ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

Σχεδίαση και υλοποίηση ενός συστήματος υπομνηματισμού εγγράφων

Σαράντης Τούλης

Μεταπτυχιακή Εργασία

Ηράκλειο, Σεπτέμβριος 1996

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

Σχεδίαση και υλοποίηση ενός συστήματος υπομνηματισμού εγγράφων

Εργασία που υποβλήθηκε από τον
Σαράντη Τούλη
ως μερική εκπλήρωση των απαιτήσεων
για την απόκτηση
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Σαράντης Τούλης
Τμήμα Επιστήμης Υπολογιστών

Εισηγητική Επιτροπή:

Πάνος Κωνσταντόπουλος, Αναπληρωτής Καθηγητής, Επόπτης

Βαγγέλης Μαρκάτος, Επίκουρος Καθηγητής, Μέλος

Πάνος Τραχανιάς, Επίκουρος Καθηγητής, Μέλος

Δεκτή:

Πάνος Κωνσταντόπουλος, Αναπληρωτής Καθηγητής
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Ηράκλειο, Σεπτέμβριος 1996

Σχεδίαση και υλοποίηση συστήματος υπομνηματισμού εγγράφων

Σαράντης Τούλης

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

ΠΕΡΙΛΗΨΗ

Μια βιβλιογραφική βάση περιέχει πληροφορίες για την περιγραφή εγγράφων (δηλ. τίτλο, συγγραφείς, ημερομηνία έκδοσης, εκδότη, κ.α.) και χρησιμοποιείται για την οργάνωση και αναζήτησή τους. Μια τέτοια βάση θα μπορούσε να περιέχει και σημειώσεις σχετικά με το πριεχόμενο των εγγράφων. Η αξία της καταχώρησης τους στη βάση έγκειται στο ότι θα μπορούν να τις μοιράζονται τα μέλη μιας ομάδας συνεργατών και μπορεί να γίνει αναζήτηση και αξιολόγηση των εγγράφων μέσω των σημειώσεων σε αυτά.

Η παρούσα εργασία ασχολείται με την παράσταση και διαχείριση βιβλιογραφίας και σημειώσεων σε αυτή. Σκοπός της εργασίας είναι η δημιουργία ενός συστήματος υπομνηματισμού, δηλ. ένα σύστημα που να υποστηρίζει την δημιουργία και διαχείριση σημειώσεων σε έγγραφα. Προορίζεται για χρήση από ένα άτομο ή μια ομάδα συνεργατών, ώστε να οργανώνουν, να μοιράζονται και να αναζητούν αποδοτικά τα έγγραφα και τις σημειώσεις τους. Η περιγραφή της πληροφορίας έγινε με την γλώσσα παράστασης γνώσης Telos.

Υλοποιήσαμε ένα μοντέλο παράστασης της λογικής δομής άρθρων και βιβλίων σύμφωνα με το πρότυπο ISO 12083. Επίσης δημιουργήσαμε τους μηχανισμούς για την πλοήγηση (navigate) στη λογική δομή του εγγράφου και την διασύνδεση αυτής με το περιεχόμενό του, που αποθηκεύεται εξωτερικά από την βάση. Στη συνέχεια υλοποιήσαμε

το μοντέλο παράστασης των σημειώσεων που μπορεί να γίνουν σε ένα έγγραφο ή τμήμα του (υπομνήματα). Συγκεκριμένα ορίσαμε, μετά από μελέτη της βιβλιογραφίας και συζητήσεις με ερευνητές, τις οπτικές γωνίες κάτω από τις οποίες γίνονται τα υπομνήματα (π.χ θεματική κατάταξη, κριτική, συσχέτιση με άλλα έγγραφα, συσχέτιση με μια εργασία, κ.α.) και τους τύπους του περιεχομένου αυτών.

Λειτουργικός πυρήνας του συστήματος που αναπτύξαμε είναι το Σημασιολογικό Σύστημα Ευρετηριασμού (ΣΣΕ), κατασκευασμένο στο Ινστιτούτο Πληροφορικής. Για την ανάκτηση πληροφορίας από το σύστημα, κατασκευάστηκαν προκαθορισμένες ερωτήσεις και κάρτες αναζήτησης λαμβάνοντας υπ'όψη τις βασικές οντότητες και τις μεταξύ τους σχέσεις. Ενώ για την ενημέρωση της βάσης χρησιμοποιήσαμε τα Δελτία Ενημέρωσης του ΣΣΕ, αφού πρώτα τα προσαρμόσαμε στο συγκεκριμένο μοντέλο δεδομένων της βάσης.

Επίσης κατά την διάρκεια σχεδίασης και υλοποίησης του συστήματος δεχθήκαμε την κριτική συνεργατών που το χρησιμοποίησαν και η κριτική αυτή οδήγησε στη βελτίωση του.

Επόπτης : Πάνος Κωνσταντόπουλος, Αναπληρωτής Καθηγητής
Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

Design and implementation of a document annotation system

Sarantis Toulis

Master of Science Thesis

Department of Computer Science
University of Crete

ABSTRACT

A bibliographic base contains information pertaining to a collection of documents (title, authors, date of publication, publisher, etc.) and is used for their organization and retrieval. Such a base could contain annotations relating to the documents' contents. The value of keeping such annotations in a database lies in that they can be shared by the members of a group, and can form the basis for document retrieval and evaluation.

This thesis deals with the representation and management of a bibliographic base and the annotations contained therein. The aim of the project is the development of a document annotation system that supports creation and management of annotations on documents. The system described is intended to be used by either a single user or a group of co–workers, in order to assist in the organization, sharing and retrieval of documents and their associated annotations. Information is described using the knowledge representation language Telos.

A representation model for the logical structure of articles and books was implemented in accordance with the ISO 12083 standard. Mechanisms for navigating within the logical document structure were developed, as well as methods for interconnecting the logical structure with the content which is stored externally with respect to the database. Following this a representation model for the annotations which can accompany a document or a part thereof was implemented. Specifically, following a literature study and discussions with

researchers, we define the points of view with which annotations are made (e.g. subject classification, critical assessment, relationships with other documents, relationships with projects, etc.), as well as the classes corresponding to the annotations' content.

The operational kernel of the Document Annotation System is the Semantic Index System (SIS), developed at the Institute of Computer Science, Foundation for Research and Technology - Hellas. In order to enhance the efficiency of user - system interaction, a set of predefined queries was developed, taking into consideration the basic notions and their interrelationships. For the update operations we have customized the SIS Entry Forms tool to our data base model.

Throughout the course of designing and implementing the system, feedback from individuals who used it was received, leading to the improvement of the system.

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

Ευχαριστίες

Ευχαριστώ τον επιβλέποντα καθηγητή μου κ. Πάνο Κωνσταντόπουλο που μου εμπιστεύθηκε αυτή την εργασία, καθώς και για την καθοδήγηση και τις συμβουλές του.

Θα ήθελα επίσης να ευχαριστήσω τα μέλη της επιτροπής εξέτασης της μεταπτυχιακής μου εργασίας καθηγητές κ.κ. Βαγγέλη Μαρκάτο και Πάνο Τραχανιά για τις χρήσιμες παρατηρήσεις τους που συνέβαλαν στην τελική μορφή του κειμένου αυτού.

Επίσης ευχαριστώ τον δρ. Martin Dörr για τις συμβουλές του καθώς και τους συνεργάτες και φίλους Γιάννη Τζίτζικα, Μάνο Θεωδοράκη, Λάμπρο Γκανά, Μάριο Συντιχάκη αλλά και όλη την ομάδα Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του Ινστιτούτου Πληροφορικής για την συνεργασία και την υποστήριξή τους.

Ακόμη ευχαριστώ τους γονείς μου , Αφροδίτη και Δημήτριο, και την Πολυτίμη Σούρβα για τη βοήθεια και την συμπαράστασή τους σε όλη τη διάρκεια των σπουδών μου.

Τέλος, θα ήθελα να ευχαριστήσω το Ινστιτούτο Πληροφορικής του Ιδρύματος Τεχνολογίας Έρευνας για την υλικοτεχνική υποδομή και την οικονομική ενίσχυση που μου παρείχε κατά τη διάρκεια των μεταπτυχιακών μου σπουδών.

Περιεχόμενα

Περίληψη **i**

Abstract **iii**

Ευχαριστίες **v**

1 Εισαγωγή **1**

1.1 Ορισμός του προβλήματος 1

1.2 Οι στόχοι της εργασίας 3

1.3 Η οργάνωση της γραπτής εργασίας 4

1.4 Το περιβάλλον υλοποίησης 5

2 Βιβλιογραφική έρευνα και επιλογές για την προσέγγιση μας **7**

2.1 Τύποι και είδη σημειώσεων 8

2.2 Τα σημεία σημειώσεων σε ένα έγγραφο 11

2.3 Σχέσεις μεταξύ σημειώσεων 12

2.4 Η χρήση των σημειώσεων 13

2.5 Αξιολόγηση σημειώσεων 13

2.6 Δημιουργία και συντήρηση σημειώσεων 14

2.7 Αποθήκευση των σημειώσεων 15

2.8 Προηγούμενα Συστήματα Σημειώσεων 17

3 Η γλώσσα Telos και το σύστημα SIS **21**

3.1 Σημαντικά χαρακτηριστικά της Telos 21

3.2 Παράσταση γνώσης στη γλώσσα Telos 22

3.3 Το Σύστημα Σημασιολογικού Ευρετηριασμού 27

3.3.1 Σύντομη περιγραφή της Βιβλιοθήκης Στοιχειωδών Συναρτήσεων (ΒΣΕΣ) 28

4 Το μοντέλο παράστασης εγγράφων με την γλώσσα Telos **31**

4.1 Η μοντελοποίηση εγγράφων 32

4.1.1 Standard Generalized Markup Language (SGML) 32

4.1.2 Open/Office Document Architecture (ODA) 35

4.1.3 Σύγκριση ODA και SGML 35

4.2 Το γενικό μοντέλο του εγγράφου 37

4.3 Το μοντέλο παράστασης του άρθρου σύμφωνα με το ISO 12083 38

4.3.1 Το SGML DTD 38

4.3.2 Η παράσταση του ISO 12083 με την γλώσσα Telos 39

4.4 Εισαγωγή δεδομένων 44

4.4.1 Αυτόματη εισαγωγή δεδομένων 44

4.4.2 Διαλογική εισαγωγή δεδομένων 46

4.5 Παρατηρήσεις 48

5 Το μοντέλο παράστασης σημειώσεων σε έγγραφα **51**

5.1 Το περιεχόμενο της σημείωσης 51

5.2 Το γενικό μοντέλο παράστασης σημειώσεων 53

5.3 Οι κατηγορίες σημειώσεων 56

5.3.1 Κατηγορίες σημειώσεων που εκφράζουν ταξινόμηση 57

5.3.2 Σημειώσεις σχετικά με τις εργασίες 59

5.3.3 Σημειώσεις των κύριων σημείων του εγγράφου 61

5.3.4 Σημειώσεις που συσχετίζουν έγγραφα μεταξύ τους 63

5.4 Οι τύποι των περιεχομένων της σημείωσης 64

5.4.1 Ο θεματικός θησαυρός όρων (SubjectThesaurus) 65

5.4.2 Οι τίτλοι εργασίας (WorkTitle) 67

5.4.3 Η φύση του εγγράφου (DocumentNature) 69

5.4.4 Η κριτική του εγγράφου (Criticism) 69

5.5 Η ενημέρωση σημειώσεων 70

5.6 Παρατηρήσεις 72

6 Η Χρήση του συστήματος **75**

6.1 Σύντομη περιγραφή της επαφής χρήσης του ΣΣΕ και του μοντέλου των προκαθορισμένων ερωτήσεων 76

6.2 Ερωτήσεις 1ης τάξεως 78

6.2.1 Η επιλογή Articles 81

6.2.2 Η επιλογή Notes 83
6.2.3 Η επιλογή ‘‘Notes by Subject’’ 84
6.2.4 Η επιλογή ‘‘Notes by Work Title’’ 85
6.2.5 Η επιλογή ‘‘Main Point Notes’’ 86
6.2.6 Η επιλογή ‘‘Notes by Existing Document’’ 87
6.2.7 Η επιλογή ‘‘Notes by Criticism’’ 87
6.2.8 Η επιλογή ‘‘Notes by Document Nature’’ 88
6.2.9 Η επιλογή Proposals 88
6.2.10 Η επιλογή Subject Terms 89
6.2.11 Η επιλογή Work Titles 90
6.2.12 Η επιλογή Persons 91
6.3 Αναζήτηση με πολλές παραμέτρους 92
6.4 Αναδρομικές προκαθορισμένες ερωτήσεις 95
6.5 Τα εργαλεία υποστήριξης του συστήματος 98

7 Συμπεράσματα και μελλοντικές κατευθύνσεις 101
7.1 Η λειτουργία του συστήματος 101
7.2 Αξιολόγηση του συστήματος 103
7.3 Μελλοντικές επεκτάσεις 108
7.4 Επίλογος 109

A Ερωτήσεις για την αξιολόγηση του συστήματος 111

B To Document Type Definition (DTD) του ISO12083 για τα άρθρα 115

Βιβλιογραφία 124

Κατάλογος Σχημάτων

3.1 Οι κλάσεις του συστήματος στην Telos 23
3.2 Ο μηχανισμός ταξινόμησης 24
3.3 Περιορισμός στις περιπτώσεις γνωρισμάτων στην Telos 25
3.4 Ο μηχανισμός γενίκευσης 26
3.5 Περιορισμός στις υποκλάσεις γνωρισμάτων 27

4.1 Παράδειγμα SGML DTD 33
4.2 Το SGML μοντέλο επεξεργασίας 34
4.3 Το μεταμοντέλο παράστασης εγγράφων 37
4.4 Το μεταμοντέλο παράστασης Συλλογών εγγράφων 38
4.5 Η παράσταση του κανόνα *part* $->$ *(title?, section_subelement*, chapter+)* . 40
4.6 Παραδείγματα για τους δύο εναλλακτικούς τρόπους παράστασης του κανόνα:*authorgroup* $->$ *(author* | *corporation_author)* 41
4.7 Η παράσταση του κανόνα:*biblist* $->$ *(head, citation)* * 42
4.8 Μια περίπτωση άρθρου 43
4.9 Το γενικό μοντέλο μετάφρασης από SGML σε O_2 44
4.10 Εισαγωγή της λογικής δομής εγγράφου 47
4.11 Το δελτίο ενημέρωσης άρθρου και το δελτίο ενημέρωσης του κυρίως κορμού (body) του συγκεκριμένου εγγράφου 48

5.1 Το γενικό μοντέλο παράστασης σημειώσεων 53
5.2 Πρώτη προσέγγιση για τις εξειδικεύσεις της κατηγορίας *annotation_type* . . 54
5.3 Δεύτερη προσέγγιση για τις εξειδικεύσεις της κατηγορίας *annotation_type* . 55
5.4 Η κατηγορία *annotation* 56
5.5 Οι κατηγορίες και μετακατηγορίες σημειώσεων 57
5.6 Οι κατηγορίες σημείωσης που δηλώνουν ταξινόμηση 58
5.7 Οι κατηγορίες σημείωσης που συσχετίζουν ένα έγγραφο με μια εργασία . . 60

5.8 Οι τύποι των περιεχομένων της σημείωσης. 65
5.9 Η δομή του θεματικού θησαυρού 66
5.10 Όροι θεματικού θησαυρού και οι συσχετίσεις τους 67
5.11 Ένα παράδειγμα τίτλων εργασίων 68
5.12 Ενημέρωση σημειώσεων σε ένα τμήμα του εγγράφου 71
5.13 Ενημέρωση της λογικής δομή και των σημειώσεων σε ένα τμήμα του εγγράφου 72

6.1 Η επαφή χρήσεως του ΣΣΕ. 76
6.2 Προκαθορισμένες ερωτήσεις σε μορφή γράφου 78
6.3 Αποτέλεσμα προκαθορισμένης ερώτησης σε μορφή κειμένου 79
6.4 Ένα υποσύνολο του μοντέλου των προκαθορισμένων ερωτήσεων 80
6.5 Ένα παράδειγμα παρουσίασης αποτελεσμάτων 84
6.6 Η κάρτα αναζήτησης εγγράφων 93
6.7 Τα λογικά τμήματα του *Article* 96
6.8 Η λογική δομή του άρθρου SIB 97
6.9 Τα εργαλεία υποστήριξης του συστήματος 99

Κεφάλαιο 1

Εισαγωγή

1.1 Ορισμός του προβλήματος

Τα έγγραφα αποτελούν ένα σημαντικό μέσο για την οργάνωση και μετάδοση των πληροφοριών. Στο παρελθόν τα έγγραφα υπήρχαν αποκλειστικά σε έντυπη μορφή αλλά στις τελευταίες δεκαετίες, με την ανάπτυξη των ηλεκτρονικών επεξεργαστών κειμένου, εμφανίζονται σε ηλεκτρονική μορφή, αφού με αυτόν τον τρόπο γίνεται ευκολότερη η δημιουργία και εκτύπωση τους. Αλλά με την ανάπτυξη και επέκταση των δικτύων, στις περισσότερες γεωγραφικές περιοχές του πλανήτη μας, γίνεται ευκολότερη και η μετάδοση των ηλεκτρονικών εγγράφων. Έτσι εμφανίζεται η ανάγκη για την δημιουργία και διαχείριση μιας ηλεκτρονικής βιβλιοθήκης. Άρα στο προσεχές μέλλον οι αναγνώστες θα έχουν τα βιβλία και τα άρθρα των περιοδικών τους και σε ηλεκτρονική μορφή, δηλαδή η προσωπική τους βιβλιοθήκη θα είναι ηλεκτρονική.

Βέβαια ένα ηλεκτρονικό έγγραφο μπορεί να μην αποτελείται μόνο από απλό κείμενο αλλά μπορεί να είναι με την μορφή μη γραμμικού κειμένου (hypertext) και να περιέχει πολύμορφα (multimedia) δεδομένα (ήχος, εικόνα, βίντεο, κ.α.). Επομένως παρουσιάζεται η ανάγκη για την δημιουργία εργαλείων που διευκολύνουν την αναζήτηση και περιπλάνηση (browsing) στα έγγραφα της προσωπικής ηλεκτρονικής βιβλιοθήκης. Για την διευκόλυνση αυτής της ανάγκης απαιτείται τουλάχιστον η γνώση της λογικής δομής του εγγράφου. Όπως είναι γνωστό κάθε έγγραφο έχει μια λογική δομή, δηλαδή τα λογικά τμήματα από τα οποία αποτελείται (π.χ. κεφάλαιο, ενότητα, παράγραφος, κ.α.), έχει μια ή περισσότερες μορφές εμφάνισης (layout) και έχει ένα περιεχόμενο (content).

Όμως όταν κάποιος (π.χ. ένας ερευνητής) έχει μια μεγάλη συλλογή από έγγραφα

παρουσιάζεται η ανάγκη οργάνωσης της προσωπικής ηλεκτρονικής βιβλιοθήκης. Δηλαδή θα πρέπει να έχει την δυνατότητα να κατατάσσει τα έγγραφα σε κατηγορίες (π.χ. θεματική κατηγορία) ή η κατάταξη να γίνεται αυτόματα με μια από τις μεθόδους που υπάρχουν (π.χ. [MR88]), και να έχει την δυνατότητα να κάνει σημειώσεις, π.χ. να συσχετίζει ένα τμήμα ενός εγγράφου με μια συγκεκριμένη εργασία του ή με μια ενότητα ενός άλλου εγγράφου. Με τον τρόπο αυτό θα μπορεί να βρίσκει γρήγορα και με ακρίβεια τα έγγραφα, και συγκεκριμένα τα τμήματα του εγγράφου, που έχει μελετήσει και τον ενδιαφέρουν για ένα συγκεκριμένο θέμα. Άρα εμφανίζεται η ανάγκη για την συγκρότηση μιας βάσης δεδομένων με τις σημειώσεις που έχουν γίνει πάνω στα έγγραφα.

Η μεταπτυχιακή αυτή εργασία ασχολείται με τον υπομνηματισμό, δηλαδή την δημιουργία και διαχείριση σημειώσεων σε έγγραφα. Συγκεκριμένα ασχοληθήκαμε με τα εξής ζητήματα:

- Την μοντελοποίηση και την αυτόματη εξαγωγή της λογικής δομής των εγγράφων.
- Την μοντελοποίηση των υπομνημάτων που μπορεί να γίνουν σε ένα επιστημονικό έγγραφο. Δηλαδή τον διαχωρισμό των υπομνημάτων σε κατηγορίες και τον προσδιορισμό των τύπων του περιεχομένου του υπομνήματος.
- Την προσαρμογή του Συστήματος Σημασιολογικού Ευρετηριασμού (Semantic Index System ή SIS) και των Δελτίων Ενημέρωσης (Entry Forms) στο συγκεκριμένο μοντέλο παράστασης των εγγράφων και των σημειώσεων σε αυτά, ώστε να γίνεται η εύρεση και εισαγωγή πληροφοριών (έγγραφα και σημειώσεις) που αποθηκεύονται σε μια βάση γνώσεων σύμφωνα με το οντοκεντρικό μοντέλο.

Το Σύστημα Σημασιολογικού Ευρετηριασμού (ΣΣΕ) μπορεί να διαχειρίζεται αποδοτικά μεγάλους πληθυσμούς από δεδομένα πολλαπλώς διασυνδεδεμένα. Χρησιμοποιείται ως επαφή χρήσης (interface) με βάσεις δεδομένων που δημιουργούνται σύμφωνα με την γλώσσα παράστασης γνώσης Telos [MBJK90].Η γλώσσα αυτή υποστηρίζει ένα δομικά οντοκεντρικό μοντέλο δεδομένων στο οποίο οι βασικές μονάδες είναι οι *Οντότητες*, και τα *Γνωρίσματα*. Οι *Οντότητες* είναι οι έννοιες της θεματικής περιοχής που περιγράφεται ενώ τα *Γνωρίσματα* είναι οι συσχετίσεις μεταξύ των *Οντοτήτων*.

Το οντοκεντρικό μοντέλο επιτρέπει την ‘‘φυσική’’ παράσταση της δομής και της σημασίας των εννοιών του αυτών μέσω των μηχανισμών ταξινόμησης, γενίκευσης και απόδοσης γνωρίσματος που χρησιμοποιεί (η λειτουργία των μηχανισμών αυτών περιγράφεται στο υποκεφάλαιο 3.2). Βασικό χαρακτηριστικό των οντοτήτων είναι ότι σε καθεμιά αντιστοιχεί ένα αναγνωριστικό όνομα, ως στοιχείο ταυτότητας. Έτσι ενώ είναι

δυνατόν να αλλάξουν τα γνωρίσματά της, όσο η σημασία της παραμένει η ίδια, όλες οι αναφορές σε αυτήν την οντότητα παραμένουν σωστές ακόμη και αν αναδιοργανωθεί λογικά ή και φυσικά η βάση δεδομένων. Αυτό παρέχει υψηλού βαθμού αφαίρεση και ανεξαρτησία στα δεδομένα. Γι' αυτό το λόγο οι οντοκεντρικές βάσεις δεδομένων χρησιμοποιούνται σε συστήματα CAD/CAM, γεωγραφικά πληροφοριακά συστήματα, συστήματα αυτοματισμού γραφείου [KLK89], όπου οι έννοιες που χρησιμοποιούνται είναι αρκετά πολύπλοκες για να περιγραφούν από τις εγγραφές μιας σχεσιακής βάσης δεδομένων. Για παράδειγμα τα κεφάλαια ενός εγγράφου δεν έχουν τον ίδιο αριθμό από ενότητες και οι ενότητες δεν έχουν τον ίδιο αριθμό από παραγράφους, πίνακες, λίστες και εικόνες. Επίσης δεν γνωρίζουμε πόσες και τί τύπου σημειώσεις θα ήθελε να κάνει κάποιος σχολιαστής σε ένα σημείο ενός εγγράφου.

1.2 Οι στόχοι της εργασίας

Το πληροφοριακό σύστημα που δημιουργήσαμε παρέχει αυτόματους και διαλογικούς μηχανισμούς για την δημιουργία μιας βιβλιογραφικής βάσης καθώς και σημειώσεων πάνω στα έγγραφα που περιέχει. Επίσης αν υπάρχει η λογική δομή του εγγράφου, η οποία αποθηκεύεται στην βάση, και το περιεχόμενο του, που αποθηκεύεται εξωτερικά, τότε υπάρχουν οι μηχανισμοί για την περιπλάνηση στην λογική δομή και το περιεχόμενο του εγγράφου ώστε να βρούμε τις πληροφορίες που ζητάμε και να κάνουμε σημειώσεις σε αυτές.

Η βιβλιογραφική βάση, εμπλουτισμένη με τις σημειώσεις των μελετητών, αποτελεί ένα ισχυρό μέσο για την οργάνωση και διαχείριση της προσωπικής τους βιβλιοθήκης, αλλά και για την συνεργασία των μελών της ομάδας στην οποία ανήκουν. Ένα βασικό σημείο για την αξία αυτής της βάσης δεδομένων είναι η επιλογή της κατηγορίας συσχέτισης ενός τμήματος του εγγράφου με μια σημείωση. Η κατηγορία αυτή δηλώνει την οπτική γωνία παρατήρησης κάτω από την οποία έγινε η σημείωση. Άρα για μια σημείωση δεν έχουμε μόνο μια φράση ή ένα κείμενο ή ένα σχέδιο αλλά έχουμε πληροφορία πριν καν τη διαβάσουμε, δηλαδή προσδιορίζεται η οπτική γωνία κάτω από την οποία έγινε η σημείωση. Με τον τρόπο αυτό γίνεται ευκολότερη η αναζήτηση εγγράφων που έχουν σχολιαστεί για ένα συγκεκριμένο θέμα και κάτω από μια οπτική γωνία παρατήρησης.

