεδιάστηκε ένα μοντέλο (με τη χρήση της γλώσσας Telos) παράστασης της λογικής δομής και του τρόπου παρουσίασης εγγράφων που ακολουθούν το πρότυπο ODA.

Kabiria

Η εργασία αυτή ξεκίνησε το 1989 και αποσκοπούσε στην ανάπτυξη ενός συστήματος για την ανάκτηση εγγράφων βάσει τόσο του περιεχομένου όσο και του σκοπού που αυτά επιτελούν σε ένα περιβάλλον γραφείου κατά τον κύκλο ζωής τους.
Το μοντέλο παράστασης των εγγράφων αποτελείται από το εννοιλογικό μοντέλο στο οποίο παριστάνεται η σημασιολογία τους και το μοντέλο ανάκτησης στο οποίο παριστάνονται οι σχέσεις των εγγράφων με τους ρόλους που επιτελούν σε έναν οργανισμό.
Το σύστημα παρέχει τη δυνατότητα εισαγωγής και ταξινόμησης των εγγράφων (σε προδηλωμένους τύπους εγγράφων) καθώς επίσης και ανάκτηση εγγράφων σύμφωνα με το περιεχόμενο, τη δομή και το ρόλο τους σε έναν οργανισμό. Για την ανάκτηση των εγγράφων παρέχεται η δυνατότητα σχηματισμού ερωτήσεων (SQL) και η δυνατότητα πλοήγησης.
Τα έγγραφα αποθηκεύονται σε ένα σύστημα διαχείρισης βάσεων δεδομένων (Oracle) και η αρχιτεκτονική του συστήματος ακολουθεί το μοντέλο πελάτη εξυπηρετητή. Επίσης χρσιμοποιείται ένα σύστημα λεξικογραφικής αναζήτησης (BRS)

2.2.2 Εμπορικά Συστήματα

Astoria

Το προϊόν αυτό [1] επιτρέπει την εύκολη διαχείριση εγγράφων SGML και τμημάτων τους καθώς επίσης και μη δομημένων εγγράφων, εικόνων κτλ.
Επιτρέπει την εισαγωγή, εξαγωγή, τροποποίηση εγγράφων καθώς επίσης και την αναχρησιμοποίηση των τμημάτων των εγγράφων του συστήματος για δημιουργία νέων. Η ανάκτηση των εγγράφων (ή τμημάτων τους) γίνεται βάσει περιεχομένου, χαρακτηριστικών ή ονομάτων που αποδίδονται στα τμήματά τους.
Το σύστημα παρέχει δυνατότητα πλοήγησης στη δομή των εγγράφων, παρουσίαση των

εγγράφων (με χρήση συνεργαζόμενου Editor) και προγραμματική διεπαφή χρήσης.
Η αρχιτεκτονική του συστήματος ακολουθεί το μοντέλο πελάτη - εξυπηρετητή και χρησιμοποιείται μια οντοκεντρική βάση δεδομένων για την αποθήκευση των εγγράφων.

BasisPlus

Το σύστημα αυτό [**?**] σχεδιάστηκε για την αποδοτική αποθήκευση μεγάλου όγκου πληροφορίας.
Προσφέρει μηχανισμούς επεξεργασίας κειμένου (text manipulation), πλοήγηση στη λογική δομή εγγράφων που ακολουθούν το πρότυπο της SGML και μηχανισμούς που απορρέουν από τη χρήση συστημάτων διαχείρισης βάσεων δεοδομένων όπως μηχανισμοί ελέγχου ταυτόχρονης πρόσβασης, αδιεξόδων(deadlock), ανάνηψης (recovery), ελεγχόμενη πρόσβαση χρηστών στο σύστημα κτλ.
Ο χειρισμός της πληροφορίας γίνεται σε επίπεδο λεπτομέρειας τμημάτων των εγγράφων. Επίσης, η λειτουργία της αναζήτησης υποστηρίζεται από ένα σύστημα λεξικογραφικής αναζήτησης και από την ύπαρξη ενός θησαυρού όρων.
Το σύστημα παρέχει προγραμματική διεπαφή χρήσης (API) για τη σύνδεσή του με άλλα συστήματα.
Η αρχιτεκτονική του συστήματος ακολουθεί το μοντέλο πελάτη-εξυπηρετητή και το σύστημα διαχείρισης βάσεων δεδομένων που χρησιμοποιείται είναι σχεσιακό.

DocMan

Το σύστημα αυτό [2] υποστηρίζει τη διαχείριση εγγράφων SGML καθώς επίσης και τη ροή εργασίας (workflow management) στο σύστημα.
Αποτελείται από δύο κύρια τμήματα (modules). Το πρώτο τμήμα είναι υπεύθυνο για την δημιουργία του σχήματος της βάσης αποθήκευσης και την εισαγωγή των εγγράφων. Το δεύτερο τμήμα είναι υπεύθυνο για την διαχείριση των εργασιών του συστήματος.
Παρέχει τη δυνατότητα εισαγωγής και εξαγωγής, ευρετηριασμού, τροποποίησης, αποθήκευσης (archiving), ανάκτησης των εγγράφων και αναχρησιμοποίηση του περιεχομένου των εγγράφων για τη δημιουργία νέων εγγράφων. Επίσης παρέχεται προγραμματική διεπαφή χρήσης για σύνδεση με άλλα συστήματα (πχ. συστήματα λεξικογραφικής ανάλυσης κειμένων)
Το επίπεδο λεπτομέρειας για τον ευρετηριασμό και την ανακτηση των εγγράφων καθορίζεται από το χρήστη και φτάνει έως και τα τμήματα του εγγράφου. Επίσης καθοριζόμενα από τον χρήστη ειναι και τα κλειδιά αναζήτησης που χρησιμοποιούνται για τον ευρετηριασμό

της πληροφορίας.
Για την υλοποίηση του συστήματος χρησιμοποιήθηκε μια σχεσιακή βάση δεδομένων, η γλώσσα ερωτήσεων SQL και οι απομακρυσμένες κλήσεις ερωτήσεων (rpcs) του UNIX (για την διαχείριση της ροής εργασίας του συστήματος).

Information Manager

Βασικό χαρακτηριστικό του συστήματος [**?**] είναι η δυνατότητα ταυτόχρονης επεξεργασίας εγγράφων SGML και η αναχρησιμοποίηση του περιεχομένου των για δημιουργία νέων εγγράφων. Το προϊόν αυτό προσφέρει μηχανισμούς ερωτήσεων (με χρήση SQL, προκαθορισμένων ερωτήσεων και χρήση εντύπων), πλοήγησης στη δομή των εγγράφων, τροποποίησης, εισαγωγής και εξαγωγής εγγράφων, έλεγχου πρόσβασης στο σύστημα, χειρισμού ροής εργασιών (workflow) και ιστορικού δημιουργίας εκδόσεων εγγράφων. Επίσης παρέχεται προγραμματική διεπαφή χρήσης (API) για σύνδεση του συστήματος με άλλα συστήματα.
Το σύστημα αποθήκευσης που χρησιμοποιείται είναι ένα ιβρυδικό σχήμα αφού χρησιμοποιείται μια σχεσιακή βάση δεδομένων για την αποθήκευση των εγγράφων και μια οντοκεντρική βάση δεδομένων για την παράσταση πληροφορίας γι αυτά (πληροφορίες για τη δομή τους).

Life*CDM

Το σύστημα αυτό [10] είναι κατάλληλο για τη διαχείριση εγγράφων SGML (οποιουδήποτε DTD) στα οποία γίνονται συχνές τροποποιήσεις και δημιουργούνται νέες εκδόσεις. Επίσης είναι κατάλληλο για εφαρμογές εκδόσεων (publishing) στις οποίες αναχρησιμοποιείται η πληροφορία που υπάρχει στο σύστημα για δημιουργία νέων εγγράφων. Το προϊόν αυτό παρέχει μηχανισμούς εισαγωγής, εξαγωγής και τροποποίησης εγγράφων, αποθήκευσης (archiving), ευρετηριασμού και ανάκτησης. Επίσης παρέχεται προγραμματική διεπαφή χρήσης για σύνδεση του συστήματος με άλλα συστήματα.
Για την αποθήκευση των εγγράφων χρησιμοποιείται μια σχεσιακή βάση δεδομένων (Oracle).

PCDocs

Το προϊόν αυτό [36] παρέχει αποτελεσματική οργάνωση πληροφορίας ώστε να είναι αποδοτική η αναζήτηση εγγράφων που διαχειρίζεται.
Παρέχει προγραμματική διεπαφή χρήσης για σύνδεση με άλλα συστήματα. Επίσης παρέχει μηχανισμό ελέγχου πρόσβασης στην πληροφορία που διαχειρίζεται, δυνατότητα αποθήκευσης (archiving), δημιουργίας εκδόσεων. Το σύστημα επαφής-χρήσης είναι

προσαρμόσιμο στις επιλογές του τελικού χρήστη. Η αναζήτηση των εγγράφων γίνεται βάσει χαρακτηριστικών που αποδίδονται στα έγγραφα (πχ. συγγραφέας) και με τη χρήση συστήματος λεξικογραφικής αναζήτησης (Infolink).
Συνεργάζεται με ένα συνολο από σχεσιακά συστήματα διαχείρισης βάσεων δεδομένων για την αποθήκευση των εγγράφων (Oracle, Sybase, Microsoft).

Saros

Το σύστημα αυτό [36] αποβλέπει στην παροχή υπηρεσιών βιβλιοθήκης στα έγγραφα τα οποία διαχειρίζεται. Οι υπηρεσίες αυτές αφορούν στην αποδοτική αναζήτηση, έλεγχο πρόσβασης, έλεγχο δημιουργίας εκδόσεων, διατήρηση αντιγράφων (backups), ασφάλεια και έλεγχο φόρτου εργασίας στο σύστημα.
Το σύστημα παρέχει όλα τα πλεονεκτήματα συστημάτων που χρησιμοποιούν συστήματα διαχείρισης βάσεων δεδομένων (ανάληψη, έλεγχο ταυτόχρονης πρόσβασης κτλ.). Επίσης υποστηρίζει την αποθήκευση, ευρετηριασμό και την αναζήτηση εγγράφων. Η αναζήτηση μπορεί να γίνει με χρήση λέξεων κλειδιών, ιδιοτήτων που αποδίδονται στα έγγραφα (πχ. περιγραφή, που έχουν τοποθετηθεί κτλ.) και λεξικογραφική αναζήτηση (με τη χρήση του λεξικογραφικού αναλυτή Fulcrum).
Η αρχιτεκτονική του συστήματος ακολουθεί το μοντέλο πελάτη - εξυπηρετητή, χρησιμοποιείται ένα σχεσιακό σύστημα διαχείρισης βάσεων δεδομένων και παρέχεται προγραμματική διεπαφή χρήσης για τη σύνδεση του συστήματος με άλλα συστήματα.

SIM

Πρόκειται για ένα ολοκληρωμένο σύστημα [**?**] διαχείρισης εγγράφων SGML που παρέχει σύνδεση με το WWW, γλώσσα προγραμματισμού για μορφοποίηση εγγράφων SGML κτλ.
Το σύστημα αυτό υποστηρίζει μηχανισμούς εισαγωγής, ευρετηριασμού, τροποποίησης και διαγραφής εγγράφων, δημιουργίας εκδόσεων, διαχείριση ροής εργασίας, πλοήγηση στη δομή των εγγράφων, λεξικογραφική αναζήτηση. Επίσης, παρέχει προγραμματική διεπαφή χρήσης για σύνδεση με άλλα συστήματα και παρουσίαση εγγράφων SGML σε ένα σύνολο μορφών (πχ. HTML). Τέλος, η αποθήκευση των εγγράφων γίνεται σε ένα σύστημα διαχείρισης βάσεων δεδομένων. Το σχήμα της βάσης δεδομένων καθορίζεται από τον τύπο εγγράφων (DTD).

2.3 Σύγκριση συστημάτων

Η αποθήκευση των εγγράφων στα εμπορικά συστήματα γίνεται κυρίως σε σχεσιακές βάσεις δεδομένων (Oracle, Sybase). Αντίθετα τα ερευνητικά συστήματα που μελετήσαμε κάνουν χρήση οντοκεντρικών βάσεων δεδομένων. Εξαίρεση αποτελεί το σύστημα Kabiria το οποίο χρησιμοποιεί μια σχεσιακή βάση δεδομένων.
Με τη χρήση σχεσιακών ΒΔ είναι δυσκολότερη η προσαρμογή και τροποποίηση του σχήματος της βάσης σε σύγκριση με συστήματα που χρησιμοποιούν ευέλικτα σχήματα για την παράσταση της δομής των εγγράφων. Επίσης η χρήση σχεσιακών μοντέλων για την παράσταση των εγγράφων SGML οδηγεί σε περίπλοκα σχήματα [34] και είναι επίπονη σε σχέση με τη χρήση οντοκεντρικών μοντέλων (GMD, Inria, EDS, Multos) διότι συχνά απαιτείται αλλαγή τόσο του σχήματος όσο και των δεδομένων της βάσης.

Τα περισσότερα συστήματα διαχειρίζονται πολλαπλούς τύπους εγγράφων. Εξαίρεση αποτελούν το σύστημα Kabiria που χειρίζεται ορισμένους προδηλωμένους τύπους εγγράφων καθώς επίσης και το σύστημα που σχεδιαστηκε απο την εργασία DLI που χειρίζεται έναν τύπο εγγράφου.

Οι τρόποι αναζήτησης πληροφορίας στα έγγραφα γίνεται με χρήση ερωτήσεων και με πλοήγηση στη βάση δεδομένων των συστημάτων διαχείρισης εγγράφων. Τα εμπορικά συστήματα που μελετήσαμε επιτρέπουν πλοήγηση στα έγγραφα και στη δομή τους μεσα από ειδικό γραφικό περιβάλλον. Το σύστημα που αποτελεί προϊόν της εργασίας DLI επιτρέπει πλοήγηση στη δομή των εγγράφων καθώς και στο θησαυρό που χρησιμοποιεί το σύστημα. Το σύστημα multos επιτρέπει πλοήγηση τόσο στη δομή όσο και στην σημασιολογική πληροφορία των εγγράφων.

Βασική διαχειριστική απαίτηση είναι η αναζήτηση εγγράφων βάσει της δομής τους ή του περιεχομένου τους. Για την ικανοποίηση της απάιτησης αυτής υπάρχουν πολλές προσεγγίσεις και αντίστοιχοι μηχανισμοί. Τα περισσότερα συστήματα επιτρέπουν ερωτήσεις δομικού χαρακτήρα (ερωτήσεις που αφορούν στη δομή των εγγράφων). Η εργασία που περιγράφεται στο [49] περιγραφει την επέκταση της γλώσσας ερωτήσεων της O_2 για την υποστήριξη τέτοιων ερωτήσεων.
Η αναζήτηση βάσει του περιεχομένου των εγγράφων συνήθως υποστηρίζεται με ερωτήσεις βάσει λέξεων ή προτάσεων που μπορεί να περιέχονται στα περιεχόμενα των εγγράφων

(λεξικογραφική ανάλυση των εγγράφων) και εξυπηρετούνται από ανάλογα συστήματα (full text retrieval systems). Στην κατηγορία αυτή ανήκουν το σύστημα που αναπτύχθηκε απο την εργασία DLI και τα εμπορικά συστήματα Astoria, BasisPlus, Saros, SIM. Η αναζήτηση βάσει του περιεχομένου γίνεται επίσης με χρήση σημασιολογικής πληροφορίας η οποία παριστάνεται σε ένα σημασιολογικό δίκτυο ή αποθηκεύεται σε μία βάση δεδομένων. Η παράσταση (ή αποθήκευση) σημασιολογικής πληροφορίας γίνεται από ερευνητικά κυρίως συστήματα. Αναλυτικότερα, η παράσταση της σημασιολογίας των εγγράφων γίνεται αυτόματα με τη χρήση ειδικών αλγορίθμων όπως στην περίπτωση της εργασίας DLI ή κατόπιν επεξεργασίας των εγγράφων από ειδικούς (multos, Kabiria, Κώδιξ). Βέβαια η χρήση αυτόματων αλγορίθμων στην εργασία DLI υποστηρίχθηκε από ειδικούς υπερυπολογιστές, κάτι που για την ώρα είναι απαγορευτικό για εφαρμογές μικρής ή μεσαίας κλίμακας. Επίσης το σύστημα Kabiria επικεντρώνει στην παράσταση σημασιολογικής πληροφορίας που αφορά στους ρόλους που επιτελεί ένα έγγραφο μέσα σε έναν οργανισμό και στις σχέσεις (ταξινόμηση) μεταξύ εγγράφων βάσει των ρόλων τους. Ανάλογα, το σύστημα multos δίνει ιδιαίτερη έμφαση στην ταξινόμηση των εγγράφων ανάλογα με τη δομή και τη χρησιμότητά τους. Ομως το μοντέλο παράστασης των εγγράφων ακολουθεί το πρότυπο ODA το οποίο έχει σημαντικά μειονεκτήματα σε σχέση με την SGML [8].

Κεφάλαιο 3

Το πρότυπο SGML

Η SGML είναι ένα πρότυπο παράστασης της λογικής δομής των εγγράφων. Παρέχει την δυνατότητα εμπλουτισμού των εγγράφων με πληροφορία που αφορά στη σημασιολογία του, στον τρόπο επεξεργασίας του και στη δομή τους. Η πληροφορία αυτή είναι ανεξάρτητη του συστήματος το οποίο διαχειρίζεται τα έγγραφα, επιτρέποντας την μεταφορά και χρήση των εγγράφων σε οποιοδήποτε περιβάλλον εφαρμογής. Στο κεφάλαιο αυτό γίνεται μία ιστορική αναδρομή στους τρόπους κωδικοποίησης που έχουν χρησιμοποιηθεί για την περιγραφή εγγράφων. Κατόπιν αναλύεται η χρησιμότητα της SGML για εφαρμογές που χρησιμοποιούν δομημένα έγγραφα και περιγράφεται το περιβάλλον μιας τέτοιας εφαρμογής. Τέλος, αναλύονται οι βασικές έννοιες που χρησιμοποιεί η SGML για την περιγραφή δομημένων εγγράφων.

3.1 Ιστορική Αναδρομή

Στο παρελθόν υιοθετήθηκαν τρόποι κωδικοποίησης εγγράφων οι οποίοι χρησιμοποιούνταν για την απόδοση γνωρισμάτων που σχετίζονταν με την δομή, το περιεχόμενο και τη μορφή τους. Μια πρωτόγονη κωδικοποίηση εγγράφων που σκοπό έχει να διαχωρίσει το περιεχόμενο και τη σημασία των εγγράφων αυτών είναι πχ. η χρήση κεφαλαίων γραμμάτων μέσα στο κείμενο.
Εκτός από τον παραπάνω απλό τρόπο κωδικοποίησης εγγράφων χρησιμοποιήθηκαν κι άλλοι τρόποι πιό εκφραστικοί αλλά και πιό περίπλοκοι. Τέτοιοι τρόποι κωδικοποίησης είναι οι εξής:

- Η **Κωδικοποίηση Λειτουργίας** (procedural markup) χρησιμοποιεί εντολές συγκεκριμένων εργαλείων (devices) που αποσκοπούν στην μορφοποίηση του κειμένου των εγγράφων.

Οι εντολές αυτές συνήθων είναι ρήματα που δηλώνουν τη συγκεκριμένη πράξη μορφοποίησης που θα εφαρμοστεί στο κείμενο πχ. το γνωστό πρόγραμμα **Tex** χρησιμοποιεί την εντολή (ρήμα) **centerline** για να τοποθετήσει στο κέντρο της γραμμής μία πρόταση:

```
\centerline {\bf This is the title of the Book}
```

- Η **Γενικευμένη Κωδικοποίηση** (γύρω στα τέλη του 1960) είναι μια βελτίωση του παραπάνω τρόπου διότι δημιουργεί γενικούς κώδικες, **macros** (λέξεις) για τον συνδιασμό πολλών εντολών που αποδίδουν χαρακτηριστικά για την μορφοποίηση και τη δομή των εγγράφων. Οι λέξεις είναι κατάλληλα επιλεγμένες ώστε το έγγραφο να είναι ευανάγνωστο και να είναι κατανοητή η λειτουργία των κωδικών λέξεων αυτών από τον τελικό χρήστη. Το μειονέκτημα της κωδικοποίησης αυτής είναι ότι τα έγγραφα που κωδικοποιούνται με αυτόν τον τρόπο δεν είναι προσαρμόσιμα σε όλα τα συστήματα.

- Το 1969 οι Goldfarb, Mosher και Lorie δημιούργησαν την GML (Generalized Markup Language). Η GML [23] εισήγαγε την έννοια του τυπικά ορισμένου τύπου εγγράφου (Document Type). Από το 1978 ο Goldfarb εργάστηκε για τη δημιουργία ενός προτύπου βασισμένο στην GML (SGML). Το 1980 εκδόθηκε η πρώτη τεχνική αναφορά για το πρότυπο αυτό.

3.2 Σκοπιμότητα και Χρήση

Το πρότυπο της SGML επιτρέπει κωδικοποίηση της λογικής δομής και του τρόπου επεξεργασίας ηλεκτρονικών εγγράφων. Η κωδικοποίηση αυτή είναι ανεξάρτητη του λογισμικού και του υλικού που χρησιμοποιείται με συνέπεια α) τα έγγραφα SGML μπορούν να επεξεργασθούν από διαφορετικές εφαρμογές κατά διαφορετικούς τρόπους δίχως να απαιτείται τροποποίηση των περιεχομένων τους και β) τα έγγραφα που είναι εναρμονισμένα με το διεθνές αυτό πρότυπο μπορούν να ανταλλάσονται και να διαχειρίζονται από διαφορετικά υπολογιστικά συστήματα.

Επίσης τα έγγραφα ενός οργανισμού δεν είναι φυλακισμένα στο συγκεκριμένου τύπου σύστημα στο οποίο δημιουργήθηκαν. Το γεγονός αυτό επιτρέπει στις επιχειρήσεις αλλαγές στην τεχνολογία που επιθυμούν δίχως πρόσθετο κόστος τροποποίησης των εγγράφων που αυτά διαχειρίζονται.

Η κωδικοποίηση της λογικής δομής των εγγράφων επιτρέπει την απόδοση σημασιολογίας σε αυτά. Η λογική δομή ενός εγγράφου αναφέρεται στα λογικά (νοηματικά) τμήματα που αποτελούν τα περιεχόμενα του (πχ. συγγραφέας, τίτλος, περίληψη, συμπεράσματα κτλ.). Τα τμήματα αυτά μπορεί να μην είναι διακριτά στην (εξωτερική) μορφή του εγγράφου. Η χρήση της λογικής δομής στη διαχείριση εγγράφων και ιδιαίτερα στην αναζήτηση είναι ιδιαίτερα επιθυμητή.

Τα εργαλεία που έχουν αναπτυχθεί για να εξυπηρετήσουν την διαχείριση εγγράφων SGML είναι στην πλειονότητά τους εύχρηστα και προσαρμοσμένα στον οικείο τρόπο εργασίας του χρήστη. Επίσης υποστηρίζουν ένα διεθνές πρότυπο και για το λόγο αυτό δεν υπόκεινται σε συνεχείς αναβαθμίσεις.

Τα παραπάνω πλεονεκτήματα δικαιολογούν την υιοθέτηση της SGML από μεγάλους εκδοτικούς οργανισμούς για την έκδοση εγκυκλοπαιδιών, λεξικών κτλ. Επίσης την χρήση του προτύπου απο μεγάλες εταιρείες (όπως αεροναυπηγικές κτλ.) για παραγωγή τεχνικών αναφορών (technical documentation) μεγάλου όγκου (τμήματα των οποίων αναχρησιμοποιούνται ή τροποποιούνται συχνά) και διακινούνται στην αγορά.

3.3 Τα στοιχειώδη εργαλεία της SGML

Μια *Εφαρμογή SGML* ορίζεται [23] ως μια *εφαρμογή επεξεργασίας κειμένου*. Μια εφαρμογή επεξεργασίας κειμένου ορίζεται ως ενα σύνολο σχετικών επεξεργασιών που εφαρμόζονται σε έγγραφα σχετικών τύπων.

Για κάθε εφαρμογή SGML απαιτούνται τουλάχιστον τα εξής εργαλεία:

- Συντακτικός Αναλυτής της SGML (SGML parser)
  Πρόκειται για έναν μεταφραστή ο οποίος αναγνωρίζει το συντακτικό εγγράφων SGML και ελέγχει την συντακτική ορθότητά τους.

- Διαχειριστής Εικονικών Αντικειμένων Αποθήκευσης (SGML Entity manager)
  Είναι ένα εργαλείο το οποίο παρέχει πρόσβαση στα Εικονικά Αντικείμενα Αποθήκευσης (ΕΑΑ) που υπάρχουν. Τα ΕΑΑ είναι μια έννοια η οποία χρησιμοποιείται στην SGML για την αποδοτική αποθήκευση των εγγράφων SGML.

- Ενας κειμενογράφος
  Για τη δημιουργία των Τύπων Εγγράφων και των εγγράφων SGML.

3.4 Το Περιβάλλον μιας Εφαρμογής SGML

Τα εργαλεία που έχουν αναπτυχθεί (και υπάρχουν στην αγορά) για υποστήριξη μιας εφαρμογής SGML ταξινομούνται ως εξής :

- Εργαλεία δημιουργίας εγγράφων SGML. Τα εργαλεία αυτά επιτρέπουν την ευχάριστη καθοδηγούμενη και γρήγορη δημιουργία εγγράφων πχ. **SGML Editors, DTD Builders**. Επίσης στην κατηγορία αυτή ανήκουν και εργαλεία (translators) που χρησιμοποιούνται για τη μετατροπή εγγράφων SGML από άλλες μορφές κωδικοποίησης πληροφορίας (μη έγγραφα SGML).

- Εργαλεία που χρησιμοποιούνται για την αποθήκευση (repositories) όλης της πληροφορίας που σχετίζεται με τη διαχείριση εγγράφων SGML. Η πληροφορία αυτή περιλαμβάνει τα έγγραφα SGML, πληροφορία για τη ροή εργασίας (workflow), εικόνες κτλ.

- Εργαλεία εξαγωγής πληροφορίας. Τα εργαλεία αυτά χρησιμοποιούνται για την εξαγωγή πληροφορίας από τα έγγραφα SGML και την ηλεκτρονική ή έντυπη παράστασή της, την παράστασή της σε οπτικούς δίσκους (CD-ROM)κτλ.

Παρακάτω δίνουμε αναλυτική περιγραφή των εργαλείων που ανήκουν στις προαναφερθείσες κατηγορίες.

3.4.1 Δημιουργία εγγράφων SGML

SGML Editors

Η δημιουργία (και τροποποίηση) εγγράφων SGML μπορεί να γίνει με τη χρήση ενός απλού κειμενογράφου (πχ. vi, Word κτλ.).
Ωστόσο, έχουν αναπτυχθεί συστήματα που κάνουν την διαδικασία αυτή απλούστερη με τη χρήση εντύπων (για κάθε τύπο εγγράφου) και διαχωρισμό των κωδικών λέξεων από το περιεχόμενο των εγγράφων. Επίσης τα συστήματα αυτά είναι πιό φιλικά διότι παρέχουν καθοδήγηση στον τελικό χρήστη με τη χρήση καταλόγου επιτρεπόμενων κωδικών λέξεων και βοήθειας σε κάθε βήμα της ανάπτυξης ενός εγγράφου, διαδικασίες αυτόματης επικύρωσης κτλ.
Αρκετά από τα συστήματα αυτά παρέχουν γραφική αναπαράσταση της δομής του εγγράφου. Τα συστήματα αυτά απαιτούν συνήθων γραφικό περιβάλλον και παρέχουν ένα οικείο σύστημα επαφής-χρήσης.

DTD Builders

Η δημιουργία του τύπου εγγράφου μπορεί να γίνει με τη χρήση ειδικών εργαλείων που χρησιμοποιούν γραφικό περιβάλλον ανάπτυξης για την παράσταση, την τροποποίηση και τη δημιουργία της δομής των εγγράφων.

3.4.2 Αποθήκευση εγγράφων SGML

Τα εργαλεία που παρέχουν αποθήκευση δομημένων εγγράφων χωρίζονται σε δύο κατηγορίες:

- Συστήματα που χρησιμοποιούν Βάσεις Δεδομένων για την αποθήκευση των εγγράφων.
- Συστήματα που χρησιμοποιούν Βάσεις Δεδομένων για αποθήκευση πληροφορίας σχετική με τα έγγραφα (meta-information). Τα έγγραφα στα οποία αποθηκεύονται τα πραγματικά δεδομένα βρίσκονται στο σύστημα αρχείων του συστήματος.

Για την αξιολόγηση και επιλογή του κατάλληλου συστήματος για την αποθήκευση των εγγράφων καθοριστικό ρόλο παίζει η φύση της εφαρμογής για την οποία αναπτύσονται [**?**] .

Συστήματα που χρησιμοποιούν Βάσεις Δεδομένων για την αποθήκευση των εγγράφων.

- Για την αποθήκευση των εγγράφων έχουν χρησιμοποιηθεί **Σχεσιακές βάσεις** δεδομένων. Η τεχνολογία αυτή παρέχει τη δυνατότητα αποθήκευσης μεγάλων αντικειμένων (blobs). Για την καλύτερη διαχείριση των εγγράφων γίνεται χρήση ευρετηρίου (index) το οποίο έχει πληροφορίες για το περιεχόμενο κάθε εγγράφου.
- Επίσης, για την αποθήκευση εγγράφων έχουν χρησιμοποιηθεί και **Οντοκεντρικές βάσεις** δεδομένων. Η προσέγγιση αυτή είναι μεταγενέστερη και στηρίζεται στο γεγονός ότι υπάρχει ομοιότητα της τεχνολογίας της SGML με τις οντοκεντρικές βάσεις δεδομένων διότι και οι δύο τεχνολογίες χειρίζονται αντικείμενα και χαρακτηριστικά αντικειμένων (objects, attributes) και κάνουν χρήση της έννοιας της ιεραρχίας [**?**] .
- Τέλος, έχουν χρησιμοποιηθεί **υβριδικά συστήματα βάσεων δεδομένων** στα οποία γίνεται συνδιασμός σχεσιακών και οντοκεντρικών βάσεων δεδομένων. Στα συστήματα αυτά οι σχεσιακές βάσεις χρησιμοποιούνται κυρίως για την αποθήκευση των εγγράφων και οι οντοκεντρικές βάσεις δεδομένων παρέχουν μέτα-πληροφορίες για τα έγγραφα αυτά.

Συστήματα που χρησιμοποιούν Βάσεις Δεδομένων για την αποθήκευση πληροφορίας για τα έγγραφα .

Τα συστήματα της κατηγορίας αυτής χρησιμοποιούν συστήματα βάσεων δεδομένων για την διαχείριση πληροφορίας σχετικής με τα έγγραφα. Τα πραγματικά έγγραφα αποθηκεύονται στο σύστημα αρχείων (file system) με μορφή αρχείου. Ενα έγγραφο μπορεί να είναι αποθηκευμένο σε ένα ή περισσότερα αρχεία.

3.4.3 Εξαγωγή πληροφορίας από έγγραφα SGML

SGML Viewers

Το πρότυπο της SGML δίνει την δυνατότητα εμπλουτισμού του περιεχομένου ενός εγγράφου με πληροφορία που αφορά στη μορφή του.
Η πληροφορία αυτή χρησιμοποιείται κατόπιν από ειδικά εργαλεία παρουσίασης (Viewers) που σκοπό έχουν την (ηλεκτρονική) παρουσίαση των εγγράφων. Τα εργαλεία αυτά διαθέτουν γλώσσες ανάπτυξης (εξαρτόμενες από το συγκεκριμένο περιβάλλον ανάπτυξης) για τη μετάφραση της πληροφορίας αυτής σε συγκεκριμένους τρόπους μορφοποίησης των εγγράφων.
Τέλος, συνήθως παρέχονται τύποι εγγράφων (που χρησιμοποιούνται συχνά, κλασσικοί) με έτοιμες οδηγίες μορφοποίησης (templates).

SGML Browsers

Τα εργαλεία αυτά παρέχουν δύο λειτουργίες:

- Παρουσιάζουν ένα έγγραφο σύμφωνα με κάποιες οδηγίες μορφοποίησης.
- Δίνουν τη δυνατότητα πλοήγησης στη δομή του εγγράφου χρησιμοποιώντας γραφικό περιβάλλον.

3.5 Περιγραφή της SGML

Ενα **έγγραφο SGML** είναι ένα σύνθετο έγγραφο το οποίο περιέχει

- Μια **δήλωση SGML** η οποία δηλώνει τους χαρακτήρες του εγγράφου SGML.
- Ενα **πρόλογο** ο οποίος ορίζει τύπους εγγραφών .

- Το **σύνολο των εγγράφων** που περιέχουν τα πραγματικά δεδομένα και στα οποία θα αναφερόμαστε με το όνομα **Βασικά Εγγραφα** ή BE.

3.5.1 Δήλωση SGML

Είναι τμήμα κάθε εγγράφου SGML το οποίο σκοπό έχει να περιγράψει το σύνολο των χαρακτήρων που το έγγραφο χρησιμοποιεί. Η SGML κάνει τον εξής διαχωρισμό στους χαρακτήρες :

- Χαρακτήρες ελέγχου (function chars): Καθορίζουν την αρχή και το τέλος γραμμής σε έγγραφα SGML κτλ.
- Χαρακτήρες ονομάτων (name chars) : Χρησιμοποιούνται για την ονοματοδοσία (π.χ. γράμματα, ψηφία κτλ.)
- NSGML Χαρακτήρες: Χαρακτήρες που δεν περιέχονται σίγουρα στο έγγραφο (π.χ. εικόνες, εκτελέσιμα αρχεία κτλ.)
- Χαρακτήρες Δεδομένων (data chars): Χαρακτήρες που μπορεί να χρησιμοποιηθούν για την κωδικοποίηση οποιασδήποτε πληροφορίας. (Ολοι οι μή NSGML χαρακτήρες)

Επίσης στο ίδιο τμήμα περιγράφονται κανόνες ονοματοδοσίας (Naming rules) ,το σύνολο των διαχωριστικών (delimiters) και τα δεσμευμένα ονόματα .

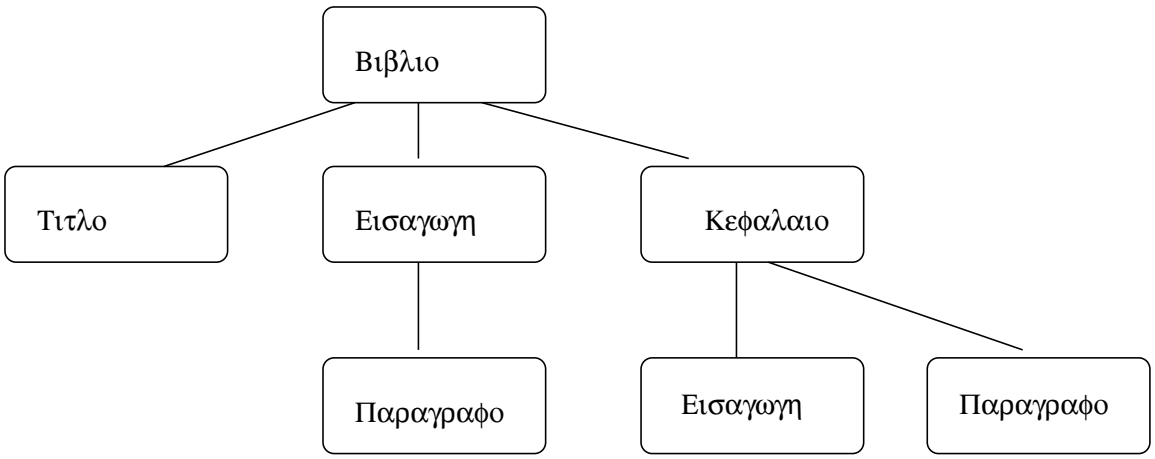
3.5.2 Πρόλογος

Το τμήμα αυτό δηλώνει το τι επιτρέπεται να υπάρχει στο έγγραφο SGML . Ειδικότερα πρόκειται για τρεις δηλώσεις.

Δήλωση Τύπου Εγγράφου (DTD)

Ορίζει έναν τύπο εγγράφου (DTD από δω και στο εξής). Κάθε τύπος εγγράφου παριστάνεται με μία **ιεραρχική δομή**. Ρίζα της δομής αυτής είναι ένας κομβος που παριστάνει ολόκληρο το έγγραφο και ονομάζεται **βασικό τμήμα** του εγγράφου. Τα παιδιά αυτού του κόμβου παριστάνουν τη λίστα περιεχομένων (υποτμήματα) του εγγράφου. Οι εσωτερικού κόμβοι της ιεραρχικής δομής παριστάνουν τμήματα του εγγράφου που αποτελούνται από άλλα υποτμήματα και γι αυτό ονομάζονται **μη τελικά τμήματα**. Τα φύλλα της ιεραρχικής δομής είναι τα τμήματα του εγγράφου που περιέχουν τα δεδομένα του εγγράφου και για

το λόγο αυτό ονομάζονται **τελικά τμήματα**(βλέπε σχ. 3.1.



Σχήμα 3.1: Ιεραρχική δομή παράστασης του τύπου εγγράφου "Βιβλιο"

Ο ορισμός ενός DTD μπορεί να περιέχει τεσσάρων ειδών δηλώσεις:

- **Δηλώσεις Τύπων Τμημάτων**
  Τμήμα (element) είναι το μικρότερο δομικό συστατικό ενός ΒΕ. Ενα τμήμα μπορεί να περιέχει αδόμητα δεδομένα ή υποτμήματα (subelements).
  Κάθε τμήμα έχει ένα όνομα και μια λίστα περιεχομένων. Κάθε ένα από τα τμήματα της λίστας περιεχομένων μπορεί με τη σειρά του να αποτελείται από υποτμήματα. Για την δήλωση ενός τμήματος χρησιμοποιείται μια κανονική έκφραση. Κάθε υποτμήμα συνδέεται με δύο τελεστές: **ο τελεστής εμφάνισης** και ο **τελεστής ύπαρξης**.

  -- Ο τελεστής εμφάνισης δηλώνει πόσες φορές εμφανίζεται ένα υποτμήμα στη λίστα περιεχομένων ενός τμήματος. Υπάρχουν τρεις τελεστές εμφάνισης:

  * ***** : Χρησιμοποιείται για να δηλώσει ότι ένα υποτμήμα μπορεί να εμφανίζεται καμία ή περισσότερες φορές.
  * **+** : Χρησιμοποιείται για να δηλώσει ότι ένα υποτμήμα μπορεί να εμφανίζεται μία ή περισσότερες φορές.
  * **?** : Χρησιμοποιείται για να δηλώσει ότι ένα υποτμήμα μπορεί να εμφανίζεται καμία ή μια φορά.

  π.χ. <!ELEMENT letter date? section+ signature* >
  στο παραπάνω παράδειγμα έχουμε μία δήλωση ενός στοιχείου (element) με όνομα "letter". Η λίστα περιεχομένων του στοιχείου αυτού αποτελείται από το

υποστοιχείο ”date” που εμφανίζεται μία ή καμία φορά, το υποστοιχείο ”section” που εμφανίζεται μία ή περισσότερες φορές και το υποστοιχείο ”signature” το οποίο μπορεί να εμφανίζεται καμία , μία ή περισσότερες φορές.

-- **Ο τελεστής θέσης** δηλώνει σε ποιά θέση της λίστας περιεχομένων βρίσκεται το υποστοιχείο αυτό.
Υπάρχουν δύο τελεστές θέσης:

* **,** : Χρησιμοποιείται για να δηλώσει ότι ένα υποτμήμα ακολουθεί (εμφανίζεται αμέσως μετά) το υποτμήμα που προηγείται στη λίστα περιεχομένων ενός τμήματος.
* **&** : Χρησιμοποιείται για να δηλώσει ότι ένα υποτμήμα μπορεί να εμφανίζεται οπουδήποτε στη λίστα περιεχομένων ενός τμήματος.

π.χ. <!ELEMENT letter (date,section+,(name|signature)) & figure >
στο παραπάνω παράδειγμα έχουμε μια δήλωση ενός στοιχείου με όνομα ”letter”. Η λίστα περιεχομένων του στοιχείου αυτού αποτελείται από το υποστοιχείο ”date” το οποίο εμφανίζεται πρώτο στη λίστα περιεχομένων του στοιχείου. Ακολουθεί το υποστοιχείο ”section” μια ή περισσότερες φορές. Στη λίστα περιεχομένων υπάρχει επίσης ένα από τα υποστοιχεία name και signature. Τέλος, στη λίστα περιεχομένων υπάρχει και το υποστοιχείο ”figure” το οποίο μπορεί να εμφανίζεται οπουδήποτε μέσα στη λίστα περιεχομένων.

Δύο ή περισσότερα τμήματα με το ίδιο μοντέλο περιεχομένων μπορούν να δηλωθούν μαζί (group).
π.χ. <!ELEMENT (introduction|section) paragraph+ >

- **Δηλώσεις Γνωρισμάτων**
Σε κάθε τμήμα μπορεί να αποδίδεται ένα ή περισσότερα χαρακτηριστικά. Κάθε χαρακτηριστικό έχει ένα όνομα, ένα πεδίο τιμών (υπάρχουν δεκάξι προκαθορισμένοι τύποι τιμών) και μία τιμή εξ’ ορισμού (default value).

-- Καθορισμένα χαρακτηριστικά.
Ειναι τα χαρακτηριστικά εκείνα τα οποία παίρνουν πάντα σταθερή τιμή (την εξ’ ορισμού που έχει δοθεί σε αυτό το χαρακτηριστικό).Η ύπαρξή τους εξυπηρετεί ειδικές εφαρμογές που απαιτούν από ένα στοιχείο να έχει ορισμένα χαρακτηριστικά (π.χ. ODA).

- **Δηλώσεις Εικονικών Αντικειμένων Αποθήκευσης,ΕΑΑ**
  Η δήλωση ενός ΕΑΑ περιέχει ένα όνομα και ένα κείμενο (το οποίο περιγράφεται είται αυτούσιο, είτε μέσω λογικής αναφοράς σε αρχείο του λειτουργικού).
  Κάθε ΕΑΑ επίσης δηλώνεται ότι ανήκει σε μία από τις παρακάτω τέσσερις κατηγορίες [23] :

  -- Εσωτερικά ΕΑΑ (Internal entities)
  Κάποιο τμήμα του εγγράφου που επαναλαμβάνεται συχνά μέσα στο έγγραφο μπορεί να δηλωθεί σαν ΕΑΑ και να του αποδωθεί ένα όνομα. Ετσι το τμήμα αυτό θα υπάρχει μια μόνο φορά μεσα στο έγγραφο αλλά θα μπορεί να αναφέρεται πολλές φορές.

  -- Παραμετρικά ΕΑΑ (Parameter entities)
  Χρησιμοποιούνται για συντομογραφίες (όπως ακριβώς και τα εσωτερικά ΕΑΑ) αλλά εμφανίζονται σε λίστες περιεχομένων των τμημάτων.

  -- Εξωτερικά ΕΑΑ (External entities)
  Μερικά αρχεία του συστήματος, βιβλιοθήκες, macros(system dependent) μπορούν να κρατηθούν έξω από το σώμα του εγγράφου και να γίνεται αναφορά σε αυτά με ένα αναγνωριστικό συστήματος π.χ.
  <!ENTITY part1 SYSTEM ”file1.txt”>, file1.txt : αναγνωριστικό συστήματος.

  -- Δημόσια ΕΑΑ (Public entities)
  Ενα ΕΑΑ που βρίσκεται πέρα από το έγγραφο ή το περιβάλλον του συστήματος λέγεται δημόσιο (public) ΕΑΑ. Σε αυτό αποδίδεται ένα αναγνωριστικό που είναι γνωστό σε μια κοινότητα ανθρώπων και μπορεί να κοθορίζεται από ένα διεθνές ή εθνικό πρότυπο.
  Χαρη σε αυτά τα ΕΑΑ μπορεί μια κοινότητα ανθρώπων να μοιράζεται οποιαδήποτε πληροφορία .

  -- Ενα ΕΑΑ που περιέχει οδηγίες επεξεργασίας (processing instruction) για την αλλαγή του περιβάλλοντος επεξεργασίας (π.χ. αλλαγή font) μπορεί να δηλωθεί σαν ΕΑΑ τυπου PI (Processing Instruction).

Δηλώσεις Σχολίων Εγγράφου

Ενα σχόλιο είναι ένα αλφαριθμητικό (string) από SGML χαρακτήρες που χρησιμοποιούνται μέσα στο έγγραφο SGML ώστε να το επεξηγούν.

Οδηγίες Επεξεργασίας (Processing Instructions)

Η SGML παρέχει δυνατότητα κωδικοποίησης πληροφορίας που αφορά στην επεξεργασία ενός εγγράφου (π.χ. μορφοποίηση). Η κωδικοποίηση αυτή μπορεί να γίνει δίνοντας οδηγίες επεξεργασίας στο σύστημα (σε γλώσσα που την καταλαβαίνει) με έναν από τους παρακάτω τρεις τρόπους:

- Ορισμός ενός ΕΑΑ τύπου **PI**. Χρησιμοποιούνται ΕΑΑ αυτού του τύπου για οδηγίες επεξεργασίας που έχουν ως αποτέλεσμα την αλλαγή του περιβάλλοντος επεξεργασίας (π.χ. χρησιμοποίηση ειδικού τύπου χαρακτήρων για την επόμενη παράγραφο κτλ.).
- Ορισμός ενός ΕΑΑ τύπου **SDATA**. Χρησιμοποιούνται ΕΑΑ αυτού του τύπου για οδηγίες επεξεργασίας που έχουν ως αποτέλεσμα να "παράγουν" δεδομένα π.χ. εισαγωγή εικόνας σε συγκεκριμένη θέση στο έγγραφο.
- Αμεσες Οδηγίες επεξεργασίας. Πρόκειται για ένα τρόπο αποστολής οδηγιών επεξεργασίας σε μία εφαρμογή, σε γλώσσα που περιορίζεται από το σύστημα. (π.χ. οδηγίες για αλλαγή σελίδας στη μορφοποίηση ενος εγγράφου). Τετοιες οδηγίες επεξεργασίας μπορεί να βρίσκονται οπουδήποτε στο έγγραφο. π.χ. <? system data>

3.5.3 Περιπτώσεις εγγράφων (Document Instance Set)

Στο τμήμα αυτό υπάρχουν τα "πραγματικά έγγραφα" (BE) που ακολουθούν τον τύπο εγγράφου που περιγράφηκε στο δεύτερο τμήμα (prolog).

3.5.4 Κωδικοποίηση Εγγράφων SGML.

Η κωδικοποίηση (markup) των εγγράφων SGML προσθέτει στα έγγραφα πληροφορία που αφορά στη δομή και τη σημασιολογία τους (λογική δομή) καθώς επίσης και τον τρόπο επεξεργασίας τους.

- Για την κωδικοποίηση πληροφορίας (σύμφωνα με το συντακτικό της SGML) απαιτούνται δύο κωδικές λέξεις:
  - -- Μια κωδική λέξη που δηλώνει την αρχή μιας κωδικοποιημένης πληροφορίας (start-tag)
  - -- Μια κωδική λέξη που δηλώνει το τέλος μιας κωδικοποιημένης πληροφορίας (end-tag)

Τα είδη κωδικοποίησης είναι τα εξής :

- Descriptive Markup (tags)
  Ορίζουν τη δομή του BE.

- entity reference
  Απαίτηση για ένα ΕΑΑ να περιληφθεί σε ένα έγγραφο.

- parameter entity reference
  Απαίτηση για ένα παραμετρικό ΕΑΑ να περιληφθεί σε κάποιο τμήμα του DTD.

- Markup Declaration
  Η κωδικοποίηση αυτή χρησιμοποιείται στο DTD για την περιγραφή του τύπου εγγράφου.

- Οδηγίες επεξεργασίας
  Ειναι η κωδικοποίηση που χρησιμοποιείται για τις οδηγίες επεξεργασίας (system dependent) .

3.5.5 Αρχικοί τύποι στην SGML

Στο πρότυπο της SGML ορίζονται οι αρχικοί τύποι (primitives) που χρησιμοποιούνται για τον χαρακτηρισμό του περιεχομένου τελικών τμημάτων εγγράφων και των κωδικών λέξεων που χρησιμοποιούνται σε ένα έγγραφο SGML. Οι τύποι αυτοί είναι οι εξής:

- **Είδη ονομάτων**

  -- NAME
  αποτελείται από χαρακτήρες ονομάτων(γράμματα, ψηφία, τελεία, hyphen) και ο πρώτος χαρακτήτας(name start char) ακολουθεί τους περιορισμούς που περιγράφονται στην Δήλωση SGML (SGML Declaration).

  -- NAMES
  αποτελείται από ένα ή περισσότερα NAME.

  -- NUMBER
  αποτελείται από έναν αριθμο(σύνολο ψηφίων).

  -- NUMBERS
  αποτελείται από έναν ή περισσότερους αριθμούς.

-- NMTOKEN
αποτελείται από ένα name token(όπως το name αλλά αρχίζει με οποιοδήποτε χαρακτήρα).

-- NMTOKENS
αποτελείται από έναν ή περισσότερα name tokens

-- NUTOKEN
αποτελείται από ένα number token(όπως το name token αλλά πρέπει να αρχίζει με ψηφίο.

-- NUTOKENS
αποτελείται από έναν ή περισσότερα number tokens(όπως τα name tokens αλλά πρέπει να αρχίζουν με ψηφίο.

- **Είδη περιεχομένου τελικών τμημάτων εγγράφων.**
Τα τελικά τμήματα των εγγράφων περιέχουν πληροφορία η οποία περιγράφεται στον τύπο εγγράφου με τους τύπους **CDATA, RCDATA, PCDATA**. Η διαφορά των προηγούμενων τύπων έγκειται στον τρόπο με τον οποίο

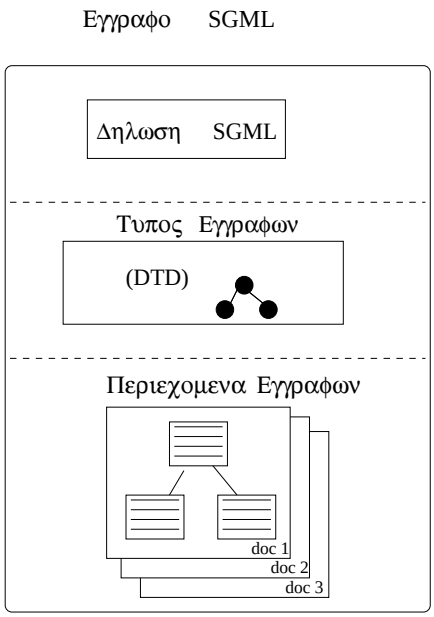
- **Είδη περιεχομένου τελικών τμημάτων εγγράφων.** θα γίνει η συντακτική ανάλυση (parsing) των τμημάτων αυτών. Αναλυτικότερα, οι τύποι αυτοί περιγράφονται παρακάτω:

-- CDATA
Στον τύπο αυτό ανήκουν τα περιεχόμενα τμημάτων που αποτελούνται μόνο από χαρακτήρες δεδομένων. Δεν επιτρέπονται κωδικές λέξεις.

-- RCDATA
Ειναι μια παραλλαγή των CDATA στα οποία αναγνωρίζεται κωδικοποίηση αναφοράς σε EAA.

-- PCDATA
Είναι χαρακτήρες δεδομένων οι οποίοι πρέπει να αναλυθούν συντακτικά (parsing) ώστε να αναγνωρισθούν οι κωδικές λέξεις που περιέχουν.

-- SDATA Ειναι χαρακτήρες δεδομένων με ειδική μεταχείριση από τον συντακτικό αναλυτή. Δεν αναγνωρίζονται τα περισσότερα διαχωριστικά (start tag, end tag).



Σχήμα 3.2: Δομή ενός εγγράφου SGML.

Κεφάλαιο 4

Η γλώσσα SIS-Telos και το σύστημα Σύστημα Σημασιολογικού Ευρετηριασμού

Η SIS-Telos είναι μια γλώσσα παράστασης γνώσης που υιοθετεί οντοκεντρική προσέγγιση στο μοντέλο δεδομένων που υποστηρίζει. Τα εκφραστικά μέσα που χρησιμοποιεί η γλώσσα αυτή για την παράσταση γνώσης είναι οι *οντότητες* και οι **σχέσεις** (γνωρίσματα) μεταξύ οντοτήτων. Η Telos προτάθηκε από τον κ. Μυλόπουλο [40] και μια υλοποίησή της (SIS-Telos) έχει γίνει από την ομάδα Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του ΙΠ-ΙΤΕ [27].

Η γλώσσα SIS-Telos αποτελεί τον πυρήνα ανάπτυξης του Συστήματος Σημασιολογικού Ευρετηριασμού (από δω και στο εξής ΣΣΕ). Το σύστημα αυτό χρησιμοποιείται για την αποθήκευση και διερεύνηση δεδομένων πολλαπλώς συνδεδεμένων.

Στην παρούσα εργασία έχει χρησιμοποιηθεί το σύστημα ΣΣΕ και η γλώσσα SIS-Telos για την υλοποίηση της βάσης γνώσης του συστήματος **Κώδιξ** στην οποία γίνεται η παράσταση πληροφορίας σχετικής με τη δομή, τη σημασία και τον τρόπο παρουσίασης των δομημένων εγγράφων του συστήματος.

Στη συνέχεια του κεφαλαίου δίνονται τα βασικά χαρακτηριστικά και τα πλεονεκτηματα της γλώσσας παράστασης γνώσης SIS-Telos. Επίσης γίνεται μια αναλυτική περιγραφή της γλώσσας αυτής και του ΣΣΕ.

4.1 Παράσταση γνώσης στη γλώσσα SIS-Telos

Η γλώσσα παράστασης γνώσης SIS-Telos παριστάνει τη γνώσση με τη μορφή σημασιολογικού δικτύου. Το μοντέλο αυτό έχει τα εξής χαρακτηριστικά:

- Κάθε αντικείμενο του πραγματικού κόσμου παριστάνεται στο μοντέλο αυτό σαν μια ανεξάρτητη οντότητα.
- Κάθε οντότητα έχει μοναδικό όνομα.
- Σε κάθε οντότητα μπορούν να αποδωθούν γνωρίσματα.
- Οντότητες με κοινά εγγενή χαρακτηριστικά ταξινομούνται σε μια κλάση οντοτήτων. Κάθε κλάση οντοτήτων αποτελεί μια οντότητα και μπορεί με τη σειρά της να ταξινομηθεί σε κλάσεις οντοτήτων κτλ.
- Μια κλάση μπορεί να είναι εξειδίκευση (υποσύνολο ή υποκλάση) μιας άλλης κλάσης της οποίας κληρονομεί τα γνωρίσματα.

Η γλώσσα SIS-Telos εξασφαλίζει τα παραπάνω χαρακτηριστικά (που έχει κάθε σημασιολογικό δίκτυο) με τους παρακάτω μηχανισμούς :

- **Μηχανισμός Ονοματοδοσίας αντικειμένων**
  Κάθε οντότητα του πραγματικού κόσμου παριστάνεται από μια ξεχωριστή οντότητα, *Object*. Η οντότητα αυτή έχει ένα μοναδικό όνομα που παράγεται από το σύστημα (SYSID). Επίσης σε κάθε οντότητα μπορεί να αποδωθεί και ένα *λογικό* όνομα από τον χρήστη του συστήματος.
- **Μηχανισμός Ταξινόμησης**
  Κάθε οντότητα του σημασιολογικού δικτύου της SIS-Telos ταξινομείται σε μία ή περισσότερες κλάσεις οντοτήτων και κληρονομεί τα χαρακτηριστικά τους. Η κλάση είναι κι αυτή οντότητα (άρα μπορεί να ταξινομηθεί κι αυτή). Η SIS-Telos απαιτεί κάθε οντότητα να είναι περίπτωση μιας τουλάχιστον κλάσης. Με τον τρόπο αυτό δημιουργείται μια ανω μη φραγμένη ιεραρχία. Στην κατώτερη στάθμη της ιεραρχίας αυτής βρίσκονται τα ατομικά αντικείμενα (σταθμη Token). Το σύστημα έχει έναν αρχικό πληθυσμό κλάσεων τις *κλάσεις συστήματος*. Με τον τρόπο αυτό ικανοποιείται η απαιτηση του συστήματος κάθε οντότητα να ταξινομείται σε μια τουλάχιστον κλάση. Κάθε οντότητα που παριστάνεται στη γλώσσα SIS-Telos ταξινομείται στη κλάση συστήματος *Object*. Υποσύνολα της κλάσης αυτής είναι

οι κλάσεις συστήματος *Individual, Attribute, Class, Token.* Στην κλάση *Individual* ταξινομούνται οι οντότητες, οι κλάσεις οντοτήτων, οι κλάσεις από κλάσεις οντοτήτων κτλ. Στην κλάση *Attribute* ταξινομούνται τα γνωρίσματα των οντοτήτων, οι κλάσεις γνωρισμάτων (κατηγορίες όπως συνηθίζονται να λέγονται), οι κλάσεις κλάσεων γνωρισμάτων κτλ.

Οι κλάσεις οντοτήτων που ορίζονται από τον χρήστη ταξινομούνται κάτω από τη κλάση *Class.* Οσες από τις κλάσεις αυτές ανήκουν σε απλή στάθμη ταξινομούνται στη κλάση συστήματος *S_Class.* Οι κλάσεις που ανήκουν στη στάθμη M1 (μετακλάσεις) ταξινομούνται στη κλάση συστήματος *M1_Class.* Οι κλάσεις που ανήκουν στη στάθμη M2 (μετα-μετακλάσεις) ταξινομούνται στη κλάση συστήματος *M2_Class* κτλ.

Τα γνωρίσματα των οντοτήτων θεωρούνται κι αυτά οντότητες και έτσι ταξινομούνται κι αυτά σε κλάσεις γνωρισμάτων (κατηγορίες από δω και στο εξής). Τα γνωρίσματα τα οποία δέχονται μοναδική τιμή ταξινομούνται στη κατηγορία συστήματος *Unique.* Τα γνωρίσματα που πιθανό να έχουν τιμή ταξινομούνται στη κατηγορία συστήματος *Possible.* Τα γνωρίσματα που απαραιτητα δέχονται τιμή (μία ή περισσότερες) ταξινομούνται στη κατηγορία συστήματος *Necessary*. Τελος, για τις τιμές ακεραίων, πραγματικών αριθμών, και αλφαριθμητικών χρησιμοποιούνται αντίστοιχα οι κλάσεις συστήματος *Telos_Integer, Telos_Real, Telos_String.*

- **Μηχανισμός Γενίκευσης-Εξειδίκευσης.**
  Με το μηχανισμό αυτό ορίζεται μια σχέση υποσυνόλου μεταξύ κλάσεων ίδιας στάθμης. Μία κλάση μπορεί να είναι εξειδίκευση (υποσύνολο-υποκλάση) μιας ή περισσοτέρων κλάσεων των οποίων κληρονομεί τα γνωρίσματα. Επίσης μια κλάση μπορεί να είναι γενίκευση (υπερσύνολο-υπερκλάση) μιας ή περισσότερων κλάσεων. Τέλος, μια κλάση που είναι γενίκευση ή εξειδίκευση μιας κλάσης μπορεί επίσης να έχει δικά της γνωρίσματα.
  Ο μηχανισμός γενίκευσης-εξειδίκευσης δεν εφαρμόζεται σε ατομικές οντότητες (Tokens).

- **Μηχανισμός Απόδοσης Γνωρίσματος**
  Σε κάθε οντότητα που παριστάνεται στο σύστημα μπορεί να αποδοθούν γνωρίσματα. Στα γνωρίσματα αυτά μπορούν να δοθούν παραπάνω από μια τιμές (πλειότιμα) ή να μη δοθεί καμία (προαιρετικά). Κάθε γνώρισμα μπορεί να ταξινομηθεί, να αποτελεί γενίκευση ή εξειδίκευση κάποιου άλλου γνωρίσματος και να του απο δοθουν γνωρίσματα (ιδιότητες που προκύπτουν από την όμοια μεταχείριση οντοτήτων και

γνωρισμάτων).

Μια τυπική περιγραφή της SIS-Telos υπάρχει στο [4].

4.2 Βασικά χαρακτηριστικά και Πλεονεκτήματα της SIS-Telos

Η γλώσσα SIS-Telos δεν κάνει διαχωρισμό στον τρόπο παράστασης του σχήματος και των δεδομένων της εφαρμογής που διαχειρίζεται. Αυτό επιτρέπει την εύκολη και γρήγορη τροποποίηση και επέκταση του σχήματος της βάσης.
Τα γνωρίσματα αντιμετωπίζονται σαν οντότητες από τη γλώσσα SIS-Telos κάνοντας ομοιόμορφο τον τρόπο διαχείρισης οντοτήτων και γνωρισμάτων.

Στην SIS-Telos οι σύνδεσμοι (ταξινόμησης, γενίκευσης, γνωρισμάτων) αποθηκεύονται ως σύνδεσμοι διπλής κατεύθυνσης προσφέροντας καλές επιδόσεις στις διασχίσεις και ερωτήσεις, μεγαλώνοντας όμως το μέγεθος της βάσης [**?**].
Οι οντότητες που παριστάνονται με τη γλώσσα SIS-Telos μπορούν να ανήκουν σε πολλαπλές ιεραρχίες ταξινόμησης. Επίσης μπορούν να αποτελούν εξειδικεύσεις ή γενικεύσεις πολλών οντοτήτων.

4.3 Περιγραφή του Συστήματος Σημασιολογικού Ευρετηριασμού - SIS .

Το Σημασιολογικό Σύστημα Ευρετηριασμού (**S**emantic **I**ndex **S**ystem) χρησιμοποιείται για την αποθήκευση και διερεύνηση μεγάλου πληθυσμού δεδομένων πολλαπλώς συνδεδεμένων. Για το λόγο αυτό είναι κατάλληλο για εφαρμογές σχεδιαστικής ή κατασκευαστικής φύσεως.
Παρέχει τους παρακάτω μηχανισμούς διαχείρισης:

- **Μηχανισμός αποθήκευσης αντικειμένων**
  Ο μηχανισμός αυτός υποστηρίζει δοσοληψίες και ταυτόχρονη πρόσβαση από πολλούς χρήστες

- **Μηχανισμός ερωτήσεων**
  Το ΣΣΕ συνοδεύεται από ένα μηχανισμό ερωτήσεων ([**?**]), ο οποίος παρέχει ένα σύνολο από ερωτηματικές εντολές, με τις οποίες είναι δυνατή η πλοήγηση στο σύνολο της πληροφορίας που έχει παρασταθεί στη βάση καθώς και η διατύπωση

αναδρομικών ερωτήσεων με πολλαπλά κριτήρια. Οι ερωτήσεις μπορούν να κληθούν από ένα διαλογικό εργαλείο, τον answerer ή μέσα από άλλες εφαρμογές με τη βοήθεια του προγραμματικής διεπαφής χρήσης (PQI). [30, 5].

- **Μηχανισμός Εισαγωγής Δεδομένων** Η εισαγωγή των δεδομένων στο ΣΣΕ γίνεται με δύο τρόπους α) ασύγχρονη μαζική εισαγωγή δεδομένων με χρήση του συντακτικού αναλυτή της SIS-Telos και β) χρήση των Δελτίων Εισαγωγής Δεδομένων για σύγχρονη εισαγωγή και τροποποίηση των δεδομένων με αλληλεπίδραση χρήστη-ΣΣΕ. Τα Δελτία Εισαγωγής Δεδομένων προσαρμόζονται κατάλληλα στις ανάγκες εισαγωγής δεδομένων κάθε εφαρμογής που χρησιμοποιεί το ΣΣΕ. Αναλυτική περιγραφή του τρόπου χρήσης και προσαρμογής των δελτίων αυτών υπάρχει στα [31, 32].

- **Μηχανισμός Προγραμματικής Διεπαφής Χρήσης**
  Η διασύνδεση και επικοινωνία του ΣΣΕ με εξωτερικές εφαρμογές υποστηρίζονται από ένα σύνολο συναρτήσεων (Application Programming Interface, API) με τις οποίες είναι δυνατή η άντληση πληροφορίας που έχει παρασταθεί στο ΣΣΕ από τις εφαρμογές αυτές. Παρακάτω δίνεται μια αναλυτική περιγραφή των συναρτήσεων αυτών.

Τέλος, το ΣΣΕ παρέχει ένα γραφικό περιβάλλον [28] με τη χρήση του οποίου είναι δυνατή η επικοινωνία του τελικού χρήστη με αυτό. Στο σημασιολογικό μοντέλο (Βάση Γνώσης) του ΣΣΕ παριστάνεται το γραφικό αυτό περιβάλλον και οι κατάλογοι επιλογών (menus) που παρέχει. Επίσης παριστάνονται οι προκαθορισμένες ερωτήσεις και τα δελτία εισαγωγής δεδομένων.

Η προσαρμογή του γραφικού περιβάλλοντος, των προκαθορισμένων ερωτήσεων και των Δελτίων Εισαγωγής Δεδομένων του ΣΣΕ σε νέες εφαρμογές έγκειται στην προσαρμογή του μοντέλου αυτού στις απαιτήσεις της νέας εφαρμογής. Αναλυτική παρουσίαση του μοντέλου και του τρόπου προσαρμογής του δίνεται στα [6],

4.3.1 Γραφικό περιβάλλον του ΣΣΕ

Η επικοινωνία του ΣΣΕ με το χρήστη γίνεται με ένα προσαρμόσιμο εργαλείο (βλέπε σχήμα 5.2) το οποίο επιτρέπει διερεύνηση σε σχήμα και δεδομένα καθώς και παρουσιάσεις πολύμορφων δεδομένων (multimedia) εξωτερικά αποθηκευμένων. Υπάρχει η δυνατότητα ορισμού και ενεργοποίησης προκαθορισμένων ερωτήσεων, οι οποίες μπορούν να οργανωθούν ιεραρχικά, καθώς και η δυνατότητα διατύπωσης ερωτήσεων με δελτία (retrieval card). Επίσης υποστηρίζονται πολλαπλές και προκαθοριζόμενες γραφικές παρουσιάσεις (όψεις)

των περιεχομένων της βάσης. Τέλος, είναι δυνατή η κλήση εξωτερικών εφαρμογών (προγραμμάτων) μέσα από το γραφικό περιβάλλον του ΣΣΕ.

4.3.2 Προγραμματική διεπαφή χρήσης του ΣΣΕ

Το σύστημα Σημασιολογικού Ευρετηριασμού προσφέρει ένα σύνολο πρωτογενών (primitive) συναρτήσεων πρόσβασης και επικοινωνίας της βάσης γνώσης με εξωτερικές εφαρμογές. Το σύνολο των συναρτήσεων αυτών ονομάζεται **PQI** (Programmatic Query Interface) και εφαρμόζεται σε κόμβους του σημασιολογικού δικτύου (οντότητες του μοντέλου) και έχουν ως αποτέλεσμα ένα σύνολο οντοτήτων.

Σενάριο χρήσης του PQI

Η επικοινωνία μιας εξωτερικής εφαρμογής και του συστήματος ΣΣΕ γίνεται ως εξής:

- Εγκαθίσταται η επικοινωνία της εφαρμογής και του ΣΣΕ με τη συνάρτηση *begin_query*
- Τίθεται ο κόμβος στον οποίο θα εφαρμοστεί κάποια εκ των ερωτηματικών συναρτήσεων του PQI με τη συνάρτηση *set_current_node*.
- Χρησιμοποιούνται οι κατάλληλες ερωτηματικής συναρτήσεις του PQI για αναζήτηση πληροφορίας στη Βάση Γνώσης (ΒΓ) του ΣΣΕ.
- Επαναλαμβάνονται τα δύο παραπάνω βήματα (προαιρετικά) και χρησιμοποιούνται κατάλληλες συναρτήσεις διαχείρισης συνόλων απάντησης του PQI.
- Τερματίζεται η επικοινωνία της εφαρμογής και του ΣΣΕ με τη χρήση της συνάρτησης *end_query*.