Η οργάνωση της προσωπικής βιβλιοθήκης ενός ερευνητή κάνει πιο εύκολη και ακριβή την αναζήτηση πληροφορίας σχετικά με τα άρθρα που έχει μελετήσει. Οι σημειώσεις που κάνει σε ένα άρθρο είναι σχετικές με τα θέματα που τον ενδιαφέρουν και παρέχουν

σημαντική βοήθεια στην συγγραφή των δικών του άρθρων. Άρα ο συγγραφέας θα έχει τις σημειώσεις, στα άρθρα που έχει μελετήσει, και θα μπορεί να αναζητήσει και να βρει εύκολα, μέσω της ταξινόμησης και των λέξεων-κλειδιών, αυτές που τον ενδιαφέρουν ώστε να γράψει το άρθρο του. Δηλαδή μπορεί να συγκεντρώσει τις σημειώσεις για ένα συγκεκριμένο θέμα ώστε να διευκολυνθεί στην δημιουργία του σκελετού του άρθρου αλλά και στην ανάπτυξη και τεκμηρίωση του. Επίσης μπορεί να οργανώνει καλύτερα τις παραπομπές του σε άλλα άρθρα γιατί, μέσω των σημειώσεων που έχει κάνει, θα μπορεί να βρίσκει και να αιτιολογεί τις παραπομπές του.

Όμως κάθε ερευνητής ανήκει σε μια ομάδα ερευνητών και όλοι μαζί συνεργάζονται σε κάποιες θεματικές περιοχές εργασίας. Επομένως οι σημειώσεις των άρθρων συγκεντρώνονται σε μια βάση δεδομένων, διατηρώντας την ταυτότητα αυτού που τα σημείωσε, και βρίσκονται στην διάθεση των μελών της ομάδας. Έτσι τα μέλη τις ομάδας μπορούν να βλέπουν τις σημειώσεις ενός άλλου μέλους , και αφού γνωρίζουν τα ενδιαφέροντα και την ωριμότητά του, να επιλέγουν ποιά άρθρα να διαβάσουν, να διευκολύνονται στην κατανόηση του άρθρου, να κάνουν σημειώσεις σχετικά με τις σημειώσεις των άλλων (οι οποίες πυροδοτούν μια συζήτηση που θα γίνει σε κάποια συνάντηση), να προβληματίζονται σχετικά με τις μελλοντικές εργασίες. Έτσι θα δημιουργηθεί ένα μέσο για την συνεργασία των μελών της ομάδας στην μελέτη άρθρων , η οποία βοηθάει στην εξέλιξη της εργασίας τους.

1.3 Η οργάνωση της γραπτής εργασίας

Στο *δεύτερο κεφάλαιο* εντάσσεται το πρόβλημα της εργασίας αυτής στην ευρύτερη ερευνητική περιοχή που ανήκει. Παρουσιάζονται τα θέματα που απασχολούν την περιοχή αυτή και αξιολογούνται διάφορες προσεγγίσεις. Επίσης για κάθε θέμα παρουσιάζεται και η προσέγγιση που επιλέξαμε ή επινοήσαμε για την υλοποίηση του συστήματος που περιγράφεται σε αυτήν την εργασία.

Στο *τρίτο κεφάλαιο* παρουσιάζεται συνοπτικά η γλώσσα παράστασης γνώσης Telos και το Σύστημα Σημασιολογικού Ευρετηριασμού (ΣΣΕ). Περιγράφονται οι βασικοί μηχανισμοί παράστασης γνώσης που χρησιμοποιούνται στην Telos καθώς και οι δυνατότητες του ΣΣΕ ως επαφή χρήσης (interface) για αναζήτηση και παρουσίαση των πληροφοριών.

Στο *τέταρτο κεφάλαιο* παρουσιάζεται το μοντέλο της λογικής δομής των εγγράφων και συγκεκριμένα επιστημονικών άρθρων και βιβλίων. Επισημαίνονται οι δυνατότητες για

αυτόματη άντληση της λογικής δομής του εγγράφου και η συσχέτιση του μοντέλου με τον τρόπο αποθήκευσης των εγγράφων. Επίσης περιγράφονται οι διαδικασίες εισαγωγής των εγγράφων με αυτόματο τρόπο ή με την χρήση των Δελτίων Ενημέρωσης.

Στο *πέμπτο κεφάλαιο* παρουσιάζεται το μοντέλο παράστασης των σημειώσεων. Περιγράφονται οι κατηγορίες των συσχετίσεων ενός εγγράφου με μια σημείωση καθώς και οι τύποι του περιεχομένου των σημειώσεων που υποστηρίζει το μοντέλο. Επίσης περιγράφονται οι διαδικασίες εισαγωγής σημειώσεων στα έγγραφα με την χρήση των Δελτίων Ενημέρωσης.

Στο *έκτο κεφάλαιο* παρουσιάζεται η λειτουργία του συστήματος υπομνηματισμού εγγράφων. Περιγράφεται το μοντέλο ερωτήσεων για την αναζήτηση και παρουσίαση των πληροφοριών με την χρήση του Συστήματος Σημασιολογικού Ευρετηριασμού [CD94].

Στο *έβδομο κεφάλαιο* περιγράφεται η αξιολόγηση του συστήματος και τα συμπεράσματα που προέκυψαν. Πέντε άτομα χρησιμοποιήθηκαν σε μια διαδικασία που περιλάμβανε τόσο ευρηματική όσο και εμπειρική αξιολόγηση. Επίσης αναφέρονται και οι προτάσεις για μελλοντικές επεκτάσεις του συστήματος.

1.4 Το περιβάλλον υλοποίησης

Το μοντέλο των πληροφοριών, που διαχειρίζεται το σύστημα, υλοποιήθηκε στην γλώσσα Telos που περιγράφεται στο [MBJK90]. Η συγκεκριμένη υλοποίηση της γλώσσας έγινε στο Ινστιτούτο Πληροφορικής του Ιδρύματος Τεχνολογίας και Έρευνας (Ι.Τ.Ε.). Η επαφή χρήσης (interface) του ΣΣΕ προσαρμόστηκε στο μοντέλο που υλοποιήσαμε (βλ. κεφάλαιο 6).

Κεφάλαιο 2

Βιβλιογραφική έρευνα και επιλογές για την προσέγγιση μας

Τα έγγραφα αποτελούν ένα σημαντικό μέσο για την οργάνωση και διάδοση των πληροφοριών. Όταν διαβάζουμε ένα έγγραφο (π.χ. βιβλίο, άρθρο περιοδικού, εφημερίδα, μια σελίδα HTML στο World Wide Web, κ.α.) κάνουμε ορισμένες παρατηρήσεις για τα θέματα και τις απόψεις που αναπτύσσει (π.χ. ίσως να αποκτούμε κάποιες νέες πληροφορίες σχετικά με τα ενδιαφέροντα μας, να επιβεβαιώνουμε κάποιες σκέψεις μας σχετικά με ένα θέμα, κ.α.). Αυτές οι παρατηρήσεις μπορούν να σημειωθούν σε κάποιο σημείο του εγγράφου, για να μην τις ξεχάσουμε. Όμως πώς θα μπορέσουμε να βρούμε τις παρατηρήσεις μας (π.χ. για ένα θέμα) όταν το θελήσουμε; Μήπως θα πρέπει να θυμόμαστε σε ποιά έγγραφα έχουν γίνει ή θα πρέπει να έχουμε φυλάξει όλα τα έγγραφα σε γνωστή ή ευδιάκριτη θέση του γραφείου μας ώστε να ξεφυλλίσουμε όλα εκείνα που ίσως έχουν κάποια σχέση; Αυτές οι λύσεις είναι δύσκολο να εφαρμοστούν με επιτυχία. Άρα δημιουργείται η ανάγκη για την διαχείριση των σημειώσεων μας ώστε να μπορούμε να βρούμε γρήγορα αυτό ακριβώς που μας ενδιαφέρει για να μας βοηθήσει στο θέμα εργασίας μας.

Πριν την εμφάνιση και εξάπλωση των υπολογιστών τα συστήματα διαχείρισης σημειώσεων λειτουργούσαν με την χρήση δελτίων[1] για κάθε έγγραφο. Σε ένα δελτίο καταγράφονται οι παρατηρήσεις για ένα έγγραφο και όλα τα δελτία ταξινομούνται σύμφωνα με κάποιο κριτήριο (π.χ. θεματικά, αλφαβητικά , κ.α.). Με την εξάπλωση όμως της χρήσης των υπολογιστών, η διαχείριση των σημειώσεων άρχισε να γίνεται

[1]Το δελτίο αυτό είναι συνήθως ένα μικρό σκληρό χαρτί (καρτέλα).

μέσω υπολογιστών. Έχουν δημιουργηθεί μερικά τέτοια συστήματα και είχαν καλά αποτελέσματα. Τα θέματα που απασχολούν αυτά τα συστήματα αναλύονται στις επόμενες ενότητες.

2.1 Τύποι και είδη σημειώσεων

Υπάρχουν διαφορετικά είδη και τύποι σημειώσεων που γίνονται σε ένα έγγραφο. Τα τρία βασικά είδη είναι τα εξής:

- Ένα έγγραφο που δημιουργείται για να εκφράσει την παρατήρηση. Αυτό είναι ένα έγγραφο με την γενικότερη έννοια, δηλαδή μπορεί να είναι κείμενο ή εικόνα ή ήχος ή βίντεο ή ένας συνδυασμός των προηγουμένων.

- Ένας όρος, που αποτελείται από μία ή περισσότερες λέξεις, ενός ελεγχόμενου λεξιλογίου [Sve89] που εκφράζει τις έννοιες μιας θεματικής περιοχής. Επίσης μεταξύ των όρων υπάρχουν συσχετίσεις. Έτσι ένα σημείο του εγγράφου μπορεί να σημειώνεται με έναν όρο που έχει μια σημασία για αυτούς που αποδέχονται το συγκεκριμένο ελεγχόμενο λεξιλόγιο.

- Ένας σύνδεσμος, ο οποίος συνδέει το τμήμα του εγγράφου που θέλουμε να σημειώσουμε με ένα άλλο υπαρκτό έγγραφο το οποίο βρίσκεται σε μια βάση δεδομένων ή γενικότερα κάπου στο δίκτυο (π.χ. ένα έγγραφο που έχει το ίδιο θέμα, το αρχείο του πηγαίου κώδικα του σχετικού προγράμματος, μια συλλογή από ηλεκτρονικά μηνύματα που σχολιάζουν το έγγραφο, κ.α.). Αυτό το είδος ονομάζεται αναφορική σημείωση (reference annotation).

Δύο άλλα λιγότερο σημαντικά είδη, που περιγράφονται στη εργασία του Nielsen [Nie84], είναι η υπογράμμιση τμημάτων του εγγράφου και η σημείωση της δομής ενός τμήματος του εγγράφου, π.χ. μιας παραγράφου. Η υπογράμμιση έχει ως στόχο να ξεχωρίσει ορισμένα τμήματα του εγγράφου και μπορεί να γίνει είτε με διαφορετικές γραμμές είτε με διαφορετικά χρώματα. Η διαφορετική υπογράμμιση είτε προσδιορίζει διαφορετικό βαθμό σπουδαιότητας (π.χ. το πράσινο για πολύ σημαντικό και το κίτρινο για σημαντικό θέμα) είτε διαφορετικό είδος πληροφορίας (π.χ. το κόκκινο για τους ορισμούς και το μπλε για τα θεωρήματα). Ενώ η σημείωση της δομής ενός τμήματος του εγγράφου έχει ως στόχο είτε να ξεκαθαρίσει το νόημα του κειμένου είτε να συσχετίσει τα πιο ενδιαφέροντα σημεία. Ο προσδιορισμός της δομής γίνεται είτε με την αρίθμηση των δομικών σημείων είτε με την απόδοση λέξεων κλειδιών σε αυτά.

Το είδος της σημείωσης ασχολείται με τον τρόπο με τον οποίο γίνεται η σημείωση, ενώ ο τύπος της σημείωσης αναφέρεται στο νόημα της σημείωσης. Ένα σημαντικό θέμα σχετικά με τους τύπους σημειώσεων είναι η πληρότητα του συνόλου των τύπων. Μπορούμε να αφήσουμε τον σχολιαστή να διαλέξει ή να ορίσει τον τύπο της σημείωσης του (open ended set) ή να τον υποχρεώνουμε να εντάξει την σημείωση του σε ένα τύπο από ένα σύνολο τύπων που έχουμε ορίσει (fixed set). Στην πρώτη περίπτωση το πλεονέκτημα είναι ότι μπορούμε κάθε φορά που κάνουμε μια σημείωση να ορίζουμε, αν δεν υπάρχει, τον κατάλληλο τύπο, αλλά το μειονέκτημα είναι ότι δεν μπορούμε να κάνουμε επιλογή σημειώσεων με βάση τους τύπους που υπάρχουν, δηλαδή υπάρχει πρόβλημα στην οργάνωση και αναζήτηση τους. Ενώ στην δεύτερη περίπτωση, αν επιλέξουμε κατάλληλα ένα σύνολο τύπων, θα κάνουμε το σύστημα πιο εύχρηστο. Ένας τύπος σημείωσης είναι η *Απάντηση* (Response) που χρησιμοποιείται από το σύστημα Hypernews [LaL94]. Σε αυτό το σύστημα για να κάνουμε μια σημείωση τύπου *Απάντησης* σε μια είδηση (News) που διαβάσαμε στέλνουμε μια δική μας είδηση που είναι απάντηση σε αυτήν που διαβάσαμε. Επίσης στο HyperNews χρησιμοποιούνται και οι εξής τύποι σημειώσεων: ερώτηση, συμφωνία, διαφωνία, γενικό σχόλιο. Ένα άλλο παράδειγμα τύπων συναντάμε στο WIT [Mar], το οποίο είναι ένα σύστημα για την συζήτηση προτάσεων. Συγκεκριμένα στο σύστημα WIT ορίζονται κάποια *Θέματα* πάνω στα οποία γίνονται ορισμένες *Προτάσεις*. Για κάθε πρόταση γίνονται σημειώσεις, που ονομάζονται *Επιχειρήματα* και είναι τύπου *Αρνητικό* ή *Θετικό* ή *Ουδέτερο*. Ένα άλλο παράδειγμα είναι η μελέτη της Principia Cybernetica[2], που προτείνει τους εξής τύπους:

- Απόρριψη(refutation). Είναι μια απόρριψη της άποψης που εκφράζεται σε αυτό το σημείο του άρθρου.

- Συμφωνία/Αποδοχή (confirmation). Είναι μια αποδοχή της άποψης που εκφράζεται σε αυτό το σημείο του άρθρου.

- Διόρθωση (correction). Είναι μια διόρθωση της άποψης που εκφράζεται σε αυτό το σημείο του άρθρου.

- Εικονογράφηση (illustration). Είναι μια εικονογράφηση αυτού του σημείου του άρθρου.

- Σχόλιο (comment). Είναι μια σημείωση που δεν ανήκει στις παραπάνω κατηγορίες.

[2]Η σχετική σελίδα στο WWW είναι: URL:http://pespmc1.vub.ac.be/Default.html

Οι τύποι σημειώσεων αποδίδονται είτε στην σύνδεση που συσχετίζει το έγγραφο με την σημείωση είτε στο έγγραφο που αποτελεί την ίδια την σημείωση. Συγκεκριμένα στο τρίτο είδος σημείωσης (αναφορική σημείωση) ο τύπος αποδίδεται στην σύνδεση διότι το έγγραφο που χρησιμοποιείται ως σχόλιο έχει ήδη μια υπόσταση. Οι Harrison και Munson μελέτησαν αυτό το είδος σημείωσης [HM93] και πρότειναν βελτιώσεις στο BIBTEX [Che86], το οποίο εκτός από την αναφορική σημείωση παρέχει την δυνατότητα να γίνονται σχόλια σε κάθε αναφορά σε άλλο έγγραφο, δηλαδή σε κάθε αναφορά υπάρχει ένα πεδίο που ονομάζεται ANNOTE στο οποίο σχολιάζεται η αναφορά. Αλλά και στα hypermedia συστήματα κάθε σύνδεση μπορεί να θεωρηθεί ως μία σημείωση (τύπου επεξήγησης) με ένα άλλο έγγραφο, εικόνα, ήχο, βίντεο. Ακόμη οι σημειώσεις που γίνονται με συνδέσεις μας βοηθούν να εκφράσουμε τις συσχετίσεις ενός εγγράφου με το περιβάλλον του [CFP95], ώστε να μπορούμε να εντοπίσουμε τα έγγραφα προσδιορίζοντας το περιβάλλον τους.

Το δεύτερο είδος σημείωσης, δηλαδή σύναψη όρου (term reference), είναι ουσιαστικά μια ταξινόμηση ενός εγγράφου ή τμήματος του. Όπως και στο τρίτο είδος έτσι και στο δεύτερο ο τύπος της σημείωσης αποδίδεται στον σύνδεσμο που συσχετίζει το έγγραφο με τον όρο. Υπάρχουν αρκετά συστήματα που κάνουν ταξινόμηση εγγράφων είτε με στατιστικές μεθόδους [Gor88] είτε με την βοήθεια ειδικών ανθρώπων που γνωρίζουν την θεματική περιοχή ταξινόμησης. Βέβαια το σημαντικότερο θέμα σε αυτό το είδος σημείωσης είναι η δημιουργία και συντήρηση των όρων του ελεγχόμενου λεξιλογίου και των συσχετίσεων που υπάρχουν μεταξύ τους [Sve89]. Ο Διεθνής Οργανισμός Τυποποίησης ISO έχει εκδόσει ένα σχετικό πρότυπο [ISO86]. Επίσης οι Milstead και Berger έχουν δημοσιεύσει μια εργασία τους [MB93] για την δημιουργία ενός θησαυρού όπου περιγράφουν τα θέματα που τους απασχόλησαν, τις αποφάσεις που πήραν και την υλοποίηση στη οποία κατέληξαν. Ενώ άλλες εργασίες περιγράφουν στρατηγικές για την αύξηση ενός θησαυρού [RM87], για την συνεργασία ενός θησαυρού με έναν άλλο μέσω κανόνων μετάφρασης [SK93] και την συγχώνευση θησαυρών [MR88]. Οι όροι των ελεγχόμενων λεξιλογίων μπορεί να εκφράζουν γενικές έννοιες μιας επιστήμης, τα θέματα εργασίας μιας ομάδας ανθρώπων, τις έννοιες για τον χαρακτηρισμό της φύσης ενός εγγράφου, τις έννοιες για την κριτική ενός εγγράφου, κ.α..

Σε αυτή την μεταπτυχιακή εργασία μελετήθηκαν τα παραπάνω είδη σημειώσεων, και ιδιαίτερα το δεύτερο είδος[3] ώστε οι σημειώσεις να γίνονται γρήγορα, να είναι εκφραστικές και να διευκολύνουν την αναζήτηση τους. Δηλαδή γίνονται γρήγορα γιατί κατά την εισαγωγή μιας σημείωσης διαλέγουμε ένα τύπο σημείωσης και έπειτα έναν όρο από το

[3]Δηλαδή σημείωση με όρο(term) ενός ελεγχόμενου λεξιλογίου

ελεγχόμενο λεξιλόγιο, είναι εκφραστικές γιατί παρέχεται η δυνατότητα να οριστούν όροι και να συσχετιστούν μεταξύ τους ώστε να εκφράσουν τα θέματα που μας ενδιαφέρουν και διευκολύνουν την αναζήτηση τους γιατί μπορούμε να εκφράσουμε ερωτήσεις αναζήτησης συνδυάζοντας τους όρους του λεξιλογίου.

2.2 Τα σημεία σημειώσεων σε ένα έγγραφο

Όταν διαβάζουμε ένα έγγραφο ίσως να θέλουμε να κάνουμε σημειώσεις σε οποιοδήποτε σημείο του. Πώς όμως αυτές οι σημειώσεις γίνονται και εμφανίζονται σε σχέση με το έγγραφο; Οι τεχνικές που χρησιμοποιούνται είναι οι εξής:

- Οι σημειώσεις γράφονται και εμφανίζονται στο τέλος κάθε εγγράφου. Αυτή η τεχνική ίσως είναι ικανοποιητική για ορισμένους.
- Οι σημειώσεις γίνονται σε σημεία που ο συγγραφέας του εγγράφου το επιτρέπει. Δηλαδή ορίζει σημεία, στα οποία μπορούν να γίνουν σημειώσεις ή να εμφανιστούν σημειώσεις, που έχουν γίνει σε αυτά τα σημεία. Αυτό συνήθως γίνεται τοποθετώντας ένα κουμπί στο σημείο του εγγράφου που επιτρέπεται να γίνουν σημειώσεις, ώστε επιλέγοντας το κουμπί να εμφανίζεται μια ειδική κάρτα με τους τίτλους των σημειώσεων που έχουν γίνει, τις συσχετίσεις τους και μια επιλογή για την εισαγωγή νέας σημείωσης. Συγκεκριμένα στο σύστημα CoNote [DH94] ο συγγραφέας ορίζει τα σημεία που επιτρέπει την εισαγωγή σημειώσεων τοποθετώντας μια εικόνα ενός χαρακτήρα, η οποία εμφανίζει μια φωτογραφία του πρώτου χρήστη που έγραψε σημείωση σε εκείνο το σημείο.
- Οι σημειώσεις μπορούν να γίνουν σε κάθε λογικό τμήμα του εγγράφου. Τα λογικά τμήματα μπορεί να είναι ένα βιβλίο, ένα μέρος του βιβλίου, ένα κεφάλαιο, ένα περιοδικό, ένα άρθρο, μια ενότητα, μια υποενότητα, μια παράγραφος, μια λίστα, ένα στοιχείο της λίστας, ένας πίνακας, ένα στοιχείο του πίνακα, μία εικόνα, μια πρόταση, κ.α.. Κάθε λογικό τμήμα διαπραγματεύεται ένα θέμα και άρα κάνουμε μια σημείωση για αυτό το θέμα. Το πρόβλημα σε αυτήν την περίπτωση είναι ότι αν θέλουμε να βλέπουμε το κείμενο και να βάλουμε κουμπάκια σε όλα τα λογικά τμήματα τότε θα υπερφορτώσουμε το έγγραφο με κουμπάκια. Αν βάλουμε κουμπάκι μόνο σε κάθε παράγραφο τότε ίσως να έχουμε καλύτερα αποτελέσματα.
- Οι σημειώσεις γίνονται σε τυχαία σημεία. Ένα τυχαίο σημείο θα μπορούσε να προσδιορίζεται με ένα μετρητή (π.χ. ο αριθμός του χαρακτήρα του εγγράφου που

σημειώνουμε, αρχίζοντας την μέτρηση από την αρχή του εγγράφου). Αν όμως το έγγραφο υποστεί κάποιες αλλαγές, τότε ο μετρητής θα έχει λάθος τιμή. Μια δεύτερη μέθοδος, που ξεπερνάει αυτό το πρόβλημα, είναι με την χρήση μιας χαρακτηριστικής φράσης της περιοχής του εγγράφου, για την οποία θέλουμε να κάνουμε την σημείωση.

Επίσης ένα σημαντικό θέμα σχετικά με τα σημεία στα οποία γίνονται σημειώσεις είναι και η εκτύπωση των εγγράφων με τις σημειώσεις ή μόνο των σημειώσεων. Κάθε σημείωση ή ομάδα σημειώσεων ή το αποτέλεσμα μιας αναζήτησης σημειώσεων θα πρέπει να μπορεί να εκτυπωθεί. Δηλαδή για ένα έγγραφο θα πρέπει να μπορούμε να επιλέξουμε ποιές σημειώσεις θέλουμε να εκτυπώσουμε και με ποια σειρά. Η σειρά εκτύπωσης έχει σχέση με κριτήρια όπως π.χ η σειρά εμφάνισης στο άρθρο, ο σχολιαστής, η αξία/βάρος της σημείωσης, κ.α.. Βέβαια μεταξύ των κριτηρίων υπάρχει μια ιεραρχία, δηλαδή παίρνουμε υπόψιν το πρώτο κριτήριο και δημιουργούμε κάποια σύνολα σημειώσεων και έπειτα σε κάθε σύνολο εφαρμόζουμε το δεύτερο κριτήριο και δημιουργούμε υποσύνολα των σημειώσεων κ.ο.κ. Αν δεν οριστούν κάποια κριτήρια, τότε θα ισχύει το κριτήριο της σειράς εμφάνισης της σημείωσης στο άρθρο.

2.3 Σχέσεις μεταξύ σημειώσεων

Οι σημειώσεις που γίνονται σε ένα έγγραφο ή σε περισσότερα έγγραφα μιας βάσης δεδομένων ή σε έγγραφα που βρίσκονται διασκορπισμένα σε ένα δίκτυο(on-line) έχουν κάποιες σχέσεις μεταξύ τους. Οι σχέσεις που εντοπίσαμε είναι οι εξής:

- Σημειώσεις πάνω στο ίδιο έγγραφο ή στο ίδιο σημείο του εγγράφου. Αυτές οι σημειώσεις ίσως να οργανώνονται σε συζητήσεις [Mar], αλλά ίσως και να μην έχουν καμία σχέση μεταξύ τους.

- Σημειώσεις που έχουν γίνει με τον ίδιο τρόπο ή έχουν το ίδιο νόημα. Για παράδειγμα όλες οι σημειώσεις που εκφράζουν θεματική ταξινόμηση έχουν γίνει με τον ίδιο τρόπο, δηλαδή την σύναψη όρου (βλ. ενότητα 2.1).

- Σημειώσεις του ιδίου ανθρώπου ή μιας ομάδας συνεργατών. Αυτές οι σημειώσεις σχετίζονται αυτόματα γιατί έγιναν από το ίδιο άτομο ή τα άτομα μιας ομάδας.

- Σημειώσεις που ταξινομούνται στα ίδια ή σχετικά θέματα (π.χ. θέματα που σχετίζονται με σχέση ιεραρχίας ή άλλου είδους σχέση). Αυτές οι σημειώσεις σχετίζονται γιατί έχουν το ίδιο ή σχετικό περιεχόμενο.

- Σημειώσεις που ανήκουν στη ίδια συλλογή που δημιούργησε ένας χρήστης. Αυτές οι σημειώσεις συγκεντρώνονται σε μια συλλογή γιατί έχουν κάποια κοινά σημεία σύμφωνα με την αντίληψη ενός ή περισσοτέρων χρηστών. Ένα παράδειγμα είναι οι σημειώσεις σύνδεσμοι που συνδέουν έγγραφα για να παρουσιάσουν ένα θέμα (π.χ. μια ξενάγηση σε μία πόλη, μια εισαγωγή σε ένα επιστημονικό θέμα, μια παρουσίαση των λύσεων που έχουν προταθεί για ένα επιστημονικό πρόβλημα,κ.α.).

2.4 Η χρήση των σημειώσεων

Οι σημειώσεις που γίνονται στα έγγραφα συγκεντρώνονται και αξιοποιούνται για διάφορες χρήσεις. Οι σημαντικότερες είναι οι εξής:

- Διαχείριση της προσωπικής βιβλιοθήκης. Με την χρήση ειδικών εργαλείων είναι δυνατή η αναζήτηση εγγράφων που σχολιάστηκαν σχετικά με ένα συγκεκριμένο θέμα και η αναζήτηση σημειώσεων που έγιναν για ένα θέμα [Van90].

- Άμεση δημιουργία βιβλιογραφίας που έχει σχολιαστεί, ώστε να προστεθεί σε ένα βιβλίο ή άρθρο και άμεση δημιουργία αναφορών κατά την συγγραφή ενός εγγράφου [HM93].

- Πρόσβαση των σημειώσεων από τα μέλη μιας ομάδας ώστε να μοιράζονται τις σκέψεις τους, να δημιουργούν συζητήσεις για ένα θέμα, να επιλέγουν τα έγγραφα που θα μελετήσουν (π.χ. εκείνα με τις σημαντικότερες σημειώσεις ή εκείνα που σημείωσε ένα συγκεκριμένο μέλος της ομάδας).

- Αξιολόγηση άρθρων και αναφορών εργασίας. Οι κριτικοί μπορούν να εμφανίζουν τις σημειώσεις τους , να τις συσχετίζουν και να τις συζητούν με επιχειρήματα [CBY89].

2.5 Αξιολόγηση σημειώσεων

Όταν οι σημειώσεις που γίνονται σε έγγραφα είναι διαθέσιμες στα μέλη μιας ομάδας τότε θα ήταν σκόπιμο να αξιολογούνται, ώστε να γίνεται καλύτερα η αναζήτηση και η επιλογή πληροφορίας. Μια μορφή αξιολόγησης γίνεται όταν κάθε μέλος της ομάδας χαρακτηρίζει μια σημείωση με ΝΑΙ ή ΟΧΙ, όπου το ΝΑΙ σημαίνει ότι είναι ενδιαφέρουσα και το ΟΧΙ ότι δεν είναι ιδιαίτερης σημασίας. Ανάλογα με τον αριθμό των ΝΑΙ και ΟΧΙ η κάθε σημείωση θα αξιολογείται (π.χ. με το ποσοστό επί της εκατό των ΝΑΙ). Ίσως,

για να δοθεί μεγαλύτερη άνεση αξιολόγησης, να δίνεται ένας αριθμός από το 0 έως το 10 ώστε τελικά να υπολογίζεται ένας αριθμός που εκφράζει την αξία της σημείωσης. Αλλά για κάθε άτομο ίσως να μετράει διαφορετικά η αξιολόγηση που κάνει ένα άλλο μέλος της ομάδας, δηλαδή να θέλει την αξιολόγηση μιας σημείωσης σύμφωνα με μια υποομάδα ατόμων με διαφορετικό βάρος για κάθε άτομο. Βέβαια με τον ίδιο τρόπο που γίνεται η αξιολόγηση των σημειώσεων μπορεί να γίνει και η αξιολόγηση των εγγράφων [RMW95], πράγμα που σε τελική ανάλυση αποτελεί και ένα τύπο σημείωσης.

2.6 Δημιουργία και συντήρηση σημειώσεων

Οι βασικές λειτουργίες για την δημιουργία και συντήρηση σημειώσεων είναι η εισαγωγή, η διαγραφή και η διόρθωση σημειώσεων καθώς και η δημιουργία και συντήρηση συλλογών σημειώσεων. Βέβαια η δημιουργία συλλογών εγγράφων ή σημειώσεων είναι ουσιαστικά η εισαγωγή μιας σημείωσης για όλα αυτά τα έγγραφα ή σημειώσεις, δηλαδή ότι παίρνουν μέρος σε μια συλλογή που έχει κάποια κοινά χαρακτηριστικά. Οι βασικές λειτουργίες είναι οι εξής:

- Κατά την εισαγωγή μιας σημείωσης θα πρέπει να προσδιοριστεί το σημείο αναφοράς της σημείωσης, η ταυτότητα αυτού που κάνει την σημείωση και η χρονική στιγμή της σημείωσης. Έπειτα γίνεται επιλογή του είδους και του τύπου της σημείωσης ώστε αυτή να γίνει με την βοήθεια ενός επεξεργαστή εγγράφου ή με την δημιουργία μιας σύνδεσης σε ένα υπαρκτό έγγραφο ή σε ένα όρο που εκφράζει μια έννοια ή θεματική περιοχή ή περιοχή εργασίας. Η περιοχή του εγγράφου στην οποία έγινε η σημείωση θα πρέπει να αποκτήσει κάποιο χαρακτηριστικό που δηλώνει την ύπαρξη σημείωσης. Επίσης ένα σημαντικό θέμα είναι η αποθήκευση των σημειώσεων (βλ. ενότητα 2.7).

- Αλλαγή σημειώσεων. Η αλλαγή σημειώσεων σχετίζεται με την θέση της σημείωσης πάνω στο άρθρο, την ταξινόμηση και το περιεχόμενο της σημείωσης.

- Διαγραφή σημειώσεων. Όταν κάποια σημείωση δεν μας ενδιαφέρει πια ή διαπιστώσουμε ότι είναι ακατάλληλη τότε απαιτείται η διαγραφή της.

- Ομαδοποίηση σημειώσεων. Όταν κάποιες σημειώσεις σχετίζονται μεταξύ τους τότε ίσως να μας ενδιαφέρει να τις έχουμε συγκεντρωμένες σε μια ομάδα. Η ομάδα αυτή σημειώσεων μπορεί να προκύψει από το αποτέλεσμα μιας ερώτησης (σε σχέση με το

θέμα ή την περιοχή της σημείωσης ή τον σχολιαστή ή τον χρόνο σχολιασμού) ή με την άμεση επιλογή των σημειώσεων.

- Ομαδοποίηση άρθρων. Ομάδες άρθρων συγκροτούνται από κάποιον αναγνώστη γιατί βρίσκει κάποια κοινά σημεία σε αυτά (π.χ θέμα, πηγή προέλευσης, έχουν παραπομπές στο ίδιο άρθρο, κ.α.).

Ένα ενδιαφέρον θέμα σχετικά με την διαχείριση των σημειώσεων είναι η εξουσιοδοτήσεις των χρηστών. Δηλαδή ποιοί χρήστες μπορούν να κάνουν σημειώσεις και σε ποιά έγγραφα, ποιοί μπορούν να διαβάζουν ή να αλλάζουν ή να διαγράφουν τις σημειώσεις ενός εγγράφου ή τις σημειώσεις ενός σχολιαστή σε κάποια έγγραφα, κ.α.. Στο σύστημα σημειώσεων CoNote [DH94] οι χρήστες έχουν κάποιους ρόλους. Συγκεκριμένα υπάρχουν οι απλοί αναγνώστες εγγράφων, οι αναγνώστες εγγράφων και σημειώσεων, οι κανονικοί χρήστες (δηλαδή μπορούν να διαβάζουν έγγραφα και σημειώσεις αλλά και να κάνουν σημειώσεις) και οι συγγραφείς που έχουν τα δικαιώματα των κανονικών χρηστών αλλά επιπλέον μπορούν να σβήνουν σημειώσεις άλλων χρηστών στα δικά τους έγγραφα.

Τέλος ένα άλλο σημαντικό θέμα για την διαχείριση των σημειώσεων είναι η ειδοποίηση των χρηστών ότι έχουν προστεθεί νέες σημειώσεις ή ότι έχουν αλλάξει οι υπάρχουσες. Ορισμένοι χρήστες που ενδιαφέρονται για κάποια έγγραφα και σημειώσεις άλλων χρηστών θα πρέπει να ειδοποιούνται (π.χ. με ένα ηλεκτρονικό μήνυμα) για την εισαγωγή ή αλλαγή σημειώσεων σε εκείνα τα έγγραφα ή θέματα που τους ενδιαφέρουν. Τα ενδιαφέροντα των χρηστών μπορούν να εκφραστούν με ερωτήσεις σχετικά με τα έγγραφα και τις σημειώσεις τους ώστε να γίνει κάποια επιλογή τους [GNOT92]. Η ειδοποίηση των χρηστών μπορεί να γίνεται κάθε φορά που εντοπίζεται μια αλλαγή ή κατά τακτά χρονικά διαστήματα (π.χ. κάθε εβδομάδα). Τα παραπάνω θέματα είναι σχετικά με την ενεργοποίηση κανόνων και απασχολούν την έρευνα σε ενεργά μοντέλα διαχείρισης δεδομένων [GRSW95].

2.7 Αποθήκευση των σημειώσεων

Τα έγγραφα στα οποία μπορούμε να κάνουμε σημειώσεις είναι συγκεντρωμένα σε έναν υπολογιστή ή είναι διασκορπισμένα σε ένα δίκτυο υπολογιστών. Στην πρώτη περίπτωση έχουμε το κείμενο ή το σχέδιο που εκφράζει τη σημείωση και μπορεί να αποθηκεύεται είτε στην βάση δεδομένων είτε σε ξεχωριστά αρχεία. Αν αποθηκεύεται στην βάση δεδομένων τότε το μέγεθος της θα γίνει πολύ μεγάλο και για να διαβάσουμε ή να αλλάξουμε το κείμενο ή το σχέδιο θα πρέπει να το κάνουμε μέσω της βάσης. Αν χρησιμοποιήσουμε

αρχεία του συστήματος που αποθηκεύονται σε ένα κατάλογο (directory) με το όνομα Notes, τότε θα μειωθεί το μέγεθος της βάσης και θα είναι πιο εύκολη η ανάγνωση ή αλλαγή της σημείωσης. Ένα πρόβλημα που μπορεί να προκύψει με αυτήν την μέθοδο είναι η μεγάλη συγκέντρωση αρχείων στον ίδιο κατάλογο. Μια λύση σε αυτό το πρόβλημα, που εφαρμόζεται στο BiblioText [Van90], είναι η δημιουργία μιας θεματικής ιεραρχίας από καταλόγους ώστε οι σημειώσεις, ανάλογα με το θέμα τους, να αποθηκεύονται και στον αντίστοιχο θεματικό κατάλογο.

Όταν τα έγγραφα είναι διασκορπισμένα σε ένα δίκτυο τότε τίθεται το θέμα της επιλογής του μηχανήματος στο οποίο θα αποθηκευτούν οι σημειώσεις. Μια πρώτη επιλογή είναι να αποθηκεύονται οι σημειώσεις στο μηχάνημα που είναι αποθηκευμένο το έγγραφο στο οποίο ανήκουν. Αυτή η μέθοδος συγκεντρώνει τα έγγραφα και τις σημειώσεις τους, αλλά ίσως να απαιτήσει μεγάλα ποσά χώρου για να αποθηκευτούν όλες οι σημειώσεις. Μια άλλη προσέγγιση είναι να αποθηκεύονται οι σημειώσεις στο μηχάνημα αυτού που τις έκανε, ώστε ο χώρος αποθήκευσης να διατίθεται από τον σχολιαστή. Αλλά με αυτήν την μέθοδο ίσως το σύστημα να μην μπορεί να προσπελάσει ορισμένες σημειώσεις λόγω μη λειτουργίας ορισμένων μηχανημάτων και αν δημιουργηθούν πολλές σημειώσεις από διάφορους χρήστες τότε θα παρουσιάσει καθυστέρηση στη συγκέντρωση των σημειώσεων. Τέλος μια άλλη μέθοδος είναι η χρήση ενός αντιπροσωπευτικού υπηρέτη (server) ο οποίος αναλαμβάνει να αποθηκεύει σημειώσεις για μια ομάδα εγγράφων ή μια ομάδα σχολιαστών. Αυτή η μέθοδος προτείνεται από την ομάδα εργασίας W3C[4] Annotation Working Group [Gra95] και ταιριάζει για τις σημειώσεις των μελών μιας ομάδας εργασίας σε ένα σύνολο εγγράφων.

Ένα άλλο θέμα που προκύπτει σχετικά με την αποθήκευση των σημειώσεων είναι πώς τροποποιούν τα έγγραφα στα οποία γίνονται. Οι σημειώσεις εισάγονται σε ένα έγγραφο με την εισαγωγή συνδέσμων σε αυτό. Όμως το πότε αυτοί οι σύνδεσμοι εισάγονται και τροποποιούν το έγγραφο, αποτελεί ένα σημαντικό θέμα. Μια μέθοδος είναι να εισάγονται μόλις γίνονται και να τροποποιούν το έγγραφο. Μια άλλη μέθοδος είναι να δημιουργείται ένα αντίγραφο του εγγράφου και όλες οι σημειώσεις να εισάγονται στο αντίγραφο. Τέλος μια άλλη μέθοδος, που εφαρμόζεται στο σύστημα σημειώσεων ComMentor [RMW95], δημιουργεί τους συνδέσμους δυναμικά όταν ζητηθεί η εμφάνιση κάποιων σημειώσεων που ικανοποιούν κάποια κριτήρια. Για τον σκοπό αυτό πρέπει να διατηρείται πληροφορία σχετικά με το σημείο αποθήκευσης της σημείωσης και το σημείο του εγγράφου στο οποίο έχει γίνει.

[4]World Wide Web Consortium

2.8 Προηγούμενα Συστήματα Σημειώσεων

Τα συστήματα σημειώσεων διαχωρίζονται σε δύο ομάδες με κριτήριο το αν βασίζονται ή όχι στο World−Wide Web (WWW).

Συστήματα σημειώσεων που βασίζονται στο WWW

- FRESCO, ComMentor

  Το σύστημα αυτό είναι κατανεμημένο και ακολουθεί την αρχιτεκτονική πελάτη - εξυπηρετητή (client - server) για να έχει πρόσβαση σε μια βάση σημειώσεων. Υποστηρίζει ένα μοντέλο για σημειώσεις, χρήστες και έγγραφα. Οι χρήστες ανήκουν σε μία ή περισσότερες ομάδες και οι σημειώσεις ταξινομούνται σε σύνολα. Οι σημειώσεις γίνονται σε οποιαδήποτε σημείο του εγγράφου και δηλώνουν την ύπαρξη τους με μια εικόνα, μεγέθους ενός χαρακτήρα, που επιλέγει ο σχολιαστής. Ανάλογα με τα δικαιώματα πρόσβασης και τα κριτήρια του χρήστη εμφανίζονται οι σημειώσεις που συσχετίζονται με ένα έγγραφο και είναι αποθηκευμένες σε διάφορα μηχανήματα (hosts) του δικτύου.
  http://Mjosa.Stanford.EDU:80/COMMENTOR/
  rmr@cs.Stanford.EDU (Martin Roscheisen)

- CoNote

  Είναι ένα σύστημα που επιτρέπει τη συνεργασία των χρηστών πάνω σε ένα σύνολο από έγγραφα. Οι χρήστες ανήκουν σε ομάδες και κάνουν σημειώσεις σε ένα σύνολο από έγγραφα. Οι σημειώσεις μπορεί να είναι σχόλια ή ερωτήσεις. Οι σημειώσεις μπορεί να αναφέρονται σε άλλες σημειώσεις, δηλαδή οι σημειώσεις οργανώνονται σε ”συζητήσεις”. Ο συγγραφέας του εγγράφου καθορίζει τα σημεία στα οποία μπορούν να γίνουν σημειώσεις, δηλαδή δεν γίνονται σημειώσεις σε οποιοδήποτε σημείο του εγγράφου. Τα έγγραφα και οι σημειώσεις είναι σε ASCII ή HTML μορφή για να διαβάζονται από ένα τυπικό διερευνητή (browser) του WWW.
  Jim Daves (davis@dri.cornell.edu)

- Mole

  Είναι ένα hypertext σύστημα που ακολουθεί την αρχιτεκτονική πελάτη - εξυπηρέτη (client - server) και υποστηρίζει την συνεργασία για την δημιουργία σημειώσεων πάνω στις γραμμές του κειμένου. Οι σημειώσεις είναι προσωπικές ή δημόσιες και

μπορούν και αυτές να δεχτούν σημειώσεις. Για την εμφάνιση των εγγράφων χρησιμοποιείται ένας τυπικός διερευνητής (browser) του WWW, αλλά για την εισαγωγή σημειώσεων χρησιμοποιείται ένας εξειδικευμένος πελάτης (client).
Dave Whittungton (dave@cs.strath.ac.uk)

- NetQ

  Είναι ένας συνδυασμός από ηλεκτρονικές εκδόσεις και σημειώσεις. Το σύστημα αυτό αποθηκεύει αντίγραφα των άρθρων και δίνει την δυνατότητα να γίνουν ερωτήσεις από διάφορους αναγνώστες προς τον συγγραφέα. Αν το άρθρο είναι σε HTML μορφή τότε στο τέλος κάθε παραγράφου υπάρχει ένας υπερσύνδεσμος(hyperlink) που οδηγεί στις ερωτήσεις σχετικά με αυτήν την παράγραφο. Αν το άρθρο είναι σε PostScript μορφή τότε οι ερωτήσεις είναι στο τέλος του άρθρου.
  Stewart Wilson (netq@spo.rowland.org)

- HyperNews

  Είναι ένας συνδυασμός ανάμεσα στο WWW και στο Usenet News. Η βασική ιδέα είναι ότι επιτρέπει στους αναγνώστες να αποκρίνονται σε άρθρα ή αποκρίσεις άλλων αναγνωστών που διάβασαν στο HyperNews web. Οι αποκρίσεις αυτές οργανώνονται σε δενδρική μορφή με απεριόριστο βάθος. Υπάρχουν διάφοροι τύποι απόκρισης (π.χ. απάντηση, ερώτηση, σχόλιο, συμφωνία, διαφωνία).
  http://union.ncsa.uiuc.edu/HyperNews/get/hypernews/related.html
  Daniel LaLiberte (liberte@ncsa.uiuc.edu)

- WIT

  Είναι ένα σύστημα που επιτρέπει σε πολλούς χρήστες να συγκεντρώσουν επιχειρήματα για ένα θέμα. Δηλαδή συζητούνται διάφορα Θέματα (Topics) χρησιμοποιώντας ορισμένες Προτάσεις (Proposals) και κάθε Πρόταση συζητείται με κάποια Επιχειρήματα (Arguments) που μπορεί να είναι Αρνητικά, Θετικά ή Ουδέτερα. Τα Επιχειρήματα μπορούν και αυτά να συζητηθούν με Επιχειρήματα.
  Mark – Antre Schenk, Ecole des HEC/University of Lausanne
  marc – andre.schenk@hec.unil.ch

- MOO-Based Collaborative hypermedia

  Είναι ένα σύστημα που υποστηρίζει την συνεργασία στην συγγραφή υπερ-εγγράφων (hyperdocument) και δίνει τη δυνατότητα στους χρήστες να κάνουν σημειώσεις στα

έγγραφα άλλων χρηστών και να παίρνουν μέρος σε συζητήσεις που κάνουν κριτική σε ένα έγγραφο [CFP95].

Συστήματα Σημειώσεων που δεν βασίζονται στο WWW

- BiblioText

  Είναι ένα εργαλείο για την αναζήτηση (browsing) βιβλιογραφικών δεδομένων και σχετικών σημειώσεων. Τα βιβλιογραφικά δεδομένα είναι σε μορφή BIBTEX και αποθηκεύονται σε αρχεία που οργανώνονται σε καταλόγους ώστε να δημιουργηθεί μια δομημένη ιεραρχία για τα διάφορα είδη άρθρων. Ο κάθε χρήστης έχει έναν αριθμό από βιβλιογραφικές βάσεις δεδομένων και μπορεί να δημιουργεί ομάδες εγγράφων, να γράφει σχόλια σχετικά με ένα έγγραφο ή μια ομάδα, να δημιουργεί συνδέσεις μεταξύ εγγράφων και να σχολιάζει αυτές τις συνδέσεις, να δίνει λέξεις κλειδιά σε έγγραφα και να βρίσκει τα έγγραφα που τον ενδιαφέρουν είτε με περιπλάνηση (browsing) στα δεδομένα είτε με λέξεις κλειδιά. Ένας χρήστης μπορεί να πάρει πληροφορίες από την βάση ενός άλλου χρήστη αν έχει δικαίωμα ανάγνωσης στα αρχεία αυτού του χρήστη.
  ftp://tr-ftp.cs.berkeley.edu/pub/tech-reports/csd/csd-90-603/
  Michael L. Van De Vanter , University of California, Berkeley

- Microsoft Word 6.0

  Επιτρέπει την προσθήκη σημειώσεων σε ένα έγγραφο. Οι σημειώσεις αποθηκεύονται σαν ”κρυφό κείμενο” και έτσι η εμφάνιση τους, που γίνεται σε ένα συγκεκριμένο σημείο του παραθύρου, είναι προαιρετική. Δίνεται η δυνατότητα σημείωσης σε οποιοδήποτε σημείο του εγγράφου.

- Lotus Notes

  Υποστηρίζει την δημιουργία μιας συλλογής από έγγραφα που ανήκουν σε κάποιους χρήστες. Τα έγγραφα αυτά περιέχουν πρόσθετες συνδέσεις σε άλλα έγγραφα.

- Tapestpy Project

  Είναι ένα σύστημα που διαχειρίζεται την εισροή ηλεκτρονικών εγγράφων (ηλεκτρονικά μηνύματα, news, άρθρα). Κάποιοι χρήστες κάνουν σημειώσεις σχετικά με αυτά τα έγγραφα και έπειτα είναι δυνατή η επιλογή εγγράφων με την χρήση αυτών των σημειώσεων (collaborative filtering). Δηλαδή από τις σημειώσεις ενός αναγνώστη μπορούμε να βγάλουμε συμπεράσματα για την ποιότητα του άρθρου

(πώς το χαρακτήρισε; πόσες σημειώσεις έκανε; με ποιά άλλα έγγραφα το συσχέτισε; κ.α.) και, γνωρίζοντας τα ενδιαφέροντα του αναγνώστη, να επιλέγουμε τα έγγραφα που ίσως μας ενδιαφέρουν [GNOT92].
http://www.xerox.com/PARC/dlbx/filtering.html

- Electronic Book Technologies

  Η SunSoft ασχολείται, αυτό το χρονικό διάστημα, με την δημιουργία ενός διερευνητή (browser) που μπορεί να βλέπει, να αναζητεί και να σημειώνει έγγραφα που είναι σε ένα μεγάλο πεδίο από μορφές(format).

Κεφάλαιο 3

Η γλώσσα Telos και το σύστημα SIS

Το Σημασιολογικό Σύστημα Ευρετηριασμού (Semantic Index System ή SIS) είναι ένα εργαλείο για περιγραφή και τεκμηρίωση μεγάλου πληθυσμού, ιδιόμορφων, εξελισσόμενων και πολλαπλώς συνδεδεμένων δεδομένων. Για το λόγο αυτό είναι κατάλληλο για την παράσταση επιστημονικής, σχεδιαστικής και κατασκευαστικής γνώσης. Το SIS αποτελείται από ένα μηχανισμό αποθήκευσης, που βασίζεται σε ένα συνδυασμό οντοκεντρικού μοντέλου και σημασιολογικού δικτύου, και από μία γενική διαλογική επαφή χρήσης για την εισαγωγή και ανάκτηση (retrieval) πληροφοριών. Το μοντέλο των δεδομένων περιγράφεται με την γλώσσα παράστασης γνώσης Telos. Το SIS και η ενσωματωμένη σε αυτό υλοποίηση του δομικού μέρους της γλώσσας Telos έχουν αναπτυχθεί από την Ομάδα Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του Ινστιτούτου Πληροφορικής του Ιδρύματος Τεχνολογίας Έρευνας (Ι.Τ.Ε.)

3.1 Σημαντικά χαρακτηριστικά της Telos

Η Telos είναι μια γλώσσα παράστασης γνώσης η οποία υποστηρίζει ένα οντοκεντρικό μοντέλο δεδομένων. Προσφέρει τους εξής στατικούς εκφραστικούς μηχανισμούς: μηφραγμένη ιεραρχία ταξινόμησης, πολλαπλή και αυστηρή κληρονομικότητα και πλειότιμα γνωρίσματα τα οποία μπορούν να έχουν τα δικά τους γνωρίσματα. Επίσης παρέχει μηχανισμούς εξαγωγής συμπερασμάτων και χρονικής λογικής. Η γλώσσα Telos αναπτύχθηκε από τον Μυλόπουλο και τους συνεργάτες του [MBJK90]. Η υλοποίηση της Telos που χρησιμοποιήθηκε δεν περιλαμβάνει τον μηχανισμό εξαγωγής συμπερασμάτων και χρονική λογική που περιγράφονται στο [MBJK90]. Η επιλογή αυτή έγινε για λόγους απόδοσης.

Η γλώσσα Telos παρέχει πολλές στάθμες αφαίρεσης. Άρα προσφέρει τη δυνατότητα οργάνωσης του σχήματος σύμφωνα με ένα μετασχήμα, το οποίο περιγράφει πιο αφηρημένες έννοιες και ιδιότητες, επιτρέποντας έτσι την διατύπωση και απάντηση ερωτήσεων που βασίζονται σε αφηρημένες ιδιότητες. Με τη σειρά τους, οι έννοιες του μετασχήματος μπορούν να υπάγονται σε ένα μετα-μετασχήμα, κ.ο.κ. Έτσι η Telos παρέχει ομοιόμορφη διαχείριση δεδομένων και σχήματος τόσο στην εισαγωγή στοιχείων όσο και στον μηχανισμό ερωτήσεων. Αυτό επιτρέπει δυναμικό ορισμό και ενημέρωση του σχήματος της βάσης χωρίς χρονική επιβάρυνση.

Οι οντότητες και τα γνωρίσματα θεωρούνται και χρησιμοποιούνται σαν ισότιμα αντικείμενα, τα οποία μπορούν να έχουν λογικά ονόματα. Τα λογικά ονόματα παρέχουν επιπλέον πληροφορίες για τα αντικείμενα, κάνοντας έτσι τα περιεχόμενα μιας βάσης πιο κατανοητά στο χρήστη. Επίσης αποτελούν έναν άμεσο και κατανοητό τρόπο αναφοράς των αντικειμένων της βάσης.