Το σύνολο συναρτήσεων που προσφέρονται από το PQI κατατάσονται ανάλογα με τη λειτουργικότητά τους σε πέντε κατηγορίες:

- Στην πρώτη κατηγορία ανήκουν οι συναρτήσεις που είναι υπεύθυνες για την έναρξη/ληξη της επικοινωνίας με το ΣΣΕ.
- Στη δεύτερη κατηγορία ανήκουν συναρτήσεις που χρησιμοποιούνται για τροποποίηση γενικών παραμέτρων που χρησιμοποιούνται από τις ερωτηματικές συναρτήσεις.
- Στη τρίτη κατηγορία ανήκουν όλες οι ερωτηματικές συναρτήσεις δηλ. οι συναρτήσεις που χρησιμοποιούνται για ερωτήσεις στη ΒΓ. Με τις συναρτήσεις αυτές είναι δυνατό να αντληθεί πχ. η εξής πληροφορία από τη βάση γνωσης του ΣΣΕ:

-- Οι τιμές των γνωρισμάτων μιας οντότητας.

-- Οι οντότητες των οποίων αποτελεί γνώρισμα μια οντότητα.

-- Η κλάση της οποίας μια οντότητα αποτελεί περίπτωση.

-- Οι περιπτώσεις μιας οντότητας.

-- Οι υπερκλάσεις μιας οντότητας.

-- Οι υποκλάσεις μιας οντότητας.

-- Οι τιμές των γνωρισμάτων μιας οντότητας, τα οποία αποτελούν περιπτώσεις μιας συγκεκριμένης κατηγορίας (ή μετακατηγορίας) γνωρισμάτων.

-- Οι οντότητες, των οποίων τα γνωρίσματα μιας δεδομένης κατηγορίας (ή μετακατηγορίας) έχουν ως τιμή μια δεδομένη οντότητα.

-- Αναδρομική εφαρμογή των δύο παραπάνω ερωτήσεων για μια ή περισσότερες κατηγορίες (ή μετακατηγορίες).

- Στην τέταρτη κατηγορία ανήκουν οι συναρτήσεις εκείνες που επιτρέπουν το χειρισμό των συνόλων που χρησιμοποιούνται για την αποθήκευση των αποτελεσμάτων (answer set) μιας ερώτησης στη ΒΓ του ΣΣΕ.

- Στην τελευταία κατηγορία ανήκουν οι συναρτήσεις εκείνες που επιτρέπουν ανάκτηση πληροφορίας από τα σύνολα αποθήκευσης αποτελεσμάτων.

Αναλυτική περιγραφή των συναρτήσεων αυτών βρίσκεται στο [30].

Κεφάλαιο 5

Ένα μοντέλο παράστασης εγγράφων SGML σε Telos

Στο κεφάλαιο αυτό χρησιμοποιείται η γλώσσα παράστασης γνώσης SIS-Telos για την παράσταση πληροφορίας σχετικής με την δομή και το περιεχόμενο εγγράφων SGML. Η πληροφορία αυτή αποθηκεύεται στη Βάση Γνώσης του συστήματος Κώδικας.
Η οντοκεντρική προσέγγιση της γλώσσας SIS-Telos επιτρέπει την αποδοτική παράσταση εννοιών με πολλαπλές συσχετίσεις όπως είναι οι έννοιες που περιέχονται μέσα στα έγγραφα. Επίσης, τα εκφραστικά μέσα της γλώσσας αυτής (οντότητες , σχέσεις οντοτήτων) είναι αντιστοιχα με τα εκφραστικά μέσα της SGML για την παράσταση της δομής των εγγράφων (τμήματα, χαρακτηριστικά τμημάτων). Η χρήση της γλώσσας SIS-Telos για την παράσταση της δομής ενός πολύπλοκου τύπου εγγράφου φανερώνει την ικανότητα της γλώσσας για την υποστήριξη εφαρμογών SGML που χρησιμοποιούν βάσεις δεδομένων για την παράσταση πληροφορίας για τα έγγραφα που διαχειρίζονται.
Στη συνέχεια του κεφαλαίου δίνεται μια περιγραφή του τμήματος της οντολογίας του σημασιολογικού μοντέλου της SIS-Telos που παριστάνει τις έννοιες που χρησιμοποιεί η SGML για την περιγραφή δομημένων εγγράφων. Επίσης παρουσιάζεται ένας τύπος εγγράφου κατάλληλος για την περιγραφή πολιτισμικής πληροφορίας και δίνεται η παράστασή του στο σημασιολογικό μοντέλο της SIS-Telos.

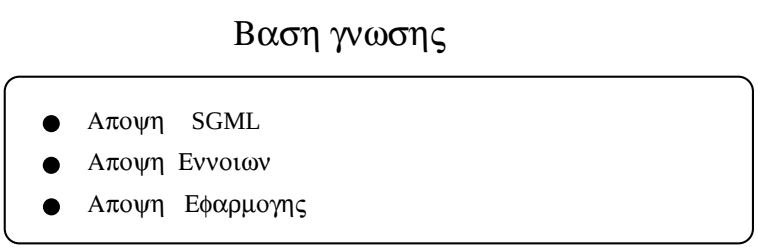
5.1 Παρουσίαση του μοντέλου

Η οντολογία του Κώδικα χωρίζεται (σημασιολογικά) σε τρεις απόψεις.

- Η **άποψη SGML** που παριστάνει έννοιες της SGML, τα DTD και τη δομή των βασικών εγγράφων που διαχειρίζεται η εφαρμογή.

- Η **Αποψη Εννοιών** που παριστάνει έννοιες που σχετίζονται με το σημασιολογικό περιεχόμενο των εγγράφων (κόσμο της εφαρμογής).

- Η **Αποψη Επεξεργασίας** που παριστάνει την πληροφορία που χειρίζεται το σύστημα για τους σκοπούς της εφαρμογής. Η πληροφορία αυτή αφορά στο σχολιασμό και ευρετηριασμό των εγγράφων, στην καταγραφή εκδόσεων (διασκευών ή παραλλαγών), στη παραγωγή προϊόντων με μορφή ηλεκτρονικών ή τυπωμένων εγγράφων, οπτικών δίσκων, παρουσιάσεων (έτοιμων ή on-line) κτλ.



Σχήμα 5.1: Οι τρεις απόψεις του Σημασιολογικού μοντέλου του Συστήματος Σημασιολογικού Ευρετηριασμού.

Στο κεφάλαιο αυτό παρουσιάζουμε την περιγραφή της **άποψης SGML** και της **άποψης Επεξεργασίας**. Η άποψη Εννοιών λόγω άμεσης εξάρτησης του με τα βασικά έγγραφα που καλείται κάθε φορά να διαχειρίζεται το σύστημα Κώδικας δεν αποτελείται από ένα σταθερό αρχικό πληθυσμό αλλά προσαρμόζεται κάθε φορά στη συγκεκριμένη εφαρμογή. Για το λόγο αυτό παρουσίαση της άποψης Εννοιών θα γίνει στο κεφ. 7 στα πλαίσια μια συγκεκριμένης εφαρμογής του Κώδικα για τα έγγραφα του Τουρκικού Αρχείου Χανίων.

5.2 Αποψη SGML

Οι οντότητες που συνιστούν την άποψη αυτή αποτελούν τις βασικές έννοιες τις οποίες χρησιμοποιεί η SGML για την περιγραφή της λογικής δομής εγγράφων.
Για κάθε μια από τις παραπάνω έννοιες υπάρχουν αντίστοιχες μετακλάσεις στο μοντέλο που τις περιγράφουν. Συγκεκριμένα , παρακάτω δίνεται η περιγραφή του μοντέλου που σχεδιάστηκε.

Το πρότυπο της SGML (βλέπε σχετική περιγραφή στο κεφ. 3) ορίζει ότι ένα έγγραφο SGML αποτελείται από τρία τμήματα.

- Το **πρώτο τμήμα** περιέχει την Δήλωση SGML (SGML Declaration). Το τμήμα αυτό δεν περιγράφεται στο μοντέλο μας διότι περιέχει πληροφορίες που αφορούν τεχνικά χαρακτηριστικά του συγκεκριμένου συστήματος που υλοποιείται η εφαρμογή SGML (πχ. σύνολο χαρακτήρων που χρησιμοποιείται κτλ.) και δεν σχετίζεται με τις έννοιες που χρησιμοποιεί η SGML για την περιγραφή της δομής εγγράφων.

- Το **δεύτερο τμήμα** κάθε εγγράφου SGML , περιέχει την δήλωση του DTD. Ενα DTD περιέχει δηλώσεις EAA που ένα έγγραφο μπορεί να έχει. Επίσης περιέχει δηλώσεις των λογικών τμημάτων τα οποία αποτελούν τη λογική δομή του τύπου εγγράφου, καθώς και χαρακτηριστικών που αποδίδονται στα τμήματα αυτά.
  Οι οντότητες και σχέσεις που παριστάνουν τις δηλώσεις αυτές περιέχονται στην οντολογία της άποψης SGML του συστήματος.

- Το τρίτο τμήμα περιέχει τα βασικά (πραγματικά) έγγραφα (από δω και στο εξής BE). Οι οντότητες και οι σχέσεις που παριστάνουν τη δομή των εγγράφων αυτών περιέχονται στην οντολογία της άποψης SGML του συστήματος.

Η οντολογία της άποψης SGML χωρίζεται σε τρείς στάθμες:

- Η **πρώτη στάθμη** (Γενικό μοντέλο, στάθμη M1_Class) περιέχει κλάσεις αντικειμένων ανεξάρτητες του τύπου των δομημένων εγγράφων που χρησιμοποιούνται από την συγκεκριμένη εφαρμογή . Οι κλάσεις αυτές περιγράφουν τις βασικές έννοιες που χρησιμοποιεί η SGML για την περιγραφή των εγγράφων και αποτελούν τον αρχικό πληθυσμό της άποψης SGML.

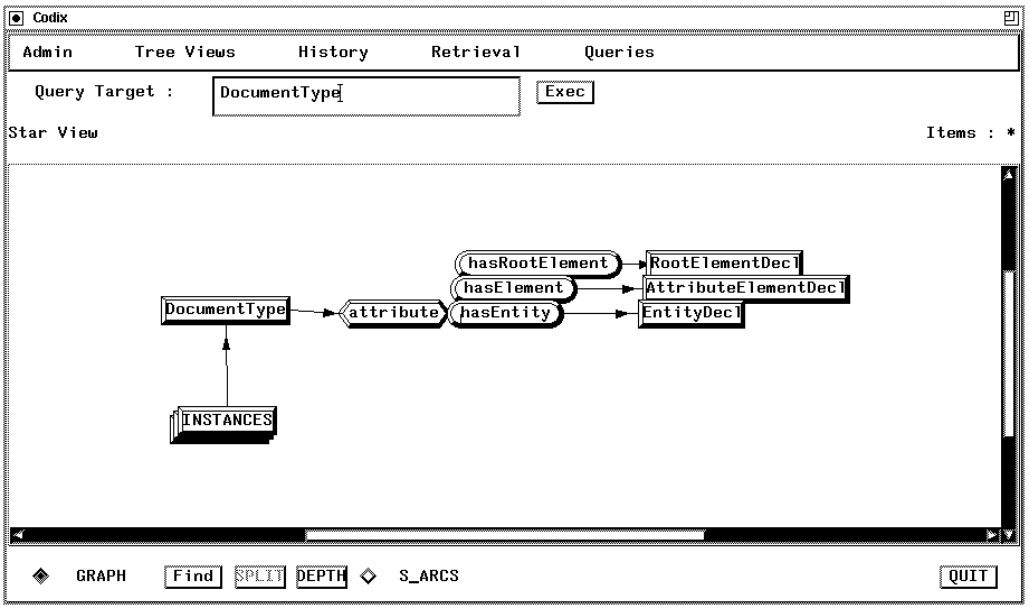
- Η **δεύτερη στάθμη** (στάθμη S_Class) περιέχει κλάσεις αντικειμένων οι οποίες εξαρτώνται κάθε φορά από τον τύπο των εγγράφων που διαχειρίζεται η εφαρμογή. Οι κλάσσεις αυτές περιγράφουν τα δομικά στοιχεία των τύπων των εγγράφων αυτών (DTDs) καθώς και τις σχέσεις μεταξύ τους και συνδέονται κατάλληλα με τις οντότητες του γενικού μοντέλου.

- Η **τρίτη στάθμη** (στάθμη Token) παριστάνει τη δομή των (βασικών) εγγράφων που χειρίζεται η εφαρμογή. Δημιουργείται αφού έχει δημιουργηθεί το τμήμα της οντολογίας της δεύτερης στάθμης που παριστάνει το DTD που ακολουθούν τα έγγραφα αυτά. δεύτερης στάθμης που παρι

Παρακάτω περιγράφονται αναλυτικά οι οντότητες που σχεδιάστηκαν και η σημασία τους.

5.2.1 Πρώτη στάθμη

Οι έννοιες που παριστάνουν ορισμούς Τύπων Εγγράφων συγκεντρώνονται κάτω από τη μετακλάση **DocumentType** . Η μετακλάση αυτη έχει τα γνωρίσματα *hasRootElement*, *hasElement*, *hasEntity* που συνδέουν αντίστοιχα τις οντότητες που παριστάνουν το Βασικό Τύπο Τμήματος, τους υπόλοιπους Τύπους Τμημάτων και τα ΕΑΑ ενός Τύπου Εγγράφου με την οντότητα που τον παριστάνει.



Σχήμα 5.2: Τμήμα της οντολογίας του μοντέλου για την περιγραφή εννοιων σχετικών με δηλώσεις Τυπων Εγγράφων.

Η μετακλάση **SimpleElementDecl** συγκεντρώνει τις δηλώσεις των απλών τύπων τμημάτων όπως αυτά εμφανίζονται στη λογική δομή ενός εγγράφου. Υποκλάσεις της είναι οι κλάσεις **TerminalElementDecl, EmptyElementDecl** που παριστάνουν αντίστοιχα τις δηλώσεις των τύπων τμημάτων που περιέχουν τελική πληροφορία (PCDATA,CDATA) και των τύπων τμημάτων που έχουν κενή λίστα περιεχομένων.
Παράδειγμα δήλωσης τελικών και κενών τύπων τμημάτων είναι οι δηλώσεις αντίστοιχα των τύπων τμημάτων **title και figure** που δίνονται παρακάτω :

```
<! ELEMENT title (#PCDATA) >
```

```
<! ELEMENT figure ( EMPTY) >
```

Η μετακλάση **GroupElementDecl** παριστάνει τους τύπους τμήματος των οποίων η λίστα περιεχομένων αποτελείται από συγκεκριμένα υποτμήματα τα οποία εμφανίζονται ακολουθιακά (όταν συνδέονται με ”,”) ή με οποιαδήποτε σειρά (όταν συνδέονται με ”&”.
Στην παρακάτω δήλωση , το περιεχόμενο της παρένθεσης, **(intro?, chapter)+** στη λίστα περιεχομένων του τύπου τμήματος **book** αποτελεί περίπτωση της μετακλάσης **GroupElementDecl** :

```
<! ELEMENT book ( (title,(intro?, chapter)+)) >
```

Η μετακλάση **OrElementDecl** συγκεντρώνει τις δηλώσεις των τύπων τμημάτων οι οποίες μπορεί να αποτελούνται από τύπους τμήματος που εμφανίζονται εναλλακτικά. Παράδειγμα τέτοιων τύπων τμημάτων αποτελεί το δεύτερο συστατικό του μοντέλου περιεχομένων του τύπου τμήματος **chapter** (που περικλείεται από την παρένθεση) όπως φαίνεται παρακάτω:

```
<! ELEMENT chapter ( title, (para | fig) + ) >
```

Υπερσύνολο των δύο παραπάνω κλάσεων αποτελεί η μετακλάση **NonTerminalElementDecl** περιπτώσεις της οποίας είναι οι τύποι τμημάτων που είναι μη τελικοί δηλ. το μοντέλο περιεχομένων τους αποτελείται από άλλους τύπους τμημάτων (subelements) . Η ύπαρξη της μετακλάσης αυτής διαχωρίζει τους **μη τελικούς** τύπους τμημάτων από αυτούς που είναι τελικοί ή κενοί.
Η μετακλάση **RootElementDecl** συγκεντρώνει τις δηλώσεις των τύπων τμημάτων που είναι βασικοί (Root Elements).
Υπερσύνολο των μετακλάσεων **RootElementDecl, NonTerminalElementDecl** είναι η **ComplexElementDecl**. Η μετακλάση αυτή ομαδοποιεί όλες τις δηλώσεις τύπων τμημάτων που το μοντέλο περιεχομένων τους αποτελείται από τύπους υποτμημάτων (είναι μή τελικό ή μή κενό).
Στην μετακλάση αυτή αποδίδεται το γνώρισμα:

- *hasContents*: Κατηγορία γνωρισμάτων που παίρνει τιμές από την κλάση ElementDecl. Οι τιμές αυτές παριστάνουν το είδος των τμημάτων από τα οποία αποτελείται αυτό το τμήμα (μοντέλο περιεχομένων του τμήματος). Τα γνωρίσματα που αποδίδονται σε αυτή τη κατηγορία γνωρισμάτων είναι το εξής:
  - *hasNext*: Κατηγορία γνωρισμάτων που παίρνει τιμές από την κλάση **ElementDecl** και δηλώνει τον τύπο τμήματος που το ακολουθεί στη λίστα περιεχομένων. Το

γνώρισμα αυτό μπορεί να δηλώνεται ως περίπτωση μιας εκ των των κλάσεων συτήματος **Unique, Necessary, Possible** δηλώνοντας αντίστοιχα ότι ο τύπος τμήματος που ακολουθεί στο μοντέλο περιεχομένων μιας δήλωσης ενός τύπου τμήματος εμφανίζεται αντίστοιχα μία ακριβώς φορά, μία ή περισσότερες φορές και καμία ή μία φορά.

Επίσης ορίζεται η μετακλάση **ElementDecl** που έχει υποσύνολα τις μετακλάσεις **SimpleElementDecl ,NonTerminalElementDecl, ParameterEntityDecl** . Κάτω από την μετακλαση **ParameterEntityDecl** συγκεντρώνονται όλα τα παραμετρικά ΕΑΑ που δηλώνονται σε ένα τυπο εγγράφου (DTD).
Ορίζεται η μετακλάση **AttributeElementDecl** υποσύνολα της οποίας είναι οι **SimpleElementDecl, ComplexElementDecl**. Σε αυτήν αποδίδεται το γνώρισμα *hasAttribute* που παίρνει τιμές από την κλάση **AttributeDecl** . Οι τιμές του γνωρίσματος αυτού δηλώνουν τις τιμές των χαρακτηριστικών που αποδίδονται σε ένα τύπο τμήματος. Η ύπαρξη της **AttributeElementDecl** εξυπηρετεί για το διαχωρισμό των δηλώσεων τύπων τμημάτων στις οποίες επιτρέπεται ή όχι να αποδίδονται χαρακτηριστικά. Παράδειγμα δήλωσης τύπου τμήματος (ElementDecl) στην οποία δεν επιτρέπεται η απόδοση χαρακτηριστικών είναι οι δηλώσεις τύπων τμήματος που είναι περιπτώσεις της μετακλάσης **ParameterEntityDecl**.

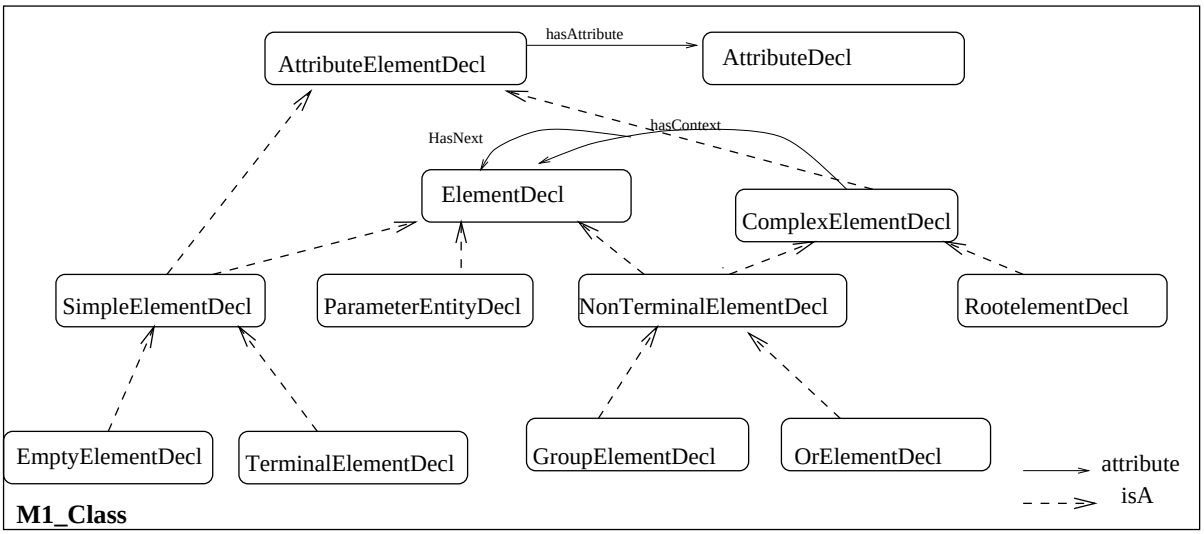
Στο Γενικό μοντέλο ορίζονται επίσης οι κλάσεις **Element** και **Document**. Οι κλάσεις αυτές αποτελούν συζυγείς κλάσεις των κλάσεων αντίστοιχα **ElementTypeDecl** και **DocumentType**. Η κλάση **Element** είναι περίπτωση της κλάσης **ElementTypeDecl** και υποκλάσεις αυτής είναι όλες οι δηλώσεις τύπων τμημάτων που είναι περιπτώσεις της ElementTypeDecl. Ανάλογα, η κλάση **Document** είναι περίπτωση της κλάσης **DocumentType** και υποκλάσεις αυτής είναι όλες οι δηλώσεις τύπων τμημάτων που είναι περιπτώσεις της DocumentType. Περισσότερα για τη χρήση των συζυγών κλάσεων στο ΣΣΕ δίνονται στο [6].

Εκτός όμως από τις μετακλάσεις που περιγράφουν δηλώσεις τύπων εγγράφων (Document types) και τύπων τμημάτων (Element types) έχουν οριστεί μετακλάσεις για την περιγραφή εννοιών της SGML που χρησιμοποιούνται για απόδοση χαρακτηρισικών (attributes). Παρακάτω περιγράφονται αναλυτικά οι μετακλάσεις αυτές.
Ορίζεται η μετακλάση **AttributeDecl** κάτω από την οποία συγκεντρώνονται τα χαρακτηριστικά που αποδίδονται στους τύπους τμημάτων. Στην κλάση αυτή αποδίδεται το εξής γνώρισμα:

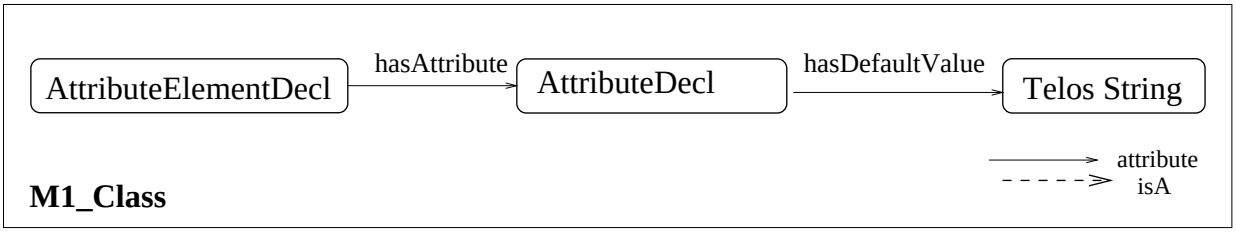
- *hasDefaultValue* : Κατηγορία γνωρισμάτων που παίρνει τιμές από την κλάση συστήματος **Telos String** και δηλώνει την τιμή εξ ορισμού που μπορεί να αποδίδεται στα χαρακτηριστικά που συγκεντρώνονται κάτω από τη μετακλάση **AttributeDecl**.

Τα πεδία τιμών των χαρακτηριστικών που αποδίδονται στα τμήματα παριστάνονται από τις

οι κλάσεις **ENTITY,NMTOKEN,PENTITY,NUMBER,CDATA, NUTOKEN,ID,IDREF,IDREFS,ORVALUE** που περιγράφουν αντίστοιχα πως το πεδίο τιμών ενός χαρακτηριστικού είναι τύπου ENTITY ,NMTOKEN,PARAMETER ENTITY ,NUMBER,CDATA, NUTOKEN,ID, IDREF, IDREFS ή μία λίστα τιμών (ORVALUE) όπως ορίζεται από το πρότυπο της SGML. Στο σχήμα 5.3 φαίνεται η οντολογία που αναπτύχθηκε για την αναπαράσταση των βασικών εννοιών που χρησιμοποιεί η SGML για την περιγραφή της δομής των εγγράφων που ακολουθούν αυτό το πρότυπο.



Σχήμα 5.3: Τμήμα της οντολογίας του μοντέλου για την περιγραφή εννοιων σχετικών με δηλώσεις Τυπων Τμημάτων.



Σχήμα 5.4: Τμήμα της οντολογίας του συστήματος για την περιγραφή εννοιών σχετικών με δηλώσεις Χαρακτηριστικών.

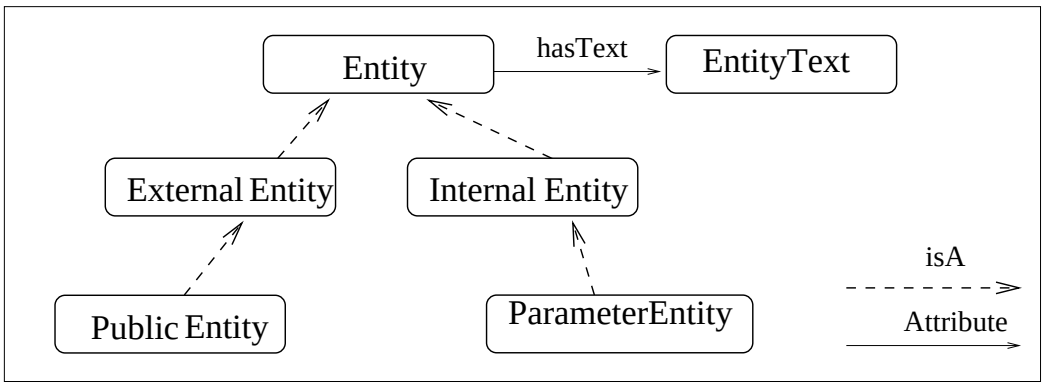
Ορίζεται η μετακλάση **EntityDecl** που περιγράφει την έννοια των εικονικών αντικειμένων αποθήκευσης(EAA). Οπως ειπώθηκε στο κεφ. 3 ,τα εικονικά αντικείμενα αποθήκευσης είναι έννοιες που υποστηρίζονται από το πρότυπο της SGML ώστε να γίνεται αποδοτικότερη

η αποθήκευση των εγγράφων SGML. Υποσύνολα αυτής είναι οι κλάσεις **InternalEntityDecl, ExternalEntityDecl** που περιγράφουν αντιστοιχα τα εσωτερικά και εξωτερικά ΕΑΑ. Στην **EntityDecl** αποδίδεται το γνώρισμα *hasText* που αναφέρεται σε κατηγορία γνωρισμάτων που παίρνει τιμές από την κλάση **EntityText** και συνδέει ένα ΕΑΑ με το κείμενο που ανιπροσωπευει.

Ορίζεται επίσης η μετακλάση **EntityText** που περιγράφει το κείμενο αντικατάστασης κάθε φορά που γίνεται αναφορά σε κάποιο ΕΑΑ. από την κλάση συστήματος **Telos_String** και περιγράφει το αναγνωριστικό

Υποσύνολο της **ExternalEntityDecl** είναι η μετακλάση **PublicEntityDecl** που περιγράφει τα ΕΑΑ των οποίων το κείμενο αντικατάστασης (όταν γίνεται αναφορά σε αυτά) βρίσκεται εξωτερικά του εγγράφου SGML , αποθηκευμένο σε κάποιο αρχείο του συστήματος.

Υποσύνολο της **InternalEntityDecl** είναι η μετακλάση **ParameterEntityDecl** που περιγράφει τα Παραμετρικά ΕΑΑ .

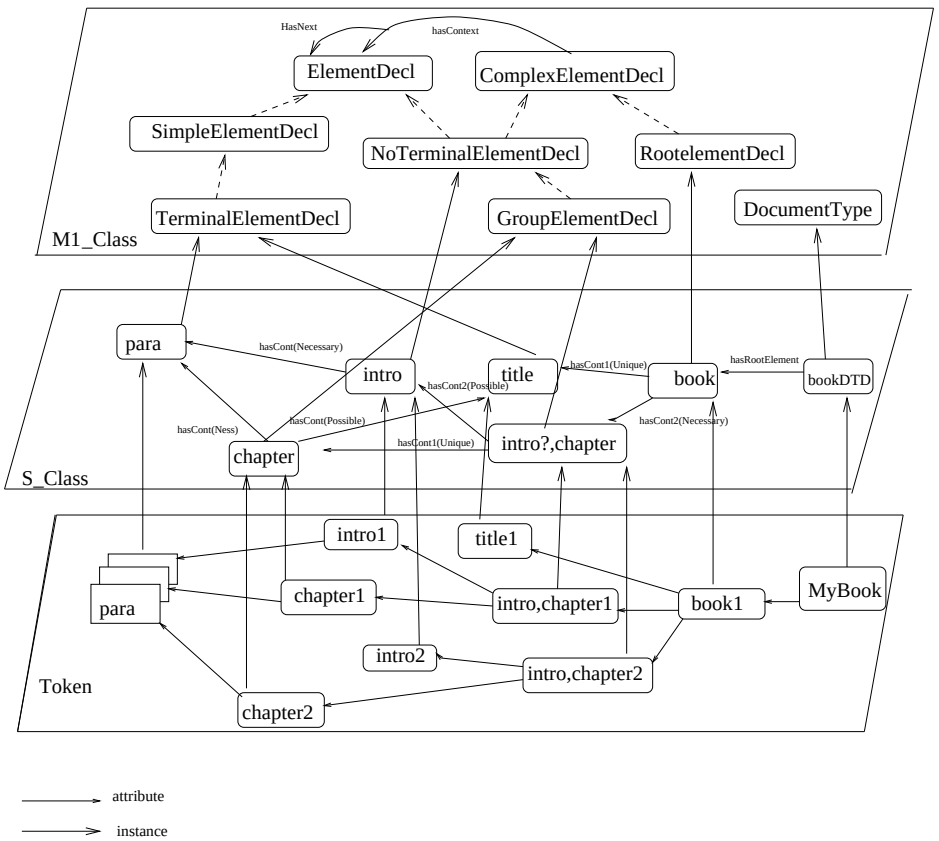


Σχήμα 5.5: Τμήμα της οντολογίας του μοντέλου για την περιγραφή εννοιων σχετικών με δηλώσεις ΕΑΑ .

Στο σχήμα 5.6 φαίνονται οι κλάσεις που δημιουργούνται για τη περιγραφή ενός τύπου εγγράφου τύπου **book** που περιγράφεται από την παρακάτω δήλωση (DTD):

```
<!DOCTYPE book[
<!ELEMENT book (title,(intro?, chapter)+)>
<!ELEMENT chapter (title, para+) >
<!ELEMENT intro (para+) >
```

```
<!ELEMENT title (#PCDATA) >
<!ELEMENT para (#PCDATA) >
]>
```



Σχήμα 5.6: Τμήμα της οντολογίας του μοντέλου μετά την εισαγωγή του book DTD.

5.3 Αποψη Επεξεργασίας

Η άποψη Επεξεργασίας παριστάνει πληροφορίες χρήσιμες για την επιτέλεση των λειτουργιών που υποστηρίζει το σύστημα Κώδιξ. Οι πληροφορίες αυτές αφορούν στον τρόπο ευρετηριασμού των εγγράφων, στο σχολιασμό τους, στις εκδόσεις και διασκευές των εγγράφων που διαχειρίζεται το σύστημα. Συγκεκριμένα, οι οντότητες και οι σχέσεις μεταξύ τους που συνιστούν την άποψη αυτή κατηγοριοποιούνται σύμφωνα με τη λειτουργία που εξυπηρετούν και είναι οι ακολουθες:

- **Παρουσίαση τμήματος**

  Για τη λειτουργία αυτή ορίζονται τα παρακάτω γνωρίσματα:

-- *hasOffset*: Αποδίδεται σε οντότητες τύπου Element και παίρνει τιμές στη κλάση συστήματος Telos_Int συμβολίζει δε τη θέση (γραμμή) που αρχίζουν τα περιεχόμενα του τμήματος ενός εγγράφου στο ηλεκτρονικό αρχείο.

-- *presentationAttribute*: Αποδίδεται σε οντότητες τύπου Element και παίρνει τιμές στη κλάση συστήματος Telos_String. Παριστάνει τα χαρακτηριστικά που αποδίδονται σε ένα τμήμα εγγράφου και αφορούν στην παρουσίασή του (πχ. fontsize).

-- *electronicFile*: Αποδίδεται σε οντότητες τύπου Document και παίρνει τιμές στη κλάση συστήματος Telos_String. Παριστάνει το όνομα του ηλεκτρονικού αρχείου στο οποίο αποθηκεύεται ένα έγγραφο.