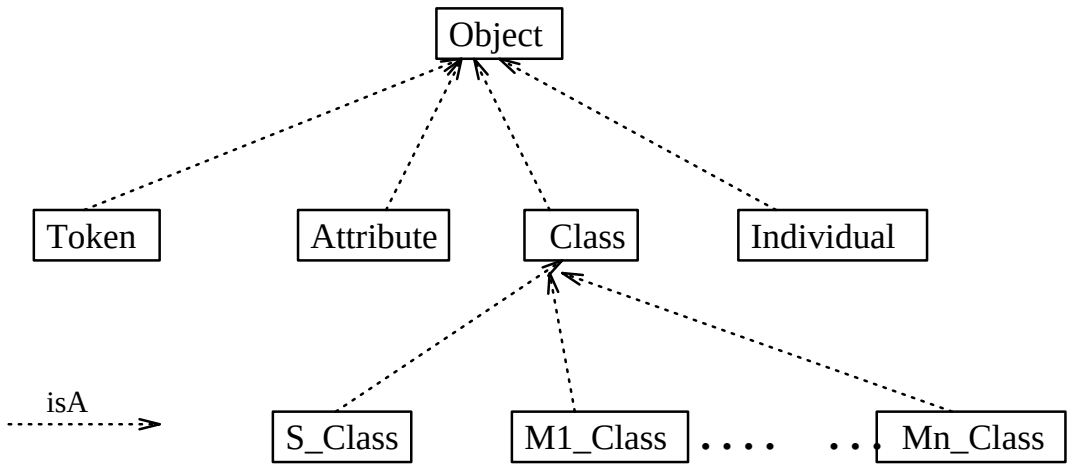
Στην υλοποίηση της Telos που χρησιμοποιήθηκε, οι σύνδεσμοι (ταξινόμησης, γενίκευσης, γνωρισμάτων) αποθηκεύονται ως σύνδεσμοι διπλής κατεύθυνσης προσφέροντας καλές επιδόσεις στις διασχίσεις και ερωτήσεις, μεγαλώνοντας όμως το μέγεθος της βάσης.

3.2 Παράσταση γνώσης στη γλώσσα Telos

Παρακάτω παρουσιάζονται συνοπτικά οι βασικοί εκφραστικοί μηχανισμοί που χρησιμοποιεί η γλώσσα Telos για την δημιουργία των μοντέλων παράστασης γνώσεων. Για λεπτομερέστερη περιγραφή των βασικών μηχανισμών παραπέμπουμε στο [MBJK90].

- *Ονοματοδοσία*
  Κάθε οντότητα έχει ένα εσωτερικό, παραγόμενο από το σύστημα, αναγνωριστικό όνομα. Επιπλέον ο χρήστης έχει τη δυνατότητα να ονομάσει ο ίδιος μια οντότητα με ένα *λογικό όνομα*. Το λογικό όνομα ενός γνωρίσματος είναι της μορφής $x.y$, όπου x είναι το λογικό όνομα της οντότητας της οποίας αποτελεί γνώρισμα και y είναι το όνομα του ίδιου.

- *Ταξινόμηση*
  Με το μηχανισμό αυτό μια ατομική οντότητα περιγράφεται ως μέλος (περίπτωση) μιας κλάσης, της οποίας κληρονομεί τα γνωρίσματα. Μια κλάση είναι και αυτή με τη σειρά της μια οντότητα, άρα μπορεί να είναι περίπτωση μιας άλλης κλάσης. Η Telos απαιτεί κάθε οντότητα να αποτελεί περίπτωση μιας κλάσης. Έτσι δημιουργείται

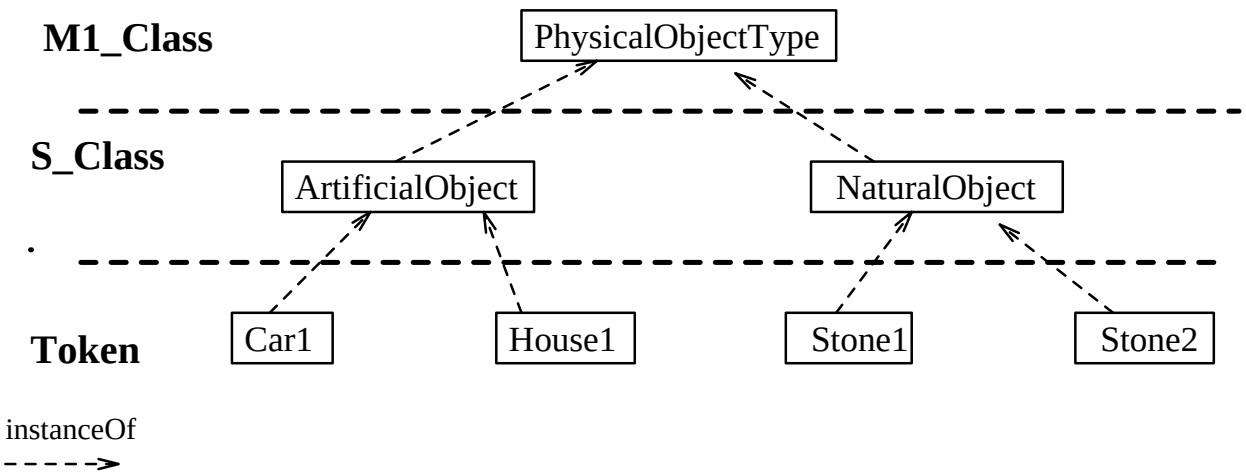


Σχήμα 3.1: Οι κλάσεις του συστήματος στην Telos

Στο σχήμα φαίνεται η σχέση που έχουν οι κλάσεις του συστήματος μεταξύ τους. Η κλάση **Object** είναι υπερκλάση όλων των άλλων κλάσεων του συστήματος αφού τόσο οι οντότητες όσο και οι σχέσεις των οντοτήτων μεταξύ τους αντιμετωπίζονται με ίδιο τρόπο από την Telos. Έτσι λοιπόν οι κλάσεις **Token, Attribute, Class** και **Individual** είναι υποκλάσεις της **Object**, ενώ οι **S_Class, M1_Class,..., Mn_Class** που περιγράφουν αν μια κλάση είναι κλάση ατόμων, κλάση κλάσεων, κλάση από κλάσεις κλάσεων, κοκ., είναι υποκλάσεις της **Class**.

μια μη φραγμένη ιεραρχία από κλάσεις. Στο κατώτερο επίπεδο τοποθετούνται οι *ατομικές οντότητες* ή *Tokens*, κατόπιν υπάρχουν οι *απλές κλάσεις* που αποτελούνται από ατομικές οντότητες, μετά οι *μετακλάσεις* που αποτελούνται από απλές κλάσεις, οι *μετα-μετακλάσεις*, κοκ. Η μη φραγμένη ιεραρχία είναι άπειρη αλλά όχι υποχρεωτικά. Η απαίτηση κάθε οντότητα να ανήκει σε μια κλάση εκπληρώνεται με την ύπαρξη ειδικών *κλάσεων του συστήματος* ή *ω-κλάσεων*. Οι κλάσεις του συστήματος δεν μπορούν να αλλαχτούν από τον χρήστη. Αποτελούν τον αρχικό πληθυσμό της βάσης και οποιαδήποτε δεδομένα που θα εισάγει στο σύστημα ο χρήστης πρέπει να συνδεθούν έμμεσα ή άμεσα με αυτές. Ακολουθεί μια σύντομη περιγραφή τους.

Όλες οι οντότητες που περιγράφονται με την Telos ταξινομούνται στην κλάση του συστήματος *Object*. Υποκλάσεις της *Object* είναι οι κλάσεις *Individual* και *Attribute*. Στην κλάση *Individual* ταξινομούνται οι οντότητες, οι κλάσεις από οντότητες, οι κλάσεις από κλάσεις από οντότητες, κοκ. Στην κλάση *Attribute* ταξινομούνται οι σχέσεις που έχουν οι οντότητες μεταξύ τους , οι κλάσεις σχέσεων, οι κλάσεις από κλάσεις σχέσεων, κοκ. Οι κλάσεις που ορίζει ο χρήστης ταξινομούνται στην κλάση του συστήματος *Class*. Τα Tokens ταξινομούνται στην κλάση *Token*. Οι



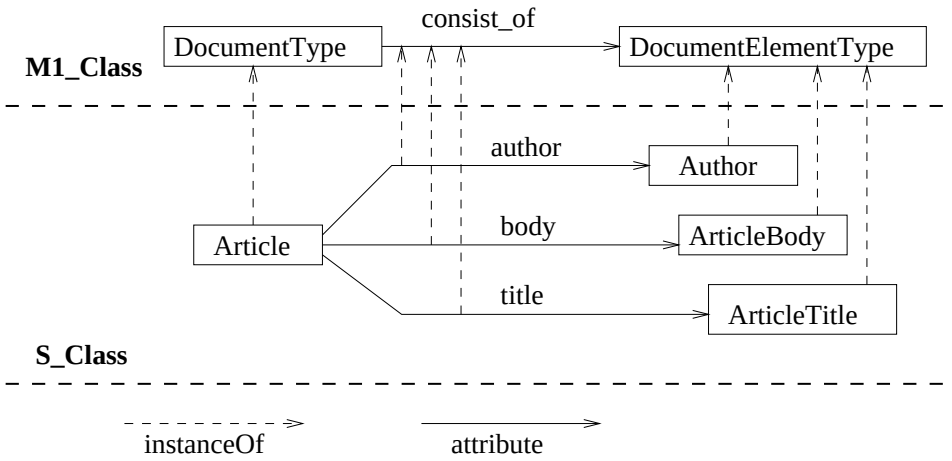
Σχήμα 3.2: Ο μηχανισμός ταξινόμησης

Στο σχήμα φαίνεται ένα παράδειγμα χρήσης του μηχανισμού της ταξινόμησης. Οι ατομικές οντότητες όπως οι **Car1, House1, Stone1, Stone2** ανήκουν σε κλάσεις οντοτήτων που περιγράφουν όχι πια συγκεκριμένα άτομα αλλά τις αφηρημένες κλάσεις ατόμων όπως οι **ArtificialObject** και **NaturalObject**. Οι κλάσεις αυτές με τη σειρά τους ανήκουν σε κλάσεις από κλάσεις που περιγράφουν αφηρημένους τύπους, όπως εδώ η **PhysicalObjectType** που περιγράφει τύπους φυσικών αντικειμένων.

απλές κλάσεις ταξινομούνται στην κλάση *S_Class* , οι μετακλάσεις στην *M1_Class*, οι μετα-μετακλάσεις στην *M2_Class*, κοκ. Επίσης υπάρχουν οι πρωτογενείς τιμές *Integer, Real, String*. Οι πρωτογενείς τιμές δεν μπορούν να δημιουργηθούν ή να καταστραφούν αλλά μόνο να γίνει αναφορά σε αυτές. Η σχέση των *κλάσεων του συστήματος* μεταξύ τους φαίνεται στο σχήμα 3.1.

Με την ύπαρξη των κλάσεων του συστήματος η ιεραρχία ταξινόμησης φτάνει μόνο ως εκεί που είναι απαραίτητο. Το πλεονέκτημα που προσφέρεται με την ύπαρξη μη φραγμένης ιεραρχίας είναι ότι το σχήμα της βάσης μπορεί να αντιμετωπιστεί όπως τα δεδομένα και άρα να αλλάζει δυναμικά.

Μία οντότητα μπορεί να ανήκει σε παραπάνω από μία κλάσεις. Με αυτόν τον τρόπο μπορεί η ταξινόμηση να υποκαταστήσει την απόδοση γνωρισμάτων (δυνατότητα που χρησιμοποιείται και στο μοντέλο που δημιουργήσαμε). Με αυτόν τον τρόπο όμως πρέπει να αποδίδονται μόνο εγγενείς ιδιότητες μιας οντότητας ενώ ο μηχανισμός απόδοσης γνωρίσματος είναι πιο κατάλληλος για την περιγραφή των επιφανειακών γνωρισμάτων που πιθανόν να έχει η οντότητα. Στο σχήμα 3.2 φαίνεται η χρήση του μηχανισμού ταξινόμησης με ένα παράδειγμα.



Σχήμα 3.3: Περιορισμός στις περιπτώσεις γνωρισμάτων στην Telos

Η λειτουργία του περιορισμού στην ταξινόμηση γνωρισμάτων φαίνεται στο παράδειγμα του σχήματος. Οι κλάση **Article** είναι περίπτωση της μετακλάσης **DocumentType**. Η **DocumentType** έχει το γνώρισμα *consist_of* που παίρνει τιμές στην κλάση **DocumentElementType**. Τα γνωρίσματα *author*, *title* και *body* του **Article** είναι περιπτώσεις του *consist_of*, άρα πρέπει και τα πεδία τιμών τους (οι κλάσεις **Author**, **ArticleTitle** και *ArticleBody*) να είναι περιπτώσεις του πεδίου τιμών του *consist_of*, δηλαδή της μετακλάσης **DocumentElementType**.

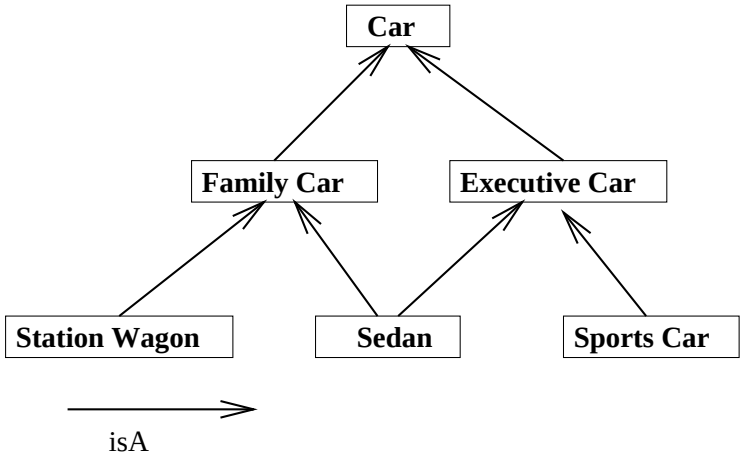
- *Απόδοση Γνωρίσματος*

  Με το μηχανισμό αυτό αποδίδονται γνωρίσματα στις οντότητες. Τα γνωρίσματα αυτά μπορούν να θεωρηθούν ως σχέσεις μεταξύ οντοτήτων μιας και έχουν ένα *πεδίο ορισμού* (την οντότητα στην οποία αποδίδονται) και ένα *πεδίο τιμών* (την οντότητα που προσδιορίζει τί τύπου είναι η τιμή του γνωρίσματος). Κάθε γνώρισμα μπορεί να έχει παραπάνω από μια ή και καμία τιμή. Επίσης μπορεί να έχει και αυτό γνωρίσματα (πράγμα που απορρέει από την ισότιμη μεταχείριση οντοτήτων και γνωρισμάτων από την Telos).

- *Περιορισμός στην ταξινόμηση των γνωρισμάτων*

  Αν ένα γνώρισμα είναι περίπτωση μιας κατηγορίας γνωρισμάτων [1] , τότε το πεδίο ορισμού και το πεδίο τιμών του πρέπει να είναι περιπτώσεις των πεδίων ορισμού και τιμών της κλάσης γνωρισμάτων στην οποία ανήκει. Στο σχήμα 3.3 φαίνεται ο περιορισμός αυτός σε κάποιες από τις οντότητες του μοντέλου που δημιουργήσαμε.

[1] Κατηγορία γνωρισμάτων ονομάζεται μια κλάση γνωρισμάτων. Συνήθως με αυτόν τον όρο αναφερόμαστε σε κλάσεις γνωρισμάτων στο μετα-επίπεδο, αλλά και ένα γνώρισμα σε επίπεδο *S_Class* αποτελεί κατηγορία γνωρισμάτων για τις περιπτώσεις του σε επίπεδο *Token*.



Σχήμα 3.4: Ο μηχανισμός γενίκευσης

Στο σχήμα φαίνεται ένα παράδειγμα χρήσης του μηχανισμού γενίκευσης/εξειδίκευσης. Η κλάση **Sedan** αποτελεί άμεση εξειδίκευση των κλάσεων *Family Car* και *Executive Car* και έμμεση εξειδίκευση της κλάσης *Car*. Η ιεραρχία αυτή βρίσκεται σε ένα επίπεδο ταξινόμησης.

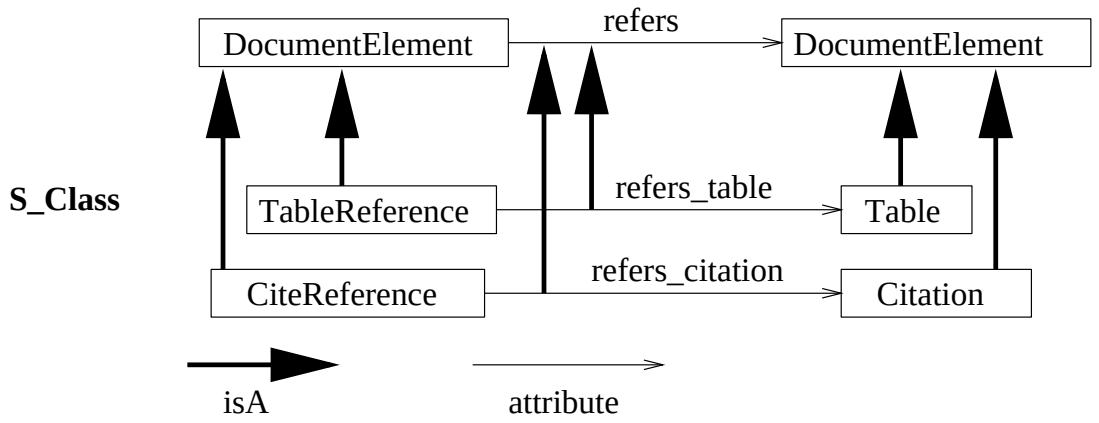
- *Γενίκευση (αντίστροφο : εξειδίκευση)*

  Ο μηχανισμός αυτός ισχύει μόνο για κλάσεις (όχι δηλαδή για άτομα) που βρίσκονται στο ίδιο επίπεδο ταξινόμησης. Ορίζει μια σχέση υποσυνόλου μεταξύ των κλάσεων που ονομάζεται *isA* . Αν A *isA* B (A και B κλάσεις), τότε η A ονομάζεται υποκλάση της B και η B υπερκλάση της A. Η A κληρονομεί όλα τα γνωρίσματα της B και είτε έχει επιπλέον γνωρίσματα είτε περιορίζει το σύνολο τιμών των γνωρισμάτων που κληρονομεί από την B. Μια κλάση μπορεί να έχει παραπάνω από μια υπερκλάσεις. Έτσι η σχέση *isA* υποστηρίζει πολλαπλή και αυστηρή κληρονόμηση. Στο σχήμα 3.4 παρουσιάζεται ένα παράδειγμα χρήσης της γενίκευσης.

  Ο μηχανισμός γενίκευσης/εξειδίκευσης επιτρέπει την οργάνωση των κλάσεων σε διάφορες ιεραρχίες γενίκευσης οι οποίες προσδίδουν οικονομία και συνέπεια στο μοντέλο, αφού δεν χρειάζεται να επαναληφθεί ο ορισμός ενός γνωρίσματος που έχει ήδη αποδοθεί σε μια γνωστή υπερκλάση.

- *Περιορισμός στις υποκλάσεις γνωρισμάτων*

  Έστω AC1 και AC2 κατηγορίες γνωρισμάτων. Αν AC1 *isA* AC2, τότε τα πεδία ορισμού και τιμών της AC1 είναι υποκλάσεις των αντίστοιχων της AC2. Στο σχήμα 3.5 φαίνεται ο περιορισμός αυτός με ένα παράδειγμα από το μοντέλο παράστασης των άρθρων.



Σχήμα 3.5: Περιορισμός στις υποκλάσεις γνωρισμάτων

Ο περιορισμός στις υποκλάσεις γνωρισμάτων φαίνεται στο παράδειγμα του σχήματος. Οι κλάσεις **TableReference** και **CiteReference** είναι υποκλάσεις της **DocumentElement**, η οποία έχει το γνώρισμα *refers* που παίρνει τιμές στην ίδια την **DocumentElement**. Τα γνωρίσματα *refers_table* και *refers_citation* είναι υποκλάσεις του *refers* και γι'αυτό τα πεδία ορισμού και τιμών τους πρέπει να είναι υποκλάσεις της **DocumentElement**.

Επίσης πρέπει να πούμε ότι στην κλάση του συστήματος *Individual* έχουν αποδοθεί τα γνωρίσματα *display* και *comment*. Το γνώρισμα *display* έχει ως πεδίο τιμών την κλάση *Telos_String* και η τιμή του εκφράζει ένα εκτελέσιμο πρόγραμμα. Ενώ το γνώρισμα *comment* έχει ως πεδίο τιμών την κλάση *hyperText* και η τιμή του εκφράζει ένα σχόλιο. Αλλά οι κλάσεις ανεξάρτητων οντοτήτων και οι απλές ανεξάρτητες οντότητες αποτελούν μέλη (περιπτώσεις) της κλάσης *Individual*, επομένως σε κάθε ένα από αυτά μπορούμε να αποδώσουμε ένα ή περισσότερα εκτελέσιμα προγράμματα και να κάνουμε ένα ή περισσότερα σχόλια.

3.3 Το Σύστημα Σημασιολογικού Ευρετηριασμού

Το Σημασιολογικό Σύστημα Ευρετηριασμού (Semantic Index System ή SIS) είναι ένα εργαλείο για περιγραφή και τεκμηρίωση μεγάλου πληθυσμού, ιδιόμορφων, εξελισσόμενων και πολλαπλώς συνδεδεμένων δεδομένων ([CD94]). Για το λόγο αυτό είναι κατάλληλο για την παράσταση επιστημονικής, σχεδιαστικής και κατασκευαστικής γνώσης. Το SIS αποτελείται από ένα μηχανισμό αποθήκευσης, που βασίζεται σε ένα συνδυασμό οντοκεντρικού μοντέλου και σημασιολογικού δικτύου, και από μια γενική διαλογική επαφή χρήσης για την εισαγωγή και ανάκτηση (retrieval) πληροφοριών.

Το μοντέλο των δεδομένων περιγράφεται με την γλώσσα παράστασης γνώσης Telos.

Ο μηχανισμός αποθήκευσης των αντικειμένων της Telos , υποστηρίζει δοσοληψίες και ταυτόχρονη πρόσβαση από πολλούς χρήστες [Γεω94a]. Το SIS συνοδεύεται από ένα μηχανισμό ερωτήσεων [Ντα93], ο οποίος παρέχει ένα σύνολο από ερωτηματικές εντολές, με τις οποίες είναι δυνατή η πλοήγηση στη βάση καθώς και η διατύπωση αναδρομικών ερωτήσεων με πολλαπλά κριτήρια (βλ. ενότητα 3.3.1). Οι ερωτήσεις μπορούν να γίνουν με ένα διαλογικό εργαλείο, τον answerer, ή μέσα από άλλες εφαρμογές με τη βοήθεια του programmatic query interface [DD92].

Η επικοινωνία του SIS με το χρήστη υποστηρίζεται από ένα αποδοτικό και προσαρμόσιμο εργαλείο (βλ. σχήμα 6.1) το οποίο επιτρέπει διερεύνηση σε όλα τα επίπεδα αφαίρεσης και στα δεδομένα καθώς και παρουσιάσεις πολύμορφων δεδομένων (multimedia) εξωτερικώς αποθηκευμένων. Υπάρχει η δυνατότητα ορισμού και ενεργοποίησης προκαθορισμένων ερωτήσεων. Αυτές οι ερωτήσεις διακρίνονται σε δύο κατηγορίες: α) σε αυτές που το αποτέλεσμα τους είναι ένας γράφος και β) σε αυτές που το αποτέλεσμα τους είναι κείμενο. Οι ερωτήσεις γίνονται για μία οντότητα της βάσης, ενώ υπάρχει η δυνατότητα για διατύπωση ερωτήσεων με δελτία, δηλαδή παίρνοντας υπ' όψην δύο ή περισσότερες οντότητες. Η εισαγωγή δεδομένων γίνεται μαζικά από τον μεταφραστή της Telos ή διαλογικά από ένα Δελτίο Εισαγωγής Δεδομένων [Δασ96a], του οποίου η λειτουργία μπορεί να προσαρμοστεί ανάλογα με τις ανάγκες κάθε εφαρμογής. Η διασύνδεση και επικοινωνία του SIS με εξωτερικά εργαλεία και εφαρμογές υποστηρίζονται από ειδικούς μηχανισμούς που έχουν αναπτυχθεί.

Το SIS αποτελεί το κύριο εργαλείο για την υλοποίηση της εργασίας (βλ. κεφάλαιο 6) και έχει ήδη χρησιμοποιηθεί σε ένα εργαλείο για στατική ανάλυση λογισμικού και σε ένα πληροφοριακό σύστημα μουσείων [Χρι94b].

3.3.1 Σύντομη περιγραφή της Βιβλιοθήκης Στοιχειωδών Συναρτήσεων (ΒΣΕΣ)

Η ΒΣΕΣ αποτελείται από συναρτήσεις οι οποίες εφαρμόζονται σε κόμβους του σημασιολογικού δικτύου (οντότητες του μοντέλου) που ορίζεται από την περιγραφή των δεδομένων σε Telos, και έχουν ως αποτέλεσμα ένα σύνολο οντοτήτων. Με τις συναρτήσεις αυτές είναι δυνατόν να αντληθεί η εξής πληροφορία :

- 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 δυο παραπάνω ερωτήσεων για μια ή περισσότερες κατηγορίες (ή μετακατηγορίες).

Περισσότερα για την ΒΣΕΣ καθώς και μια αναλυτική περιγραφή των συναρτήσεων βρίσκονται στα [Ντα93], [DD92].

Οι συναρτήσεις αυτές μπορούν να χρησιμοποιηθούν σε προγράμματα σε γλώσσα C ή C++ , αλλά υπάρχει και ένας διερμηνευτής που κάνει δυνατή τη διαλογική χρησιμοποίηση των συναρτήσεων αυτών. Ο διερμηνευτής αυτός χρησιμοποιεί από την επαφή χρήσεως για την εκτέλεση ερωτήσεων οι οποίες έχουν περιγραφεί μέσω της γλώσσας Telos. Με αυτόν τον τρόπο κατασκευάστηκαν οι προκαθορισμένες ερωτήσεις που περιγράφονται στην ενότητα 6.1.

Κεφάλαιο 4

Το μοντέλο παράστασης εγγράφων με την γλώσσα Telos

Κάθε έγγραφο αποτελείται από τρία διαφορετικά τμήματα : το περιεχόμενο, την λογική δομή και την μορφή παρουσίασης (layout). Τα τμήματα αυτά είναι εξίσου σημαντικά για τον προσδιορισμό της υπόστασης του εγγράφου. Το περιεχόμενο αποτελεί την πραγματική πληροφορία που υπάρχει στο έγγραφο. Η λογική δομή αποτελεί την οργάνωση της πληροφορίας του εγγράφου σύμφωνα με τον συγγραφέα του. Τέλος η μορφή παρουσίασης (layout) καθορίζει τον τρόπο που το έγγραφο θα παρουσιαστεί στον αναγνώστη. Στην συγκεκριμένη εργασία που παρουσιάζεται, σημαντικό ρόλο διαδραματίζει η γνώση της λογικής δομής του εγγράφου. Διότι, γνωρίζοντας την λογική δομή, μπορούμε να περιπλανηθούμε (navigate) στο έγγραφο, να κάνουμε σημειώσεις σε κάποιο λογικό τμήμα του και να συσχετίσουμε λογικά τμήματα του ιδίου ή διαφορετικών εγγράφων.

Υπάρχουν διάφορα πρότυπα για την παράσταση της λογικής δομής και της μορφής παρουσίασης των εγγράφων. Σε αυτό το κεφάλαιο παρουσιάζονται τα δύο σημαντικότερα πρότυπα, το **Open/Office Document Architecture (ODA) ISO 8613:1989** και το **Standard Generalized Markup Language (SGML) ISO 8879:1986**. Στην συνέχεια παρουσιάζεται το παράστασης της λογικής δομής εγγράφων με την γλώσσα παράστασης γνώσης Telos, σύμφωνα με το πρότυπο ISO 12083:1993.

4.1 Η μοντελοποίηση εγγράφων

Η μοντελοποίηση των εγγράφων έχει ως κύριο σκοπό την παράσταση πληροφοριών για την λογική δομή και την εμφάνιση του εγγράφου. Η πληροφορία αυτή χρησιμοποιείται από διάφορα εργαλεία για την επεξεργασία και παρουσίαση (εκτύπωση ή παρουσίαση στην οθόνη του υπολογιστή) του εγγράφου. Η παράσταση της πληροφορίας γίνεται σε διάφορες γλώσσες (π.χ. LaTeX, HTML, Microsoft Word, Postscript, κ.α.) ενώ ένα εργαλείο (tool) επεξεργασίας εγγράφων συνήθως μπορεί να ερμηνεύσει ένα μικρό υποσύνολο αυτών των γλωσσών. Άρα απαιτείται η προώθηση ενός προτύπου που θα μπορεί να περιγράψει την αρχιτεκτονική των εγγράφων, ώστε να γίνεται εύκολα η διαχείριση και μετατροπή ενός εγγράφου από την μια μορφή στην άλλη. Τα δύο σημαντικότερα πρότυπα σε αυτόν τον χώρο είναι η SGML (Standard Generalized Markup Language) και το ODA (Open/Office Document Architecture). Η SGML δημιουργήθηκε για το περιβάλλον των δημοσιεύσεων από το Ερευνητικό Κέντρο της IBM στο San Jose των ΗΠΑ και χρησιμοποιείται τόσο στην Αμερική όσο και στην Ευρώπη. Ενώ το ODA αναπτύχθηκε αρχικά για το περιβάλλον γραφείου και προέρχεται από την Ευρώπη. Εξάλου μελέτες που ασχολούνται με τους τρόπους παράστασης της λογικής δομής εγγράφων συναντάμε στα άρθρα [Fur89], [Mur89], [Bai92], [FNK92], [HI93].

4.1.1 Standard Generalized Markup Language (SGML)

Η SGML είναι μια γλώσσα για την περιγραφή τύπων από έγγραφα. Σκοπός της είναι να δημιουργηθούν έγγραφα που είναι ανεξάρτητα από το λογισμικό (software) και το υλικό (hardware) που χρησιμοποιούνται, καθώς και να αυτοματοποιηθεί η επεξεργασία των εγγράφων που βασίζεται στην δομή τους. Χρησιμοποιεί *σύμβολα (Markups)* για να διαχωρίσει τα λογικά τμήματα ενός εγγράφου. Συγκεκριμένα η SGML χρησιμοποιείται για να περιγράψει τον *Ορισμό Τύπου του Εγγράφου* (*Document Type Definition* ή *DTD*). Ο DTD προσδιορίζει με μια γραμματική τους τύπους των συμβόλων(Markups) που επιτρέπονται, τα σύμβολα που απαιτούνται, πώς τα σύμβολα σχετίζονται μεταξύ τους και πώς διαχωρίζονται από το περιεχόμενο του εγγράφου.

Ένα παράδειγμα DTD παρουσιάζεται στην εικόνα 4.1(α). Ο τύπος εγγράφου που ορίζεται έχει την ονομασία *memo*[1] και αποτελείται από *παραγράφους* και *λίστες*. Το έγγραφο και κάθε τμήμα του μπορεί να έχουν κάποια γνωρίσματα (attributes). Στο συγκεκριμένο DTD, το *memo* έχει ένα γνώρισμα με το όνομα *status* που μπορεί να έχει

[1] Ένα μικρό έγγραφο στο οποίο γράφουμε ορισμένες πληροφορίες που θέλουμε να θυμόμαστε.

(α): Ορισμός ενός Τύπου Εγγράφου (Dodument Type Definition)

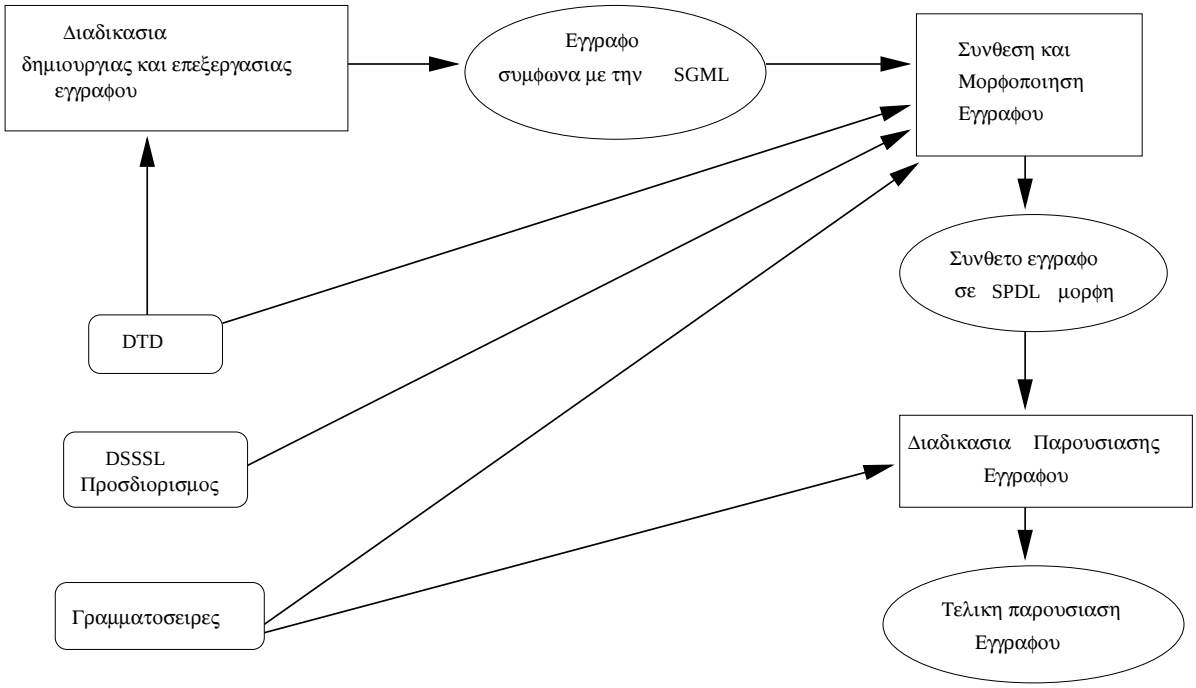
```
  < −− DTD for simple memoran data −− >
<!ENTITY %doctype "memo" −−document type generic identifier −− >
 <! − −          ELEMENTS   MIN        CONTENT (EXCEPTIONS) −− >
 <!ELEMENT       memo       −−         (title?, body) >
 <!ELEMENT       body       −()        (paragraph* | list*) >
 <!ELEMENT       paragraph  −()        (#PCDATA | emphasis) >
 <!ELEMENT       emphasis   −−         (#PCDATA) >
 <!ELEMENT       list       −−         (head?, listitem+) >
 <!ELEMENT       head       −()        (#PCDATA) >
 <!ELEMENT       listitem   −()        (#PCDATA | paragraph *) >
 <! − −                                              −− >
 <! − −          ELEMENTS   NAME       VALUE              DEFAULT −− >
 <!ATTLIST       memo       STATUS     (Private | Public) Public >
 <!ATTLIST       paragraph  id         ID                 #IMPLIED >
 <!ENTITY        SGML       "Standard  Generalized Markup Language">
```

(β): Μια περίπτωση του παραπάνω DTD

```
<memo>
<title>Το πρότυπο περιγραφής εγγράφων SGML</title>
<body>
<paragraph> Η SGML είναι μια γλώσσα περιγραφής τύπων εγγράφων.
Χρησιμοποιεί <emphasis>σύμβολα (Markups)</emphasis> για να διαχωρίσει
τα λογικά τμήματα ενός εγγράφου.  Συγκεκριμένα η SGML
χρησιμοποιείται για να περιγράψει τον <emphasis> Ορισμό Τύπου του
Εγγράφου</emphasis> ( <emphasis>Document Type Definition</emphasis> ή
<emphasis>DTD</emphasis> ).
<list>
<head>Το DTD προσδιορίζει με μια γραμματική τα εξής:
<listitem>Τους τύπους των συμβόλων(Markups) που επιτρέπονται
<listitem>Τα σύμβολα που απαιτούνται
<listitem>Πώς τα σύμβολα διαχωρίζονται από το περιεχόμενο του εγγράφου
</list>
</memo>
```

Σχήμα 4.1: Παράδειγμα SGML DTD

(α) Ένας Ορισμός Τύπου Εγγράφου (Document Type Definition), (β) ένα συγκεκριμένο έγγραφο που είναι περίπτωση (instance) αυτού του DTD.



Σχήμα 4.2: Το SGML μοντέλο επεξεργασίας

Το ορθογώνιο δηλώνει μια διαδικασία, η έλλειψη δηλώνει μια κατάσταση του εγγράφου και το ορθογώνιο με τις στρογγυλεμένες γωνίες δηλώνει τους προσδιορισμούς για την παράσταση του εγγράφου.

τιμή *Ιδιωτικό* ή *Δημόσιο*, ενώ η προκαθορισμένη τιμή του είναι *Δημόσιο*. Στο σχήμα 4.1(β) παρουσιάζεται ένα συγκεκριμένο έγγραφο που συμφωνεί με το παραπάνω SGML DTD.

Η SGML αναφέρεται μόνο στην λογική δομή του εγγράφου, ενώ η μορφή παρουσίασης προσδιορίζεται σύμφωνα με το πρότυπο DSSSL (Document Style Semantic and Specification Language). Το DSSSL ασχολείται με την μορφοποίηση του εγγράφου. Συγκεκριμένα παρέχει, σε μια πρότυπη (standard) μορφή, τις οδηγίες για την μορφοποίηση (formatting) του εγγράφου. Αλλά επειδή δεν είναι μορφοποιητής (formatter) απαιτεί την ύπαρξη μεταφραστών για τους υπάρχοντες μορφοποιητές. Ένα άλλο πρότυπο σχετικά με την SGML είναι το SPDL (Standard Page Description Language), το οποίο περιγράφει, ανεξάρτητα από την συσκευή και την διαδικασία, πώς το κείμενο και οι εικόνες του εγγράφου θα τυπωθούν στο χαρτί ή θα παρουσιαστούν σε μια άλλη συσκευή. Η περιγραφή γίνεται με ένα γενικό τρόπο ώστε να μην περιορίζεται σε συγκεκριμένη συσκευή ή διαδικασία.

Τα δύο αυτά πρότυπα χρησιμοποιούνται στο SGML μοντέλο επεξεργασίας εγγράφων που παρουσιάζεται στην εικόνα 4.2.

4.1.2 Open/Office Document Architecture (ODA)

Το ODA αναπτύχθηκε για το περιβάλλον γραφείου και παρέχει την δυνατότητα διατήρησης των εγγράφων σε μια *επεξεργάσιμη μορφή (processable form)*[2], ώστε να μπορεί να μεταφέρεται σε διαφορετικά συστήματα επεξεργασίας πληροφορίας. Αρχικά ονομαζόταν Office Document Architecture γιατί προοριζόταν ως πρότυπο για το περιβάλλον γραφείων, αλλά λόγω των δυνατοτήτων που προσφέρει δεν περιορίστηκε τελικά εκεί και μετονομάστηκε σε Open Document Architecture. Το ODA ορίζει την γενική αρχιτεκτονική παράστασης των πληροφοριών που περιέχονται σε ένα έγγραφο. Αναφέρεται τόσο στην λογική δομή όσο και στην μορφή παρουσίασης (layout). Το πρότυπο αποτελείται από 6 τμήματα:

- Την λογική όψη (logical view) : είναι η λογική δομή του εγγράφου.
- Την όψη παρουσίασης (layout view) : είναι ο τρόπος μορφοποίησης του εγγράφου.
- Το περιεχόμενο (content) : είναι το κείμενο, τα γεωμετρικά γραφικά και οι εικόνες του εγγράφου.
- Το προφίλ του εγγράφου (document profile): περιέχει τα γνωρίσματα που χαρακτηρίζουν ολόκληρο το έγγραφο. Συγκεκριμένα περιέχει πληροφορίες διαχείρισης (π.χ. ο τίτλος, ο συγγραφέας, ημερομηνία δημιουργίας, λέξεις κλειδιά, κ.α.) και τεχνικές πληροφορίες (π.χ. αν το έγγραφο περιέχει την λογική δομή του ή την μορφή παρουσίασης του ή και τα δύο).
- Γενική λογική δομή και όψη παρουσίασης (Logical and Layout Generic Structure) : είναι ένα σύνολο κανόνων που ορίζουν τον τύπο του εγγράφου.
- Το στυλ ή όψη του εγγράφου (style) : περιγράφει τον γενικό τρόπο παρουσίασης των λογικών τμημάτων και των πραγματικών πληροφοριών που περιέχει το έγγραφο.

Ένα άλλο πρότυπο, σχετικό με το ODA, είναι το ODL (Open Document Language). Αυτό το πρότυπο είναι μια SGML εφαρμογή, δηλαδή ένα σύνολο από DTDs, που προσδιορίζει την μετατροπή ενός ODA εγγράφου σε ένα SGML έγγραφο.

4.1.3 Σύγκριση ODA και SGML

Το πρότυπο ODA πλεονεκτεί σε σχέση με το SGML στα εξής σημεία:

[2]με τον όρο επεξεργάσιμη μορφή εννοούμε ότι υπάρχουν όλες οι πληροφορίες, σύμφωνα με το πρότυπο, για την λογική δομή και την μορφή παρουσίασης

- Παρέχει πληροφορία τόσο για την λογική δομή όσο και για την μορφοποίηση (layout) του εγγράφου, ενώ το SGML παρέχει πληροφορία μόνο για την λογική δομή.
- Υπάρχει το ODL πρότυπο που χρησιμοποιείται για την μετατροπή ODA εγγράφων σε SGML έγγραφα.
- Για την μετάδοση των ODA εγγράφων αρκεί ο αποδέκτης να γνωρίζει ότι πρόκειται για ένα ODA έγγραφο, ενώ για ένα SGML έγγραφο θα πρέπει αυτός που στέλνει το έγγραφο και αυτός που το λαμβάνει να συμφωνούν σε ένα συγκεκριμένο DTD.

αλλά υστερεί στα εξής σημεία:

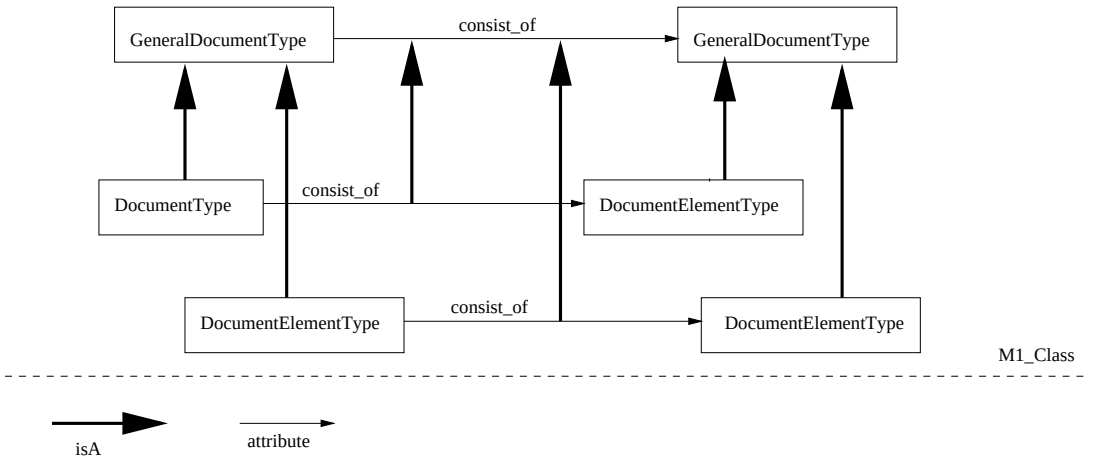
- Το ODA πρότυπο είναι αρκετά μεγάλο και πολύπλοκο σε σχέση με το SGML. Άρα είναι πιο δύσκολο στην εκμάθηση.
- Το ODA δεν μπορεί να διαχειρίζεται πλήρως πολύμορφα (multimedia) έγγραφα.
- Το μέγεθος ενός εγγράφου σε ODA είναι πολύ μεγάλο σε σχέση με το αντίστοιχο σε SGML.
- Υπάρχουν λίγες εφαρμογές που κάνουν χρήση του προτύπου ODA. Αυτή η αδυναμία του είναι ίσως η σημαντικότερη και σε συνδυασμό με την μεγάλη χρήση του SGML οδηγεί στην ακόμη μεγαλύτερη αποδοχή του SGML.

Στην εργασία που παρουσιάζεται παραστήσαμε, με την γλώσσα Telos, την λογική δομή που περιγράφεται στο πρότυπο ISO 12083, το οποίο είναι ένα SGML DTD για την δημιουργία άρθρων και βιβλίων. Το πρότυπο αυτό έχει προταθεί από τον οργανισμό Association of American Publishers (AAP) και αυτήν την περίοδο βελτιώνεται από το Electronic Publishing Special Interest Group (EPSIG)[3]. Έχει ήδη γίνει αποδεκτό από μεγάλους οργανισμούς έκδοσης επιστημονικών περιοδικών (π.χ American Physical Society[4], IEEE Computer Society, American Society of Civil Engineers, American Society of Agricultural Engineers, American Institute of Aeronautics and Astronautics) και χρησιμοποιείται από πολλά πρόσφατα συστήματα διαχείρισης και αναζήτησης επιστημονικών άρθρων (π.χ. [SMC+96]). Διαδίδεται με ταχείς ρυθμούς λόγω της επιρροής του EPSIG που ανέλαβε την βελτίωση και προώθηση του.

Στις επόμενες ενότητες παρουσιάζεται το μοντέλο που υλοποιήσαμε για την παράσταση των εγγράφων.

[3]Πληροφορίες για το EPSIG υπάρχουν στο URL:http://www.sil.org/sgml/epsig.html

[4]Για περισσότερες πληροφορίες βλέπε την σελίδα στο WWW
URL: http://publish.aps.org/SGML/sgml.html



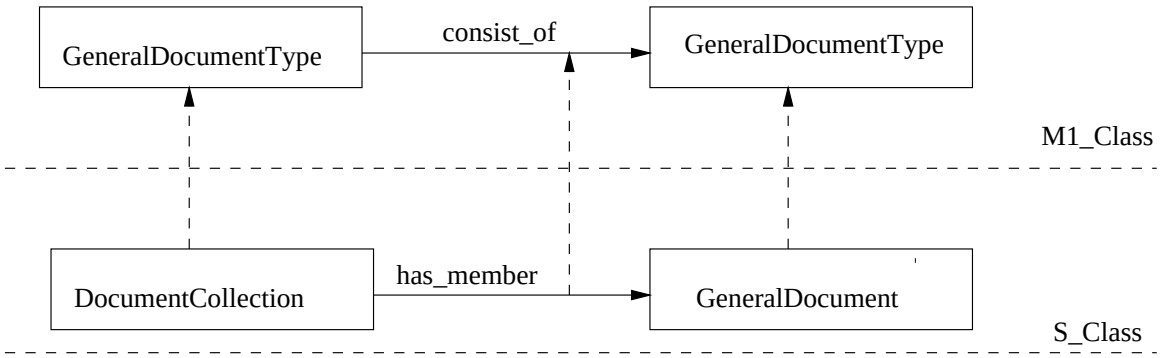
Σχήμα 4.3: Το μεταμοντέλο παράστασης εγγράφων

4.2 Το γενικό μοντέλο του εγγράφου

Στο μοντέλο του εγγράφου που υλοποιήσαμε σημαντικό ρόλο διαδραματίζει η λογική δομή των πληροφοριών του εγγράφου. Σύμφωνα με την ορολογία του προτύπου SGML, κάθε DTD ορίζει ένα τύπο εγγράφου (Document Type) που αποτελείται από κάποια λογικά τμήματα (Document Elements), τα οποία μπορεί να αποτελούνται από άλλα λογικά τμήματα, κ.ο.κ. Το μεταμοντέλο του εγγράφου παρουσιάζεται στην εικόνα 4.3.

Η ανάγκη να ομαδοποιήσουμε όλα τα τμήματα εγγράφων και τους τύπους εγγράφων σε μια κλάση οδήγησε στον ορισμό μιας κλάσης με το όνομα GeneralDocumentType, η οποία είχε ως υποκλάσεις τις DocumentType και DocumentElementType (εικόνα 4.3). Η κλάση GeneralDocumentType χρησιμοποιείται όταν θέλουμε να αναφερθούμε σε ένα τύπο εγγράφου ή ένα τύπο τμήματος του και είναι πολύ σημαντική για την διατύπωση γενικών ερωτήσεων. Η κλάση DocumentType έχει ως περιπτώσεις όλους τους τύπους εγγράφων. Ενώ το DocumentElementType έχει ως περιπτώσεις όλους τους τύπους τμημάτων των εγγράφων.

Για κάθε μια από τις παραπάνω τρεις κλάσεις έχουμε δημιουργήσει τις συζυγείς τους κλάσεις. Με την έκφραση ”το ζεύγος DocumentType και Document είναι συζυγές” εννοούμε ότι όλες οι περιπτώσεις (instances) της κλάσης DocumentType είναι υποκλάσεις της κλάσης Document. Ο μηχανισμός αυτός των συζυγών οντοτήτων περιγράφεται στο 4ο κεφάλαιο του [Χρι94b] και έχει ως στόχο τον ορισμό εννοιών σε διαφορετικά επίπεδα αφαίρεσης ώστε να είναι πιο κατανοητό το μοντέλο και να γίνονται ευκολότερα προκατασκευασμένες ερωτήσεις που βασίζονται στις ”γενικές” έννοιες.



Σχήμα 4.4: Το μεταμοντέλο παράστασης Συλλογών εγγράφων

Στο συγκεκριμένο μοντέλο οι πραγματικοί τύποι εγγράφων που υπάρχουν είναι το άρθρο και το βιβλίο. Ενώ ένας μη πραγματικός, αλλά χρήσιμος, τύπος εγγράφου είναι η συλλογή από έγγραφα ή τμήματα εγγράφων που περιγράφεται στην εικόνα 4.4. Η κλάση των συλλογών ορίζεται στο S_Class επίπεδο και άρα οι περιπτώσεις της θα ανήκουν στο Token επίπεδο. Ένας φυσιολογικός ορισμός θα ήταν να ορίζαμε την κλάση των συλλογών στο M1_Class επίπεδο ώστε οι συγκεκριμένες συλλογές να ορίζοταν στο S_Class επίπεδο και τα μέλη της συλλογής να ήταν περιπτώσεις (instances) της συλλογής. Αλλά με αυτόν τον τρόπο δεν θα μπορούσαμε να κάνουμε σημειώσεις στις συλλογές διότι ορίσαμε τις σημειώσεις να γίνονται σε δεδομένα, δηλαδή σε οντότητες που ανήκουν στο Token επίπεδο.

Βέβαια αυτός είναι ένας μη δυναμικός τρόπος να ορίζουμε συλλογές. Στην ενότητα 5.5 παρουσιάζεται πως μπορούμε να δημιουργούμε δυναμικά συλλογές εγγράφων ή τμημάτων τους με κριτήριο τα γνωρίσματα τους (π.χ. με βάση τον συγγραφέα του εγγράφου, την ημερομηνία εγγραφής, το είδος σημείωσης που έχουν, το περιεχόμενο της σημείωσης, κ.α.)

4.3 Το μοντέλο παράστασης του άρθρου σύμφωνα με το ISO 12083

4.3.1 Το SGML DTD

Το ISO 12083 είναι ένα σύνολο από SGML DTDs για την δημοσίευση και ηλεκτρονική μετάδοση άρθρων και βιβλίων. Δημιουργήθηκε το 1993 από τον οργανισμό Association of American Publishers (AAP) και αυτήν την περίοδο βελτιώνεται από το Electronic Publishing Special Interest Group (EPSIG) Έχει ήδη γίνει αποδεκτό από μεγάλους οργανισμούς έκδοσης επιστημονικών περιοδικών (π.χ American Physical Society, IEEE Computer Society, American

Society of Civil Engineers, American Society of Agricultural Engineers, American Institute of Aeronautics and Astronautics) και χρησιμοποιείται από πολλά πρόσφατα συστήματα διαχείρισης και αναζήτησης επιστημονικών άρθρων (π.χ. [SMC+96]). Διαδίδεται με ταχείς ρυθμούς λόγω της επιρροής του EPSIG που ανέλαβε την βελτίωση και προώθηση του.