- **Ευρετηριασμός τμήματος**

  Για τη λειτουργία αυτή ορίζονται τα παρακάτω γνωρίσματα και οντότητες:

  -- *associatedWith*: Συνδέει όρους του ευρετηρίου (οντότητες της άποψης Εννοιών) με οντότητες που παριστάνουν έγγραφα ή τμήματά τους.

  -- *see*: Συνδέει μεταξύ τους οντότητες της άποψης SGML. Η σχέση αυτή δηλώνει ότι ένα έγγραφο ή τμήμα του (αφετηρία του συνδέσμου) έχει συναφές περιεχόμενο με ένα άλλο έγγραφο ή τμήμα του (προορισμό του συνδέσμου).

  -- *seeAlso*: Συνδέει μεταξύ τους οντότητες της άποψης SGML. Η σχέση αυτή δηλώνει ότι ένα έγγραφο ή τμήμα του (αφετηρία του συνδέσμου) είναι σχετικό (ως προς το περιεχόμενο) με ένα άλλο έγγραφο ή τμήμα του (προορισμό του συνδέσμου).

  -- *detailedDescription*: Συνδέει μεταξύ τους οντότητες της άποψης SGML. Η σχέση αυτή δηλώνει ότι ένα έγγραφο ή τμήμα του (αφετηρία του συνδέσμου) έχει αναλυτικότερη παρουσίαση ενός θέματος (όσον αφορά στο περιεχόμενο) από ένα άλλο έγγραφο ή τμήμα του (προορισμό του συνδέσμου).

Για την ανάλυση της σημασιολογίας των συνδέσμων ευρετηριασμού μεταξύ εγγράφων (ή τμημάτων εγγράφων) έχουν αναπτυχθεί σχετικές εργασίες [**?**]. Στην παρούσα εργασία θεωρήθηκε ικανοποιητική η κατηγοριοποίηση των συνδέσμων στις παραπάνω βασικές κατηγορίες για δύο λόγους α) η σημασία (και η διαφοροποίησή τους) είναι προφανής και δεν οδηγούν σε αμφισημίες, παρανοήσεις κτλ. β) είναι ανάλογες με τις σημασιολογικές σχέσεις μεταξύ όρων θησαυρού όπως αυτές ορίζονται από το διεθνές πρότυπο ISO 2788 [**?**]. Επίσης, ο ευρετηριασμός εγγράφων ή τμημάτων τους με όρους

(οντότητες) του σημασιολογικού δικτύου γίνεται με τη χρήση του συνδέσμου *associatedWith*. Δεν κρίθηκε απαραίτητη η κατηγοριοποίηση του συνδέσμου αυτού διότι α) η χρήση του σημασιολογικού μοντέλου του ΣΣΕ (και συγκεκριμένα η θέση κάθε όρου ευρετηρίου στη δομή του μοντέλου) ενέχει την σημασιολογική πληροφορία (πχ ευρετηριασμός ως προς τόπο, χρόνο) που θα προσέφερε η κατηγοριοποίηση του συνδέσμου αυτού και β) ο σημασιολογικός ευρετηριασμός εξαρτάται απο τις συγκεκριμένεςαπαιτήσεις αναζήτησης κάθε εφαρμογής οπότε και μπορεί ο σύνδεσμος αυτός να εξειδικευτεί (βλέπε κεφ. 7 παράγραφο 7.1.3)
Στα παραπάνω γνωρίσματα αποδίδονται τα γνωρίσματα *date*, *indexer* και *description* που δηλώνουν αντίστοιχα το χρόνο, το πρόσωπο που έκανε τον ευρετηριασμό και μια μικρή περιγραφή (σχόλιο) από τον ευρετηριαστή. Τα δύο πρώτα γνωρίσματα παίρνουν τιμή στην κλάση συστήματος Telos_String. Το γνώρισμα *description* είναι κείμενο και παίρνει τιμή στην κλάση *hyperText*.

- **Θησαυρός όρων**
  Το σύστημα υποστηρίζει ευρετηριασμό εγγράφων ή τμημάτων τους βάσει όρων που ανήκουν σε θησαυρό. Οι σχέσεις μεταξύ των όρων του θησαυρού που υποστηρίζουμε καθορίζονται σύμφωνα με το διεθνές πρότυπο για θησαυρούς, ISO 2788 [**?**] και είναι οι εξής α) σχέσεις ευρύτερου-στενότερου όρου β) σχετικών όρων και γ) συνώνυμων όρων. Για την παράσταση των όρων του θησαυρού και των σχέσεων μεταξύ τους δημιουργήθηκαν οι παρακάτων κλάσεις αντικειμένων και σύνδεσμοι οι οποίοι ανήκουν στην άποψη επεξεργασίας:

  -- **Term**: Περιπτώσεις της κλάσης αυτής είναι όλοι οι όροι που παριστάνονται στο σύστημα Κώδικας.

  -- **PreferedTerm**: Περιπτώσεις της κλάσης αυτής είναι όλοι οι όροι που χρησιμοποιούνται από έναν θησαυρό όρων (δόκιμοι όροι). Η κλάση αυτή είναι υποκλάση της κλάσης **Term**.

  -- **NonPreferedTerm**: Περιπτώσεις της κλάσης αυτής είναι οι όροι εκείνοι που δεν είναι δόκιμοι στον θησαυρό που υποστηρίζει το σύστημα και είναι συνώνυμοι με όρους του θησαυρού με τους οποίους συνδέονται μέσω του συνδέσμου *synonymTerm*. Η κλάση αυτή είναι υποκλάση της κλάσης **Term**.

  -- **HierarchyTerm**: Περιπτώσεις της κλάσης αυτής είναι οι όροι του θησαυρού που ανήκουν σε σημασιολογικές ιεραρχίες όρων. Η κλάση αυτή είναι υποσύνολο της κλάσης **PreferedTerm**.

-- *broaderTerm*: Το γνώρισμα αυτό συνδέει έναν όρο του θησαυρού με έναν άλλο όρο του θησαυρού σημασιολογικά ευρύτερο. Αφετηρία και κατάληξη του συνδέσμου είναι η κλάση HierarchyTerm.

-- *relatedTerm*: Το γνώρισμα αυτό συνδέει έναν όρο του θησαυρού με έναν άλλο όρο του θησαυρού σημασιολογικά σχετικό. Αφετηρία και κατάληξη του συνδέσμου είναι η κλάση HierarchyTerm.

-- *synonymTerm*: Το γνώρισμα αυτό συνδέει έναν μή δόκιμο όρο με έναν συνώνυμό του όρο που ανήκει στον θυσαυρό. Αφετηρία του συνδέσμου είναι η κλάση NonPreferedTerm και κατάληξη του συνδέσμου είναι η κλάση PreferedTerm.

- **Σχολιασμός τμήματος**
  Για τη λειτουργία του σχολιασμού δημιουργήθηκαν οι παρακάτω κλάσεις αντικειμένων και σύνδεσμοι:

  -- **Comment** : Περιπτώσεις της κλάσης αυτής είναι τα σχόλια που γίνονται πάνω σε τμήματα εγγράφων.

  -- *hasAnnotation* : Το γνώρισμα αυτό συνδέει τμήματα εγγράφων με όντότητες που παριστάνουν σχόλια που έχουν γίνει πάνω σε αυτά.

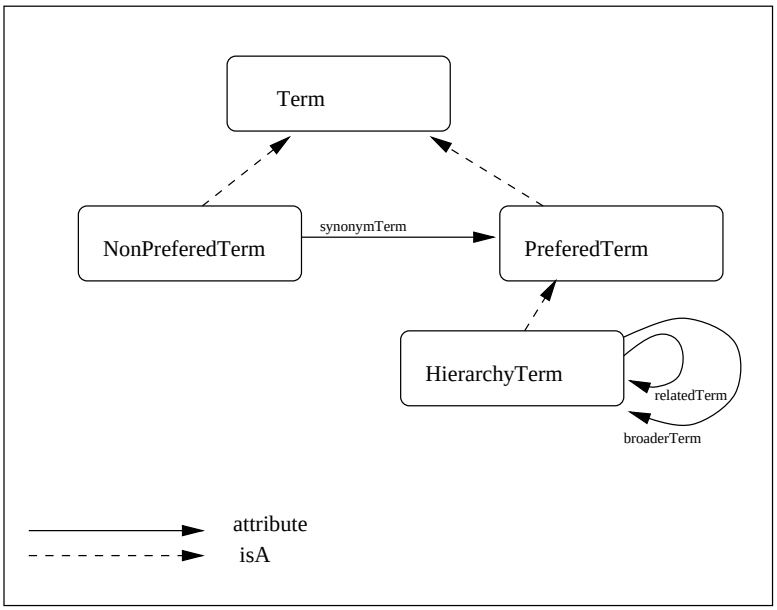
  -- *date* : Το γνώρισμα αυτό αποδίδεται σε περιπτώσεις της κλάσης Comment και δηλώνει τον χρόνο δημιουργίας του σχολίου.

  -- *annotator* : Το γνώρισμα αυτό αποδίδεται σε περιπτώσεις της κλάσης Comment και δηλώνει τον δημιουργό του σχολίου.

- **Καταγραφή εκδόσεων εγγράφων**
  Η καταγραφή εκδόσεων εγγράφων συνίσταται στην παράσταση πληροφορίας σχετικής με το ιστορικό εγγράφων που προκύπτουν απο τροποποίηση (δόρθωση, βελτίωση, ανασχεδιασμό) των ήδη αποθηκευμένων εγγράφων. Διακρίνουμε δύο δυνατές περιπτώσεις: Οταν τα νέα έγγραφα ακολουθούν την ίδια δομή με τα αρχικά (ήδη υπάρχοντα) έγγραφα και διαφέρουν μόνο κατά το περιεχόμενο τότε πρόκειται για καταγραφή **παραλλαγής** (variants) εγγράφου.
  Οταν όμως τα νέα έγγραφα διαφέρουν από τα αρχικά τόσο στη δομή όσο και στο περιεχόμενο τους τότε αναφερόμαστε στη καταγραφή **διασκευής** (revision) του εγγράφου.
  Για τη λειτουργία αυτή δημιουργήθηκαν οι παρακάτω σύνδεσμοι:

-- *variantOf* : Το γνώρισμα αυτό συνδέει δυο διαδοχικές παραλλαγές ενός εγγράφου. Αφετηρία και κατάληξη του γνωρίσματος είναι η κλάση **Document**.

-- *revisionOf* : Το γνώρισμα αυτό συνδέει δυο διαδοχικές διασκευές ενός εγγράφου και είναι εξειδίκευση του γνωρίσματος *variantOf*. Αφετηρία και κατάληξη του γνωρίσματος είναι η κλάση **Document**.

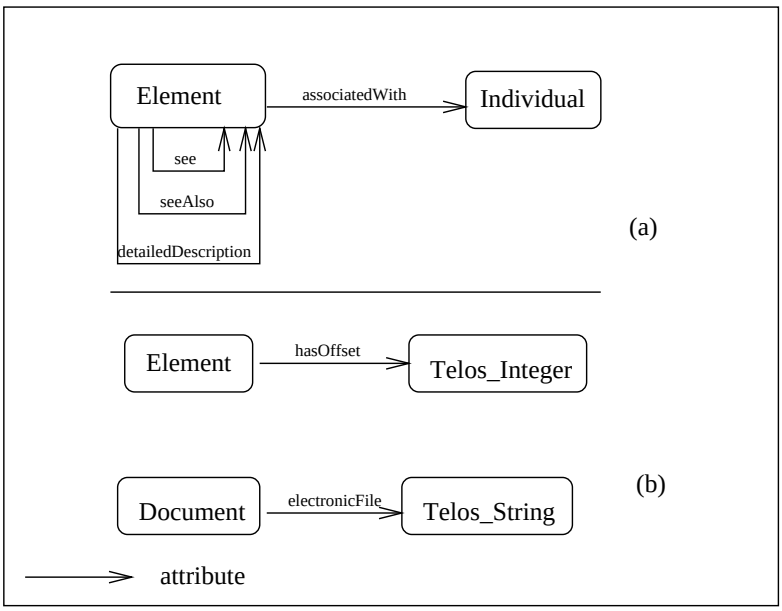


Σχήμα 5.7: Τμήμα της οντολογίας της άποψης Επεξεργασίας που αναφέρεται στην παράσταση του θησαυρού όρων.

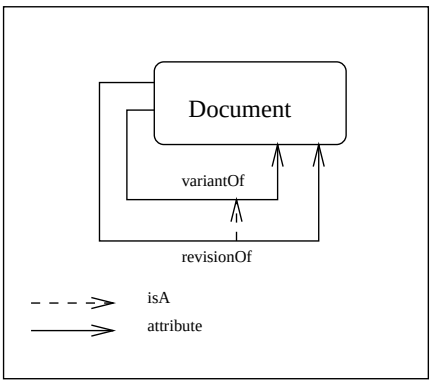
5.4 Εκτεταμένο Παράδειγμα χρήσης του Γενικού Μοντέλου της SGML

Στα πλαίσια εργασιών του προγράματος **Aquarelle**[1] . σχεδιάστηκε από τη EUROCLID ο τύπος εγγράφου **CI** (από δω και στο εξής CI DTD), κατάλληλος για έγγραφα που περιέχουν πολιτισμική πληροφορία.

[1]Το Aquarelle είναι ένα έργο στο οποίο συμμετέχει το ΙΠ-ΙΤΕ και εκτελείται στα πλαίσια του ευρωπαϊκού προγράμματος Τηλεματικής (ESPRIT). Αποσκοπεί στη δημιουργία ενός κατανεμημένου συστήματος πρόσβασης σε πολιτισμική πληροφορίας προς χρήση οργανισμών και επαγγελματιών)[**?**]



Σχήμα 5.8: Τμήμα της οντολογίας της άποψης Επεξεργασίας για τον ευρετηριασμό (a) και την παρουσίαση (b) των διαχειριζομένων εγγράφων.



Σχήμα 5.9: Τμήμα της οντολογίας της άποψης Επεξεργασίας που αναφέρεται στην καταγραφή των εκδόσεων ενός εγγράφου.

5.4.1 Περιγραφή του CI DTD.

Η δομή του τύπου εγγράφου CI που σχεδιάστηκε αποτελείται από δύο κατηγορίες τύπων τμημάτων.

- Πληροφοριακοί τύποι τμημάτων.
- Τύποι τμημάτων ταξινόμησης.

Οι **πληροφοριακοί τύποι τμημάτων** είναι οι τύποι τμημάτων του εγγράφου που περιέχουν πολιτισμική πληροφορία (πληροφορία για πολιτισμικά αντικείμενα, σύνολα αντικειμένων, συλλογές αντικειμένων κ.ο.κ.)
Παραδείγματα τέτοιων τύπων τμημάτων είναι οι εξής:

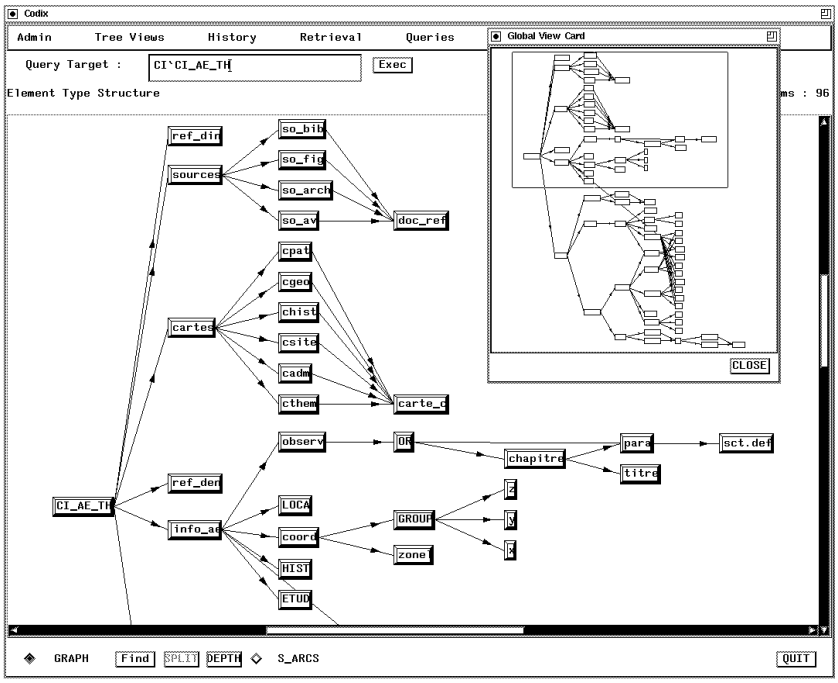
- **dco** : Συλλογές αντικειμένων (collections).
- **din** : Ανεξάρτητα αντικείμενα (individuals) όπως πχ. ακίνητα μνημεία.
- **dobj**: Μη ανεξάρτητα αντικείμενα (objects).
- **den**: Σύνολα αντικειμένων (sets).
- **photos, photog, local, coord, source-info k.o.k.** : Φωτογραφία ή σύνολο φωτογραφιών, τόπος, τοπικές συντεταγμένες, πηγή πληροφορίας για πολιτισμική πληροφορία.

Οι **τύποι τμημάτων ταξινόμησης** είναι τύποι εγγράφων που περιέχουν πληροφορία για την ταξινόμηση των πληροφοριακών τύπων τμημάτων με θεματικά και τοπογραφικά κριτήρια. Οι τύποι τμημάτων ταξινόμησης είναι οι εξής:

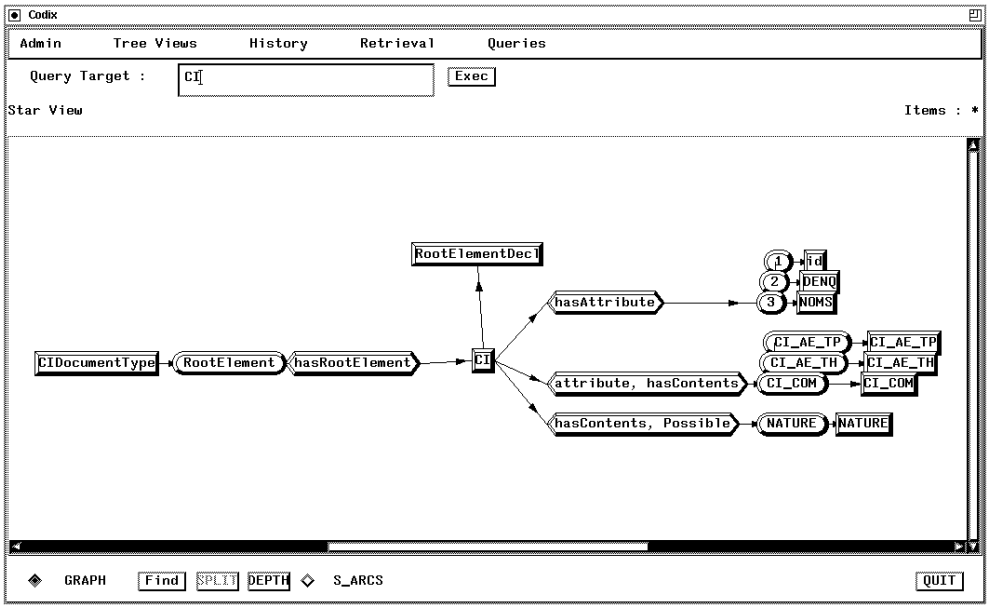
- CI_AE_TH
- CI_AE_TP
- CI_COM

5.4.2 Εισαγωγή του τύπου εγγράφου CI στο ΣΣΕ

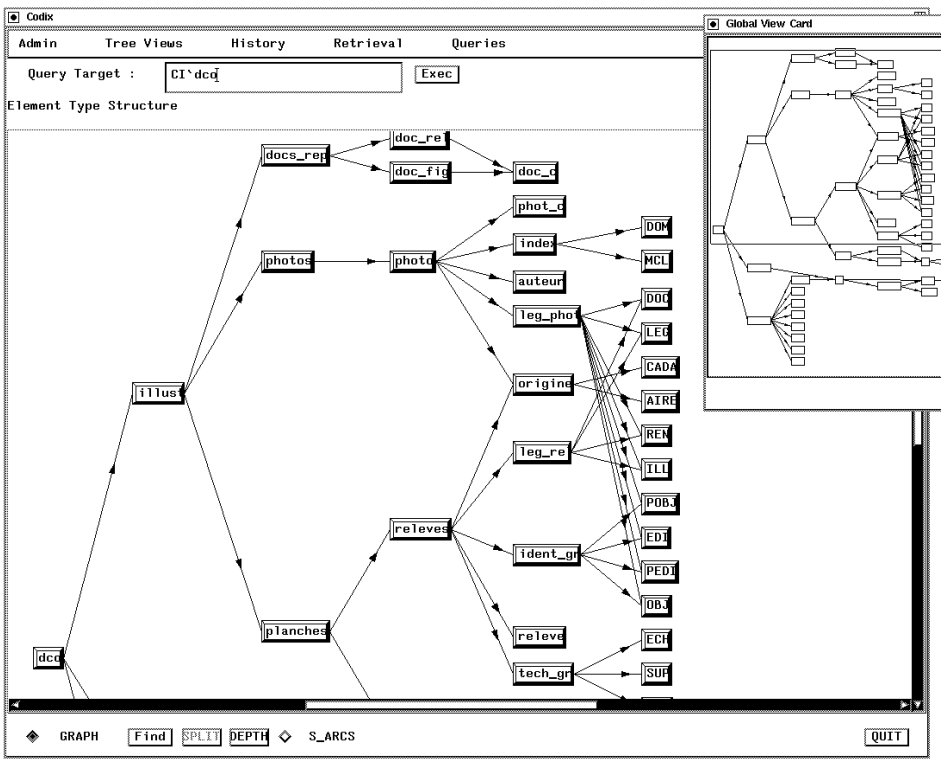
Κατα την εισαγωγή του τύπου εγγράφου που παρουσιάστηκε παραπάνω στο ΣΣΕ, δημιουργούνται κλάσεις αντικειμένων για κάθε τύπο τμήματος που περιέχεται στην αναλυτική περιγραφή του CI. Παρακάτω βλέπουμε τμήματα της δομής του με τη βοήθεια του γραφικού αναλυτή του ΣΣΕ (Gain). Το σχήμα 5.11 απεικονίζει τη δομή του βασικού τμήματος του CI DTD. Επίσης οι εικόνες 5.10 και 5.12 απεικονίζουν ένα τμήμα της δομής του DTD.



Σχήμα 5.10: Τμήμα της οντολογίας του μοντέλου για την περιγραφή του CI_AE_TH τμήματος το οποίο περιέχει πληροφορίες για τη θεματική ταξινόμηση της πολιτισμικής πληροφορίας.



Σχήμα 5.11: Τμήμα της οντολογίας του μοντέλου για την περιγραφή της δομής του βασικού τύπου τμήματος CI .



Σχήμα 5.12: Τμήμα της οντολογίας του μοντέλου για την περιγραφή της δομής του τύπου τμήματος dco που περιέχει πληροφορίες για συλλογές αντικειμένων.

5.4.3 Ονοματολογία των οντοτήτων της άποψης SGML

Οντότητες πρώτης και δεύτερης στάθμης

Οι οντότητες της πρώτης στάθμης της άποψης SGML αποτελούν το γενικό μοντέλο του συστήματος και έχει γίνει αναλυτική περιγραφή στην παράγραφο 5.2. Οι οντότητες της δεύτερης στάθμης της άποψης SGML παριστάνουν τη δομή ενός τύπου εγγράφου. Για τις οντότητες αυτές υιοθετείται η παρακάτω ονοματολογία:

- Το όνομα της οντότητας που παριστάνει το DTD αποτελείται από δύο τμήματα.
  - -- Το πρώτο τμήμα ειναι το ίδιο με το όνομα του DTD όπως αυτό αναφέρεται στο αρχείο που το περιέχει.
  - -- Το δεύτερο τμήμα είναι το αλφαριθμητικό ”DTD”.
- Το όνομα της οντότητας που παριστάνει το βασικό τύπο τμήματος ειναι το ίδιο με το όνομα του βασικού τύπου τμήματος όπως αυτό αναφέρεται στο αρχείο που το περιέχει.
- Το όνομα οντοτήτων που παριστάνουν τύπους τμημάτων (τελικών ,μή τελικών και κενών) ενός DTD αποτελείται από δύο μέρη:
  - -- Το πρώτο μέρος του ονόματος των οντοτήτων αυτών είναι το όνομα της οντότητας που παριστάνει το βασικό τύπο τμήματος του DTD στο οποίο ανήκουν.
  - -- Το δεύτερο μέρος του ονόματος τους είναι ίδιο με το όνομα των αντίστοιχων τύπων τμημάτων του DTD όπως αυτό αναφέρεται στο αρχείο περιγραφής του DTD.
- Τα ονόματα οντοτήτων που παριστάνουν σύνολα (group) τύπων τμημάτων αποτελούνται από δύο τμήματα:
  - -- το πρώτο τμήμα είναι αλφαριθμητικό που αποτελείται από τους τρεις πρώτους χαρακτήρες του ονόματος των τύπων τμημάτων έτσι όπως αυτές αναφέρονται στο αρχείο αποθήκευσης του DTD.
  - -- Το δεύτερο τμήμα είναι το αλφαριθμητικό ”SEQ”, ”AND”, ”OR” εαν οι τύποι τμήματος εμφανίζονται με τελεστή θέσης **","","&"** ή είναι εναλλακτικά (”|”) αντίστοιχα.

- Τα ονόματα οντοτήτων που παριστάνουν χαρακτηριστικά τύπων τμημάτων αποτελούνται από δύο τμήματα:

  -- Το πρώτο τμήμα είναι το όνομα της οντότητας του τύπου τμήματος στην οποία αποδίδεται το χαρακτηριστικό.

  -- Το δεύτερο τμήμα του ονόματος ειναι το όνομα του χαρακτηριστικού όπως αυτό αναφέρεται στο αρχείο αποθήκευσης του DTD.

Οντότητες τρίτης στάθμης

Οι οντότητες τρίτης στάθμης παριστάνουν τη δομή των ΒΕ. Τα ονόματα των οντοτήτων αυτών έχουν σαν πρόθεμα τα ονόματα των οντοτήτων της δεύτερης στάθμης των οποίων είναι περίπτωση. Το επίθεμα του ονόματός τους αποτελείται από δύο αριθμούς. Ο πρώτος αριθμός δηλώνει το πλήθος των ΒΕ που παραστάθηκαν στη ΒΓ και αποτελούν περίπτωση του ίδιου DTD. Ο δεύτερος αριθμός δηλώνει τον αύξοντα αριθμό των όμοιων τμημάτων που παριστάνονται στη δομή του ίδιου ΒΕ.
πχ. οι οντότητες **book.para.1.1, book.para.1.2** παριστάνουν αντίστοιχα την πρώτη και δεύτερη παράγραφο του πρώτου ΒΕ που παραστάθηκε στη ΒΓ του ΣΣΕ και ακολουθεί το DTD book.

Κεφάλαιο 6

Περιγραφή του συστήματος διαχείρισης Εγγράφων SGML, Κώδιξ

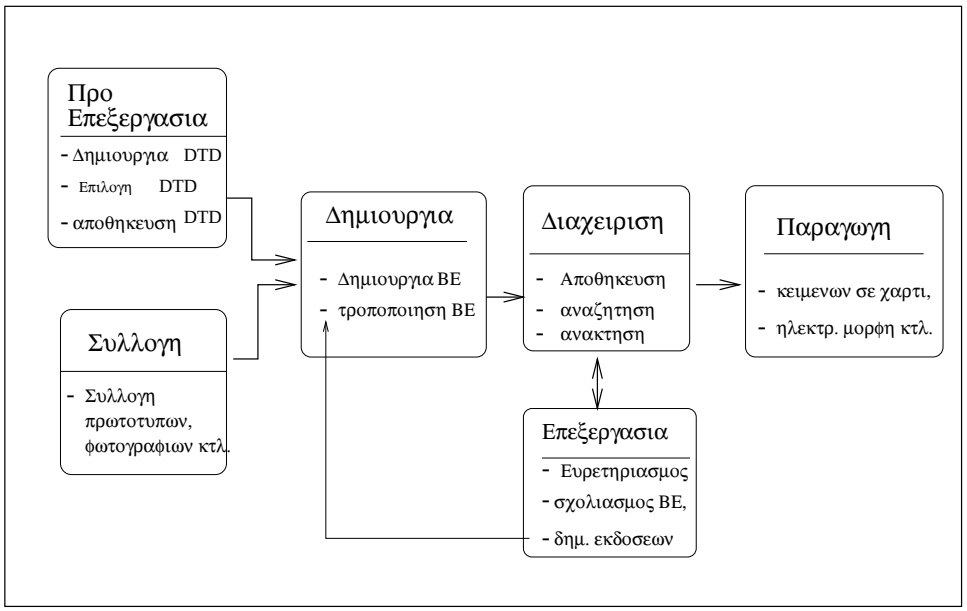
Μία εφαρμογή SGML διαχειρίζεται έγγραφα σ' όλο τον κύκλο ζωής τους. Το σύστημα που θα περιγραφεί παρακάτω επιτρέπει την αποδοτική διαχείριση τέτοιων εγγράφων χρησιμοποιώντας το περιβάλλον του Συστήματος Σημασιολογικού ευρετηριασμού, εργαλεία δημιουργίας και παρουσίασης εγγράφων SGML και έναν απλό εξυπηρετητή εγγράφων.

Στο κεφάλαιο αυτό περιγράφεται ο κύκλος ζωής εγγράφων σε μία εφαρμογή SGML. Ο κύκλος αυτός αποτελείται από τη συλλογή, την προ-επεξεργασία, τη δημιουργία, τη διαχείριση, την επεξεργασία εγγράφων καθώς και την παραγωγή προϊόντων από αυτά. Επίσης, στο κεφάλαιο αυτό γίνεται μια περιγραφή της αρχιτεκτονικής, των τεχνικών χαρακτηριστικών και των λειτουργιών του συστήματος διαχείρισης εγγράφων Κώδιξ. Τέλος, παρουσιάζεται αναλυτικά το σύνολο των προκαθορισμένων ερωτήσεων που χρησιμοποιούνται από τη λειτουργία της αναζήτησης για άντληση πληροφορίας από το σημασιολογικό μοντέλο που αφορά στα έγγραφα που διαχειρίζεται το σύστημα.

6.1 Κύκλος ζωής εγγράφου σε μια εφαρμογή SGML

Στο παρακάτω σχήμα φαίνονται τα στάδια που συνιστούν τον κύκλο ζωής (από δω και στο εξής KZ) των εγγράφων σε ένα σύστημα διαχείρισης εγγράφων SGML. Οι λειτουργίες που συνθέτουν κάθε στάδιο καθορίζονται βέβαια από τους στόχους και τις ιδιαιτερότητες της συγκεκριμένης εφαρμογής ωστόσο μπορούμε να δώσουμε ένα γενικό πλαίσιο λειτουργιών που συναντώνται στα περισσότερα συστήματα διαχείρισης εγγράφων SGML που έχουμε

μελετήσει.



Σχήμα 6.1: Κύκλος Ζωής των εγγράφων SGML σε ένα σύστημα διαχείρισης εγγράφων.

6.1.1 Συλλογή εγγράφων

Το πρώτο στάδιο του ΚΖ αφορά στη διαδικασία της συλλογής των εγγράφων της εφαρμογής. Τα έγγραφα αυτά μπορεί να βρίσκονται στην πρωτότυπη μορφή τους ή να είναι διαθέσιμα με τη μορφή φωτοτυπίας, φωτογραφίας κτλ. Επίσης μπορεί να έχει προηγηθεί πληκτρολόγηση του περιεχομένου τους και να βρίσκονται σε ηλεκτρονική μορφή.

6.1.2 Προ-επεξεργασία

Οι λειτουργίες που συνθέτουν το στάδιο αυτό αφορούν στο σχεδιασμό και/ή την επιλογή και αποθήκευση του/των τύπων εγγράφων που διαχειρίζεται η εφαρμογή. Για τη δημιουργία των τύπων εγγράφων μπορεί να χρησιμοποιούνται ειδικά εργαλεία (DTD Builders) ή απλοί κειμενογράφοι. Επίσης η αποθήκευση των τύπων εγγράφων μπορεί να αφορά μόνο στην αποθήκευση του αρχείου στο οποίο έχουν περιγραφεί και/ή στην παράστασή τους στο σχήμα μιας βάσης δεδομένων.

6.1.3 Δημιουργία Βασικών Εγγράφων

Το στάδιο αυτό αφορά στη δημιουργία των βασικών εγγράφων του συστήματος. Τα έγγραφα αυτά δημιουργούνται με τη χρήση (καθοδήγηση) απλών επεξεργαστών κειμένου ή έξυπνων εργαλείων (SGML Editors). Στο στάδιο αυτό γίνεται ουσιαστικά χρήση της πληροφορίας που συλλέχθηκε κατά το πρώτο στάδιο του KZ των εγγράφων και του συντακτικού της SGML για τη δημιουργία εγγράφων των οποίων η δομή (DTD) έχει γίνει γνωστή στο στάδιο της προ-επεξεργασίας.

6.1.4 Διαχείριση βασικών Εγγράφων

Το στάδιο αυτό αναφέρεται στη αποθήκευση των βασικών εγγράφων της εφαρμογής. Η αποθήκευση των εγγράφων μπορεί να αναφέρεται στην αποθήκευση του ηλεκτρονικού αρχείου σαν ολότητα ή στον κατακερματισμό του σε μικρότερα τμήματα αποθήκευσης (τα οποία καθορίζονται αυτόματα από το σύστημα ή τα επιλέγει ο χρήστης). Επίσης στο στάδιο αυτό έχουμε τις λειτουργίες της αναζήτησης και ανάκτησης των εγγράφων της εφαρμογής. Η αναζήτηση μπορεί να γίνεται βάσει σημασιολογικών, δομικών, λεξικογραφικών κ.α. κριτηρίων και να παρέχεται ειδικό σύστημα διεπαφής για διατύπωση ερωτήσεων, πλοήγηση κτλ. καθώς επίσης συστήματα υποστήριξης του χρήστη στις λειτουργίες αυτές (θησαυροί όρων κτλ.) Τέλος, κατατάσουμε εδώ λειτουργίες που παρέχονται από τη χρήση συστημάτων διαχείρισης βάσεων δεδομένων σε εφαρμογές SGML όπως έλεγχο ταυτόχρονης πρόσβασης, ανάνηψη πληροφορίας (recovery) κτλ.