Σύμφωνα με την περιγραφή του DTD, το άρθρο ορίζεται στην SGML γλώσσα ως εξής:

<!	ELEMENT	MIN	CONTENT	(EXCEPTIONS)
<!ELEMENT	article	--	(front, body, appmat?, back)	+(figgrp|footnote|note)

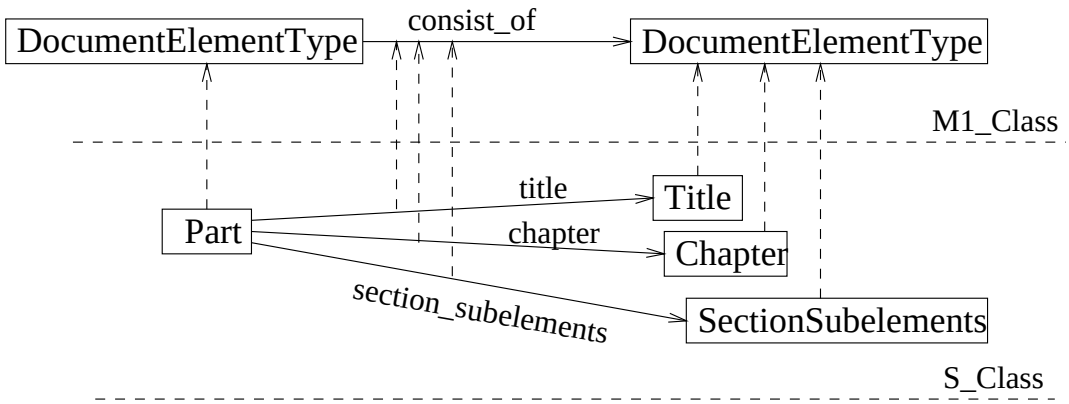
Ο συμβολισμός ”--” στο πεδίο MIN σημαίνει ότι το άρθρο αρχίζει μετά το σύμβολο <article> και τελειώνει πριν από το σύμβολο </article>. Το άρθρο αποτελείται από την κεφαλή (front), που περιλαμβάνει πληροφορίες για το άρθρο (π.χ. συγγραφέας, τίτλος, ημερομηνία έκδοσης, περίληψη, κ.α.) , τον κορμό (body), που περιλαμβάνει όλα τα λογικά τμήματα του άρθρου, ένα ή περισσότερα παραρτήματα (appmat) και την ουρά (back), που περιλαμβάνει της πληροφορίες που τοποθετούνται στο τέλος του άρθρου (π.χ. λίστα βιβλιογραφικών αναφορών, ευρετήριο όρων, κ.α.). Επίσης σε οποιοδήποτε σημείο του άρθρου μπορεί να υπάρχει μια εικόνα ή υποσημείωση ή γενική σημείωση. Για κάθε τμήμα του άρθρου υπάρχουν κανόνες που δηλώνουν την λογική τους δομή. Το πλήρες περιεχόμενο του DTD περιγράφεται στο παράρτημα Β.

4.3.2 Η παράσταση του ISO 12083 με την γλώσσα Telos

Η παράσταση του ISO 12083 με ένα Telos μοντέλο έχει ως στόχο την αποθήκευση της πληροφορίας για την λογική δομή του άρθρου, ώστε να γίνεται η περιπλάνηση (navigation) στα περιεχόμενα του άρθρου και την δημιουργία σημειώσεων σε αυτά. Κατά τη διαδικασία παράστασης αφαιρέθηκαν ορισμένες ενδιάμεσες λογικές οντότητες επειδή κρίθηκε ότι δεν είχαν λειτουργικότητα[5]. Για παράδειγμα το *άρθρο* σύμφωνα με το ISO 12083 έχει μια λογική οντότητα που ονομάζεται *πρόσοψη (front)*, η οποία αποτελείται από τον τίτλο, τους συγγραφείς, μια ημερομηνία και μια λογική ενότητα που περιλαμβάνει πληροφορίες για την δημοσίευση του άρθρου. Στην συγκεκριμένη υλοποίηση αφαιρέθηκε η *πρόσοψη* και τα λογικά της τμήματα αποδόθηκαν άμεσα ως λογικά τμήματα του *άρθρου*.

Η γενική τακτική που ακολουθήσαμε για την κατασκευή του μοντέλου συνοψίζεται στα παρακάτω βασικά σημεία, που δίνονται με την μορφή παραδείγματος.

[5]με αυτό εννοούμε ότι δεν προσφέρει σημαντικά στην οργάνωση της δομής και δεν υπάρχει η ανάγκη δημιουργίας σημειώσεων σε αυτό το λογικό τμήμα.



Σχήμα 4.5: Η παράσταση του κανόνα *part* $->$ *(title?, section_subelement*, chapter+)*

1. Για τον κανόνα της μορφής:

ELEMENT	CONTENT
part	(title?, section_subelement∗, chapter+)

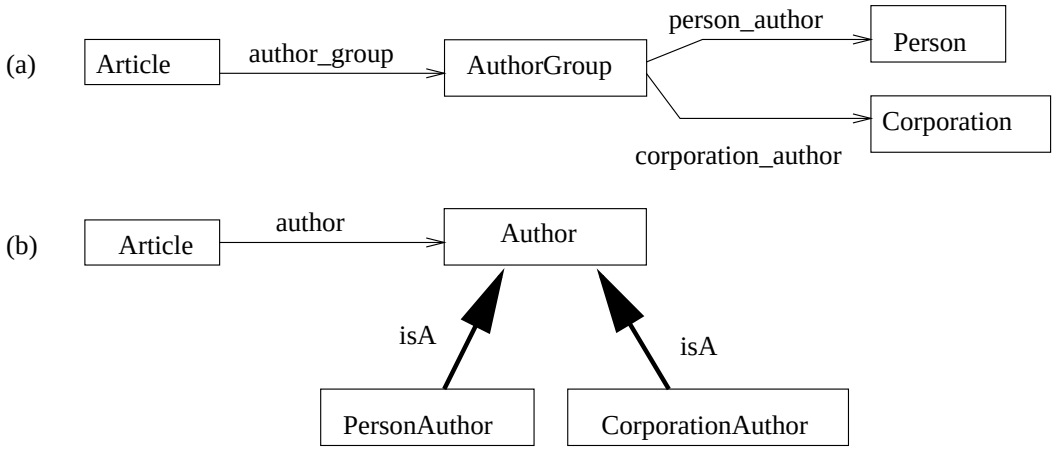
δημιουργήσαμε μια οντότητα *Part* με γνωρίσματα τα title, section_subelement, chapter όπως φαίνεται στην εικόνα 4.5. Τα γνωρίσματα αυτά του Part είναι περιπτώσεις της κατηγορίας consist_of της μετακλάσης DocumentElementType.

Με την υλοποίηση της γλώσσας Telos που χρησιμοποιήσαμε δεν μπορούμε να θέσουμε περιορισμούς στις κατηγορίες και στον αριθμό των γνωρισμάτων μιας οντότητας. Άρα δεν μπορούμε σε μια περίπτωση (instance) της κλάσης Part να δώσουμε αναγκαστικά ένα ή κανένα τίτλο, καένα ή κάποιο αριθμό από τμήματα ενότητας και ένα ή περισσότερα κεφάλαια. Αλλά δίνεται η *δυνατότητα* να οριστεί σωστά η λογική δομή του, αφού ένα γνώρισμα μιας οντότητας μπορεί να μην πάρει τιμή ή να πάρει μία ή περισσότερες τιμές (βλ. ενότητα 3.2). Ο έλεγχος των περιορισμών, που θέλουμε να ισχύουν, γίνεται κατά την διάρκεια της εισαγωγής των δεδομένων (βλ. ενότητα 4.4.1).

2. Για τον κανόνα της μορφής:

ELEMENT	CONTENT
authorgroup	(person_author | corporation_author) ∗

υπάρχουν δύο τρόποι παράστασης που μας απασχόλησαν:



Σχήμα 4.6: Παραδείγματα για τους δύο εναλλακτικούς τρόπους παράστασης του κανόνα:*authorgroup* $->$ *(author* | *corporation_author)*

α). Αποδίδουμε στην οντότητα AuthorGroup τα δύο γνωρίσματα person_author και corporation_author, σχήμα 4.6(a), που παίρνουν τιμές από τις κλάσεις Person και Corporation αντίστοιχα.

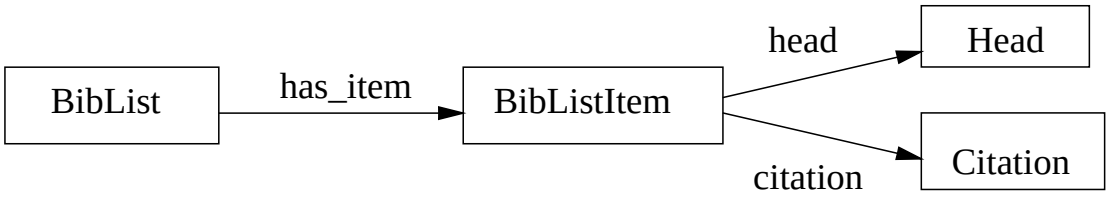
β). Ορίζουμε τις κλάσεις PersonAuthor και CorporationAuthor ως εξειδικεύσεις της κλάσης Author, ώστε οι συγγραφείς (άνθρωποι, οργανισμοί) να συνδέονται άμεσα με το *άρθρο*, σχήμα 4.6(b), και όχι μέσω του λογικού τμήματος *Author-Group*.

Στην υλοποίηση μας επιλέξαμε την δεύτερη λύση διότι προσφέρει οικονομία χώρου, όπως φαίνεται στο σχήμα 4.6. Ενώ η πρώτη λύση δεν προσφέρει σημαντικά στην οργάνωση της λογικής δομής. Σε παρόμοιους κανόνες ακολουθήσαμε την ίδια διαδικασία, εκτός αν κρίναμε ότι δεν μπορούμε να παραλείψουμε το ενδιάμεσο λογικό τμήμα του εγγράφου για λόγους οργάνωσης.

3. Για τον κανόνα της μορφής:

ELEMENT	CONTENT
biblist	(head, citation) $*$

ορίζουμε τις κλάσεις BibList, Head και Citation. Αλλά δεν μπορούμε να ορίσουμε στην κλάση BibList δύο γνωρίσματα head και citation γιατί η κλάση αυτή θα πρέπει να έχει ένα γνώρισμα που παίρνει τιμή τύπου (Head, Citation). Άρα ορίζουμε μια νέα οντότητα με όνομα BiblistItem που έχει ως γνωρίσματα το head και citation που



Σχήμα 4.7: Η παράσταση του κανόνα:*biblist* $->$ *(head, citation)* *

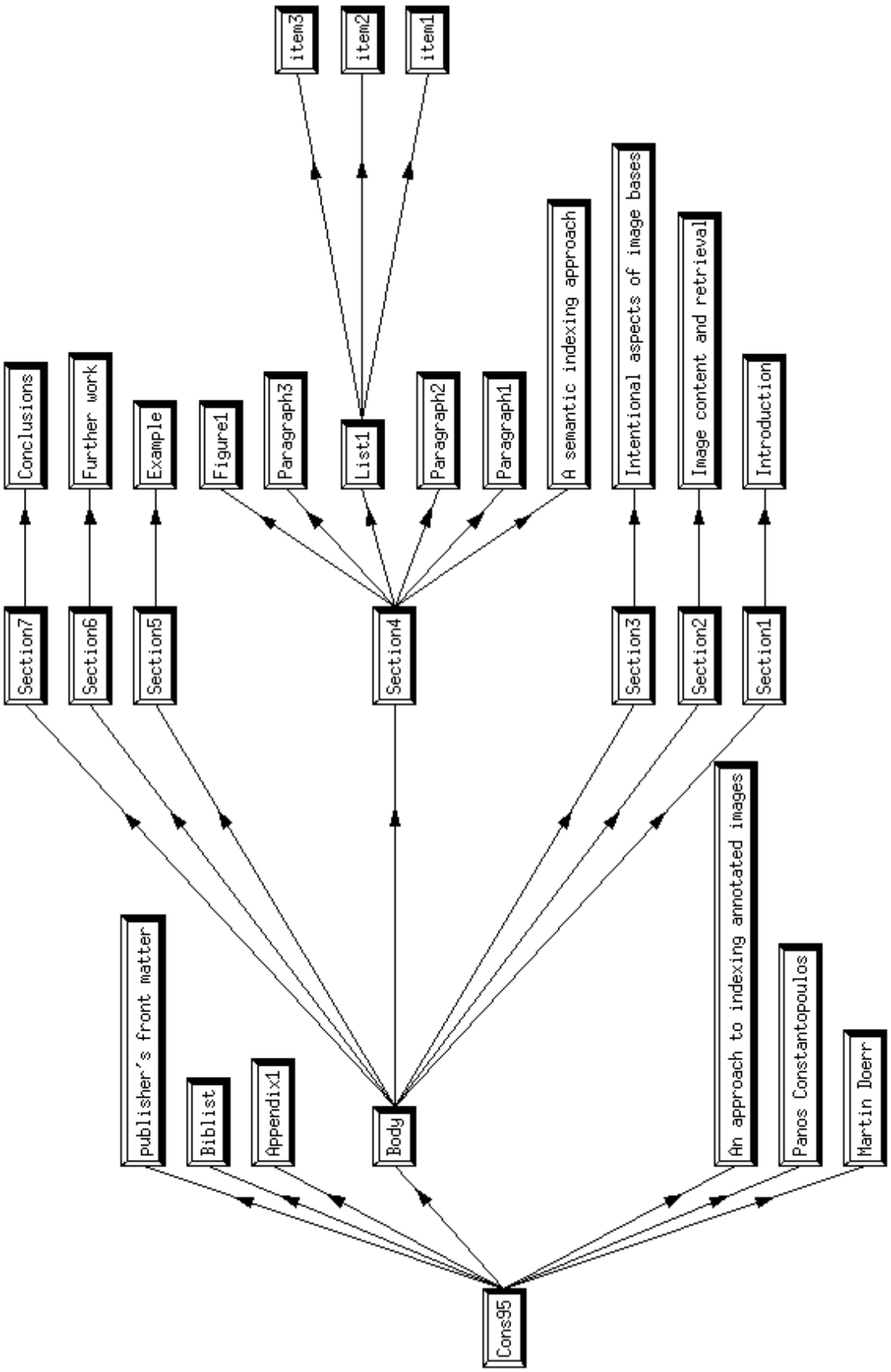
παίρνουν τιμές από τις κλάσεις Head και Citation αντίστοιχα, όπως φαίνεται στο σχήμα 4.7.

Ένα άλλο σημαντικό σημείο που μας απασχόλησε, κατά την δημιουργία του μοντέλου παράστασης του DTD με την γλώσσα Telos, είναι ο τρόπος σύνδεσης του λογικού τμήματος με το περιεχόμενο του, δηλαδή την πληροφορία που υπάρχει σε αυτό το λογικό τμήμα. Η λύση που δώσαμε είναι ότι στην περίπτωση που το περιεχόμενο προβλέπεται να είναι μικρότερο από το μήκος των λογικών ονομάτων στην Telos[6] , όπως στην περίπτωση του συγγραφέα και της ημερομηνίας, τότε βάζουμε το περιεχόμενο ως λογικό όνομα της οντότητας. Διαφορετικά αποδίδουμε μια τιμή στο γνώρισμα *display* (βλ. ενότητα 3.2) της οντότητας που παριστάνει το λογικό τμήμα. Η τιμή αυτή είναι ένα εκτελέσιμο πρόγραμμα που εμφανίζει ένα επεξεργαστή κειμένου με το έγγραφο στο σημείο που αρχίζει το συγκεκριμένο λογικό τμήμα. Μια εξαίρεση γίνεται στις περιπτώσεις της κλάσης *Title* που βάζουμε ως λογικό όνομα το ίδιο το περιεχόμενο του τίτλου και όχι ένα όνομα της μορφής *Article'niel95'Title*. Πιστεύουμε ότι είναι χρήσιμο να έχουμε αυτήν την πληροφορία ώστε να μην χρειάζεται να ανατρέχουμε στο ηλεκτρονικό αρχείο του εγγράφου.

Ένα παράδειγμα περίπτωσης εγγράφου σύμφωνα με το ISO 12083 είναι στην εικόνα 4.8. Σε αυτή την εικόνα φαίνονται τα ονόματα των λογικών τμημάτων του εγγράφου και οι μεταξύ τους συσχετίσεις. Όπως περιγράφεται στο κεφάλαιο 6, παρέχονται οι δυνατότητες για την περιπλάνηση στα λογικά τμήματα του εγγράφου, τη παρουσίαση της πληροφορίας που περιέχουν και τη δημιουργία σημειώσεων σε αυτά.

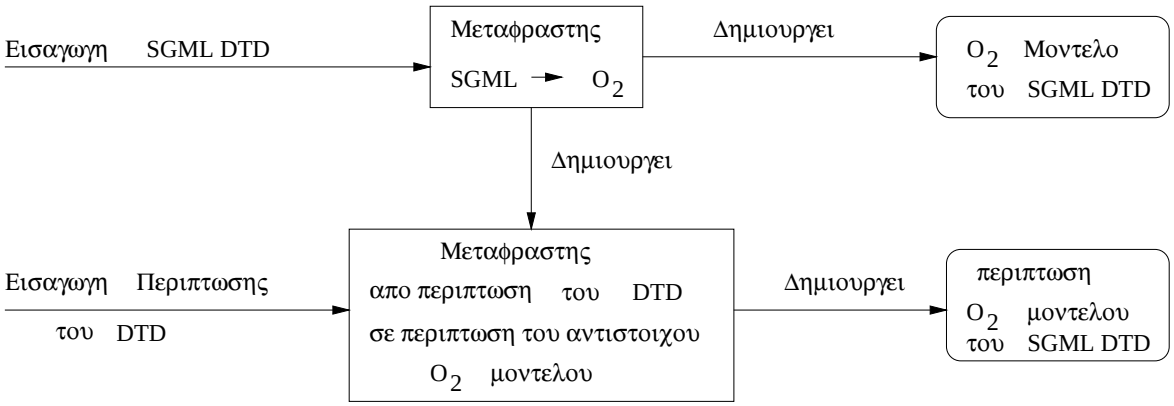
Η εργασία της παράστασης αυτού του SGML DTD με την γλώσσα Telos αποτελεί περίπτωση μιας γενικότερης εργασίας, που γίνεται παράλληλα, με σκοπό την δημιουργία μιας απεικόνισης της γλώσσας SGML σε Telos [Τρα96b]. Με βάση αυτή την απεικόνιση θα μπορούν να δημιουργηθούν τα εργαλεία για την αυτόματη παραγωγή του Telos

[6]Το μήκος των λογικών ονομάτων στη Telos είναι μικρότερο από 96 χαρακτήρες



Σχήμα 4.8: Μια περίπτωση άρθρου

Στο σχήμα φαίνονται τα λογικά τμήματα του άρθρου. Το άρθρο έχει ένα τίτλο, δυο συγγραφείς, ένα τμήμα που έχει πληροφορίες για την δημοσίευση του άρθρου, ένα παράρτημα, μια βιβλιογραφική λίστα και τον κυρίως κορμό (body). Ο κυρίως κορμός αποτελείται από ενότητες που αποτελούνται από παραγράφους,λίστες και εικόνες.



Σχήμα 4.9: Το γενικό μοντέλο μετάφρασης από SGML σε O_2

Μετάφραση του SGML DTD σε ένα μοντέλο της βάσης δεδομένων O_2 και μετάφραση περίπτωσης του SGML DTD στο μοντέλο της O_2 που παριστάνει αυτή τη περίπτωση

μοντέλου ενός συγκεκριμένου SGML DTD. Παρόμοιες εργασίες έχουν αναφερθεί στο [CR94] για την μετάφραση ενός SGML DTD σε ένα μοντέλο της βάσης δεδομένων O_2 με ταυτόχρονη δημιουργία ενός μεταφραστή από μια περίπτωση του DTD σε μια περίπτωση του αντίστοιχου O_2 μοντέλου (βλ. εικόνα 4.9).

4.4 Εισαγωγή δεδομένων

Η εισαγωγή δεδομένων αποτελεί ένα σημαντικό σημείο απασχόλησης σε κάθε πληροφοριακό σύστημα. Η επιθυμία όλων των χρηστών είναι να εισάγονται τα δεδομένα αυτόματα, αλλά όταν αυτό είναι αδύνατο θα πρέπει τουλάχιστον να υπάρχει ένα εργαλείο (tool) εισαγωγής δεδομένων το οποίο να διευκολύνει την εργασία. Στο σύστημα που δημιουργήσαμε έχουμε υλοποιήσει ορισμένα εργαλεία για την αυτόματη εισαγωγή δεδομένων που υπάρχουν σε ορισμένες μορφές (format). Επίσης παρέχουμε και ένα εργαλείο που κατευθύνει και βοηθάει στην εισαγωγή δεδομένων.

4.4.1 Αυτόματη εισαγωγή δεδομένων

Στα πλαίσια αυτής της εργασίας κατασκευάστηκε ένα φίλτρο για την αυτόματη εξαγωγή της λογικής δομής ενός άρθρου που περιγράφεται σύμφωνα με το πρότυπο ISO12083. Το φίλτρο έχει ως είσοδο ένα άρθρο που περιγράφεται σύμφωνα με το πρότυπο

ISO 12083 και έχει ως έξοδο ένα αρχείο στο οποίο περιγράφεται η λογική δομή του άρθρου και η σύνδεση με το περιεχόμενο του, σύμφωνα με το μοντέλο στη γλώσσα Telos. Έπειτα το αρχείο αυτό διοχετεύεται στο μεταφραστή της Telos και αποθηκεύεται στη βάση δεδομένων. Η διαδικασία γίνεται με την βοήθεια μιας επαφής χρήσης (interface) (βλ. εικόνα 6.9) όπου δίνουμε το λογικό όνομα του άρθρου (με το οποίο θα είναι γνωστό στη βάση) και το ηλεκτρονικό αρχείο του άρθρου. Βέβαια πρώτα ελέγχουμε με τον yasp SGML μεταφραστή αν το ηλεκτρονικό αρχείο είναι πράγματι περίπτωση (instance) του DTD του προτύπου ISO 12083.

Η λογική δομή του άρθρου της εικόνας 4.8 έχει δημιουργηθεί αυτόματα με την βοήθεια του φίλτρου αυτού από ISO 12083 σε Telos. Πρέπει να επισημάνουμε ότι τα πλήρη όνοματα των οντοτήτων δεν είναι αυτά που εμφανίζονται, δηλαδή το πλήρες όνομα του 2ου κεφαλαίου του άρθρου *Cons95* είναι *Article'Cons95'Body'Chapter2* αλλά στην εμφάνιση της λογικής δομής παρουσιάζεται μόνο το όνομα *Chapter2* για να είναι ο γράφος πιο μικρός και ευκολονόητος. Έτσι τα ονόματα των τμημάτων του εγγράφου, που δίνονται μέσω του φίλτρου, προσδιορίζουν και την θέση τους στο έγγραφο. Βέβαια τα ονόματα αυτά θα μπορούσαν να είναι πολύ πιο μικρά αν υπήρχαν οι μηχανισμοί απόδοσης ονομάτων σε σχέση με το περιβάλλον[7] που παρουσιάζονται σε μια μελέτη για την ονοματοδοσία των οντοτήτων [Θεο95].

Γενικά όταν στο έγγραφο εμπεριέχεται η πληροφορία για την λογική δομή (π.χ. σύμφωνα με ένα SGML DTD) τότε είναι δυνατή η δημιουργία ενός φίλτρου για να την περιγράψει σε μια άλλη γλώσσα παράστασης. Τα τελευταία χρόνια έχουν δημοσιευτεί ορισμένες μελέτες ([MTT+91], [PR94], [HI93]) για την αναδημιουργία της λογικής δομής ενός εγγράφου με βάση την μορφή παρουσίασης (layout) του, π.χ. από postscript να προσδιορίσουμε την λογική δομή του εγγράφου. Μια τέτοια μέθοδος εξαγωγής λογικής δομής δεν υλοποιήθηκε γιατί δεν είναι στα άμεσα ενδιαφέροντα αυτής της εργασίας και το χρονικό κόστος εκτιμήσαμε ότι θα ήταν μεγάλο.

Ακόμη υλοποιήσαμε ένα φίλτρο για την μεταφορά BIBTEX βιβλιογραφικών δεδομένων στο Telos μοντέλο των άρθρων και αντίστροφα. Η BIBTEX [Pat88] μορφή περιλαμβάνει τα στοιχεία για την αναγνώριση και διαχείριση του εγγράφου (π.χ. τον τίτλο, τους συγγραφείς, το περιοδικό έκδοσης, την ημερομηνία έκδοσης, κ.α.) και αποτελεί το μέσο για την δημιουργία βιβλιογραφίας με το LATEX [Lam86]. Υπάρχουν βιβλιογραφικές βάσεις σε όλο τον κόσμο που έχουν βιβλιογραφικά δεδομένα σε BIBTEX μορφή. Επίσης

[7] Η υλοποίηση των μηχανισμών απόδοσης ονομάτων σε σχέση με το περιβάλλον στη γλώσσα Telos βρίσκεται σε εξέλιξη στην ομάδα Τεχνολογίας Λογισμικού του Ινστιτούτου Πληροφορικής στο Ηράκλειο Κρήτης.

υπάρχουν μεταφραστές από άλλες μορφές (π.χ. MARC[8], Refer, BIDS, BIDS, EMBL, NSPEC, MathSci, MedLine , Melvyl, Papyrus, Procite, RIS) σε BIBTEX και αντίστροφα. Επομένως βιβλιογραφικά δεδομένα που παριστάνονται στις παραπάνω μορφές μπορούν να εισάγονται αυτόματα στο Telos μοντέλο. Ενώ τα βιβλιογραφικά δεδομένα που ακολουθούν το Telos μοντέλο μπορούν να μετατρέπονται στη BIBTEX μορφή και έπειτα σε κάποια άλλη μορφή μέσω των φίλτρων που υπάρχουν.