6.1.5 Επεξεργασία

Στο στάδιο της επεξεργασίας κατατάσουμε τις λειτουργίες του ευρετηριασμού των εγγράφων, του σχολιασμού, της δημιουργίας εκδόσεων/διασκευών, της μόνιμης αποθήκευσης (archiving). Τα συστήματα διαχείρισης εγγράφων SGML μπορεί να υποστηρίζουν ή όχι λειτουργίες μόνιμης αποθήκευσης πολλαπλών εκδόσεων για κάθε έγγραφο και βέβαια ευρετηριασμό εγγράφων με όρους που προτείνει ο χρήστης ή δίνονται αυτόματα από το σύστημα.

6.1.6 Παραγωγή προϊόντων

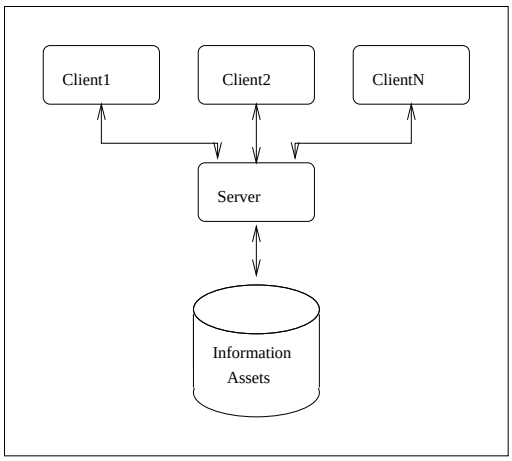
Το στάδιο παραγωγής προϊόντων συγκεντρώνει τις λειτουργίες μορφοποίησης και ηλεκτρονικής παρουσίασης εγγράφων, δημιουργίας παρουσιάσεων από αυτά, δημιουργία οπτικών δίσκων (CD-ROMs), παραγωγή εγγράφων από ήδη αποθηκευμένα έγγραφα κτλ.

6.2 Δομικά τμήματα της αρχιτεκτονικής του συστήματος Κώδιξ

Στο υποκεφάλαιο αυτό παρουσιάζεται η αρχιτεκτονική που υιοθετήθηκε για το σχεδιασμό και την υλοποίηση του συστήματος Κώδιξ. Στην παρούσα έκδοση του συστήματος δεν είναι διαθέσιμο το τμήμα SGML της αρχιτεκτονικής του συστήματος που διευκολύνει τη δημιουργία και παρουσίαση των βασικών εγγράφων SGML. Αντι αυτού χρησιμοποιείται ένας απλός κειμενογράφος (vi). Επίσης, σε μεταγενέστερη έκδοση του συστήματος θα υποστηρίζεται η αυτόματη εισαγωγή της δομής των εγγράφων και των DTDs που διαχειρίζεται το σύστημα (βλέπε παράγραφο 6.3.8)

6.2.1 Αρχιτεκτονική client-server

Η αρχιτεκτονική του συστήματος ακολουθεί το μοντέλο πελάτη - εξυπηρετητή (client - server). Ο τελικός χρήστης επισκέπτεται το σύστημα για να διαχειρισθεί και επεξεργαστεί τα έγγραφα. Το σύστημα ανταποκρίνεται ικανοποιώντας τις απαιτήσεις που δέχεται.
Οπως βλέπουμε και στο παρακάτω σχήμα υπάρχει ένας μόνο εξυπηρετητής **Server** που αναλαμβάνει να ικανοποιήσει τις απαιτήσεις που δρομολογεί ένας ή περισσότεροι πελάτες **Client1, Client2, ..., ClientN**. Ο εξυπηρετητής επικοινωνεί με την Βάση Πληροφοριών **Information Base** για να αντλήσει την πληροφορία που του χρειάζεται.
Η επιλογή της αρχιτεκτονικής αυτής κάνει εφικτή την εξυπηρέτηση απαιτήσεων πολλών χρηστών ταυτόχρονα.



Σχήμα 6.2: Μοντέλο Πελάτη - Εξυπηρετητή.

Περιγραφή του πελάτη

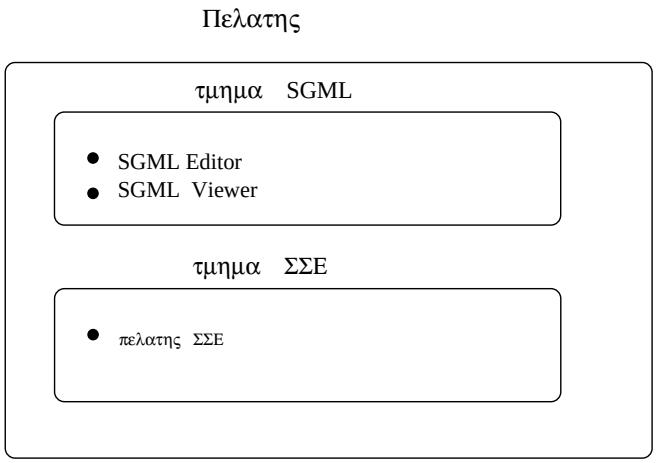
Ο πελάτης αποτελείται από δύο τμήματα, το **τμήμα SGML** και το **ΣΣΕ τμήμα**.

- Το τμήμα **SGML** φιλοξενεί εργαλεία της SGML από τα οποία ξεκινούν οι λειτουργίες της δημιουργίας και τροποποίησης εγγράφων SGML. Τα εργαλεία αυτά είναι τα εξής:

  -- Ενας SGML Editor.

  -- Ενας SGML Viewer.

- Το **ΣΣΕ τμήμα** φιλοξενεί τον ΣΣΕ πελάτη (SIS Client). Από το γραφικό περιβάλλον αυτού γίνεται η λειτουργία της αναζήτησης και ανάκτησης των εγγράφων. Επίσης από το ΣΣΕ τμήμα γίνεται ο ευρετηριασμός, ο σχολιασμός και η καταγραφή των εκδόσεων των εγγράφων. Τέλος, από το τμήμα αυτό γίνεται η εισαγωγή της δομής των εγγράφων και των DTD που διαχειρίζεται το σύστημα.



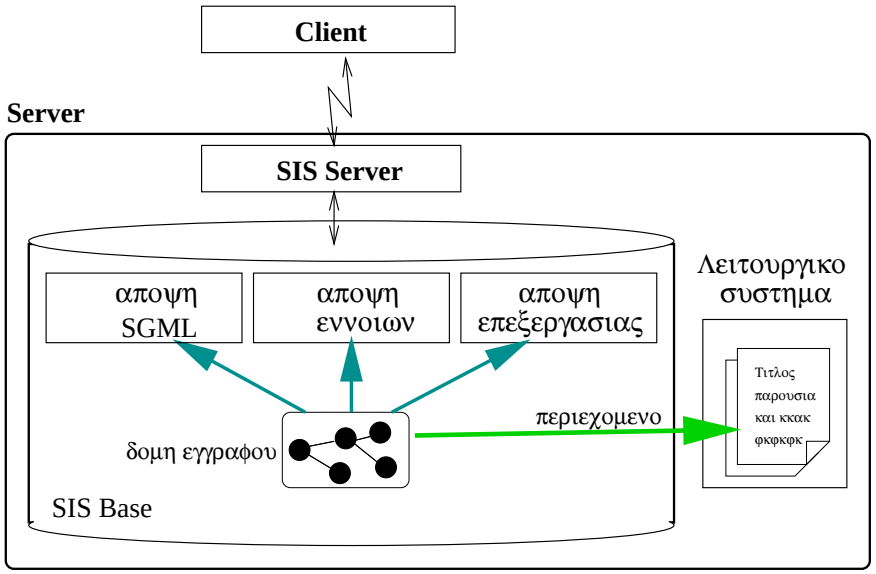
Σχήμα 6.3: Περιγραφή του Πελάτη.

Περιγραφή του εξυπηρετητή

Ο εξυπηρετητής του συστήματος αποτελείται από δύο βασικά τμήματα τα οποία αναλαμβάνουν την ικανοποίηση των απαιτήσεων του πελάτη του συστήματος:

- Το τμήμα που περιέχει τον **ΣΣΕ Εξυπηρετητή**
  Στο τμήμα αυτό δρομολογούνται οι απαιτήσεις του πελάτη που αναφέρονται σε πληροφορίες της βάσης γνώσης του συστήματος. Ο ΣΣΕ εξυπηρετητής παρέχει πρόσβαση παρέχει πρόσβαση στη βάση γνώσης του συστήματος.

- Το τμήμα που περιέχει τον **Εξυπηρετητή αρχείων** του συστήματος.
  Στο τμήμα αυτό δρομολογούνται απαιτήσεις πρόσβασης σε αρχεία του συστήματος όπως εικόνες, EAA, αρχεία SGML κτλ.



Σχήμα 6.4: Περιγραφή του εξυπηρετητή του συστήματος.

6.3 Κύκλος ζωής εγγράφου στο σύστημα Κώδιξ

Παρακάτω περιγράφουμε αναλυτικά τις λειτουργίες διαχείρισης εγγράφων SGML που υποστηρίζει το σύστημα Κώδιξ σ' όλο τον κύκλο ζωής τους στο σύστημα.

6.3.1 Συλλογή εγγράφων

Το πρώτο στάδιο του KZ αφορά στη διαδικασία της συλλογής των εγγράφων του συστήματος. Ορισμένα από τα έγγραφα αυτά είναι διαθέσιμα με τη μορφή φωτοτυπίας ενώ σε άλλα έχει προηγηθεί πληκτρολόγηση του περιεχομένου τους και βρίσκονται σε ηλεκτρονική μορφή.

6.3.2 Προ-επεξεργασία

Στο στάδιο αυτό επιλέγονται οι τύποι εγγράφων(DTDs) που θα υιοθετηθούν για την εφαρμογή του συστήματος. Το σύστημα είναι ικανό για τη διαχείριση οποιουδήποτε τύπου εγγράφου. Το πρόθεμα του ονόματος του αρχείου στο οποίο περιγράφεται το DTD

επιλέγεται κατάλληλα ώστε να δηλώνει τον τύπο εγγράφου που περιγράφει. Ακολουθείται η σύμβαση η επέκταση του ονόματος του αρχείου να είναι **.dtd** π.χ. η ονομασία **book.dtd** να αναφέρεται στο αρχείο που περιγράφει το DTD ενός βιβλίου.
Στο παρόν σύστημα η δημιουργία του αρχείου αυτού γίνεται με τη χρήση ενός απλού κειμενογράφου (πχ. vi, Word) όμως δεν αποκλείεται στο μέλλον η χρήση ενός ειδικού εργαλείου με το οποίο γίνεται πιό εύκολη και ευχάριστη η δημιουργία του DTD (DTD Builder).
Το στάδιο αυτό αποτελείται επίσης από τη λειτουργία εισαγωγής των DTD των εγγράφων στο σύστημα. Η αποθήκευση των τύπων εγγράφων γίνεται σε δύο βήματα:

- **1ο βήμα:** Εισαγωγή του DTD του ΒΕ στον εξυπηρετητή μεγάλων αντικειμένων του συστήματος. Το αρχείο που περιγράφει το DTD του εγγράφου, αφού δημιουργηθεί στο περιβάλλον εκτέλεσης του πελάτη, αποθηκεύεται στο σύστημα αρχείων (file system) του συστήματος. Στο μέλον προβλέπεται η χρήση ειδικών εξυπηρετητών μεγάλων αντικειμένων οι οποιοι θα αναλλαμβάνουν την αποδοτική και ασφαλή αποθήκευση των αρχείων αυτών

- **2ο βήμα:** Παράσταση του τύπου εγγράφου στη βάση γνώσης του ΣΣΕ.

  Η παράσταση της δομής του εγγράφου ,όπως περιγράφεται στο DTD, γίνεται στην άποψη SGML του σημασιολογικού δικτύου. Δημιουργούνται δηλαδή οι κατάλληλες οντότητες και σχέσεις μεταξύ τους που παριστάνουν τη δομή των εγγράφων του συστήματος. Οι οντότητες αυτές αποτελούν περιπτώσεις των οντοτήτων και σχέσεων του γενικού μοντέλου του συστήματος. Η λειτουργία αυτή δρομολογείται από το περιβάλλον του πελάτη (SIS client) με αλληλεπίδραση του χρήστη με το σύστημα (βλέπε παρ. 6.3.9).

6.3.3 Δημιουργία Βασικών Εγγράφων SGML

Το τρίτο στάδιο του ΚΖ των εγγράφων του συστήματος περιλαμβάνει τη δημιουργία και τροποποίηση των βασικών εγγράφων SGML, ΒΕ. Τα έγγραφα αυτά δημιουργούνται με τη βοήθεια ενός κειμενογράφου (vi). Σε επόμενη έκδοση του συστήματος η δημιουργία των ΒΕ θα γίνεται με χρήση ενός SGML Editor.

6.3.4 Διαχείριση βασικών Εγγράφων

Η διαχείριση ΒΕ περιλαμβάνει την αποδοτική αποθήκευση , την αναζήτηση και ανάκτησή τους.

Αποθήκευση Βασικών εγγράφων

Η αποθήκευση των ΒΕ γίνεται με τα εξής βήματα:

- **Παράσταση της δομής των ΒΕ** στην άποψη SGML του σημασιολογικού δικτύου.
  Η παράσταση της δομής των ΒΕ στη ΒΓ του ΣΣΕ γίνεται με τη δημιουργία των κατάλληλων οντοτήτων και σχέσεων στην άποψη SGML με αλληλεπίδραση του χρήστη με το σύστημα (βλέπε παρ. 6.3.9). Οι οντότητες και οι σχέσεις αυτές αποτελούν περιπτώσεις των (αντίστοιχων) οντοτήτων και σχέσεων που παριστάνουν το DTD ενός βασικού εγγράφου (και βρίσκονται σε στάθμη Token).

- **Αποθήκευση του αρχείου** που περιέχει το ΒΕ στο σύστημα αρχείων του συστήματος. Τηρείται η σύμβαση τα βασικά έγγραφα να αποθηκεύονται σε αρχεία των οποίων το όνομα αποτελείται από δύο τμήματα.

  -- Το **πρώτο τμήμα** αποτελείται από ένα πρόθεμα που είναι το ίδιο με το πρόθεμα του αρχείου που περιγράφει το DTD του βασικού εγγράφου. Η κατάληξη του πρώτου τμήματος είναι ένας αύξοντας αριθμός που δηλώνει τα εισαγόμενα ΒΕ που υπακούουν στο ίδιο DTD.

  -- Το **δεύτερο τμήμα** είναι το αλφαριθμητικό **.doc**
  πχ. το αρχείο **book1.doc** είναι ένα ΒΕ με τα περιεχόμενα ενός βιβλίου.

- Τα περιεχόμενα των εγγράφων που δεν είναι κείμενο (πχ. εικόνες) και αποτελούν ΕΑΑ αποθηκεύονται σε ειδικά αρχεία στο σύστημα αρχείων με το αναγνωριστικό συστήματος που τους αποδίδεται στο DTD.

Σε μεταγενέστερη έκδοση του συστήματος και αν το πλήθος των εγγράφων είναι μεγάλο ενδείκνυται η χρήση ειδικών εξυπηρετητών μεγάλων αντικειμένων αντί του συστήματος αρχείων για αποδοτική διαχείριση μεγάλων αντικειμένων (BLOBS).

Αναζήτηση - Ανάκτηση Βασικών Εγγράφων

Το σύστημα δίνει τη δυνατότητα αναζήτησης και ανάκτησης εγγράφων που έχουν αποθηκευτεί στον εξυπηρετητή εγγράφων. Η λειτουργία αυτή ενεργοποιείται από τον πελάτη (από το τμήμα του ΣΣΕ).
Το σύνολο απαντήσης (answer set) εμφανίζεται στο γραφικό περιβάλλον του ΣΣΕ (στο περιβάλλον που εκτελείται ο πελάτης) γραφικά ή με τη μορφή κειμένου. Η αναζήτηση μπορεί να γίνει με τους εξής τρόπους:

- Χρήση κάρτας αναζήτησης (retrieval card) για αναζήτηση εγγράφων βάσει πολλών παραμέτρων (αναλυτική περιγραφή της κάρτας αναζήτησης υπάρχει στην παρ. 6.4.3)

- Χρήση **προδηλωμένων ερωτήσεων** στο ΣΣΕ. (αναλυτική περιγραφή των ερωτήσεων υπάρχει στην παρ. 6.4)

- Χρήση του γραφικού αναλυτή του ΣΣΕ, gain για **πλοήγηση** σε όλο τον πληθυσμό του σημασιολογικού μοντέλου.

6.3.5 Επεξεργασία

Το στάδιο της επεξεργασίας περιλαμβάνει τον **ευρετηριασμό** (ομαδοποίηση), τον **σχολιασμό** και την **καταγραφή εκδόσεων** των ΒΕ.

Ευρετηριασμός-Ομαδοποίηση.

Ο ευρετηριασμός των εγγράφων (ή τμημάτων τους) μπορεί να γίνει βάσει άλλων εγγράφων (ή τμημάτων τους) που ανήκουν στην άποψη SGML ή βάσει οντοτήτων που ανήκουν στην άποψη Εννοιών.

- Ο ευρετηριασμός μεταξύ οντοτήτων της άποψης SGML μπορεί να γίνει με τη χρήση τριών κατηγοριών συνδέσμων. Η κατηγοριοποίηση των συνδέσμων επιλέχθηκε για την απόδοση σημασιολογίας στη συσχέτιση των οντοτήτων αυτών, κάτι που ασφαλώς βελτιώνει τους ποιοτικούς δείκτες αναζήτησης **??**. Οι κατηγορίες συνδέσμων που χρησιμοποιούνται για τον ευρετηριασμό μιας οντότητας με κάποιο όρο του ευρετηρίου που ανήκει επίσης στην άποψη SGML είναι οι εξής:

  -- *detailedDescription* : Συνδέει όρους του ευρετηρίου με οντότητες που παριστάνουν έγγραφα ή τμήματα εγγράφων που έχουν αναλυτική περιγραφή πάνω σε ένα θέμα (όσο αφορά στο περιεχόμενο).

  -- *seeAlso* : Συνδέει όρους του ευρετηρίου με οντότητες που παριστάνουν έγγραφα ή τμήματα εγγράφων που είναι σχετικά (όσο αφορά στο περιεχόμενο).

  -- *see* : Συνδέει όρους του ευρετηρίου με οντότητες που παριστάνουν έγγραφα ή τμήματα εγγράφων συναφή (ως προς το περιεχόμενό τους).

- Ο ευρετηριασμός οντοτήτων της άποψης SGML με οντότητες της άποψης Εννοιών γίνεται με τη χρήση συνδέσμων κατηγορίας *associatedWith*.

Η λειτουργία αυτή γίνεται με την καθοδήγηση του συστήματος από τον χρήστη-ειδικό στον συγκεκριμένο τομέα εφαρμογής. Οπως είναι φυσικό αυτή η λειτουργία δεν αυτοματοποιείται διότι εξαρτάται από την συγκεκριμένη εφαρμογή και τα έγγραφα που διαχειρίζεται το σύστημα.
Με τη λειτουργία αυτή συνδέονται οντότητες της άποψης SGML (έγγραφα ή τμήματά τους) με οντότητες της άποψης Εννοιών. Για τη βοήθεια και υποστήριξη του χρήστη στον ευρετηριασμό των εγγράφων προσαρμόστηκε κατάλληλα το γραφικό περιβάλλον του ΣΣΕ (στο περιβάλλον του πελάτη του συστήματος) ώστε να παρέχεται ένας θησαυρός όρων που αποτελείται από οντότητες της άποψης Εννοιών που χρησιμοποιούνται ως όροι του ευρετηρίου.

Κατά τη διαδικασία ευρετηριασμού ο χρήστης ακολουθεί την παρακάτω διαδικασία:

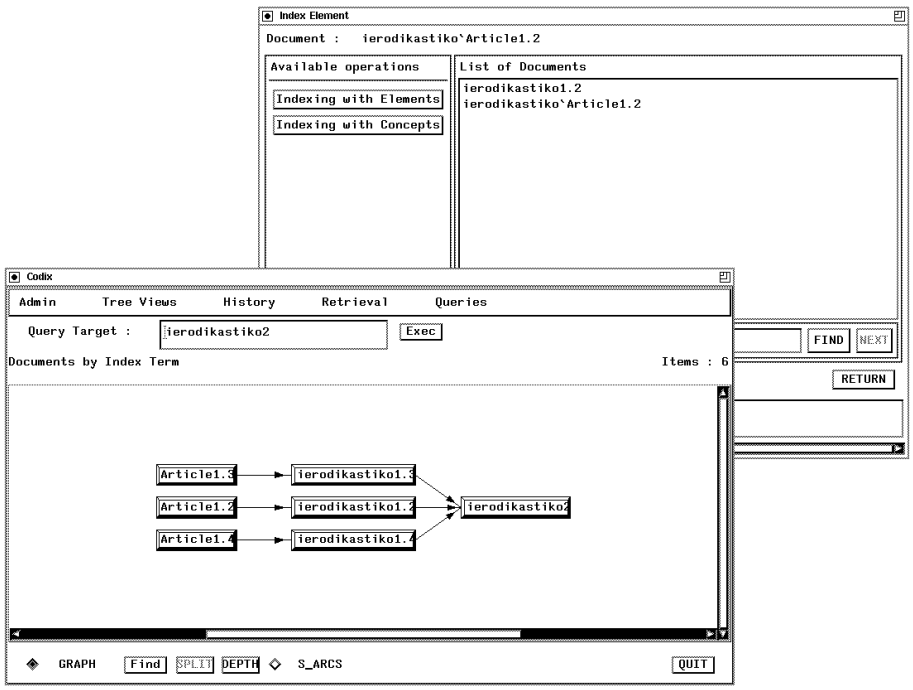
- Ο χρήστης επιλέγει το έγγραφο που πρόκειται να ευρετηριαστεί.
- Ο χρήστης επιλέγει το τμήμα του εγγράφου που τον ενδιαφέρει. Εαν πρόκειται να γίνει ευρετηριασμός ολόκληρου του εγγράφου επιλέγεται το βασικό τμήμα του εγγράφου.
  - -- Εαν πρόκειται να γίνει ευρετηριασμός βάσει κάποιου άλλου εγγράφου, εμφανίζονται οι δυνατοί τρόποι σύνδεσης (τρεις κατηγορίες γνωρισμάτων).
  - -- Εαν πρόκειται να γίνει ευρετηριασμός βασει κάποιας οντότητας της άποψης Εννοιών εμφανίζεται ως επιλέξιμη η επιτρεπτή κατηγορία γνωρίσματος (associatedWith).
- Ο χρήστης επιλέγει την κατηγορία του συνδέσμου που θα δημιουργηθεί και την οντότητα βάσει της οποίας γίνεται ο ευρετηριασμός.

Στο σχήμα 6.5 βλέπουμε ένα παράδειγμα ευρετηριασμού εγγράφων. Τα έγγραφα **Article1.2, Article1.3, Article1.4** είναι σχετικά με το **ierodikastiko2**.

Σχολιασμός Βασικών Εγγράφων.

Η λειτουργία αυτή δρομολογείται από το περιβάλλον εργασίας του πελάτη. Το σχόλιο (σε μορφή κειμένου) συμπληρώνεται μέσα σε ένα γραφικό περιβάλλον.

Η συστηματική μελέτη εγγράφων από ειδικούς συχνά έχει ως αποτέλεσμα την συγγραφή σχολίων που αφορούν στο περιεχόμενο (σημασιολογία), στη μορφή, στη διαχείρησή τους κτλ.
Τα σχόλια αυτά σχετίζονται με όλο το έγγραφο ή κάποιο τμήμα του. Για την παράσταση



Σχήμα 6.5: Παράδειγμα ευρετηριασμού σχετικών εγγράφων. Τα έγγραφα Article1.2, Article1.3, Article1.4 είναι σχετικά με το ierodikastiko2

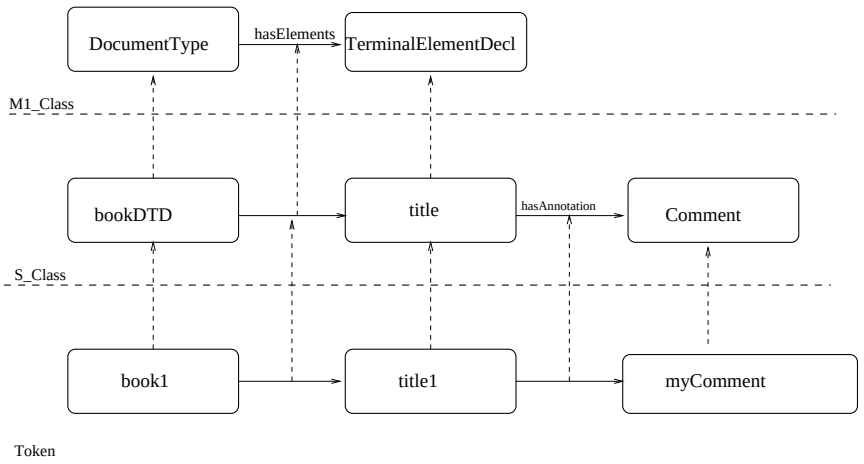
των σχολίων που δημιουργούνται για ένα BE (και την ταξινόμησή τους) δημιουργούνται κατάλληλες οντότητες στην άποψη της Επεξεργασίας. Οι οντότητες αυτές συνδέονται με σχέσεις (links) τύπου *hasAnnotation* με οντότητες της άποψης SGML που παριστάνουν τα σχολιασμένα έγγραφα (ή τμήματά τους). Εχει προηγηθεί σχετική εργασία [8] συστηματικής μελέτης του σχολιασμού των εγγράφων με σημασιολογική ταξινόμηση των σχόλιων που αποδίδονται σε αυτά.

Για το σχολιασμό των εγγράφων προσαρμόστηκαν κατάλληλα τα δελτία εισαγωγής δεδομένων. Η διαδικασία που ακολουθεί ο χρήστης για τον σχολιασμό ενός εγγράφου (ή τμήματός του) είναι η ακόλουθη:

- Ο χρήστης επιλέγει τη λειτουργία σχολιασμού από τον κατάλογο εργασιών των δελτίων εισαγωγής δεδομένων.
- Ο χρήστης επιλέγει το έγγραφο που πρόκειται να σχολιάσει.
- Ο χρήστης επιλέγει το τμήμα του εγγράφου που θα σχολιαστεί (για τον σχολιασμό ολόκληρου του εγγράφου επιλέγεται ο βασικός τύπος του εγγράφου).
- Ο χρήστης δημιουργεί ένα σύνδεσμο κατηγορίας *hasAnnotation* μεταξύ της επιλεχθείσας οντότητας της άποψης SGML και μιας οντότητας που παριστάνει το σχόλιο που

έχει ήδη δημιουργηθεί (με την επιλογή του πλήκτρου ”OLD”) ή που πρόκειται να δημιουργηθεί (με την επιλογή του πλήκτρου ”NEW”).

- Τα σχόλια (σε μορφή κειμένου) συμπληρώνονται μέσα σε ένα γραφικό περιβάλλον.
- Ο χρήστης επιβεβαιώνει τις παραπάνω επιλογές και η διαδικασία σχολιασμού τερματίζεται.



Σχήμα 6.6: Παράδειγμα σχολιασμού ενός τμήματος βασικού εγγράφου.

Ενα παράδειγμα της οντολογίας του μοντέλου για το σχολιασμό εγγράφου φαίνεται στο παρακάτω σχήμα. Το Βασικό Εγγραφο **book1** περιγράφεται από το **bookDTD**. Το έγγραφο αυτό έχει ένα τίτλο (title1) για το οποίο έχει δημιουργηθεί ένα σχόλιο. Το σχόλιο αυτό είναι τύπου **hasAnnotation** και έχει παρασταθεί από την οντότητα **myComment**.

Καταγραφή Εκδόσεων Βασικών Εγγράφων.

Η λειτουργία αυτή επιτρέπει την ταξινόμηση εγγράφων κάτω από κλάσεις οντοτήτων που παριστάνουν εκδόσεις εγγράφων. Υποθέτουμε εδώ ότι η δομή του εγγράφου έχει ήδη παρασταθεί στη βάση γνώσης του συστήματος Κώδιξ (με τη λειτουργία εισαγωγής ΒΕ).
Η λειτουργία αυτή εκτελείται στο σύστημα Κώδιξ με την χρήση των δελτίων εισαγωγής δεδομένων. Η διαδικασία που ακολουθείται είναι η εξής:

- Ο χρήστης επιλέγει την λειτουργία που θέλει να εκτελέσει(καταγραφή έκδοσης).
- Ο χρήστης επιλέγει το έγγραφο που τον ενδιαφέρει.

  -- Καταγραφή νέας παραλλαγής

  Ο χρήστης επιλέγει το έγγραφο του οποίου αποτελεί παραλλαγή το αρχικά

επιλεχθέν έγγραφο. Κατόπιν δημιουργείται σύνδεσμος τύπου *variantOf* μεταξύ των δύο εγγράφων.

-- Καταγραφή νέας διασκευής
Ο χρήστης επιλέγει το έγγραφο του οποίου αποτελεί διασκευή το αρχικά επιλεχθέν έγγραφο. Κατόπιν δημιουργείται σύνδεσμος τύπου *revisionOf* μεταξύ των δύο εγγράφων.

- Ο χρήστης επιβεβαιώνει την επιθυμία του για εκτέλεση της παραπάνω διαδικασίας.

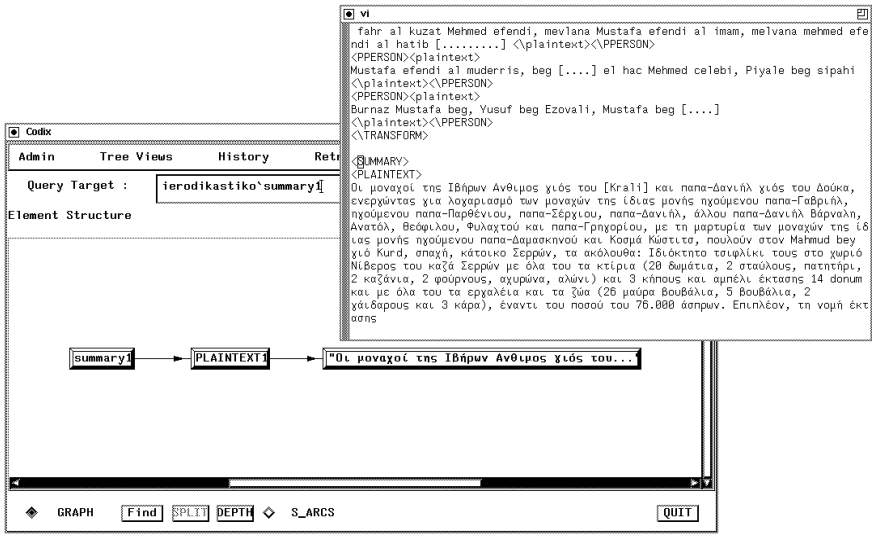
Απόδοση ονόματος στα αρχεία που περιέχουν διασκευές ή παραλλαγές.
Τηρείται η σύμβαση το όνομα του αρχείου στο οποίο αποθηκεύονται διασκευές ή παραλλαγές να αποτελείται από τρία τμήματα. Το **πρώτο τμήμα** είναι ίδιο με το όνομα του βασικού εγγράφου. Το **δεύτερο τμήμα** είναι ένας αριθμός που δηλώνει τη διασκευή του εγγράφου. Το **τρίτο τμήμα** δηλώνει την παραλλαγή του εγγράφου. Οταν πρόκειται για νέα διασκευή το τρίτο τμήμα έχει τιμή εξ ορισμού 0.

Απόδοση ονόματος στις οντότητες που παριστάνουν τη δομή εκδόσεων.
Το όνομα κάθε οντότητας αποτελείται από τρία μέρη. Το πρώτο μέρος του ονόματος κάθε οντότητας καθορίζεται με τον ίδιο τρόπο που καθορίζονται τα ονόματα των οντοτήτων που παριστάνουν τη δομή ενός BE(βλέπε κεφ.6). Το δεύτερο και τρίτο τμήμα του ονόματος κάθε οντότητας ορίζεται με τον ίδιο τρόπο όπως ορίζονται τα αντίστοιχα τμήματα στην ονοματολογία των αρχείων που αποθηκεύονται οι διασκευές ή οι παραλλαγές

6.3.6 Παραγωγή προϊόντων

Η παραγωγή προϊόντων αποτελεί το τελευταίο στάδιο του KZ των εγγράφων SGML του συστήματος Κώδιξ.
Η λειτουργία που επιτελείται στο στάδιο αυτό είναι η *παρουσίαση* ενός BE και ενεργοποιείται από το γραφικό περιβάλλον του ΣΣΕ που εκτελείται στο περιβάλλον του πελάτη (βλέπε παρ. 6.3.9). Αποτέλεσμα της επιλογής αυτής είναι η εμφάνιση του περιεχομένου ενός BE σε γραφικό περιβάλλον. Στο εμφανιζόμενο έγγραφο δεν επιτρέπεται τροποποίηση των περιεχομένων του (read only). Για την εκτέλεση της λειτουργίας αυτής χρησιμοποιείται ένας απλός κειμενογράφος (vi) (βλέπε εικόνα 6.7).
Σε μεταγενέστερη έκδοση του συστήματος ενδείκνυται η χρήση ειδικών εργαλείων SGML (SGML Viewer) για τη μορφοποίηση των παρουσιαζόμενων βασικών εγγράφων. Για

τη μορφοποίηση αυτή θα γίνεται χρήση της πληροφορίας που έχει παρασταθεί στο Σημασιολογικό Μοντέλο (στην άποψη Επεξεργασίας).



Σχήμα 6.7: Με τη χρήση του κειμενογράφου (vi) εμφανίζονται τα περιεχόμενα του τμήματος *summary1*

6.3.7 Τεχνικά χαρακτηριστικά του συστήματος Κώδιξ

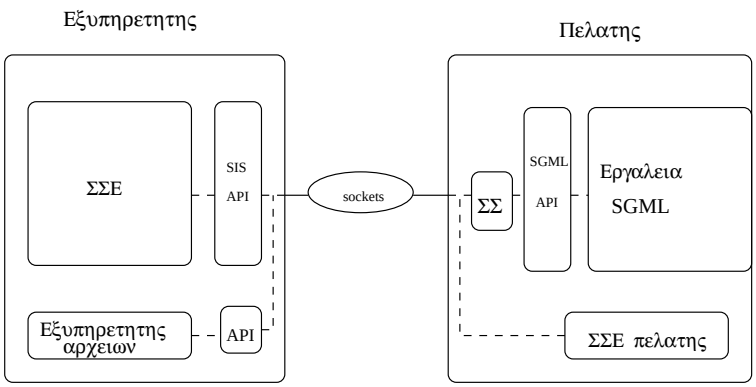
Στο παρακάτω σχήμα βλέπουμε τη γενική περιγραφή της αρχιτεκτονικής του συστήματος.

- Το περιβάλλον του **Εξυπηρετητή** εκτελείται σε σταθμό εγασίας **SUN** με λειτουργικό σύστημα **SUN OS.2** . Στο περιβάλλον αυτό υπάρχει **ο εξυπηρετητής του ΣΣΕ**. Επίσης, στο ίδιο περιβάλλον βρίσκεται και ο **εξυπηρετητής αρχείων** (file system) που αποθηκεύει τα αρχεία SGML, τις εικόνες του συστήματος τα ΕΑΑ που φυλλάσονται σε αρχεία κοκ. Τελος, στο ίδιο περιβάλλον υπάρχει η **προγραμματική διεπαφή εφαρμογής**, API του ΣΣΕ που παρεχει συναρτήσεις πρόσβασης στη βάση γνωσης του ΣΣΕ σε εξωτερικές εφαρμογές.

- Το περιβάλλον εκτέλεσης του **Πελάτη** εκτελείται επίσης σε σταθμό εργασίας **SUN**. Επίσης, προβλέπεται μεταγενέστερη υλοποίηση του πελάτη και σε προσωπικό υπολογιστή υπό το λειτουργικό σύστημα **Windows95**. Ο πελάτης αποτελείται από τον **πελάτη του ΣΣΕ**.

6.3.8 Επεκτάσεις του συστήματος Κώδιξ

Η τωρινή υλοποίηση του συστήματος Κωδιξ χρησιμοποιεί το σύστημα αρχείων για την αποθήκευση των αρχείων που περιέχουν το DTD και τα περιεχόμενα των εγγράφων. Προβλέπεται σε μια νέα έκδοση του συστήματος η χρήση ειδικών εξυπηρετητών μεγάλων αντικειμένων για αποδοτικότερη και ασφαλέστερη διαχείριση των αρχείων του συστήματος. Επίσης, η εισαγωγή της δομής των εγγράφων και των DTDs στη βάση γνώσης του συστήματος μπορεί να γίνει με τους κλασικούς τρόπους εισαγωγής δεδομένων στο ΣΣΕ [30] δηλ. με τη χρήση των δελτίων ενημέρωσης (αλληλεπίδραση χρήστη με το σύστημα) ή με τη χρήση του συντακτικού αναλυτή του ΣΣΕ. Ωστόσο, μια προβλεπόμενη επέκταση του Κώδιξ (στα πλαίσια διπλωματικής εργασίας [**?**]) αποβλέπει στην σύνδεση του ΣΣΕ με έναν SGML editor και την αυτόματη ενημέρωση της άποψης SGML της Βάσης γνώσης του ΣΣΕ σχετικά με τη δομή των εγγράφων και των DTDs που διαχειρίζεται το σύστημα.
Η προβλεπόμενη αρχιτεκτονική του συστήματος περιγράφεται στο σχήμα 6.8.



Σχήμα 6.8: Επέκταση του συστήματος Κώδιξ : Δομικά τμήματα της αρχιτεκτονικής του συστήματος (τεχνική περιγραφή).

Στην προβλεπόμενη αρχιτεκτονική του συστήματος, το τμήμα ΣΣ θα είναι υπεύθυνο για την αυτόματη εισαγωγή των DTDs και των εγγράφων στη ΒΓ του ΣΣΕ. Η επικοινωνία του πελάτη και του εξυπηρετητή γίνεται με sockets. Οι παρεχόμενες λειτουργίες του Κώδικα τροποποιούνται ως εξής:

- **Εισαγωγή εγγράφων και τύπων εγγράφων**
  Μετά τη δημιουργία των εγγράφων και των τύπων τους (στο περιβάλλον του πελάτη με τη χρήση εργαλείων SGML) το Σύστημα Σύνδεσης (με τη χρήση των συναρτήσεων του SGML API) δημιουργεί ένα αρχείο εντολών στη γλώσσα SIS-Telos για τη δημιουργία των κατάλληλων οντοτήτων και σχέσεων της άποψης SGML. Επίσης, τα αρχεία που

περιέχουν τα έγγραφα και τα DTDs μεταφέρονται (μέσω socets) στον εξυπηρετητή μεγάλων αντικειμένων.

- **Αυτόματος ευρετηριασμός.**
  Εκτός από τη λειτουργία του ευρετηριασμού που απαιτεί αλληλεπίδραση του χρήστη-ειδικού με το σύστημα θα παρέχεται η δυνατότητα αυτόματου ευρετηριασμού των εγγράφων. Οι τύποι τμημάτων που θα μπορούν να ευρετηριάζονται αυτόματα, θα πρέπει (στην περιγραφή του DTD τους) να τους αποδίδεται το χαρακτηριστικό *Term*. Η τιμή του χαρακτηριστικού αυτού θα παριστάνει την οντότητα (όρο ευρετηρίου) με την οποία ευρετηριάζεται το συγκεκριμένο τμήμα. Στην περίπτωση αυτή το σύστημα (αυτόματα) θα δημιουργεί σχέσεις (links) τύπου **associatedWith** μεταξύ των τμημάτων αυτών και των όρων ευρετηρίου. Αν η οντότητα που παριστάνει τον όρο του ευρετηρίου προϋπάρχει στη ΒΓ τότε δημιουργείται μόνο ο σύνδεσμος. Διαφορετικά, τόσο η σχέση όσο και η οντότητα δημιουργούνται και μάλιστα η νέα οντότητα ταξινομείται κάτω από την κλάση *Term*.

Περιγραφή της άποψης επεξεργασίας σε Τύπο εγγράφου

Παρακάτω δίνουμε μια σύντομη περιγραφή ενός τύπου εγγράφου που σχεδιάστηκε για τον ορισμό των λειτουργιών του συστήματος Κώδιξ σύμφωνα με το συντακτικό της SGML. Αυτός ο τύπος εγγράφου μπορεί να χρησιμοποιηθεί από την λειτουργία αυτόματης εισαγωγή DTDs στο Σύστημα Σημασιολογικού ευρετηριασμού για την αυτόματη παράσταση της άποψης εφαρμογής. Ο τύπος εγγράφου αποτελείται από τους παρακάτω τύπους τμημάτων:

- *document*: Περιέχει το όνομα του εγγράφου στο οποίο αναφέρονται οι πληροφορίες επεξεργασίας .
- *PresentationRule*: Περιέχει χαρακτηριστικά που αποδίδονται σε τμήματα εγγράφων και αφορούν στην παρουσίασή τους. Αποτελείται από δύο στοιχεία
  - -- *elem*Το τμήμα του εγγράφου στο οποίο αποδίδεται ένα χαρακτηριστικό.
  - -- *presentationAtt* Το χαρακτηριστικό που αποδίδεται. Αποτελείται από δύο υποτμήματα α) το όνομά του *name*, την τιμή του *value*.
- *IndexRule*: Περιέχει πληροφορία για τον ευρετηριασμό του εγγράφου (ή των τμημάτων του). Αποτελείται από δύο στοιχεία

-- *elem* Το τμήμα του εγγράφου που ευρετηριάζεται

-- *index* Το τμήμα αυτό δηλώνει τον όρο ευρετηρίου (με τον οποίο ευρετηριάζεται το τμήμα *elem* και αποδίδεται σε αυτό το χαρακτηριστικό (attribute) *type* που είναι ο τύπος ευρετηριασμού και παίρνει μια από τις τιμές *associatedWith, see, seeAlso, detailedDescription*

- *AnnotationRule*: Περιέχει πληροφορία για το σχολιασμό ενός τμήματος και αποτελείται από τα στοιχεία

  -- *elem* Το σχολιαζόμενο τμήμα

  -- *annotation* Το σχόλιο

- *VersioningRule*: πρόκειται για τελικό τμήμα που περιέχει το όνομα του εγγράφου που είναι προγενέστερο.

Στο παράρτημα Β υπάρχει ο τύπος εγγράφου σύμφωνα με το συντακτικό της SGML.

6.3.9 Χρήση Εξωτερικών Εργαλείων

Για την υποστήριξη ορισμένων λειτουργιών του συστήματος Κώδιξ χρησιμοποιήθηκαν εξωτερικές εφαρμογές οι οποίες καλούνται από το γραφικό περιβάλλον του ΣΣΕ. Συγκεκριμένα η κλήση των εφαρμογών αυτών γίνεται από τον κατάλογο επιλογών *Admin* της λίστας επιλογών (menu) του gain. Αναλυτικά ο κατάλογος επιλογών της επιλογής *Admin* αποτελείται (βλέπε παρακάτω σχήμα) από τις εξής επιλογές:

- *Global View*: Εμφανίζει σε γραφικό περιβάλλον το σύνολο της πληροφορίας της ΒΓ που αποτελεί απάντηση σε μια ερώτηση του τελικού χρήστη. Στο περιβάλλον αυτό περιορίζεται με παραλληλόγραμμο (μετακινούμενο) πλαίσιο το τμήμα της πληροφορίας που είναι εκείνη τη στιγμή ορατό από το κύριο παράθυρο του gain [28]
  .

- Η δεύτερη και τρίτη επιλογή του καταλόγου επιλογών *Admin* αναφέρεται σε λειτουργίες που εκτελούνται με τη χρήση του εργαλείου εισαγωγής δεδομένων *Entry Forms*. Για την ασφαλέστερη χρήση του συστήματος οι λειτουργίες αυτές ταξινομήθηκαν σε δύο κατηγορίες. Στην πρώτη κατηγορία ανήκουν λειτουργίες που εκτελούνται από χρήστες-ειδικούς οι οποίοι μπορούν να διαχειρίζονται όλη την πληροφορία της βάσης γνώσης του συστήματος ενώ στη δεύτερη κατηγορία ανήκει ένα υποσύνολο λειτουργιών της πρώτης και αναφέρεται σε απλούς χρήστες του συστήματος. Αναλυτικότερα οι κατηγορίες λειτουργιών περιγράφονται παρακάτω:

-- *ADMIN: Entry Forms* : Καλείται το εργαλείο Δελτίων Εισαγωγής Δεδομένων για λειτουργίες που επιτρέπεται να εκτελούνται απο τον διαχειριστή (ειδικό) του συστήματος. Οι επιτρεπτές λειτουργίες είναι *Δημιουργία DTD, Τροποποίηση DTD, Δημιουργία εγγράφου, Τροποποίηση εγγράφου, Σχολιασμός, Ευρετηριασμός, Καταγραφή Εκδόσεων, Γενικές τροποποιήσεις*. Στην παράγραφο 6.3.9 περιγράφουμε αναλυτικά το γραφικό περιβάλλον των δελτίων εισαγωγής δεδομένων του συστήματος Κώδιξ.

-- *USER: Entry Forms* : Καλείται το εργαλείο Δελτίων Εισαγωγής Δεδομένων για λειτουργίες που επιτρέπεται να εκτελούνται απο τον τελικό χρήστη (μη ειδικό) του συστήματος. Οι επιτρεπτές λειτουργίες είναι *Δημιουργία εγγράφου, Τροποποίηση εγγράφου, Σχολιασμός, Ευρετηριασμός, Καταγραφή Εκδόσεων*. Στην παράγραφο 6.3.9 γίνεται αναλυτική περιγραφή χρήσης των δελτίων ενημέρωσης για την εκτέλεση των λειτουργιών αυτών.

- *Document Presentation* : Εμφανίζει τα περιεχόμενα ενός βασικού εγγράφου ή ενός τμήματος του.

Χρήση Δελτίων Εισαγωγής Δεδομένων

Τα δελτία εισαγωγής δεδομένων εκτελούν τις παρακάτω εργασίες οι οποίες εμφανίζονται στον κατάλογο επιλογών *TaskList*:

- *Create DTD* : Με την επιλογή αυτή ο χρήστης δημιουργεί ένα νέο τύπο εγγράφου. Η επιλογή αυτή παραπέμπει σε μια λίστα πράξεων:

  -- *Create*: Δημιουργία νέου τύπου εγγράφου.

  -- *Delete*: Διαγραφή τύπου εγγράφου.

  -- *Rename*: Μετονομασία τύπου εγγράφου.

- *Edit DTD* : Με την επιλογή αυτή ο χρήστης τροποποιεί ένα τύπο εγγράφου. Αφού ο χρήστης επιλέξει τον τύπο εγγράφου που τον ενδιαφέρει, εμφανίζονται οι τύποι τμήματός του στους οποίους μπορούν να εφαρμοστούν οι παρακάτω πράξεις:

  -- *Delete*: Διαγραφή τύπου τμήματος.

  -- *Rename* : Μετονομασία τύπου τμήματος.

  -- *Edit Classes* : Τροποποίηση ταξινόμησης τύπου τμήματος.

  -- *Edit Attributes* : Τροποποίηση χαρακτηριστικών τύπου τμήματος.

- *Create Document* : Με την επιλογή αυτή ο χρήστης δημιουργεί ένα νέο έγγραφο. Ο χρήστης επιλέγει ένα DTD και εμφανίζονται τα έγγραφα του συστήματος που ακολουθούν αυτό το DTD. Οι επιτρεπτές πράξεις είναι οι εξής:

  -- *Create*: Δημιουργία εγγράφου.

  -- *Delete*: Διαγραφή εγγράφου.

  -- *Rename*: Μετονομασία εγγράφου.

- *Edit Document* : Με την επιλογή αυτή ο χρήστης τροποποιεί ένα έγγραφο. Αφού ο χρήστης επιλέξει το έγγραφο που τον ενδιαφέρει, εμφανίζονται τα τμήματά του στα οποία μπορούν να εφαρμοστούν οι παρακάτω πράξεις:

  -- *Delete*: Διαγραφή τμήματος.

  -- *Rename* : Μετονομασία τμήματος.

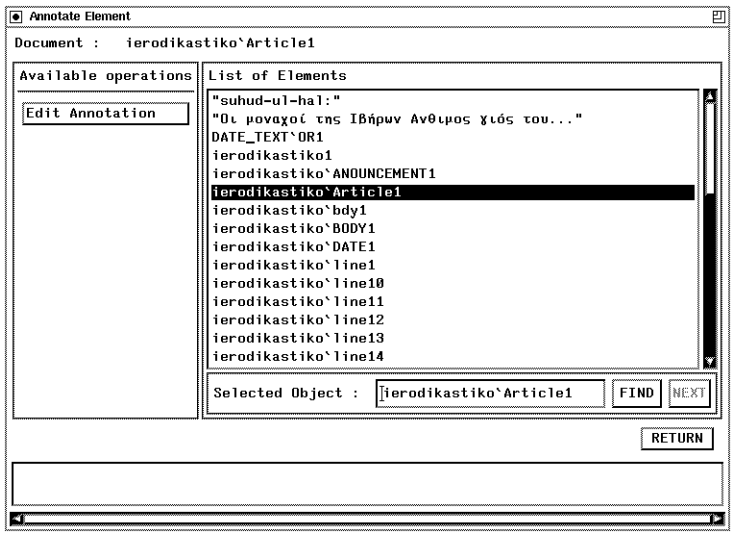
  -- *Edit Classes* : Τροποποίηση ταξινόμησης τμήματος.

  -- *Edit Attributes* : Τροποποίηση χαρακτηριστικών τμήματος.

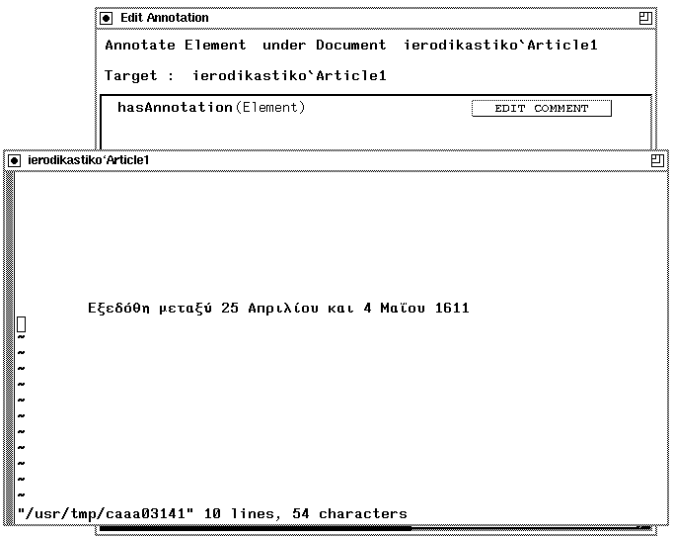
- *Annotate Element* : Απο την επιλογή αυτή ο χρήστης δημιουργεί σημειώσεις στα βασικά έγγραφα. Αρχικά (από την επιλογή *Document*) επιλέγει το έγγραφο που τον ενδιαφέρει και στη συνέχεια (από τη λίστα τμημάτων του εγγράφου που εμφανίζεται) επιλέγει το τμήμα στο οποίο θα γίνει η σημείωση. Στη συνέχεια (αφού επιλέξει την διαθέσιμη λειτουργία "Edit Annotation") εισάγει τα σχόλια στο γραφικό περιβάλλον που εμφανίζεται. Τέλος, επικυρώνει τη λειτουργία σχολιασμού με την επιλογή commit.

- *Index Element* : Απο την επιλογή αυτή ο χρήστης ευρετηριάζει βασικά έγγραφα. Αρχικά επιλέγει το έγγραφο που τον ενδιαφέρει και στη συνέχεια (από τη λίστα τμημάτων του εγγράφου που εμφανίζεται) επιλέγει το τμήμα το οποίο θα ευρετηριάσει. Στη συνέχεια επιλέγει μία από τις δύο διαθέσιμες πράξεις *Indexing with Elements* ή *Indexing with Concepts* και παραπέμπεται σε έναν κατάλογο τύπων συνδέσμων (attributes) ευρετηριασμού. Απο τον κατάλογο αυτό επιλέγει τους συνδέσμους που επιθυμεί να δημιουργήσει. Τέλος, επικυρώνει τη λειτουργία ευρετηριασμού (commit). Στα σχήματα 6.9 και 6.10 βλέπουμε την προσαρμογή των δελτίων ενημέρωσης για τη λειουργία του σχολιασμού.

- *Describe Version* : Απο την επιλογή αυτή ο χρήστης επιλέγει αρχικά το DTD που επιθυμεί και κατόπιν παραπέμπεται σε έναν κατάλογο εγγράφων που ακολουθούν



Σχήμα 6.9: 1ο βήμα: Από τη λίστα των προσφερόμενων τμημάτων του εγγράφου ο χρήστης επιλέγει το τμήμα που θα σχολιάσει



Σχήμα 6.10: 2ο βήμα: Με την επιλογή *EDIT COMMENT* εμφανίζεται ένα παράθυρο στο οποίο ο σχολιαστής συμπληρώνει το σχόλιο σε μορφή κειμένου με τη χρήση του κειμενογράφου (vi)

αυτό το DTD. Από τον καταλογο αυτό καλείται να επιλέξει το έγγραφο που θα χαρακτηρίσει ως έκδοση και θα το συνδέσει με το έγγραφο του οποίου είναι μεταγενέστερη έκδοση. Οι δυνατές πράξεις είναι *Document Revision* και *Document Variant* για τη καταγραφή αντίστοιχα διασκευών ή παραλλαγών εγγράφων.

- *General Updates* : Απο την επιλογή αυτή ο χρήστης μπορεί να επέμβει στη βάση γνώσης του συστήματος και να εκτελέσει τις εξής πράξεις για οποιαδήποτε οντότητα

  -- *Create*: Δημιουργία νέας οντότητας.

  -- *Delete*: Διαγραφή οντότητας.

  -- *Rename*: Μετονομασία οντότητας.

  -- *Edit Classes*: Τροποποίηση ταξινόμησης.

  -- *Edit Superclasses*: Τροποποίηση σχέσης εξειδίκευσης.

  -- *Edit Attributes*: Τροποποίηση γνωρισμάτων.

Στο παρακάτω σχήμα βλέπουμε τον αρχικό κατάλογο εργασιών των δελτίων εισαγωγής δεδομένων και την ακολουθία των ενεργειών για τον ευρετηριασμό ενός εγγράφου.

6.4 Παρουσίαση των Προκαθορισμένων Ερωτήσεων

Η αναζήτηση πληροφορίας στο σύστημα **Κώδιξ** γίνεται είτε με πλοήγηση του χρήστη στη βάση γνώσης του συστήματος από το περιβάλλον του πελάτη (λειτουργία που παρέχεται από το γραφικό περιβάλον του ΣΣΕ) είτε με τη χρήση προκαθορισμένων ερωτήσεων (και πάλι από το γραφικό περιβάλλον του ΣΣΕ) είτε με τη χρήση της κάρτας αναζήτησης που περιγράφεται στην παράγραφο 6.4.3.
Οι προκαθορισμένες ερωτήσεις εκτελούν αναζητήσεις πληροφορίας που πολύ συχνά έχει ανάγκη ο χρήστης του συστήματος. Επίσης, καλύπτουν αναζητήσεις για τις βασικές έννοιες του μοντέλου. Οι ερωτήσεις αυτές διακρίνονται σε **ερωτήσεις πρώτης τάξεως** και **αναδρομικές ερωτήσεις**.

6.4.1 Ερωτήσεις πρώτης τάξεως

Οι ερωτήσεις αυτές δεν είναι αναδρομικές και έχουν συνήθως αποτέλεσμα σε μορφή κειμένου. Οι ερωτήσεις αυτές εμφανίζονται με μορφή καταλόγου επιλογών από την επιλογή *Queries* του γραφικού περιβάλλοντος του ΣΣΕ. Κάθε επιλογή του καταλόγου μπορεί να είναι εκτελέσιμη εντολή (της οποίας το αποτέλεσμα εμφανίζεται στο γραφικό

περιβάλλον του ΣΣΕ με μορφή κειμένου) ή να παραπέμπει σε νέο κατάλογο επιλογών. Οι νέες αυτές επιλογές μπορεί με τη σειρά τους να είναι εκτελέσιμες εντολές ή να παραπέμπουν σε άλλο κατάλλογο επιλογών.

Οι δύο πρώτες επιλογές του αρχικού καταλόγου του συστήματος **Κώδιξ** αναφέρονται σε δύο κατηγορίες ερωτήσεων που είναι ιδιαίτερα χρήσιμες. Η πρώτη επιλογή που επιγράφεται *General Queries* παραπέμπει σε έναν κατάλογο εκτελέσιμων εντολών οι οποίες χρησιμοποιούνται πολύ συχνά κατά τη χρήση του συστήματος και ανακαλούν πληροφορία με τις βασικότερες σχέσεις που δίνει η γλώσσα SIS-telos. Η δεύτερη επιλογή που επιγράφεται *Structural Queries* αναφέρεται σε ερωτήσεις που αφορούν σε δομικά χαρακτηριστικά (λογική δομή και χαρακτηριστικά) των εγγράφων και των τμημάτων εγγράφων. Οι υπολοιπες επιλογές του αρχικού καταλόγου αφορούν τα βασικά έγγραφα και γενικές έννοιες του μοντέλου που σχετίζονται με αυτά. Οι έννοιες αυτές είναι τα βασικά έγγραφα, τα σχόλια, οι διασκευές, οι εκδόσεις, το ευρετήριο αυτών και ο θησαυρός σημασιολογικών όρων.

Παρακάτω δίνεται αναλυτική περιγραφή του αρχικού καταλόγου επιλογών καθώς επίσης των υποκαταλόγων και των εκτελέσιμων ερωτήσεων που περιέχονται σε αυτόν.

Η επιλογή "General Queries"

Η επιλογή *General Queries* είναι η πρώτη επιλογή του αρχικού καταλόγου. Παραπέμπει σε έναν κατάλογο επιλογών με τις βασικότερες σχέσεις που χαρακτηρίζουν τη γλώσσα SIS-telos και αφορούν στην εξειδίκευση (επιλογή *Special Kinds*), γενίκευση (επιλογή *General Kinds*) και ταξινόμηση (επιλογή *Object/Kind* ενός αντικειμένου. Αναλυτικότερα, οι παραπάνω επιλογές παραπέμπουν στις παρακάτω εκτελέσιμες εντολές:

- Επιλογή *Special Kinds*

  Οι ερωτήσεις στις οποίες παραπέμπει η παραπάνω επιλογή είναι οι εξής:

  -- *List All Subclasses*: Παραθέτει το μεταβατικό περίβλημα των υποκλάσεων μιας οντότητας ως προς τη σχέση εξειδίκευσης.

  -- *List Imediate Subclasses*:

  -- *List Objects of Kind* : Παραθέτει τις υποκλάσεις μιας οντότητας.

- Επιλογή *General Kinds*

  Οι ερωτήσεις στις οποίες παραπέμπει η παραπάνω επιλογή είναι οι εξής:

  -- *List All Superclasses*: Παραθέτει το μεταβατικό περίβλημα των υπερκλάσεων μιας οντότητας ως προς τη σχέση γενίκευσης.

-- *List Imediate Superclasses*: Παραθέτει τις υπερκλάσεις μιας οντότητας.

- Επιλογή *Object/Kind*
  Οι ερωτήσεις στις οποίες παραπέμπει η παραπάνω επιλογή είναι οι εξής:

  -- *List Objects of Kind*: Παραθέτει τις περιπτώσεις μιας οντότητας.

  -- *List Kinds of Object*: Παραθέτει τις κλάσεις των οποίων αποτελεί περίπτωση μια οντότητα.

Η επιλογή "Structural Queries"

Η επιλογή *Structural Queries* παραπέμπει σε έναν κατάλογο με τα βασικά στοιχεία που χρησιμοποιεί η SGML στη παράσταση της λογικής δομής των εγγράφων. Ο κατάλογος αυτός έχει τις εξής επιλογές:

- *Document Structure* : Η επιλογή αυτή παραπέμπει στον παρακάτω κατάλογο εκτελέσιμων εντολών :

  -- *List Kinds of Document Types* : Ανεύρεση των τύπων εγγράφων που υπάρχει στα δεδομένα του συστήματος.

  -- *List Elements by Document Type* : Ανεύρεση των τύπων τμημάτων ενός τύπου εγγράφου.

  -- *List Entities by Document Type* : Ανεύρεση των ΕΑΑ ενός τύπου εγγράφου.

  -- *Root Element by Document Type* : Ανεύρεση του βασικού τύπου τμήματος (Root Element Type) ενός τύπου εγγράφου.

  -- *Terminal Elements by Document Type* : Ανεύρεση των τελικών τύπων τμήματος(terminal elements) ενός τύπου εγγράφου.

  -- *Non Terminal Elements by Document Type* : Ανεύρεση των μη τελικών τύπων τμήματος ενός τύπου εγγράφου.

  -- *Empty Elements by Document Type* : Ανέυρεση των κενών τύπων τμήματος ενός τύπου εγγράφου.

  -- *List Objects by Document Type* : Ανεύρεση των περιπτώσεων ενός τύπου εγγράφου.

  -- *List All Document Types* : Ανεύρεση των περιπτώσεων τύπων εγγράφων στο σύστημα.

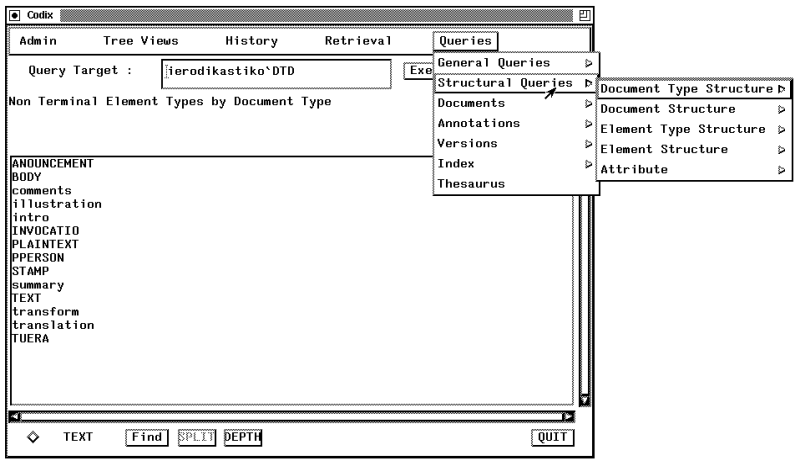
- *Element Structure*
  - -- *List Subelements of Element Types* : Ανεύρεση των υποτμημάτων ενός τύπου τμήματος.
  - -- *List Attributes of Element Types* : Ανεύρεση των χαρακτηριστικών ενός τύπου τμήματος.
  - -- *List Objects of Element Type* : Ανεύρεση των περιπτωσεων ενός τύπου τμήματος.
- *Attribute*
  - -- *List Objects of Attribute Type* : Ανεύρεση των περιπτώσεων ενός τύπου χαρακτηριστικού.

Στο σχήμα 6.11 βλέπουμε τον αρχικό κατάλογο επιλογών ερωτήσεων πρώτης τάξης και (δίπλα του) τον κατάλογο επιλογών δομικών ερωτήσεων. Στο σχήμα 6.12 βλέπουμε ερωτήσεις που αφορούν στη δομή ενός τύπου εγγράφου.

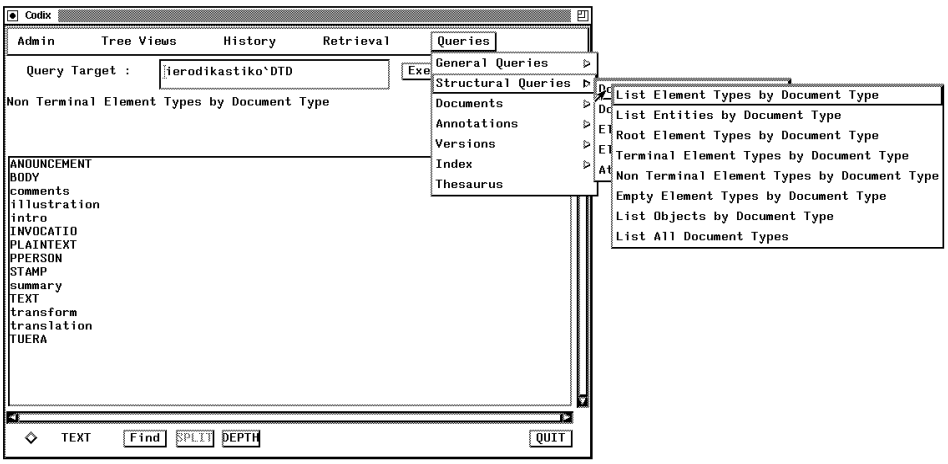
Η επιλογή "Documents"

Η επιλογή *Documents* μας παραπέμπει σε ένα κατάλογο εκτελέσιμων ερωτήσεων. Το αποτέλεσμα των ερωτήσεων αυτών είναι ένα σύνολο λογικών ονομάτων που παριστάνουν έγγραφα. Οι εκτελέσιμες ερωτήσεις είναι οι εξής:

- *Documents by Author* : Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα ενός συγγραφέα και έχει ως αποτέλεσμα όλα τα έγγραφα που έγραψε.
- *Documents by Annotator* : Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα του μελετητή που έκανε σημειώσεις (μία ή περισσότερες) σε έγγραφα και έχει ως αποτέλεσμα τα έγγραφα αυτά.
- *Documents by Indexer* Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα του μελετητή που έκανε τον ευρετηριασμό σε έγγραφα και έχει ως αποτέλεσμα τα έγγραφα που αυτός ευρετηρίασε.
- *Documents by Version Document* Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα μιας έκδοσης και έχει ως αποτέλεσμα τα έγγραφα για τα οποία υπάρχει αυτή η έκδοση.
- *Documents by Index Term* Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα μιας οντότητας (όρο ευρετηρίου) και έχει ως αποτέλεσμα τα έγγραφα που έχουν ευρετηριασθεί με αυτό τον όρο.



Σχήμα 6.11: Ο πρώτος (ανοικτός) κατάλογος επιλογών δείχνει τις κατηγορίες πρώτης τάξεως με τη βοήθεια των οποίων είναι δυνατή η εκτέλεση ερωτήσεων που αφορούν στις βασικές σχέσεις που υποστηρίζει η SIS-Telos και η αναζήτηση πληροφορίας για τη δομή των εγγράφων, έγγραφα, σχόλια, εκδόσεις κτλ. Ο δέυτερος κατάλογος παραπέμπει σε δομικές ερωτήσεις



Σχήμα 6.12: Η πρώτη επιλογή του δεύτερου καταλόγου του προηγούμενου σχήματος παραπέμπει σε ένα σύνολο ερωτήσεων που αφορά στη δομή τύπων εγγράφων. Απο τις ερωτήσεις αυτές, η επιλογή *Non Terminal Element by Document Type* επιστρέφει σε μορφή κειμένου το σύνολο των μή τελικών τμημάτων του *ierodikastikoDTD* τύπου εγγράφου

- *Documents by Related Document* Η ερώτηση αυτή δεχεται ως παράμετρο ένα έγγραφο και επιστρέφει τα έγγραφα που θεωρούνται σχετικά (ως προς το περιεχόμενο) με αυτό.
- *Documents by Referenced Document* Η ερώτηση αυτή δεχεται ως παράμετρο ένα έγγραφο και επιστρέφει τα έγγραφα που αναφέρονται (στο περιεχόμενο) σε αυτό.
- *Documents by more Detailed Document* Η ερώτηση αυτή δεχεται ως παράμετρο ένα έγγραφο και επιστρέφει τα έγγραφα που θεωρούνται πιο αναλυτικά (ως προς το περιεχόμενο) με αυτό.
- *List Hot Documents* Η ερώτηση αυτή επιστρέφει τα έγγραφα που θεωρούνται (από τους ειδικούς στο χώρο της εφαρμογής) ότι χρησιμοποιούνται συχνότερα.
- *List All Documents* Η ερώτηση αυτή επιστρέφει όλα τα έγγραφα που έχουν εισαχθεί στο σύστημα.