Ένα πρόβλημα που αντιμετωπίσαμε ήταν η ονοματοδοσία των κωδικών για τα άρθρα. Δεν θα πρέπει να βάζουμε στη βάση το ίδιο άρθρο με διαφορετικούς κωδικούς γιατί αυτό θα δημιουργούσε σύγχυση. Για να το ξεπεράσουμε αυτό εφαρμόζουμε τον αλγόριθμο 4.1, όπου η απάντηση στην ερώτηση ''Υπάρχει το άρθρο στη βάση;'' είναι θετική όταν υπάρχει στη βάση άρθρο με τον ίδιο τίτλο, την ίδια ημερομηνία έκδοσης και έχει ένα τουλάχιστον συγγραφέα κοινό με το άρθρο που θέλουμε να εισάγουμε. Ενώ η διαδικασία ''Εύρεση νέου κατάλληλου κωδικού'' κάνει τα εξής: στο τέλος του κωδικού βάζουμε τον χαρακτήρα 'a', έπειτα ελέγχουμε αν ο νέος κωδικός υπάρχει στη βάση και αν υπάρχει τότε βγάζουμε τον χαρακτήρα 'a' και προσθέτουμε τον χαρακτήρα 'b', κ.ο.κ. έως ότου βρούμε ένα κωδικό που δεν υπάρχει στην βάση.

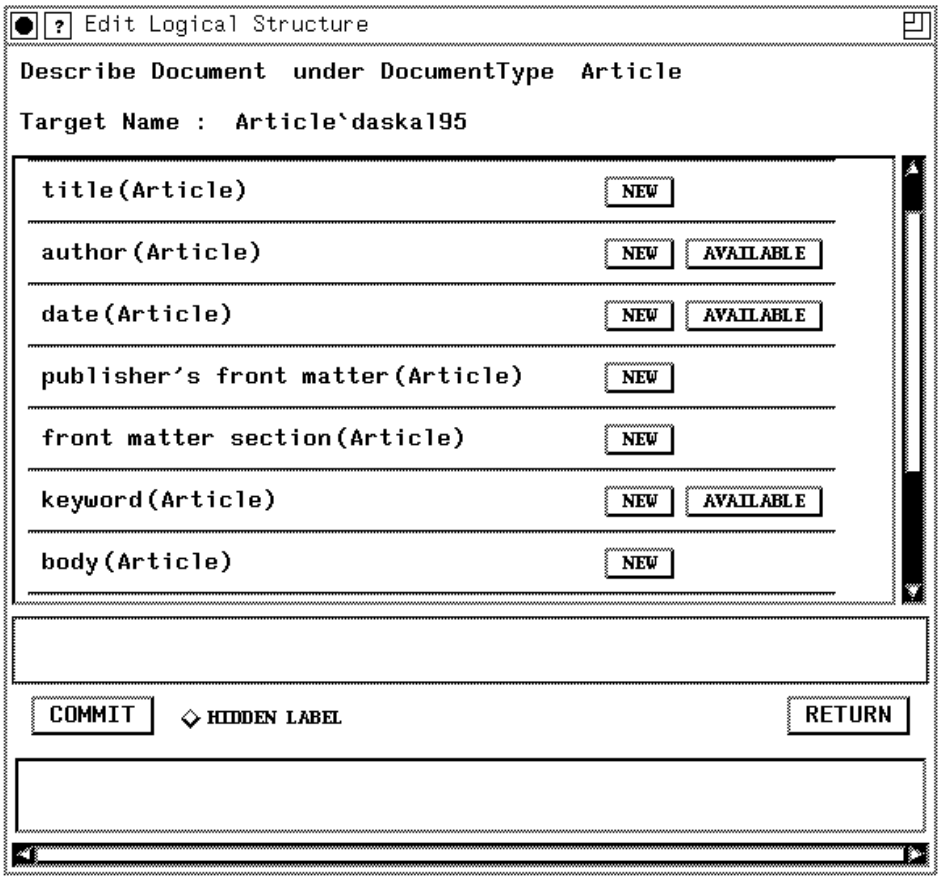
```
Αν   (Υπάρχει το άρθρο στη βάση)   τότε
   Κάνε την ενημέρωση του άρθρου με τον κωδικό που είναι ήδη γνωστό στη βάση
Αλλιώς
      Αν   (Υπάρχει ο ίδιος κωδικός στη βάση)   τότε
                     Βρές νέο κατάλληλο κωδικού
                     Κάνε την εισαγωγή του άρθρου με τον νέο κωδικό
      Αλλιώς
                     Κάνε την εισαγωγή του άρθρου με τον αρχικό κωδικό
```

Αλγόριθμος 4.1 : Επιλογή κωδικού για την εισαγωγή άρθρου.

4.4.2 Διαλογική εισαγωγή δεδομένων

Η αυτόματη εισαγωγή δεδομένων δεν είναι πάντοτε δυνατή. Για τον λόγο αυτό θα πρέπει να δίνεται η δυνατότητα στο χρήστη να εισάγει δεδομένα με ένα φιλικό και διαλογικό τρόπο. Για να ικανοποιήσουμε αυτή την απαίτηση προσαρμόσαμε τα Δελτία

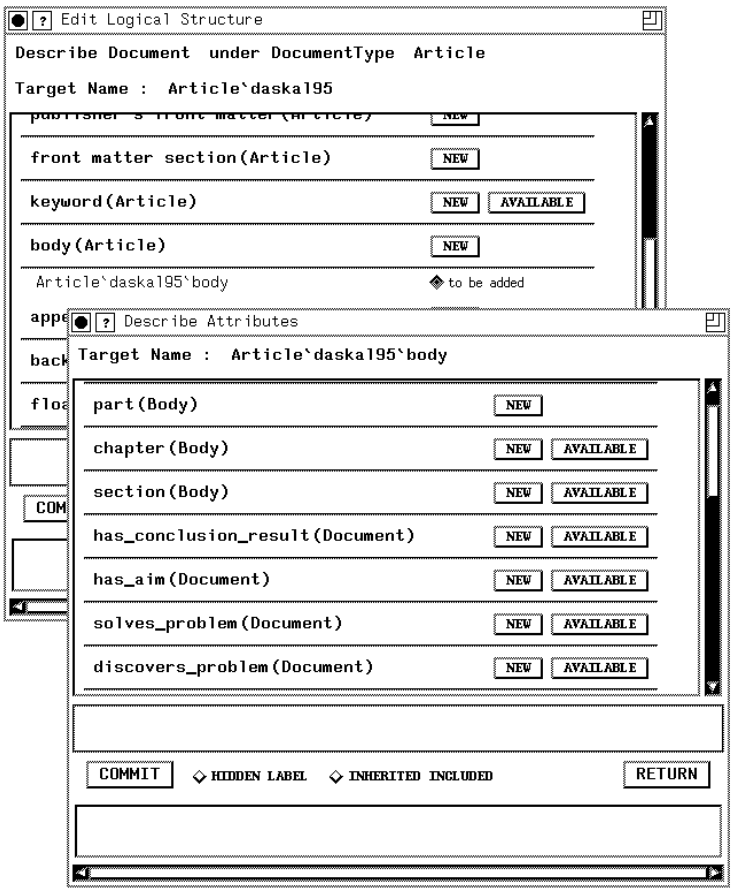
[8]Είναι η πρότυπη βιβλιογραφική μορφή που χρησιμοποιείται από την βιβλιοθήκη του Κογκρέσου των ΗΠΑ.



Σχήμα 4.10: Εισαγωγή της λογικής δομής εγγράφου

Ενημέρωσης (Entry Forms) [Δασ96a] στο συγκεκριμένο μοντέλο. Με τον τρόπο αυτό μπορούμε, για κάθε άρθρο ή λογικό τμήμα του, να χρησιμοποιούμε ένα δελτίο ενημέρωσης που περιέχει όλες τις κατηγορίες σύνδεσης με τα τμήματα από τα οποία αποτελείται. Για παράδειγμα στην εικόνα 4.10 υπάρχει το δελτίο ενημέρωσης του άρθρου με το λογικό όνομα Article‘daskal95, όπου φαίνεται ότι στο άρθρο μπορούμε να δώσουμε τίτλο, συγγραφείς, στοιχεία για την δημοσίευση του, την λογική δομή του (body), κ.α..

Όταν δίνουμε τιμή σε ένα γνώρισμα του, τότε εμφανίζεται ένα νέο δελτίο για να περιγράψουμε την τιμή που δώσαμε. Για παράδειγμα (βλ. εικόνα 4.11) όταν δίνουμε στο γνώρισμα *body* την τιμή *Article‘body* εμφανίζεται ένα νέο δελτίο για να περιγράψουμε τον κύριο κορμό του εγγράφου. Επίσης, μέσω αυτού του δελτίου μπορούμε να εισάγουμε και σημειώσεις σε αυτό το λογικό τμήμα. Όταν τελειώσουμε την ενημέρωση τότε με την εκτέλεση του *commit* του τελευταίου δελτίου γίνεται αυτόματα η ενημέρωση για όλες τις



Σχήμα 4.11: Το δελτίο ενημέρωσης άρθρου και το δελτίο ενημέρωσης του κυρίως κορμού (body) του συγκεκριμένου εγγράφου

αλλαγές που είχαμε κάνει μέχρι τότε.

Ακόμη με την βοήθεια των Δελτίων Ενημέρωσης (Entry Forms) μπορούμε να θέσουμε περιορισμούς ακεραιότητας στην εισαγωγή δεδομένων. Για παράδειγμα θέτουμε ότι το άρθρο θα πρέπει να έχει υποχρεωτικά ένα τίτλο και μπορεί να έχει μόνο ένα κυρίως κορμό (body). Αυτό επιτυγχάνεται με την κατάταξη των κατηγοριών σε προκαθορισμένες κλάσεις κατηγοριών, όπως περιγράφεται στο [Δασ96a].

4.5 Παρατηρήσεις

Το μεταμοντέλο παράστασης εγγράφων δημιουργήθηκε για να εκφράσει την γενική

δομή των τύπων από έγγραφα και ακολούθησε την ορολογία της SGML . Ενώ η παράσταση της λογικής δομής των άρθρων και βιβλίων έγινε σύμφωνα με το δημοφιλή πρότυπο για την δημοσίευση εγγράφων ISO12083, που είναι ένα σύνολο από SGML DTDs. Επίσης δημιουργήθηκαν οι μηχανισμοί για την παρουσίαση της λογικής δομής και της πληροφορίας των εγγράφων καθώς και οι μηχανισμοί για την εισαγωγή (με αυτόματο ή διαλογικό τρόπο) των δεδομένων σχετικά με τα έγγραφα. Βέβαια το μοντέλο είναι ανοιχτό και για τον ορισμό άλλων τύπων εγγράφων. Για παράδειγμα θα μπορούσαμε να μεταφράσουμε (με αυτόματο ή όχι τρόπο) ένα SGML DTD, που παριστάνει ένα άλλο τύπο εγγράφου, στο αντίστοιχο Telos μοντέλο και να δημιουργήσουμε ένα μεταφραστή για την αυτόματη εισαγωγή μιας περίπτωσης, αυτού του τύπου εγγράφου, στη βάση δεδομένων. Ενώ μπορούμε να χρησιμοποιήσουμε τους ίδιους μηχανισμούς για την παρουσίαση και αναζήτηση πληροφορίας σχετικά με το τύπο εγγράφου ή τις περιπτώσεις του.

Κεφάλαιο 5

Το μοντέλο παράστασης σημειώσεων σε έγγραφα

Όταν μελετούμε ένα έγγραφο κάνουμε ορισμένες σημειώσεις. Οι σημειώσεις αυτές διαχωρίζονται ανάλογα με το είδος των περιεχομένων τους και με την κατηγορία της σημείωσης, δηλαδή την οπτική γωνία παρατήρησης κάτω από την οποία γίνονται. Στην μεταπτυχιακή εργασία που παρουσιάζεται υλοποιήσαμε ένα μοντέλο παράστασης σημειώσεων που γίνονται σε έγγραφα και μελετήσαμε την λειτουργικότητά του. Το μοντέλο, που υλοποιήθηκε στην γλώσσα παράστασης γνώσης Telos, και τα ζητήματα που μας απασχόλησαν περιγράφονται σε αυτό το κεφάλαιο.

5.1 Το περιεχόμενο της σημείωσης

Ένα βασικό ερώτημα σχετικά με την μοντελοποίηση των σημειώσεων αφορά το περιεχόμενο της σημείωσης. Δηλαδή από τί αποτελείται η σημείωση. Η πιο συνήθης αντίληψη είναι ότι η σημείωση είναι ένα κείμενο που γράφει ο σχολιαστής για ένα τμήμα του εγγράφου που διαβάζει. Στην πραγματικότητα όμως υπάρχουν και άλλα είδη σημειώσεων, δηλαδή μπορούμε να ταξινομήσουμε το κείμενο που γράφει, σε ένα σύνολο (κλάση) σημειώσεων με κοινά γνωρίσματα. Η μελέτη της δομής (του περιεχομένου) των σημειώσεων μας οδήγησε στον διαχωρισμό τους στα παρακάτω είδη:

- Ένα έγγραφο που δημιουργείται για να εκφράσει την παρατήρηση. Αυτό είναι ένα έγγραφο με την γενικότερη έννοια, δηλαδή μπορεί να είναι κείμενο, εικόνα, ήχος, βίντεο ή ένας συνδυασμός των προηγουμένων.

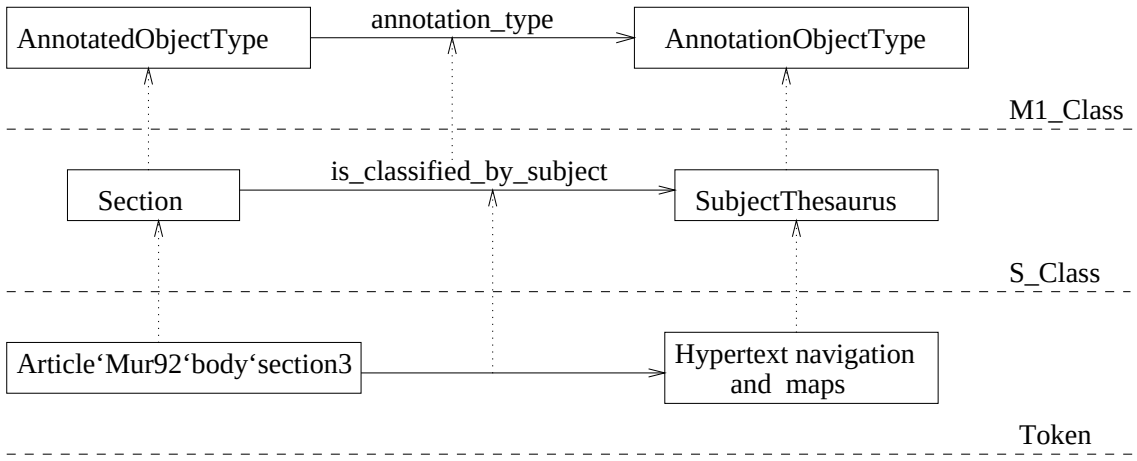
- Ένας όρος, που αποτελείται από μία ή περισσότερες λέξεις, ενός ελεγχόμενου λεξιλόγιου [Sve89] που εκφράζει τις έννοιες μιας θεματικής περιοχής. Επίσης μεταξύ των όρων υπάρχουν συσχετίσεις. Έτσι ένα σημείο του εγγράφου μπορεί να σημειώνεται με έναν όρο που έχει μια σημασία για αυτούς που αποδέχονται το συγκεκριμένο ελεγχόμενο λεξιλόγιο.

- Ένας σύνδεσμος, ο οποίος συνδέει το τμήμα του εγγράφου που θέλουμε να σημειώσουμε με ένα άλλο υπαρκτό έγγραφο το οποίο βρίσκεται σε μια βάση δεδομένων ή γενικότερα κάπου στο δίκτυο (π.χ. ένα έγγραφο που έχει το ίδιο θέμα, το αρχείο του πηγαίου κώδικα του σχετικού προγράμματος, μια συλλογή από ηλεκτρονικά μηνύματα που σχολιάζουν το έγγραφο, κ.α.). Αυτό το είδος ονομάζεται αναφορική σημείωση (reference annotation).

Δύο άλλα λιγότερο σημαντικά είδη, που περιγράφονται στη εργασία του Nielsen [Nie84], είναι η υπογράμμιση τμημάτων του εγγράφου και η σημείωση της δομής ενός τμήματος του εγγράφου, π.χ. μιας παραγράφου. Η υπογράμμιση έχει ως στόχο να ξεχωρίσει ορισμένα τμήματα του εγγράφου και μπορεί να γίνει είτε με διαφορετικές γραμμές είτε με διαφορετικά χρώματα. Η διαφορετική υπογράμμιση είτε προσδιορίζει διαφορετικό βαθμό σπουδαιότητας (π.χ. το πράσινο για πολύ σημαντικό και το κίτρινο για σημαντικό θέμα) είτε διαφορετικό είδος πληροφορίας (π.χ. το κόκκινο για τους ορισμούς και το μπλε για τα θεωρήματα). Ενώ η σημείωση της δομής ενός τμήματος του εγγράφου έχει ως στόχο είτε να ξεκαθαρίσει το νόημα του κειμένου είτε να συσχετίσει τα πιο ενδιαφέροντα σημεία. Ο προσδιορισμός της δομής γίνεται είτε με την αρίθμηση των δομικών σημείων είτε με την απόδοση λέξεων κλειδιών σε αυτά.

Όταν σημειώνουμε ένα τμήμα ενός εγγράφου με ένα όρο ενός ελεγχόμενου λεξιλόγιου (βλ. ενότητα 5.4) ή με ένα τμήμα ενός υπάρχον εγγράφου τότε η σημείωση ουσιαστικά ταυτίζεται με την σύνδεση (link) μεταξύ των δύο αντικειμένων. Διότι ο όρος ή το τμήμα του εγγράφου, που χρησιμοποιήσαμε για να εκφράσουμε την σημείωση, έχει μια υπόσταση ανεξάρτητα από την ύπαρξη ή όχι της σημείωσης. Ενώ όταν εκφράζουμε μια σημείωση δημιουργώντας ένα νέο (πολύμορφο ή μη) έγγραφο τότε αυτό το έγγραφο αποτελεί ουσιαστικά την σημείωση. Σε αυτήν την περίπτωση το έγγραφο που αποτελεί την σημείωση μπορεί να δεχτεί και αυτό σημειώσεις.

Όμως, εκτός από το περιεχόμενο της σημείωσης, σημαντικό ρόλο διαδραματίζει και η οπτική γωνία κάτω από την οποία έγινε η σημείωση. Στο μοντέλο που δημιουργήσαμε εκφράζουμε αυτήν την οπτική γωνία μέσω του ονόματος της κατηγορίας και της



Σχήμα 5.1: Το γενικό μοντέλο παράστασης σημειώσεων

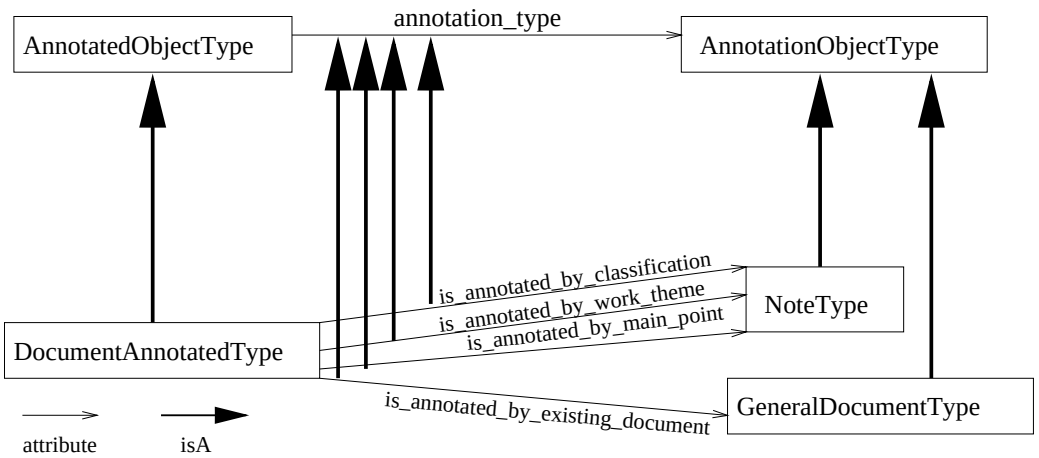
Το γενικό μοντέλο παράστασης σημειώσεων ορίζεται στο M1_Class επίπεδο ταξινόμησης της γλώσσας παράστασης γνώσης Telos. Στην εικόνα φαίνεται μια κατηγορία σημείωσης που ορίζεται στο S_Class επίπεδο και μια περίπτωση σημείωσης που ορίζεται στο Token επίπεδο.

μετακατηγορίας που συσχετίζει ένα έγγραφο με το περιεχόμενο της σημείωσης (βλ. 5.3).

5.2 Το γενικό μοντέλο παράστασης σημειώσεων

Το γενικό μοντέλο παράστασης σημειώσεων ορίζεται στο M1_Class επίπεδο ταξινόμησης της γλώσσας παράστασης γνώσης Telos και απεικονίζεται στο σχήμα 5.1. Η κλάση *AnnotatedObjectType* έχει ως περιπτώσεις (instances) όλους τους τύπους αντικειμένων στα οποία θέλουμε να κάνουμε σημειώσεις. Ενώ η κλάση *AnnotationObjectType* έχει ως περιπτώσεις όλους τους τύπους αντικειμένων που μπορεί να αποτελούν μια σημείωση (βλ. ενότητα 5.4). Τέλος ο σύνδεσμος *annotation_type* που συνδέει το *AnnotatedObjectType* με το *AnnotationObjectType* έχει ως περιπτώσεις (instances) όλες τις κατηγορίες συνδέσμων που εκφράζουν την οπτική γωνία κάτω από την οποία γίνεται η σημείωση. Ένα παράδειγμα απεικονίζεται στο σχήμα 5.1.

Σε αυτό το γενικό μοντέλο ορίσαμε ορισμένες εξειδικεύσεις, που φαίνονται στο σχήμα 5.2, ώστε να ομαδοποιήσουμε τις κατηγορίες σημειώσεων. Η κλάση *NoteType* έχει ως περιπτώσεις όλους τους τύπους ελεγχόμενων λεξιλόγιων και όλους τους τύπους (πολύμορφων ή μη) εγγράφων που δημιουργούνται για να εκφράσουν την σημείωση, δηλαδή είναι μια ομαδοποίηση του πρώτου και δεύτερου είδους σημειώσεων που παρου-



Σχήμα 5.2: Πρώτη προσέγγιση για τις εξειδικεύσεις της κατηγορίας *annotation_type*

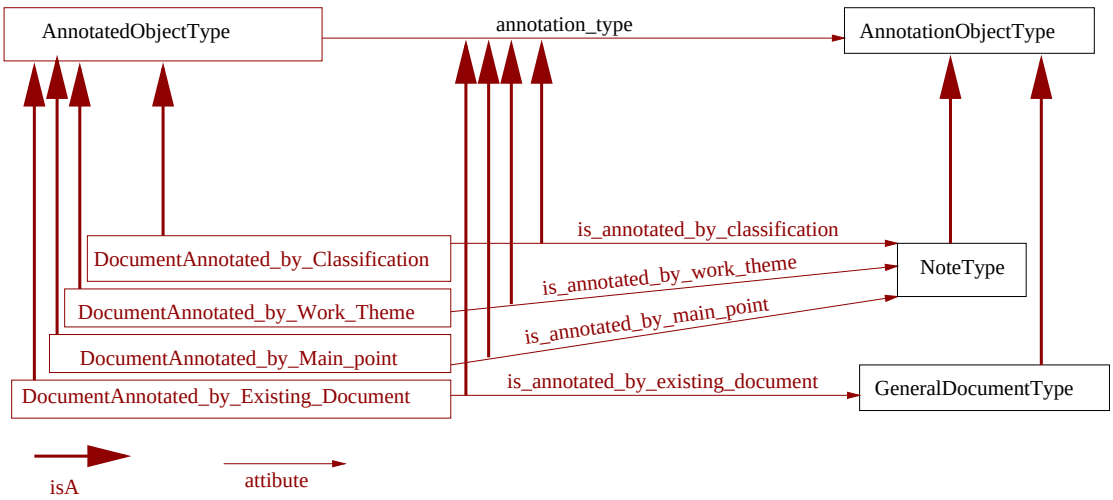
σιάζονται στην ενότητα 5.1. Ενώ το *GeneralDocumentType* έχει ως περιπτώσεις όλους τους τύπους εγγράφων και τμημάτων τους. Η σημασία των κατηγοριών του σχήματος 5.2 είναι η εξής:

- Η κατηγορία *is_annotated_by_classification* έχει ως περιπτώσεις όλες τις κατηγορίες σημειώσεων (δηλαδή οπτικές γωνίες παρατήρησης της σημείωσης) που εκφράζουν μια ταξινόμηση ενός εγγράφου ή τμήματος του σε έναν όρο ενός ελεγχόμενου λεξιλόγιου. Για παράδειγμα η θεματική ταξινόμηση ενός εγγράφου με ένα όρο του συστήματος ταξινόμησης του ACM[1] [Cou94] ή η κριτική ενός εγγράφου ως *Δεκτού* (*Accept*). Η αναλυτική περιγραφή των συγκεκριμένων κατηγοριών που επιλέξαμε γίνεται στην ενότητα 5.3.1.

- Η κατηγορία *is_annotated_by_work_theme* έχει ως περιπτώσεις όλες τις κατηγορίες σημειώσεων που εκφράζουν μια σχέση του εγγράφου με μια εργασία του σχολιαστή. Για παράδειγμα μπορούμε να σημειώσουμε αν ένα έγγραφο συμφωνεί ή διαφωνεί με μια εργασία ή αν περιγράφει παρόμοια εργασία (παρόμοια μεθοδολογία, παρόμοιες ιδέες ή αποτελέσματα).