Στο σχήμα 6.13 φαίνεται το σύνολο των ερωτήσεων πρώτης τάξης οι οποίες επιστρέφουν έγγραφα.

.

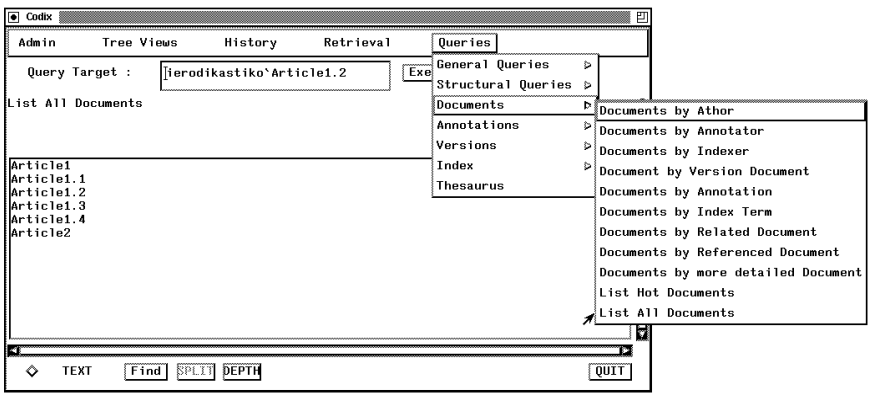
Η επιλογή "Annotations"

Η επιλογή *Annotations* μας παραπέμπει σε ένα κατάλογο εκτελέσιμων ερωτήσεων. Το αποτέλεσμα των ερωτήσεων αυτών είναι ένα σύνολο λογικών ονομάτων που παριστάνουν σημειώσεις. Οι εκτελέσιμες ερωτήσεις είναι οι εξής:

- *Annotations by Document* Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα ενός εγγράφου και επιστρέφει τις σημειώσεις που έγιναν σε αυτό.
- *Annotations by Annotator* Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα ενός συγγραφέα σημειώσεων εγγράφων και επιστρέφει τις σημειώσεις που έγιναν από αυτόν.
- *List All Annotations* Η ερώτηση αυτή επιστρέφει όλες τις σημειώσεις.

Η επιλογή "Versions"

Η επιλογή *Versions* μας παραπέμπει σε ένα κατάλογο εκτελέσιμων ερωτήσεων. Το αποτέλεσμα των ερωτήσεων αυτών είναι ένα σύνολο λογικών ονομάτων που παριστάνουν εκδόσεις εγγράφων. Οι εκτελέσιμες ερωτήσεις είναι οι εξής:



Σχήμα 6.13: Η επιλογή *Document* του αρχικού καταλόγου ερωτήσεων πρώτης τάξης παραπέμπει σε ένα σύνολο ερωτήσεων οι οποίες αναζητούν έγγραφα βάσει διαφορετικών παραμέτρων. Στο συγκεκριμένο παράδειγμα, η ερώτηση που εκτελείται επιστρέφει το σύνολο των εγγράφων που έχουν παρασταθεί στο σύστημα

- *Versions by Creator* Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα του δημιουργού εκδόσεων και επιστρέφει τις εκδόσεις εγγράφων που αυτός δημιούργησε.
- *Versions by Document* Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα ενός εγγράφου και επιστρέφει τις εκδόσεις που έχουν δημιουργηθεί γι αυτό.
- *Next Versions by Version Document* Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα ενός εγγράφου κι επιστρέφει τις μεταγενέστερες εκδόσεις αυτού.
- *List All Versions* Η ερώτηση αυτή επιστρέφει όλες τις εκδόσεις εγγράφων του συστήματος.

Η επιλογή "Index"

Η επιλογή *Index* μας παραπέμπει σε ένα κατάλογο εκτελέσιμων ερωτήσεων. Το αποτέλεσμα των ερωτήσεων αυτών είναι ένα σύνολο λογικών ονομάτων που χρησιμοποιήθηκαν για ευρετηριασμό εγγράφων. Οι εκτελέσιμες ερωτήσεις είναι οι εξής:

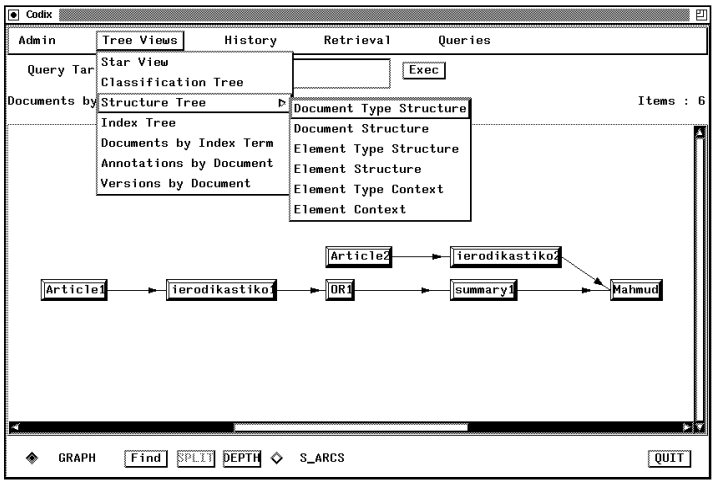
- *Index by Document* Η ερώτηση αυτή δέχεται ως παράμετρο το λογικό όνομα ενός εγγράφου κι επιστρέφει τον όρο που χρησιμοποιήθηκε για τον ευρετηριασμό του.
- *List All Index Terms* Η ερώτηση αυτή επιστρέφει όλες τις οντότητες που χρησιμοποιήθηκαν ως όροι ευρετηρίου.

Η επιλογή "Thesaurus"

Η επιλογή *Thesaurus* είναι εκτελέσιμη ερώτηση και επιστρέφει τους οντότητες του σημασιολογικού δικτύου που ανήκουν στο Επίπεδο Εννοιών και έχουν χρησιμοποιηθεί ως όροι ευρετηρίου.

6.4.2 Αναδρομικές Ερωτήσεις

Οι αναδρομικές ερωτήσεις εμφανίζονται στην επιλογή *Tree Views* του γραφικού εργαλείου (gain) του ΣΣΕ, ο κατάλογος δε στον οποίο παραπέμπουν αποτελείται από εκτελέσιμες ερωτήσεις ή υποκαταλόγους επιλογών με εκτελέσιμες ερωτήσεις. Το αποτέλεσμα των εκτελέσιμων ερωτήσεων εμφανίζεται σε γραφική μορφή στο γραφικό περιβάλλον του ΣΣΕ. Στο σχήμα 6.14 βλέπουμε το αποτέλεσμα εκτέλεσης μιας αναδρομικής ερώτησης καθως επίσης και το σύνολο των ερωτήσεων που έχουν υλοποιηθεί
Παρακάτω δίνεται αναλυτική περιγραφή του αρχικού καταλόγου επιλογών καθώς επίσης των υποκαταλόγων και των εκτελέσιμων ερωτήσεων που περιέχονται σε αυτόν.



Σχήμα 6.14: Οι αναδρομικές ερωτήσεις εμφανίζουν το αποτέλεσμα σε γραφική μορφή. Το σύνολο των αναδρομικών ερωτήσεων που έχουν υλοποιηθεί φαίνεται στους ανοικτούς καταλόγους του σχήματος

.

Η Επιλογή "Star View"

Η επιλογή *Star View* είναι η πρώτη επιλογή του αρχικού καταλόγου. Πρόκειται για εκτελέσιμη εντολή που εμφανίζει τον τρεχον κόμβο στο κέντρο του γραφικού περιβάλλοντος του ΣΣΕ και γύρω από αυτόν (σε μορφή άστρου) όλους τους κομβους του σημασιολογικού δικτύου με τους οποίους συνδέεται με μια σχέση.

Η Επιλογή "Classification Tree "

Η επιλογή *Classification Tree* είναι η δεύτερη επιλογή του αρχικού καταλόγου. Πρόκειται για εκτελέσιμη εντολή που εμφανίζει τον τρεχον κόμβο στο κέντρο του γραφικού περιβάλλοντος του ΣΣΕ και γύρω από αυτόν τους κόμβους του σημασιολογικού δικτύου με τους οποίους συνδέεται με σχέση ταξινόμησης.

Η Επιλογή "Structure Tree "

Η επιλογή *Structure Tree* είναι η τρίτη επιλογή του αρχικού καταλόγου. και συγκεντρώνει τις ερωτήσεις που αναφέρονται σε δομικά χαρακτηριστικά των εγγράφων. Παραπέμπει σε έναν κατάλογο επιλογών με εκτελέσιμες εντολές. Οι εντολές αυτές είναι οι εξής:

- *Document Type Structure*: Η εντολή αυτή εμφανίζει σε γραφική μορφή τη δομή ενός τύπου εγγράφου (DTD).
- *Document Structure*: Η εντολή αυτή εμφανίζει σε γραφική μορφή τη δομή ενός βασικού εγγράφου.
- *Element Type Structure*: Η εντολή αυτή εμφανίζει σε γραφική μορφή τη δομή ενός τύπου τμήματος.
- *Element Structure*: Η εντολή αυτή εμφανίζει σε γραφική μορφή τη δομή ενός τμήματος.

Η Επιλογή "Index Tree"

Η επιλογή *Index Tree* είναι η τέταρτη επιλογή του αρχικού καταλόγου. Πρόκειται για εκτελέσιμη εντολή που εμφανίζει σε γραφική μορφή τη δομή του ευρετηρίου(πως συνδέονται οι όροι του ευρετηρίου).

Η Επιλογή "Documents by Index Term "

Η επιλογή *Documents by Index Term* είναι εκτελέσιμη εντολή και εμφανίζει σε γραφική μορφή τα έγγραφα που ευρετηριάζονται (άμεσα ή έμμεσα) σύμφωνα με έναν όρο του ευρετηρίου.

Η Επιλογή "Annotations by Document "

Η επιλογή *Annotations by Document* είναι η έκτη επιλογή του αρχικού καταλόγου. Είναι εκτελέσιμη εντολή που έχει ως αποτέλεσμα τη γραφική παράσταση όλων των σημειώσεων που έχουν γίνει σε ένα έγγραφο (ή τμήμα του).

Η Επιλογή "Versions by Document "

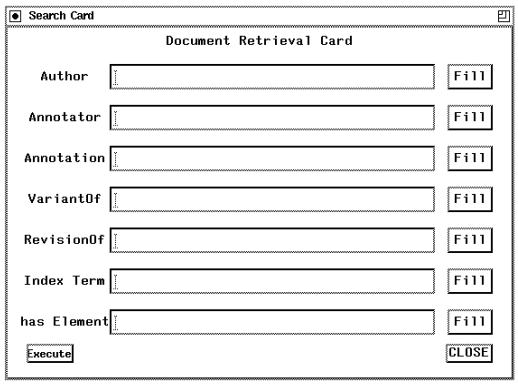
Η επιλογή *Versions by Document* είναι η τελευταία επιλογή του αρχικού καταλόγου. Εμφανίζει σε γραφική μορφή όλες τις εκδόσεις ενός εγγράφου.

6.4.3 Ερωτήσεις πολλαπλών παραμέτρων

Ο χρήστης του συστήματος έχει την δυνατότητα να διατυπώσει ερωτήσεις με τη βοήθεια ενός εντύπου (form based query). Το έντυπο αυτό ονομάζεται κάρτα αναζήτησης και παρέχεται από το γραφικό περιβάλλον του ΣΣΕ. Πρόκειται για μια κάρτα (βλέπε σχήμα 6.15 η οποία αποτελείται από πεδία στα οποία συμπληρώνει ο χρήστης τις παράμετρες αναζήτησης πληροφορίας. Κάθε πεδίο της κάρτας αντιπροσωπεύει ουσιαστικά μια ερώτηση και δέχεται ως παράμετρες μια λογική έκφραση οντοτήτων ίδιου τύπου. Οι ερωτήσεις (μία για κάθε πεδίο της κάρτας αναζήτησης) εκτελούνται χωριστά και το σύνολο απάντησης (answer set) της κάρτας αναζήτησης είναι η τομή των επιμέρους συνόλων απάντησης για κάθε ερώτηση της κάρτας αυτής. Η κάρτα αναζήτησης χρησιμοποιείται για τη διατύπωση ερωτήσεων βάσει πολλών παραμέτρων.

Για το σύστημα **Κώδιξ** προσαρμόστηκε η κάρτα αναζήτησης για την ανεύρεση εγγράφων. Οι επιμέρους ερωτήσεις (και συνεπώς τα αντίστοιχα πεδία της κάρτας) που μετέχουν στην αναζήτηση είναι οι εξής:

- *Author*: παίρνει ως παράμετρο ένα λογικό συνδιασμό από λογικά ονόματα συγγραφέων και επιστρέφει το συνολο των εγγράφων που έχουν γραφεί από αυτούς
- *Annotator*: Παίρνει ως παράμετρο ένα λογικό συνδιασμό σχολιαστών και επιστρέφει το συνολο των εγγράφων που έχουν σημειώσεις από αυτούς.



Σχήμα 6.15: Η κάρτα αναζήτησης του συστήματος Κώδιξ αναζητά έγγραφα με παραμέτρους που αφορούν στους συγγραφείς, τους σχολιαστές, τα σχόλια, τις εκδόσεις, τους όρους ευρετηρίου και των τμημάτων τους

- *Annotation*: Παίρνει ως παράμετρο ένα λογικό συνδιασμό σχολίων και επιστρέφει το συνολο των εγγράφων που έχουν αυτές τις σημειώσεις.
- *Variant*: Παίρνει ως παράμετρο ένα λογικό συνδιασμό παραλλαγών και επιστρέφει τα έγγραφα από τα οποία προήλθαν.
- *Revision*: Παίρνει ως παράμετρο ένα λογικό συνδιασμό διασκευών και επιστρέφει τα έγγραφα από τα οποία προήλθαν.
- *Index Term*: Παίρνει ως παράμετρο ένα λογικό συνδιασμό όρων ευρετηρίου και επιστρέφει το συνολο των εγγράφων που έχουν ευρετηριαστεί σύμφωνα με αυτό τον όρο.
- *Element Type*: Παίρνει ως παράμετρο ένα λογικό συνδιασμό τύπων τμημάτων και επιστρέφει τα έγγραφα που η δομή (σύνολο υποτμημάτων) τους ικανοποιούν την λογική αυτή συνθήκη.

Κεφάλαιο 7

Μια Εφαρμογή: Το Τουρκικό Αρχείο Χανίων

Στο κεφάλαιο αυτό περιγράφεται η εφαρμογή του συστήματος **Κώδιξ** για την διαχείριση των εγγράφων του **Τουρκικού Αρχείου Χανίων (ΤΑΧ).**[1] Η εφαρμογή αυτή θα επιτρέψει τη γρήγορη, εύκολη χρήση και μελέτη των βιβλίων και εγγράφων αυτών χωρίς να εμπλέκονται τα πρωτότυπα κείμενα. Ετσι, θα διασφαλισθεί η ακεραιότητα των πρωτοτύπων αφού για τη μελέτη και διαχείρισή τους θα χρησιμοποιείται η ηλεκτρονική μορφή αυτών.
Για την συγκεκριμένη εφαρμογή του συστήματος Κώδιξ αναπτύχθηκαν δύο τύποι εγγράφων (Tax dtd, Selim dtd) κατάλληλοι για την περιγραφή της δομής των ιεροδικαστικών[2] και των σουλτανικών[3] εγγράφων και έχει γίνει εισαγωγή των τύπων εγγράφων στη βάση του ΣΣΕ. Επίσης, έχει γίνει ενδεικτική παράσταση της δομής και διαχείριση ενός ιστορικού (ιεροδικαστικού) εγγράφου του Τουρκικού Αρχείου Χανίων.

7.1 Περιγραφή του Σημασιολογικού Μοντέλου

- Η πρώτη άποψη (άποψη SGML) παριστάνει πληροφορία σχετική με τη δομή των εγγράφων του ΤΑΧ.
- Η δεύτερη άποψη (άποψη Εννοιών) παριστάνει έννοιες του κόσμου της εφαρμογής.

[1] Σημείωση: Η εφαρμογή αυτή εντάσεται στα πλαίσια των εργασιών που έχει αναλλάβει το Ινστιτούτο Πληροφορικής, Ομάδα Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού, ΙΠ-ΙΤΕ, σε συνεργασία με το Ινστιτούτο Μεσογειακών Σπουδών, ΙΜΣ-ΙΤΕ, για την αρχειοθέτηση, σχολιασμό και ανάκτηση βιβλίων και ιστορικών εγγράφων του Τουρκικού Αρχείου Χανίων [11, 12].

[2] Το περιεχόμενο των εγγράφων αυτών καλύπτει μία γκάμα ιδιαίτερης ευρύτητας, γύρω από ζητήματα οικονομικής και κοινωνικής φύσης (αγοραπωλησίες ακινήτων, διαζύγια, δανεισμούς κ.ο.κ.)

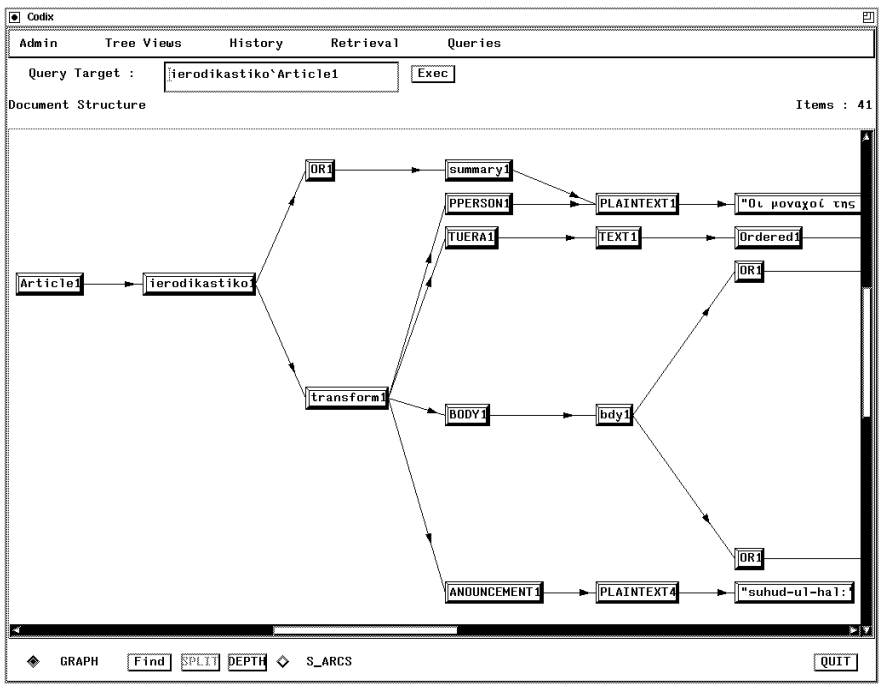
[3] Εγγραφα που εκδίδονταν κάτω από τη διαταγή του σουλτάνου

Οι έννοιες αυτές αφορούν στην ταξινόμηση των εγγράφων, τον αποστολέα, τον παραλήπτη, το περιεχόμενο των εγγράφων κτλ.

- Η τρίτη άποψη (άποψη Επεξεργασίας) παριστάνει πληροφορία σχετική με την παρουσίαση, τον ευρετηριασμό και τον σχολιασμό εγγράφων της εφαρμογής.

7.1.1 Περιγραφή της οντολογίας της άποψης SGML

Η οντολογία της άποψης SGML αποτελείται από το γενικό μοντέλο (βλέπε κεφ.5), τις οντότητες που παριστάνουν τα DTD των εγγράφων του TAX και τις οντότητες που παριστάνουν τη δομή των εγγράφων. Μερική παράσταση της δομής του τύπου εγγράφου TAX φαίνεται στο σχήμα 7.2. Επίσης στο σχήμα 7.1 βλέπουμε τμήμα της δομής του ιστορικού εγγράφου του TAX.

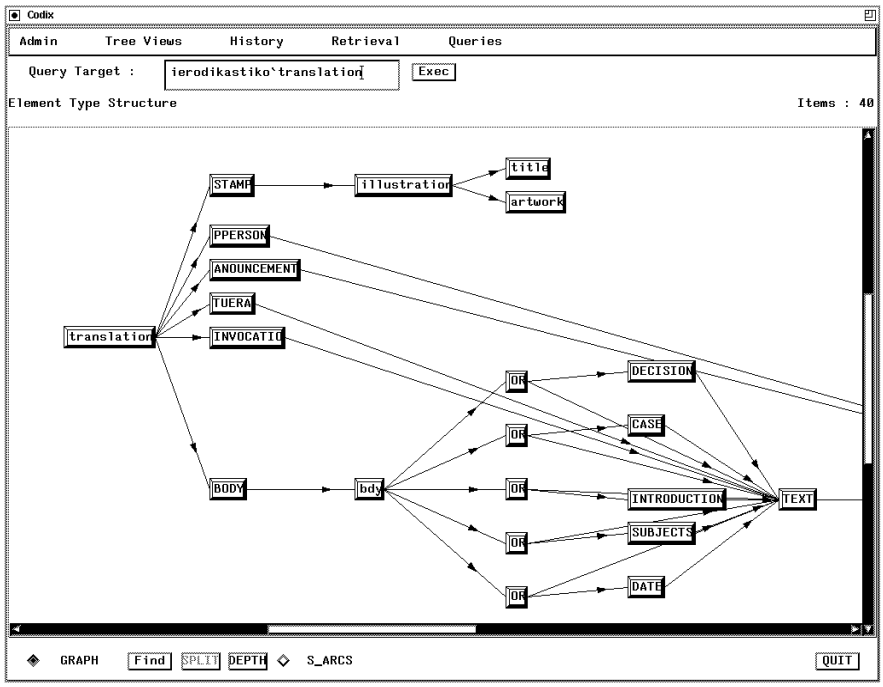


Σχήμα 7.1: Τμήμα της οντολογίας της άποψης SGML που παριστάνει τμήμα της δομή του ιεροδικαστικού εγγράφου

.

. Παρακάτω περιγράφονται αναλυτικά οι οντότητες που παριστάνουν τη δομή του Tax DTD. Στο παράρτημα Α δίνονται (σύμφωνα με το συντακτικό της SGML) οι τύποι εγγράφων Tax και Selim.

- **ierodikastiko'DTD** : Το DTD των ιεροδικαστικών εγγράφων.



Σχήμα 7.2: Τμήμα της οντολογίας της άποψης SGML που παριστάνει τμήμα της δομής του Tax DTD

- **ierodikastiko** : Ο βασικός τύπος τμήματος του TAX DTD.
- **ierodikastiko‘intro** : Μη τελικό τμήμα του εγγράφου (που παριστάνει την εισαγωγή που υπάρχει στην αρχή ενός ιεροδικαστικού εγγράφου)
- **ierodikastiko‘translation** : Μη τελικό τμήμα του εγγράφου (που παριστάνει την μετάφραση ενός ιεροδικαστικού εγγράφου). Οι οντότητες που παριστάνουν τη δομή του είναι οι εξής : **ierodikastiko‘INVOCATIO, ierodikastiko‘TUERA, ierodikastiko‘BODY, ierodikastiko‘DATIO, ierodikastiko‘ELKAB, ierodikastiko‘DUA, ierodikastiko‘INSCRIPTIO, ierodikastiko‘PLACE, ierodikastiko‘STAMP**
- **ierodikastiko‘transform** : Μη τελικό τμήμα του εγγράφου (που παριστάνει την μεταγραφή ενός ιεροδικαστικού εγγράφου) και έχει την ίδια δομή με την μετάφραση.
- **ierodikastiko‘comments** : Μη τελικό τμήμα του εγγράφου (που παριστάνει τα σχόλια που σημειώνονται σε ένα τμήμα του εγγράφου)
- **ierodikastiko‘figure** : Μη τελικό τμήμα του εγγράφου (που παριστάνει την φωτογραφία ενός ιεροδικαστικού εγγράφου)

- **ierodikastiko'illustration**: Μη τελικό τμήμα που παριστάνει εικόνα με τίτλο Η δομή του τμήματος αυτού αποτελείται από ένα τίτλο (που παριστάνεται από την οντότητα **ierodikastiko'title**) και την εικόνα (που παριστάνεται από την οντότητα **ierodikastiko'artwork**).
- **ierodikastiko'TEXT**: Μη τελικό τμήμα που παριστάνει απλό κείμενο.
- bf ierodikastiko'line : Μη τελικό τμήμα που παριστάνει μια γραμμή κειμένου.
- **ierodikastiko'word** : Τελικό τμήμα κειμένου που παριστάνει μια λέξη.
- **ierodikastiko'bdy** : ΕΑΑ που περιέχει ένα αλφαριθμητικό.

Επίσης δημιουργήθηκαν οι οντότητες που παριστάνουν τα χαρακτηριστικά που αποδόθηκαν σε κάθε τύπο τμήματος του TAX DTD. Τέτοιες οντότητες είναι πχ. οι **Root'ierodikastiko'index, Root'ierodikastiko'id** κτλ. Αναλυτική περιγραφή των χαρακτηριστικών αυτών υπάρχει στην παράγραφο 7.2.2.
Τέλος, στην άποψη αυτή δημιουργήθηκαν οντότητες που παριστάνουν το ΒΕ ιεροδικαστικού DTD. Οι οντότητες αυτές δεν αναφέρονται στο παρόν κείμενο για λόγους συντομίας.

7.1.2 Περιγραφή της οντολογίας της άποψης Εννοιών

Η πληροφορία που κρατείται για κάθε έγγραφο αφορά

- Στην ταξινόμηση του ΒΕ (πχ. Κώδικας, ΜεμονωμένοΕγγραφο).
- Στην έκδοση επικύρωσης ενός ΒΕ και υπηρεσίες που εμπλέκονται στη διαχείριση των πρωτοτύπων εγγράφων.
- Στην περιγραφή του εγγράφου ως αντικείμενο.
  Για την περιγραφή του εγγράφου παριστάνεται πχ. το **είδος της γραφής** που χρησιμοποιήθηκε (πχ. Talik, Divani, Rika κοκ.), **πρόσωπα** σχετιζόμενα με το έγγραφο (πχ. Υποκείμενο του ΒΕ που μπορεί να είναι αγοραστής-πωλητής, εκμισθωτής-μισθωτής, υποθηκεύων-δανειστής), **πράξεις, γεγονότα** που είναι σχετικά με το έγγραφο κτλ.

7.1.3 Περιγραφή της οντολογίας της άποψης Επεξεργασίας

Η άποψη Επεξεργασίας παριστάνει οντότητες που χρησιμοποιήθηκαν για τη επεξεργασία του ιεροδικαστικού εγγράφου του TAX.

Συγκεκριμένα δημιουργήθηκαν οντότητες για τον ενδεικτικό σχολιασμό του εγγράφου. Επίσης παραστάθηκε στη βάση γνώσης του συστήματος πληροφορία για την παρουσίαση του εγγράφου (ηλεκτρονικό αρχείο αποθήκευσης του εγγράφου, θέση των τμημάτων του εγγράφου στο αρχείο). Για την καταγραφή εκδόσεων και τον ευρετηριασμό του εγγράφου με άλλα έγγραφα έγινε παράσταση εικονικών εγγράφων τα οποία σχετίστηκαν με αυτό. Για τον ευρετηριασμό του εγγράφου με όρους που αφορούν στο νοηματικό περιεχόμενό του εξειδικεύτηκε ο σύνδεσμος ευρετηριασμού *associatedWith* σε τρεις κατηγορίες συνδέσμων:

- *associatedSubject* Συσχετισμός του εγγράφου (ή τμημάτων του) με το υποκείμενο της πράξης που περιγράφεται στο έγγραφο (πχ. αγοραστής βλέπε σχήμα 6.14).

- *associatedObject* Συσχετισμός του εγγράφου (ή τμημάτων του) με το αντικείμενο της πράξης που περιγράφεται στο έγγραφο (πχ. ακίνητη περιουσία)

- *associatedEvent* Συσχετισμός του εγγράφου (ή τμημάτων του) με την πράξη που περιγράφεται στο έγγραφο (πχ. αγοραπωλησία).

Οι κατηγορίες αυτές συσχετίσεων επιλέχθηκαν διότι εκφράζουν τα βασικά κριτήρια αναζήτησης ιεροδικαστικών εγγράφων.

7.2 Περιγραφή του Τύπου Εγγράφου που υιοθετήθηκε

Ο τύπος εγγράφου που σχεδιάστηκε(από δώ και στο εξής TAX dtd) επιτρέπει την περιγραφή της δομής των ιεροδικαστικών εγγράφων που διαχειρίζεται το Τουρκικό Αρχείο Χανίων. Η κωδικοποίηση των εγγράφων αυτών με το πρότυπο της SGML εξυπηρετεί δύο σκοπούς:

- Εναρμόνιση με τα διεθνή πρότυπα .

- Διευκόλυνση στην αναχρησιμοποίηση και ανταλλαγή των εγγράφων μεταξύ διαφορετικών περιβαλλόντων επεξεργασίας.

7.2.1 Δομή των Ιεροδικαστικών Εγγράφων

Στα ιεροδικαστικά έγγραφα διακρίνουμε την **Εξωτερική** και την **Εσωτερική** δομή.

- Η εξωτερική δομή των εγγράφων αναφέρεται στη μορφή που παρουσιάζουν τα έγγραφα(πχ. παράγραφοι, γραμμές).

- Η εσωτερική δομή αναφέρεται στο περιεχόμενο των εγγράφων, αφορά το κυρίως κείμενο και δεν είναι διακριτή ως προς τη μορφή της (κύριο θέμα του εγγράφου, επιβεβαίωση της επιθυμίας του σουλτάνου, δήλωση της αυθεντικότητας του εγγράφου κ.ο.κ.).

Παρακάτω δίνεται η περιγραφή του TAX dtd που σχεδιάσθηκε για την περιγραφή των ιεροδικαστικών εγγράφων.

7.2.2 Αναλυτική περιγραφή του TAX dtd

Κατά την ανάλυση ενός ιεροδικαστικού εγγράφου γράφεται ένα εισαγωγικό σημείωμα που περιέχει βασικά στοιχεία και χαρακτηριστικά του εγγράφου αυτού. Επίσης εκπονείται η μετάφραση και η μεταγραφή του με λατινικούς χαρακτήρες.
Τα βασικά αυτά προϊόντα της ανάλυσης ενός ιεροδικαστικού εγγράφου δηλ. **εισαγωγή, μετάφραση, μεταγραφή, σχόλια** αποτελούν τα τρία ουσιαστικά τμήματα που αποτελούν τη δομή του TAX dtd.

Η μεταγραφή και η μετάφραση του εγγράφου ακολουθούν την εξωτερική δομή (μορφή) του πρωτότυπου εγγράφου . Επίσης, φωτοτυπίες ή φωτογραφίες των πρωτότυπων εγγράφων αποτελούν το τέταρτο ουσιαστικό τμήμα κάθε αναλυμένου ιεροδικαστικού εγγράφου.

Στο παράρτημα Α δίνεται η περιγραφή του TAX dtd σύμφωνα με το συντακτικό της SGML. Παρακάτω δίνεται αναλυτικά η δομή του TAX dtd.

- **ierodDoc**

  -- *Περιγραφή*
  Βασικό τμήμα (ρίζα της ιεραρχικής δομής) που περιγράφει το ιεροδικαστικό έγγραφο και περιέχει μια πιθανή εισαγωγή για το έγγραφο, την μεταγραφή του εγγράφου, την μετάφραση του εγγράφου ή μια περίληψη, πιθανά σχόλια που γίνονται από τους μελετητές και βρίσκονται προς στο τέλος του εγγράφου και μια (πιθανά) εικόνα (φωτοτυπία, φωτογραφία) του πρωτότυπου κειμένου.

  -- *Αριθμός εμφάνισης*
  Υπάρχει μία φορά στη δομή του εγγράφου.

-- *Χαρακτηριστικά*

Index: Η τιμή του χαρακτηριστικού αυτού είναι ο όρος ευρετηρίου με τον οποίο θα ευρετηριαστεί αυτό το τμήμα.

- **PPERSON**

  -- *Περιγραφή*

  Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **TEXT** Πρόκειται για μια φράση που έχει τα ονόματα των μαρτύρων.

  -- *Αριθμός εμφάνισης*

  Υπάρχει μια τουλάχιστον φορά στη δομή του εγγράφου.

- **PLAINTEXT**

  -- *Περιγραφή*

  Τελικό τμήμα του εγγράφου

- **ANOUNCEMENT**

  -- *Περιγραφή*

  Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **PLAINTEXT** Πρόκειται για μια φράση που αναγγέλει τα ονόματα των μαρτύρων.

  -- *Αριθμός εμφάνισης*

  Υπάρχει μια φορά στη δομή του εγγράφου.

- **translation**

  -- *Περιγραφή*

  Μη τελικό τμήμα του εγγράφου που αποτελείται από τα υποτμήματα **INVOCATIO** (προαιρετικό), **TUERA** (μία ή περισσότερες φορές),STAMP (προαιρετικό), **BODY, ANOUNCEMENT, PPERSON** (μία ή περισσότερες φορές) . Περιέχει την μετάφραση του εγγράφου (αγγλικά ή ελληνικά).

  -- *Αριθμός εμφάνισης*

  Υπάρχει μια φορά στη δομή του εγγράφου.

- **transformation**

  -- *Περιγραφή*
  Μη τελικό τμήμα του εγγράφου που αποτελείται από τα υποτμήματα **INVOCATIO** (προαιρετικό), **TUERA** (μία ή περισσότερες φορές),**STAMP** (προαιρετικό), **BODY, ANOUNCEMENT, PPERSON** (μία ή περισσότερες φορές). Περιέχει την μετάφραση του εγγράφου (αγγλικά ή ελληνικά). Περιέχει την μεταγραφή του εγγράφου με λατινικούς χαρακτήρες.

  -- *Αριθμός εμφάνισης*
  Υπάρχει μια φορά στη δομή του εγγράφου.

- **DECISION**

  -- *Περιγραφή*
  Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **TEXT** Περιγράφει την απόφαση του καδή.

  -- *Αριθμός εμφάνισης*
  Μπορεί να υπάρχει μία φορά στη δομή του εγγράφου.

- **SUBJECTS**

  -- *Περιγραφή*
  Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **TEXT** Περιγράφει την κατάσταση και τη νομική θέση όσων εμπλέκονται στην υπόθεση.

  -- *Αριθμός εμφάνισης*
  Μπορεί να εμφανίζεται μία φορά στη δομή του εγγράφου.

- **CASE**

  -- *Περιγραφή*
  Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **TEXT** Περιγράφει την κύρια υποθεση του εγγράφου.

-- *Αριθμός εμφάνισης*
Μπορεί να υπάρχει μία φορά στη δομή του εγγράφου.

- **INTRODUCTION**

-- *Περιγραφή*
Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **TEXT** Περιγράφει μια εισαγωγική φράση (ανήκει στην εσωτερική δομή του κυρίως κειμένου)

-- *Αριθμός εμφάνισης*
Μπορεί να υπάρχει μία φορά στη δομή του εγγράφου.

- **BODY**

-- *Περιγραφή*
Μη τελικό τμήμα του εγγράφου που αποτελείται από τα υποτμήματα **INTRODUCTION ή text** (προαιρετικά), **SUBJECTS ή text** (προαιρετικά), **CASE ή text** (προαιρετικά), **DECISION ή text** (προαιρετικά), **DATE ή text** (προαιρετικά). Περιέχει το κυρίως σώμα του εγγράφου.

-- *Αριθμός εμφάνισης*
Υπάρχει μία φορά στη δομή του εγγράφου.

-- *Χαρακτηριστικά*
Index: Η τιμή του χαρακτηριστικού είναι ο όρος ευρετηρίου με τον οποίο θα ευρετηριαστεί αυτό το τμήμα.

- **summary**

-- *Περιγραφή*
Μη τελικό τμήμα του εγγράφου που αποτελείται το υποτμήμα **TEXT**. Περιέχει μια μικρή περίληψη με πληροφορίες για το έγγραφο

-- *Αριθμός εμφάνισης*
Υπάρχει προαιρετικά στην αρχή του εγγράφου.

- **intro**

-- *Περιγραφή*

Μη τελικό τμήμα του εγγράφου που αποτελείται το υποτμήμα **PLAINTEXT**. Περιέχει μια μικρή εισαγωγή με πληροφορίες για το έγγραφο(πχ. είδος εγγράφου, γραφή κ.ο.κ.)

-- *Αριθμός εμφάνισης*

Υπάρχει προαιρετικά στην αρχή του εγγράφου.

- **comments**

-- *Περιγραφή*

Τελικό τμήμα του εγγράφου. Περιέχει σχόλια που έχουν γραφεί από μελετητές.

-- *Αριθμός εμφάνισης*

Υπάρχει προαιρετικά στο τέλος του εγγράφου.

-- *Χαρακτηριστικά*

id : Αναγνωριστικό ταυτότητας του σχόλιου.
author: Συγγραφέας του σχόλιου.
date : Ημερομηνία που συντάχθηκε το σχόλιο.
type : Είδος σχόλιου.
visible: Η τιμή του χαρακτηριστικού αυτού δηλώνει εαν το σχόλιο αυτό θα είναι ορατό και θα εμφανίζεται μαζί με τα περιεχόμενα του εγγράφου.

- **INVOCATIO**

-- *Περιγραφή*

Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **TEXT** . Περιέχει επίκληση στο θεό και βρίσκεται στο πάνω μέρος του εγγράφου.

-- *Αριθμός εμφάνισης*

Υπάρχει μία φορά στη δομή του εγγράφου.

-- *Χαρακτηριστικά*

id : Αναγνωριστικό ταυτότητας του σχόλιου.
Index: Η τιμή του χαρακτηριστικού αυτού είναι ο όρος ευρετηρίου με τον οποίο θα ευρετηριαστεί αυτό το τμήμα.

- **TUERA**

  -- *Περιγραφή*

  Μη τελικό τμήμα του εγγράφου που αποτελείται από τα υποτμήματα **TEXT**. Περιέχει την υπογραφή του καδή (ονομα, τίτλο, πόλη στην οποία βρίσκεται και μια αραβική προσευχή).

  -- *Αριθμός εμφάνισης*

  Υπάρχει μία φορά στη δομή του εγγράφου.

  -- *Χαρακτηριστικά*

  id : Αναγνωριστικό ταυτότητας του σχόλιου.

  Index: Η τιμή του χαρακτηριστικού αυτού είναι ο όρος ευρετηρίου με τον οποίο θα ευρετηριαστεί αυτό το τμήμα.

- **DATE**

  -- *Περιγραφή*

  Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **TEXT**. Βρίσκεται κάτω αριστερά του εγγράφου και περιέχει την χρονολόγηση του εγγράφου.

  -- *Αριθμός εμφάνισης*

  Υπάρχει μία φορά στη δομή του εγγράφου.

  -- *Χαρακτηριστικά*

  id : Αναγνωριστικό ταυτότητας.

  Index: Η τιμή του χαρακτηριστικού αυτού είναι ο όρος ευρετηρίου με τον οποίο θα ευρετηριαστεί αυτό το τμήμα.

- **STAMP**

  -- *Περιγραφή*

  Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **illustration**. Το τμήμα αυτό έχει την σφραγίδα του καδή.

  -- *Αριθμός εμφάνισης*

  Υπάρχει καμία ή μια φορά στο έγγραφο.

-- *Χαρακτηριστικά*
id : Αναγνωριστικό ταυτότητας του σχόλιου.
Index: Η τιμή του χαρακτηριστικού αυτού είναι ο όρος ευρετηρίου με τον οποίο θα ευρετηριαστεί αυτό το τμήμα.

- **illustration**

-- *Περιγραφή*
Μη τελικό τμήμα του εγγράφου που αποτελείται από τα υποτμήματα **artwork, title** που περιέχουν αντίστοιχα την εικόνα και τον (προαιρετικό) τίτλο της. Ο τίτλος της εικόνας είναι τελικό τμήμα του εγγράφου. Το τμήμα artwork είναι κενό και έχει τα εξής χαρακτηριστικά : id : Αναγνωριστικό ταυτότητας του σχόλιου.
name : όνομα του ηλεκτρονικού αρχείου που περιέχει την εικόνα.

- **TEXT**

-- *Περιγραφή*
Μη τελικό τμήμα του εγγράφου που αποτελείται από το υποτμήμα **line** μία ή περισσότερες φορές και το υποτμήμα **comments** μηδέν ή περισσότερες φορές. Περιγράφει το απλό κείμενο που υπάρχει στο έγγραφο.

- **line**

-- *Περιγραφή*
Τελικό τμήμα του εγγράφου

7.2.3 Ιδιαιτερότητες - Επέκταση του TAX dtd

Για τη σχεδίαση του TAX dtd υιοθετήθηκε η εξωτερική δομή των εγγράφων όπως αυτή περιγράφεται από τη βιβλιογραφία [13] . Η εσωτερική δομή των εγγράφων αυτών παριστάνεται στο TAX dtd με δύο τρόπους δηλ. είτε με προαιρετικά τμήματα με διαφορετική σημασιολογία (όσον αφορά το περιεχόμενό τους) ή με απλό κείμενο. Αυτό συμβαίνει διότι η εσωτερική δομή του εγγράφου συχνά είναι δύσκολο να διαχωριστεί.
Η ανάλυση και μελέτη ιεροδικαστικών εγγράφων επιδέχεται σχολιασμού από τους

μελετητές . Για το λόγο αυτό δίνεται η δυνατότητα πρόσθεσης σχόλιων σε διάφορες στάθμες λεπτομέριας (granularity) του εγγράφου επιτρέποντας το σχολιασμό ολόκληρων εγγράφων, παραγράφων, γραμμών.

Η ανάκτηση των εγγράφων

Στο TAX dtd τα τμήματα του εγγράφου που θεωρούνται (από τους ειδικούς) ώς περισσότερο πληροφοριακά εμπλουτίζονται με το χαρακτηριστικό Index το οποίο δηλώνει ότι κάποια πληροφορία που παρουσιάζεται στο εγγράφο θα ευρετηριασθεί. Η τιμή του χαρακτηριστικού αυτού είναι ο όρος με τον οποίο θα ευρετηριαστεί αυτό το τμήμα. Ο όρος αυτός αν δεν υπάρχει στη ΒΓ του συστήματος, δημιουργείται. Ο ευρετηριασμός μπορεί να γίνεται σε στάθμη εγγράφων, παραγράφων, λέξεων.

Η χρήση ΕΑΑ διευκολύνει την επέκταση του TAX dtd. Για παράδειγμα, όταν κριθεί απαραίτητος ο εμπλουτισμός της εσωτερικής δομής και με άλλους τύπους τμημάτων τότε αυτός αυτοί μπορούν να προστεθούν στο ΕΑΑ που περιγράφει την εσωτερική δομή των ΒΕ χωρίς να απαιτείται άλλη αλλαγή στο DTD.

Κεφάλαιο 8

Επίλογος

8.1 Συμπεράσματα

Στην παρούσα εργασία έγινε αναλυτική περιγραφή του συστήματος διαχείρισης εγγράφων SGML **Κώδιξ**. Το σύστημα αυτό χρησιμοποιεί μια βάση γνώσης που ακολουθεί το οντοκεντρικό μοντέλο και είναι οργανωμένη σε τρεις απόψεις οντοτήτων. Η δυνατότητα χρήσης (και διαχείρισης) της πληροφορίας των τριών απόψεων της οντολογίας του συστήματος ανεξάρτητα ή σε συνδιασμό επιτρέπει τη συστηματικότερη μελέτη της σημασιολογίας των εγγράφων και του πεδίου εφαρμογής των. Επίσης επιτρέπει τη σημασιολογική συσχέτιση των εγγράφων μεταξύ τους καθώς επίσης και με έννοιες που αφορούν στο περιεχόμενο και την επεξεργασία τους (παρουσίαση, καταγραφή εκδόσεων κτλ.)

Η γλώσσα SIS-Telos και το Σύστημα Σημασιολογικού Ευρετηριασμού επιτρέπει την αποδοτική παράσταση της σημασιολογικά ετερογενούς και περίπλοκα συνδεδεμένης πληροφορίας που υπάρχει στη Βάση γνώσης του συστήματος. Επίσης, η χρήση του ΣΣΕ και της SIS-Telos επιτρέπει την επέκταση του σχήματος της ΒΓ χωρίς αναδιάρθρωση της ήδη υπάρχουσας πληροφορίας. Αυτό είναι ιδιαίτερα σημαντικό διότι δεν έχει αναπτυχθεί πρότυπος τρόπος (standard) παράστασης της πληροφορίας αυτής και συχνά απαιτούνται προσθήκες και/ή τροποποιήσεις στην παράσταση τέτοιας πληροφορίας.

Στη σχεδίαση του γενικού μοντέλου παραστάθηκε όλη η πληροφορία που μπορεί να υπάρχει στη δήλωση ενός τύπου εγγράφου. Η εκφραστικότητα του γενικού μοντέλου πιστοποιείται από το γεγονός ότι κατά την εφαρμογή του για την παράσταση περίπλοκων

τύπων εγγράφων δεν χρειάστηκαν προσθήκες νέων βασικών οντοτήτων παρά μόνο αυτών που εξαρτώνται απόλυτα από τη νέα εφαρμογή. Επίσης δεν χρειάστηκαν προσθήκες ή τροποποιήσεις στις παρεχόμενες λειτουργίες των δελτίων ενημέρωσης και των προκαθορισμένων ερωτήσεων.

Η δυνατότητα δημιουργίας σημειώσεων στα έγγραφα συμπληρώνει την λειτουργικότητα του συστήματος. Επίσης, η χρήση των σημειώσεων (και των χαρακτηριστικών που τις αποδίδονται) αυξάνει τους ποιοτικούς δείκτες της ανάκτησης κατά την αναζήτηση εγγράφων. Οι δείκτες αυτοί αυξάνουν και με την κατηγοριοποίηση των συνδέσμων ευρετηριασμού (και των χαρακτηριστικών τους). Προσθέτουμε εδώ οτι η απόδοση χαρακτηριστικών στους συνδέσμους (και η δυνατότητα σχολιασμού τους) επιτρέπει τη διατήρηση ιστορικού εξέλιξης του ευρετηριασμού της πληροφορίας που διαχειρίζεται το σύστημα. Με αυτό τον τρόπο είναι δυνατός ο έλεγχος της ορθότητας, η μελέτη των επιλογών και οι αναθεωρήσεις στον ευρετηριασμό της πληροφορίας.

Τελος, μελετήθηκε η πιθανότητα χρήσης του συστήματος από διαφορετικές κατηγορίες χρηστών κατά την προσαρμογή των δελτίων ενημέρωσης. Το γεγονός αυτό οδήγησε στην κατάταξη των λειτουργιών του συστήματος (που επιτελούνται μεσω των ΔΕ) σε κατηγορίες χρηστών.

8.2 Επεκτάσεις

Κατά το σχεδιασμό της Βάσης Γνώσης εστιάσαμε την προσοχή μας κυρίως στο γενικό μοντέλο του συστήματος που βρίσκεται στην άποψη SGML. Ιδιαίτερο ενδιαφέρον θα παρουσίαζε ωστόσο η εκτενής μελέτη και ο σχεδιασμός της άποψης Επεξεργασίας. Η ανάπτυξη της άποψης αυτής περιλαμβάνει την παράσταση πληροφορίας για όλα τα δυνατά προϊόντα της εφαρμογής. Τέτοια προϊόντα είναι πχ. η δημιουργία παρουσιάσεων των εγγράφων στο WWW. Επίσης, η μορφοποίηση των εγγράφων (για ηλεκτρονική παρουσίαση ή έντυπη), η δημιουργία οπτικών δίσκων κτλ. Η επέκταση της άποψης Επεξεργασίας περιλαμβάνει την παράσταση ειδικών παραμέτρων για τη δημιουργία οπτικών δίσκων, την κατηγοριοποίηση των τμημάτων των εγγράφων (κείμενο, εικόνα κτλ.). Επίσης, παράσταση του τρόπου παρουσίασης των τμημάτων (λίστα, απλό κείμενο, αριθμός στηλών στο κείμενο) τρόποι κωδικοποίησης της εικόνας (tiff, jpg), μέγεθος και θέση εικόνας κτλ.
Ειδικότερα η παρουσίαση εγγράφων στο WWW απαιτεί την σύνδεση του συστήματος Κώδιξ με το WWW και παράσταση του τρόπου παρουσίασης των λογικών τμημάτων των

εγγράφων με τη χρήση της HTML[1] .

Μια χρήσιμη επέκταση του συστήματος θα αποτελούσε η υποστήριξη εγγράφων που ακολουθούν το πρότυπο HyTime [XX]. Το πρότυπο αυτό αποτελεί επέκταση της SGML για την υποστήριξη εγγράφων που η παρουσίαση των περιεχομένων τους εξαρτάται από χρονικές παραμέτρους πχ. κινούμενη εικόνα (video).

Ακόμη θά ήταν χρήσιμη η υποστήριξη ηχητικών σχόλιων καθώς και η σημασιολογική ταξινόμησή τους. Για τη σημασιολογική ταξινόμηση των σχολίων έχει εκπονηθεί σχετική μεταπτυχιακή εργασία [8].

Κατά την ανάπτυξη του γενικού μοντέλου δώθηκε έμφαση στην παράσταση της λογικής δομής των εγγράφων. Στο μοντέλο αυτό οι τύποι τμημάτων που αναφέρονται σε συνδέσμους (hyperlinks) αντιμετωπίζονται ομοιόμορφα με τους άλλους τύπους τμημάτων. Ιδιαίτερο ενδιαφέρον θα παρουσίαζε η παράσταση του τρόπου διασύνδεσης των λογικών τμημάτων των εγγράφων με τη χρήση τέτοιων συνδέσμων.

Η αποθήκευση της πληροφορίας που διαχειρίζεται το σύστημα βρίσκεται στον κεντρικό εξυπηρετητή του συστήματος. Χρήσιμη θα ταν η μελέτη της αρχιτεκτονικής του συστήματος για κατανεμημένο περιβάλλον αποθήκευσης τόσο της ΒΓ του συστήματος όσο και των εγγράφων της εφαρμογής.

Μια άλλη επέκταση του συστήματος είναι η σύνδεσή του με ένα σύστημα λεξικογραφικής ανάκτησης πληροφορίας (full text retrieval system) για την ικανοποίηση ερωτήσεων που αφορούν σε λέξεις/φράσεις που περιέχονται στα έγγραφα.

Τέλος, πολύ ενδιαφέρον παρουσιάζει η μελέτη της αυτόματης παράστασης του περιεχομένου των εγγράφων στην άποψη Εννοιών της βάσης γνώσης του συστήματος.

[1] Στο παρελθόν εκπονήθηκε σχετική φοιτητική εργασία [9]

Βιβλιογραφία

[1] Astoria - powerfull oo sgml document component management system. URL : http://www.xsoft.com.

[2] An information and document management system. Address : STEP Sturtz Electronic Publishing GmbH Technologiepark Wurzburg-Rimpar D-97222 Rimpar Germany.

[3] The semantic index system- a brief presentation. Foundation for Research and Technology-Hellas Institute of Computer Science.

[4] . . Views. Master's thesis, , 1995. (in Greek).

[5] . telos. Master's thesis, , 1993. (in Greek).

[6] . . Master's thesis, , 1994. (in Greek).

[7] . Entry forms. Master's thesis, , 1995. (in Greek).

[8] . . Master's thesis, , 1996. (in Greek).

[9] www, 1994. .

[10] Life∗cdm highlights, 1994. CSD/Part No. H-40000 Ver. 04.94.

[11] , 1995. . 3.

[12] , 1995. . 3.

[13] , 1995. . 3.

[14] A. Brown. *Object Oriented Databases. Applications in Sofware Engineering*. McGraw-Hill, 1988.

[15] Anne Bruggemann-Klein and Derick Wood. The validation of sgml content models. Technical report, 1993.

[16] Klemens Bohm and Karl Aberer. Storing hytime documents in an object oriented database. Technical report, GMD-IPSI.

[17] Klemens Bohm and Karl Aberer. Amplifying the scope of document handling: An object oriented database application framework for structured document storage. Technical Report 811, GMD, 1993.

[18] Klemens Bohm, Karl Aberer, and Erich Neuhold. Administering structured documents in digital libraries. Technical report, 1995.

[19] Klemens Bohm and C.Rakow. Metadata for multimedia documents. *SIGMOD RECORD*, 23(4):21--26, Dec 1994.

[20] Dale C. Waldt Brian E. Travis. *The SGML Implementation Guide*. XX Press, 1995.

[21] M. Brodie, J. Mylopoulos, and J. Schmidt, editors. *On Conceptual Modeling: Perspectives-from Artificial Intelligence, Databases and Programming Languages*. Springer-Verlag, 1984.

[22] Dionysios C. Tsichritzis and Frederic H. Lochovsky. *Data Models*, chapter 10, pages 210--224. Prentice-Hall, 1982.

[23] C.F.Goldfard. *The SGML Handbook*. Clarendon Press, 1990.

[24] Zhengxin Chen. A conceptual model for storage and retrieval of short scientific texts. *Information Processing and Management* , 29(2):209--214, 1993.

[25] Costas Dadouris and Martin Doerr. Sis- programmatic query interface reference manual. Technical Report 3, Foundation for Research and Technology-Hellas, Institute of Computer Science, 1995.

[26] Euroclid, 12 Avenue des Pres 78180 Montigny le Bretonneux. *Le parseur SGML d'Euroclid*, 1991.

[27] Foundation for Research and Technology-Hellas, Institute of Computer Science. *SIB data entry language user's manual*, Dec 1992.

[28] Foundation for Research and Technology-Hellas, Institute of Computer Science. *SIB Static Analyser-Graphical Analysis Interface User's Manual - Version 1.2*, Jun 1995.

[29] Foundation for Research and Technology-Hellas, Institute of Computer Science. *SIS-Binding with External Tools - Version 1.2*, Jun 1995.

[30] Foundation for Research and Technology-Hellas, Institute of Computer Science. *SIS-Query Interpreter: An Interactive Program to Use PQI functions - version 1.2*, Jun 1995.

[31] Foundation for Research and Technology-Hellas, Institute of Computer Science. *Customizing Entry Forms - Version 1.3*, Sep 1996.

[32] Foundation for Research and Technology-Hellas, Institute of Computer Science. *SIS-Entry Formas User's manual - Version 1.3*, Sep 1996.

[33] Gawkowski, John A., and S.A. Mamrak. A universal framework for data transformation. Technical Report 94-11, R.R Donnelley and Sons Databases Technology Services,Department of Computer and Information Science,The Ohio State University, 1993.

[34] I. A. Macleod . Storage and retrieval of structured documents. *Information Processing and Management*, 26(2):197--208, 1990.

[35] Klemens Bohm. Building a configurable database application for structured documents. Technical Report 942, GMD-IPSI, 1995.

[36] Reggie Lisle and Troy k. Bartlett. Managing the paperstorm. *Lan Times*, 12(17):2--8, 1995.

[37] Arjan Loeffen. Text databases: A survey of text models and systems. *SIGMOD RECORD*, 23(1):97--106, Mar 1994.

[38] Xin Lu. Document retrieval: A structural approach. *Information Processing and Management* , 26(2):209--218, 1990.

[39] Eve Maler and Jeanne El Andaloussi. *Developing SGML DTDs: From text to model to markup.* Prentice-Hall, 1996.

[40] John Mylopoulos, Alex Borgida, Mathias Jarke, and Manolis Koubarakis. Telos: Representing knowledge about information systems. *ACM Transactions on Information Systems*, 8(2):352--362, 1990.

[41] O_2 Technology. *The O_2 User's Manual - version 4.4*, Dec 1993.

[42] Steve Pepper. The whirlwind guide :sgml tools and vendors, Jan 1995. Anonymous ftp: ftp.ifi.uio.no/pub/SGML/SGML-Tools.

[43] Darrell Raymond, Tompa Frank, and Derick Wood. From data representation to data model: Meta-semantic issues in the evolution of sgml. Technical Report 95-17, Department of Computer Science University of Waterloo Waterloo Ontario Canada - Department of Computer Science Hong Kong University of Science and Technology Kowloon Hong Kong, 1995.

[44] Lloyd Rutledge. A hytime engine for hypermedia document presentation. Master's thesis, Department of Computer Science University of Massachusetts Lowell, 1993.

[45] Gerard Salton, James Allan, and Chris Buckley. Automatic structuring and retrieval of large text files. *Communication of the ACM* , 37(2):97--108, Feb 1994.

[46] Arijit Sengupta. Design and implementation of a database environment for the manipulation of structured documents, 1995. URL : http://www.cs.indiana.edu/hyplan/asengupt/thesis/oral/oral.html.

[47] Kazem Taghva, Allen Condit, and Julie Borsack. Autotag: A tool for creating structured document collections from printed materials. Technical Report 94-11, Information Science Research Institute University of Nevada Las Vegas, 1994.

[48] Kazem Taghva, Allen Condit, Julie Borsack, John Kilburg, Changshi Wu, and Jeff Gilbreth. The manicure document processing system. Technical Report 95-02, Information Science Research Institute University of Nevada Las Vegas, 1995.

[49] V.Christofides and A.Rizk. Querying structured documents with hypertext links using oodbms. *ACM*, 1(1), 1994.

[50] V.Christofides, S.Abiteboul, S.Cluet, and M.Scholl. From structured documents to novel query facilities. *SIGMOD*, pages 313--324, 1994.

[51] Chiradeep Vittal, M. Tamer Ozsu, Duane Szafron, and Ghada El Medani. The logical design of a multimedia database for a news-on-demand application. Technical Report 94-16, Department of Computing Science , The University of Alberta,Edmonton,Alberta,Canada, 1994.

[52] W. Klas, K. Aberer, and E. Neuhold. Object oriented modeling for hypermedia systems using the vodak modeling language (vml). In A. Dogac, T. Ozsu, A. Biliris, and T. Sellis, editors, *Advances in Object Oriented Database Systems.* Springer Verlag, 1994.

[53] Yannis Kollias. *Databases.* Symmetria Press, 1989.

[54] Jian Zhang. Application of oodb and sgml techniques in text database: An electronic dictionary system. *SIGMOD RECORD*, 24(1), March 1995.

.

Γλωσσάρι ελληνικών όρων

ΒΓ	Βάση Γνώσης
ΒΔ	Βάση Βεδομένων
ΒΕ	Βασικά έγγραφα
ΕΑΑ	Εικονικό Αντικείμενο Αποθήκευσης
ΚΖ	Κύκλος Ζωής
ΣΣΕ	Σύστημα Σημασιολογικού Ευρετηριασμού
ΣΣ	Σύστημα Σύνδεσης
ΤΑΧ	Τουρκικό Αρχείο Χανίων

Γλωσσάρι αγγλικών όρων

API	Application Programming Interface
DBMS	Data Base Management System
DMS	Document Management System
DTD	Document Type Definition
FS	File System
HTML	Hypertext Markup Language
KDBMS	Knowledge Data BAse Management System
RDBMS	Relational Data Base Management System
SIS	Semantic Index System
SGML	Standard Generalized Markup Language
OS	Operating System
PI	Processing Instruction
PQI	Programming Query Interface

Παράρτημα Α

Τύπος εγγράφου ιεροδικαστικών εγγράφων

```
<!DOCTYPE ierodDoc [
<!ELEMENT ierodDoc --(intro?,transform,(translation|summary),comments?,
illustration?)>
<!ATTLIST ierodDoc index CDATA
id IDREF   >

<!ELEMENT (translation|transform)--(INVOCATIO?,TUERA+,STAMP?,BODY,ANOUNCEMENT,PPER
>
<!ENTITY %bdy   "((INTRODUCTION|TEXT)?,(SUBJECTS|TEXT)?,(CASE|TEXT)?,
(DECISION|TEXT)?,(DATE|TEXT)?
" >

<!ELEMENT BODY - - %bdy >

<!ATTLIST BODY index CDATA
id IDREF    >

<!ELEMENT INTRODUCTION - - (TEXT) >
<!ELEMENT SUBJECTS - - (TEXT) >
<!ELEMENT CASE - - (TEXT) >
```

```
<!ELEMENT DECISION - - (TEXT) >
<!ELEMENT intro - - (plaintext) >
<!ELEMENT summary - - (TEXT) >
<!ELEMENT comments - - (PCDATA) >

<!ATTLIST comments type CDATA
id ID
author CDATA
date CDATA >
visible (yes|no) no >

<!ELEMENT INVOCATIO - - (TEXT) >
<!ATTLIST INVOCATIO index CDATA
id IDREF    >

<!ELEMENT ANOUNCEMENT - - (plaintext) >
<!ELEMENT PPERSON - - (TEXT) >
<!ELEMENT DECISION - - (TEXT) >


<!ELEMENT TUERA - - (TEXT) >
<!ATTLIST TUERA index CDATA
id IDREF    >

<!ELEMENT DATE - - (TEXT) >
<!ATTLIST DATE index CDATA
id IDREF    >
<!ELEMENT STAMP - - (illustration) >
<!ATTLIST STAMP index CDATA
id IDREF    >
<!ELEMENT illustration - - (artwork,title)  >
<!ELEMENT title - O (plaintext)  >
<!ELEMENT artwork - O (EMPTY)  >
<!ATTLIST artwork id   ID #REQUIRED
```

```
name  ENTITY #REQUIRED >

<!ELEMENT plaintext - - (PCDATA)>
<!ELEMENT TEXT - - ((plaintext,comments?)+ ) >

]>
```

Παράρτημα Β

Τύπος εγγράφου Σουλτανικών εγγράφων

```
<!DOCTYPE SelimDoc [
<!ELEMENT SelimDoc - - (intro?,translation,transform,illustration?)>
<!ATTLIST SelimDoc index CDATA
id IDREF   >

<!ELEMENT (translation | transform) - - ( INVOCATIO,TUERA,BODY,
DATIO,PLACE,STAMP? ) >
<!ENTITY %bdy    " (INTITULATIO?&INSCRIPTIO?&NARRATIO?
&DISPOSITIO?&(SANCTIO|COMMINATIO)?
&CORRABORATIO?&TEXT)" >

<!ELEMENT BODY - - %bdy >

<!ATTLIST BODY index CDATA
id IDREF    >

<!ELEMENT intro - - (TEXT) >
<!ELEMENT comments - - (plaintext) >

<!ATTLIST comments type CDATA
```

```
id ID
author CDATA
date CDATA >
visible (yes|no) no >

<!ELEMENT INTITULATIO - - (TEXT) >

<!ATTLIST INTITULATIO index CDATA
id IDREF  >

<!ELEMENT INSCRIPTIO - - ( (ELKAB?&DUA?)|TEXT) >

<!ATTLIST INSCRIPTIO index CDATA
id IDREF    >

<!ELEMENT ELKAB - - (TEXT)>
<!ATTLIST ELKAB index CDATA
id IDREF    >

<!ELEMENT DUA - - (TEXT) >
<!ATTLIST DUA index CDATA
id IDREF    >

<!ELEMENT INVOCATIO - - (TEXT) >
<!ATTLIST INVOCATIO index CDATA
id IDREF     >
<!ELEMENT TUERA - - (TEXT) >
<!ATTLIST TUERA index CDATA
id IDREF     >
<!ELEMENT DATIO - - (TEXT) >
<!ATTLIST DATIO index CDATA
id IDREF     >
<!ELEMENT PLACE - -(TEXT) >
<!ATTLIST PLACE index CDATA
```

```
id IDREF       >
<!ELEMENT STAMP - - (illustration) >
<!ATTLIST STAMP index CDATA
id IDREF    >
<!ELEMENT illustration - - (artwork,title)  >
<!ELEMENT title - O (TEXT)  >
<!ELEMENT artwork - O (EMPTY)  >
<!ATTLIST artwork id   ID #REQUIRED
name  ENTITY #REQUIRED >
<!ELEMENT plaintext - - (PCDATA)>
<!ELEMENT TEXT   - - ((plaintext,comments?)+ )        >

]>
```

Παράρτημα C

Τύπος εγγράφου Αποψης Επεξεργασίας

```
<!DOCTYPE processDoc [
<!ELEMENT processDoc - - (document, PresentationRule*, IndexRule*,
VersioningRule*, AnnotationRule*)>

<!ELEMENT document - - (#PCDATA ) >

<!ELEMENT PresentationRule - - (elem, presentationAtt+) >
<!ELEMENT elem - - (#PCDATA) >

<!ELEMENT presentationAtt - - (name, value) >
<!ELEMENT name - - (#PCDATA) >

<!ELEMENT value - - (#PCDATA) >

<!ELEMENT IndexRule - - (elem, index*) >
<!ELEMENT index - - (#PCDATA) >
<!ATTLIST index type (associatedWith | see | seeAlso
 | detailedDescription) >

<!ELEMENT VersioningRule - - (#PCDATA) >
```

```
<!ELEMENT AnnotationRule - - (elem, annotation) >

<!ELEMENT annotation - - (#PCDATA) >

]>
```

Παράρτημα D

Γενικό μοντέλο άποψης SGML

```
BEGINTRANSACTION

RETELL Attribute Unique
from : Individual
to : Individual
in M1_Class end

RETELL Individual DocumentType in M1_Class
with attribute
(hasRootElement) : RootElementDecl;
(hasElement) : AttributeElementDecl;
(hasEntity) : EntityDecl
end
RETELL Individual Element in S_Class, AttributeElementDecl
end

RETELL Individual Document in S_Class, DocumentType
end

RETELL Individual SimpleElementDecl in M1_Class
isA ElementDecl, AttributeElementDecl end

RETELL Individual TerminalElementDecl in M1_Class isA
SimpleElementDecl
```

```
with attribute
hasData : PCDATA
end

RETELL Individual EmptyElementDecl in M1_Class
isA SimpleElementDecl end

RETELL Individual GroupElementDecl in M1_Class
isA NonTerminalElementDecl end

RETELL Individual OrElementDecl in M1_Class
isA NonTerminalElementDecl end

RETELL Individual NonTerminalElementDecl in M1_Class
isA ComplexElementDecl, ElementDecl
end

RETELL Individual ParameterEntityDecl in M1_Class
isA ElementDecl, InternalEntityDecl end

RETELL Individual AttributeElementDecl in M1_Class
with attribute
(hasAttribute) : AttributeDecl
end

RETELL Individual RootElementDecl in M1_Class
isA ComplexElementDecl end

RETELL Individual ComplexElementDecl in M1_Class
isA AttributeElementDecl end

RETELL Attribute hasContents
from: ComplexElementDecl
to : ElementDecl
```

```
in M1_Class
with attribute
hasNext : ElementDecl
end

RETELL Individual ElementDecl in M1_Class end

RETELL Individual AttributeDecl in M1_Class
with attribute
(hasDefaultValue) : Telos_String
end

RETELL Individual EntityDecl in M1_Class
with attribute
(hasText) : EntityText
end

RETELL Individual InternalEntityDecl in M1_Class
isA EntityDecl end

RETELL Individual ExternalEntityDecl in M1_Class
isA EntityDecl end

RETELL Individual PublicEntityDecl in M1_Class
isA ExternalEntityDecl end

RETELL Individual EntityText in M1_Class end

RETELL Individual PCDATA in M1_Class  end


ENDTRANSACTION
```