- Η κατηγορία *is_annotated_by_main_point* έχει ως περιπτώσεις όλες τις κατηγορίες σημειώσεων που εκφράζουν τα κύρια σημεία ενός εγγράφου. Δηλαδή αν η σημείωση έχει σχέση με τα συμπεράσματα ή τα αποτελέσματα του εγγράφου, αν ανακαλύπτει

[1] Association for Computing Machinery

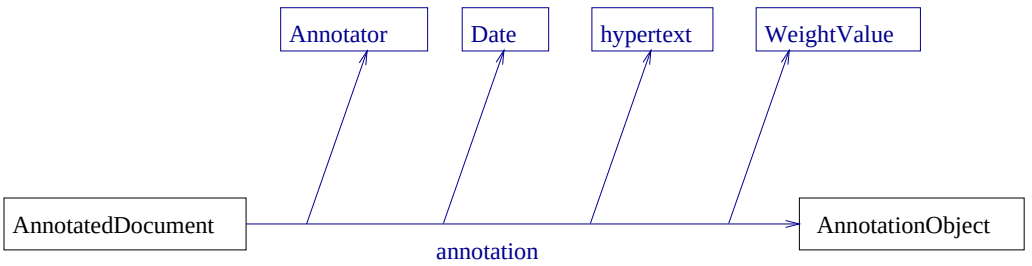


Σχήμα 5.3: Δεύτερη προσέγγιση για τις εξειδικεύσεις της κατηγορίας *annotation_type*

ή δίνει λύση σε κάποιο πρόβλημα, αν προσδιορίζει τους στόχους ή το νόημα του εγγράφου, κ.α. .

- Η κατηγορία *is_annotated_by_existing_document* έχει ως περιπτώσεις όλες τις κατηγορίες σημειώσεων που εκφράζουν μια σχέση του εγγράφου με ένα άλλο έγγραφο. Για παράδειγμα ότι ένα έγγραφο περιγράφει μια εξέλιξη εργασίας ή μια παρόμοια εργασία που περιγράφεται σε ένα άλλο έγγραφο.

Οι παραπάνω κατηγορίες χρησιμοποιούνται για την ομαδοποίηση σημειώσεων ώστε να γίνεται πιο εύκολη η εύρεση τους. Αλλά με αυτή την προσέγγιση δεν μπορούμε να κάνουμε αναζήτηση με βάση την κατηγορία της σημείωσης διότι στο εργαλείο παρουσίασης δεδομένων που χρησιμοποιήσαμε δεν μπορούμε να θέσουμε, ως σημείο αναφοράς της ερώτησης, μία κατηγορία (δηλαδή μία σύνδεση μεταξύ κλάσεων). Για τον λόγο αυτό ακολουθήσαμε μια άλλη προσέγγιση. Συγκεκριμένα ορίσαμε για κάθε μετακατηγορία σημειώσεων μια αντίστοιχη κλάση, όπως φαίνεται στο σχήμα 5.3, και ταξινομούμε την κλάση αυτή στην κλάση *AnnotationCategory* που ανήκει στο M2_Class επίπεδο ταξινόμησης της γλώσσας Telos. Με τον τρόπο αυτό μπορούμε να θέσουμε, ως σημείο αναφοράς (target) της ερώτησης, περίπτωση της *AnnotationCategory* (π.χ. DocumentAnnotated_by_Classification) και να αναφερόμαστε έμμεσα στην αντίστοιχη κατηγορία (για το προηγούμενο παράδειγμα είναι η is_annotated_by_classification).



Σχήμα 5.4: Η κατηγορία *annotation*

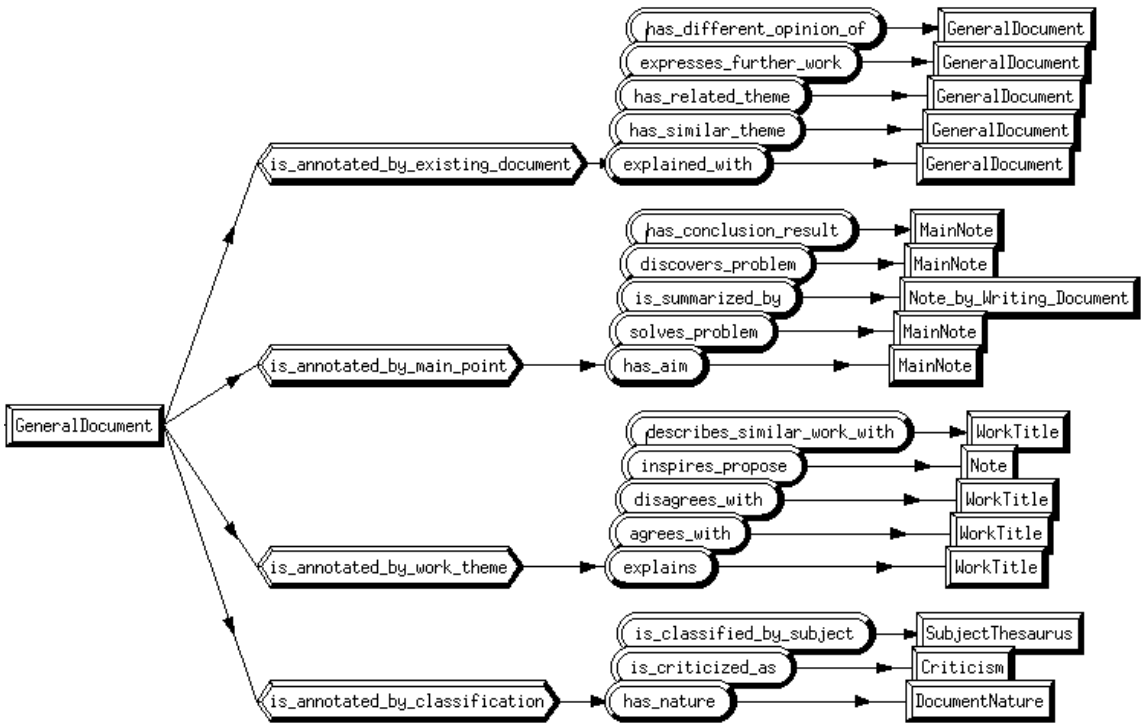
Η κατηγορία annotation έχει ως γνωρίσματα την ημερομηνία σημειώσεως, το όνομα αυτού που έκανε την σημείωση, μια τιμή που δηλώνει την αξία της σημείωσης και ένα υπερκείμενο (hypertext) σχόλιο που επεξηγεί την σημείωση.

5.3 Οι κατηγορίες σημειώσεων

Για λόγους ομαδοποίησης και για ευκολία στις ερωτήσεις αναζήτησης σημειώσεων, ορίσαμε στο S_Class επίπεδο της γλώσσας Telos το μοντέλο της σημείωσης που φαίνεται στο σχήμα 5.4. Η κλάση *AnnotatedDocument* έχει ως εξειδικεύσεις όλους τους τύπους εγγράφων και όλους τους τύπους τμημάτων των εγγράφων. Ενώ η κλάση *AnnotationObject* έχει ως εξειδικεύσεις όλους τους τύπους αντικειμένων που μπορούν να αποτελούν την σημείωση. Επομένως οι κλάσεις *AnnotatedDocument* και *AnnotationObject* είναι οι συζυγείς κλάσεις[2] των μετακλάσεων *AnnotatedObjectType* και *AnnotationObjectType* αντίστοιχα. Ενώ η κατηγορία *annotation* έχει ως εξειδικεύσεις όλες τις κατηγορίες σημειώσεων που περιγράφονται στις παρακάτω υποενότητες. Επομένως κάθε κατηγορία σημειώσεων έχει ορισμένα γνωρίσματα που κληρονομεί από την κατηγορία *annotation*. Δηλαδή σε κάθε σύνδεσμο σημείωσης μπορούμε να αποδώσουμε την ημερομηνία σημειώσεως, το όνομα αυτού που έκανε την σημείωση, μια τιμή που δηλώνει την αξία της σημείωσης και ένα υπερκείμενο (hypertext) σχόλιο που επεξηγεί την σημείωση.

Οι οπτικές γωνίες δημιουργίας σημειώσεων παριστάνονται, στο μοντέλο που υλοποιήσαμε, με τα ονόματα των κατηγοριών των συνδέσμων που συσχετίζουν το τμήμα του εγγράφου, που θέλουμε να σημειώσουμε, με το περιεχόμενο της σημείωσης. Στις παρακάτω υποενότητες παρουσιάζονται διεξοδικά οι κατηγορίες σημειώσεων που επιλέξαμε μετά από την μελέτη της βιβλιογραφίας και των συζητήσεων (συνεντεύξεων) με άτομα που κάνουν σημειώσεις στα έγγραφα που διαβάζουν. Στην εικόνα 5.5 φαίνονται οι κατηγορίες σημειώσεων καθώς και οι μετακατηγορίες στις οποίες ανήκουν. Εκτός από αυτές τις

[2]Οι συζυγείς κλάσεις περιγράφονται στην ενότητα 4.2 και πιο αναλυτικά στο [Χρι94b]

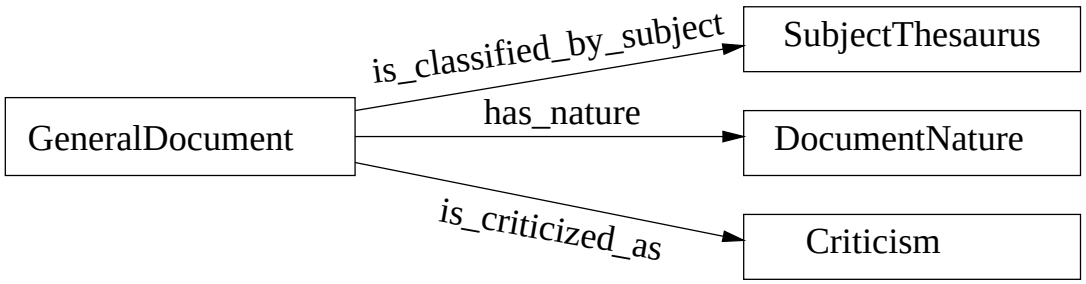


Σχήμα 5.5: Οι κατηγορίες και μετακατηγορίες σημειώσεων

κατηγορίες υπάρχει και μια άλλη κατηγορία σημείωσης, που ονομάζεται *comment*, η οποία δεν ταξινομείται σε κάποια μετακατηγορία και χρησιμοποιείται όταν θέλουμε να σχολιάσουμε ένα έγγραφο γράφοντας απλώς ένα κείμενο.

5.3.1 Κατηγορίες σημειώσεων που εκφράζουν ταξινόμηση

Οι κατηγορίες αυτές είναι περιπτώσεις της μετακατηγορίας *is_annotated_by_classification* (βλ. ενότητα 5.2) και χρησιμοποιούνται όταν θέλουμε μέσω της σημείωσης να ταξινομήσουμε ένα έγγραφο ή τμήμα του σε ένα όρο ενός ελεγχόμενου λεξιλόγιου. Η συνήθης σημείωση αυτού του είδους είναι να ταξινομήσουμε θεματικά ένα έγγραφο ή τμήμα του με ένα όρο ενός θησαυρού που εκφράζει το θέμα με το οποίο ασχολείται το έγγραφο. Ένας άλλος τρόπος ταξινόμησης είναι ανάλογα με την φύση του εγγράφου, δηλαδή να διαχωρίσουμε τα τεχνικά από τα θεωρητικά έγγραφα ή να προσδιορίσουμε ποιά έγγραφα είναι εμπορικά (προωθούν προϊόντα) και ποιά είναι ανασκόπηση της ερευνητικής εργασίας σε ένα τομέα. Επίσης η κριτική που κάνουμε σε ένα έγγραφο ή τμήμα του είναι ένα είδος ταξινόμησης, δηλαδή διαχωρίζουμε αυτά που είναι αποδεκτά από αυτά που απορρίπτουμε



Σχήμα 5.6: Οι κατηγορίες σημείωσης που δηλώνουν ταξινόμηση

ή προσδιορίζουμε ποιά επιδέχονται διορθώσεις ή περισσότερη ανάπτυξη των περιεχομένων τους.

Οι κατηγορίες σημειώσεων φαίνονται στο σχήμα 5.6 και είναι οι εξής:

- Η κατηγορία *classified_by_subject* δηλώνειτην ταξινόμηση σύμφωνα με το θέμα του εγγράφου. Το περιεχόμενο της σημείωσης είναι τύπου *SubjectThesaurus* και έχει ως περιπτώσεις (instances) τους όρους ενός θεματικού θησαυρού. Οι όροι του θησαυρού συσχετίζονται μεταξύ τους. Συγκεκριμένα, στην υλοποίηση της δομής θησαυρού που κάναμε, υπάρχουν οι σχέσεις ευρύτερου όρου, στενότερου όρου, συνωνύμων όρων και σχετικών όρων (δηλαδή όροι που δεν έχουν καμιά από τις τρεις πρώτες σχέσεις αλλά εκφράζουν θέματα σχετικά μεταξύ τους). Η πλήρης περιγραφή του θησαυρού όρων και των επιλογών που κάναμε παρουσιάζονται στην ενότητα 5.4.1. Ένα παράδειγμα περίπτωσης αυτής της κατηγορίας παρουσιάστηκε ήδη στην εικόνα 5.1.

- Η κατηγορία *has_nature* εκφράζει την ταξινόμηση σύμφωνα με την φύση του εγγράφου. Το περιεχόμενο της σημείωσης είναι τύπου *DocumentNature* και έχει ως περιπτώσεις φράσεις που δηλώνουν την φύση του εγγράφου. Στο συγκεκριμένο μοντέλο που υλοποιήσαμε δίνεται η δυνατότητα να χαρακτηρίσουμε ένα έγγραφο ως θεωρητικό, αναλυτικό, σχεδιαστικό, περιγραφικό, εμπορικό, σύνθετο και ως ανασκόπηση. Βέβαια δεν περιοριζόμαστε σε αυτούς μόνο τους χαρακτηρισμούς της φύσης του εγγράφου, αλλά μπορούν να οριστούν και νέοι ανάλογα με τις ανάγκες των χρηστών του συστήματος σημειώσεων.

- Η κατηγορία *is_criticized_as* εκφράζει την ταξινόμηση σύμφωνα με την κριτική που κάνουμε σε ένα έγγραφο. Το περιεχόμενο της σημείωσης είναι τύπου *Criticism* και

έχει ως περιπτώσεις φράσεις που εκφράζουν μια κριτική για το έγγραφο(π.χ. Accept, Reject). Οι φράσεις αυτές σχετίζονται μεταξύ τους, π.χ. αν για ένα έγγραφο είναι αποδεκτό τότε δεν το έχουμε απορρίψει.

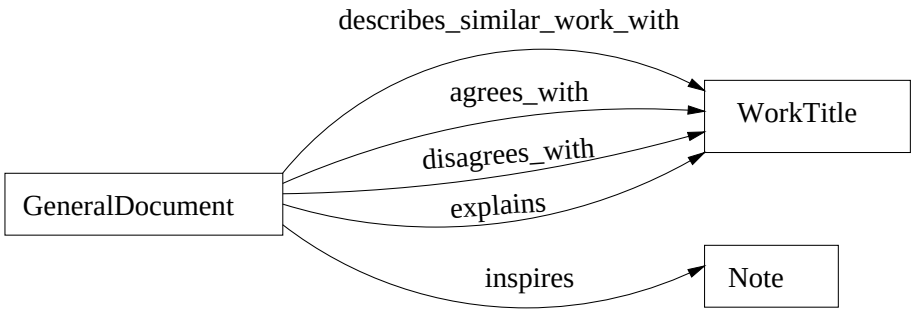
Η ταξινόμηση ενός εγγράφου, με βάση τις παραπάνω κατηγορίες που περιγράψαμε, αποτελεί ένα πολύ ισχυρό μηχανισμό για την οργάνωση των εγγράφων και των σημειώσεων σε αυτά. Δηλαδή μπορούμε να αναζητήσουμε εύκολα τα έγγραφα που έχουμε ταξινομήσει σε ένα θεματικό όρο, που έχουν μια συγκεκριμένη φύση (π.χ. θεωρητική) και για τα οποία έχουμε κάνει κριτική με μια συγκεκριμένη φράση.Επίσης μπορούμε να επιλέξουμε, από τα έγγραφα που δεν έχουμε διαβάσει, σύμφωνα με την ταξινόμηση που έκαναν άλλα άτομα[3], αφού μπορούμε σε κάθε σημείωση να προσδιορίζουμε αυτόν που την δημιούργησε.

5.3.2 Σημειώσεις σχετικά με τις εργασίες

Όταν διαβάζουμε ένα έγγραφο (π.χ. ένα άρθρο) ίσως να κάνουμε σημειώσεις σχετικά με τις εργασίες που ασχολούμαστε ή τις ιδέες για εργασίες που έχουμε. Συνήθως μας ενδιαφέρει να σημειώσουμε τα τμήματα του εγγράφου που λειτουργούν ως επιχειρήματα για την συμφωνία ή διαφωνία με την μεθοδολογία ή τα αποτελέσματα μιας εργασίας. Επίσης σημαντικό είναι να σημειώνουμε τα έγγραφα που περιγράφουν μια παρόμοια εργασία με την δική μας ή επεξηγούν την εργασία που κάνουμε. Ακόμη σημαντικό είναι να σημειώνουμε τα έγγραφα που μας έδωσαν έμπνευση για μια νέα εργασία (πρόταση για νέα εργασία).

Στο μοντέλο σημειώσεων που υλοποιήσαμε έχουμε ορίσει ορισμένες κατηγορίες που εκφράζουν τις οπτικές γωνίες συσχέτισης ενός εγγράφου με τις εργασίες ενός ατόμου (βλ. εικόνα 5.7). Η κλάση *WorkTitle*, που βλέπουμε στην εικόνα 5.7, είναι ο τύπος ενός ελεγχόμενου λεξιλόγιου και έχει ως περιπτώσεις όλες τις εργασίες των μελών μιας ομάδας. Κάθε συγκεκριμένη εργασία εκφράζεται με ένα τίτλο εργασίας και αντίστροφα ένας τίτλος εκφράζει μία μόνο εργασία. Μια εργασία μπορεί να διασπάται σε υποεργασίες και να ασχολείται με μία ή περισσότερες θεματικές περιοχές, δηλαδή ταξινομείται σε ένα ή περισσότερους όρους του θεματικού θησαυρού. Η αναλυτική περιγραφή της κλάσης *WorkTitle* και της λειτουργικότητάς της γίνεται στην υποενότητα 5.4.2. Ενώ η κλάση *Note*, που περιγράφεται στην ενότητα 5.4, είναι η γενίκευση όλων των τύπων ελεγχόμενων

[3]Συνήθως πρόκειται για τα μέλη μιας ομάδας για τα οποία γνωρίζουμε την κριτική ικανότητα τους και τους θεματικούς τομείς ενασχόλησης τους



Σχήμα 5.7: Οι κατηγορίες σημείωσης που συσχετίζουν ένα έγγραφο με μια εργασία

λεξιλόγιων και όλων των τύπων εγγράφων που δημιουργούνται για να εκφράσουν την σημείωση.

Οι κατηγορίες που φαίνονται στο σχήμα 5.7 είναι οι εξής:

- Η κατηγορία *agrees_with* δηλώνει ότι το τμήμα του εγγράφου που σημειώνουμε περιέχει επιχειρήματα ή παραδείγματα που ενισχύουν την ιδέα ή την μεθοδολογία ή τα αποτελέσματα της εργασίας με την οποία εκφράζουμε το περιεχόμενο της σημείωσης. Για παράδειγμα μπορούμε να σημειώσουμε ότι η ενότητα *Article'Xristofidi95'body'section2* συμφωνεί με την μεθοδολογία της εργασίας *''Development of a compiler from SGML DTD to Telos model''*.

- Η κατηγορία *disagrees_with* δηλώνει ότι το τμήμα του εγγράφου που σημειώνουμε περιέχει επιχειρήματα ή παραδείγματα που διαφωνούν με την ιδέα ή μεθοδολογία ή τα αποτελέσματα της εργασίας με την οποία εκφράζουμε το περιεχόμενο της σημείωσης.

- Η κατηγορία *explains* δηλώνει ότι το τμήμα του εγγράφου που σημειώνουμε επεξηγεί την μεθοδολογία ή τα αποτελέσματα της εργασίας με την οποία εκφράζουμε το περιεχόμενο της σημείωσης. Για παράδειγμα μπορούμε να σημειώσουμε ότι η ενότητα *Article'saran95'body'section3* επεξηγεί τα αποτελέσματα της εργασίας *''Collaborative filtering by annotations''*

- Η κατηγορία *describes_similar_work* δηλώνει ότι το τμήμα του εγγράφου που σημειώνουμε περιγράφει παρόμοια εργασία με αυτήν που εκφράζει ο τίτλος εργασίας με την οποία εκφράζουμε το περιεχόμενο της σημείωσης. Δηλαδή για μια εργασία μπορούμε να σημειώσουμε τα έγγραφα στα οποία παρουσιάζεται ή γίνεται μελέτη

μιας παρόμοιας εργασίας, ώστε να έχουμε συγκεντρωμένες όλες τις παρόμοιες εργασίες που έχουμε μελετήσει στην βιβλιογραφία.

- Η κατηγορία *inspires* (εμπνέει) δηλώνει ότι το τμήμα του εγγράφου που σημειώνουμε έχει βοηθήσει στην έμπνευση ή αποκρυστάλλωση μιας πρότασης για νέα εργασία. Το περιεχόμενο της σημείωσης είναι τύπου *Note*, δηλαδή μπορεί να είναι ένας τίτλος εργασίας (που εκφράζει την πρόταση εργασίας) είτε ένας όρος του θεματικού θησαυρού (που εκφράζει την θεματική περιοχή της πρότασης) είτε ένα έγγραφο (που δημιουργήθηκε για να εκφράσει την πρόταση για εργασία). Στις δύο πρώτες περιπτώσεις το περιεχόμενο της σημείωσης δεν περιγράφει πλήρως την πρόταση εργασίας, αλλά αυτό μπορεί να γίνει με το σχόλιο που έχουμε την δυνατότητα να αποδώσουμε ως γνώρισμα στο σύνδεσμο της σημείωσης (βλ. στην αρχή της ενότητας 5.3).

Οι σημειώσεις που αναφέρονται στις εργασίες των μελών μιας ομάδας είναι πολύ χρήσιμες για την συσχέτιση των εργασιών αυτών με την βιβλιογραφία που έχουν μελετήσει. Με την βοήθεια των σημειώσεων μπορούμε να αναζητήσουμε εύκολα τα έγγραφα που ασχολούνται με μία εργασία (είτε διαφωνούν είτε συμφωνούν είτε την επεξηγούν) καθώς και αυτά που ασχολούνται με παρόμοιες εργασίες. Επίσης μπορούμε να έχουμε συγκεντρωμένες όλες τις προτάσεις για νέες εργασίες και πώς αυτές σχετίζονται με τα έγγραφα που έχουμε μελετήσει. Ακόμη οι σημειώσεις σχετικά με τις εργασίες μας βοηθούν στη συγγραφή άρθρων, αφού έχουμε ήδη καλά οργανωμένη την συσχέτιση των εργασιών με την βιβλιογραφία.

5.3.3 Σημειώσεις των κύριων σημείων του εγγράφου

Οι κατηγορίες σημειώσεων που παρουσιάζονται σε αυτήν την ενότητα είναι περιπτώσεις της μετακατηγορίας *is_annotated_by_main_point* (βλ. ενότητα 5.2) και χρησιμοποιούνται όταν θέλουμε να σημειώσουμε τα κύρια και σημαντικά σημεία των εγγράφων. Συνήθως μας ενδιαφέρει να σημειώσουμε τους στόχους, τα συμπεράσματα και τα αποτελέσματα που έχει ένα έγγραφο. Επίσης ενδιαφέρον είναι να σημειώσουμε ποιά προβλήματα ανακαλύπτει και σε ποιά δίνει λύσεις. Ακόμη πολλές φορές ο αναγνώστης θέλει να γράψει μια δική του περίληψη για το έγγραφο που διάβασε, ώστε να συγκεντρώσει και να τονίσει τα σημεία του εγγράφου που τον ενδιαφέρουν.

Οι κατηγορίες σημειώσεων σχετικά με τα κύρια σημείων του εγγράφου που ορίσαμε

στο μοντέλο μας απεικονίζονται στο σχήμα 5.7. Η κλάση *MainNote* είναι η γενίκευση των κλάσεων *SubjectThesaurus*, *WorkTitle* και *Note_by_Writing_Document* (δηλαδή οι περιπτώσεις της μπορεί να είναι ένας όρος του θεματικού θησαυρού ή ένας τίτλος εργασίας ή ένα έγγραφο που δημιουργήσαμε για να εκφράσουμε την σημείωση).

Οι κατηγορίες σημειώσεων, για τα κύρια σημεία του εγγράφου, είναι οι εξής:

- Η κατηγορία *has_aim* δηλώνει ότι το τμήμα του εγγράφου που σημειώθηκε παρουσιάζει τους στόχους του εγγράφου. Οι στόχοι αυτοί περιγράφονται με ένα θεματικό όρο του θησαυρού, με ένα τίτλο εργασίας, με μια φράση ή ένα έγγραφο που γράφει ο σχολιαστής.

- Η κατηγορία *has_conclusion_result* δηλώνει ότι το τμήμα του εγγράφου που σημειώθηκε παρουσιάζει τα συμπεράσματα ή τα αποτελέσματα στα οποία καταλήγει το έγγραφο. Τα συμπεράσματα ή τα αποτελέσματα αυτά περιγράφονται με ένα θεματικό όρο του θησαυρού, με ένα τίτλο εργασίας, με μια φράση ή ένα έγγραφο που γράφει ο σχολιαστής. Επιλέξαμε να χρησιμοποιήσουμε μια μόνο κατηγορία για να δηλώνει τα συμπεράσματα και τα αποτελέσματα του εγγράφου γιατί τις πιο πολλές φορές τα συμπεράσματα και τα αποτελέσματα ταυτίζονται.

- Η κατηγορία *discovers_problem* προσδιορίζει τα προβλήματα με τα οποία ασχολείται το έγγραφο και τα οποία παρουσιάζονται στο τμήμα του εγγράφου που έγινε η σημείωση. Το πρόβλημα περιγράφεται με μια φράση ή έγγραφο που γράφει ο σχολιαστής, με ένα θεματικό όρο του θησαυρού, με ένα τίτλο εργασίας.

- Η κατηγορία *solves_problem* προσδιορίζει τα προβλήματα για τα οποία προτείνει ή υλοποιεί τρόπους επίλυσης τους και τα οποία παρουσιάζονται στο τμήμα του εγγράφου που έγινε η σημείωση. Το πρόβλημα περιγράφεται με μια φράση ή έγγραφο που γράφει ο σχολιαστής ή με ένα θεματικό όρο του θησαυρού ή με ένα τίτλο εργασίας.

- Η κατηγορία *is_summarized_by* χρησιμοποιείται όταν θέλουμε να γράψουμε μια περίληψη για το έγγραφο ή για ένα τμήμα του. Η περίληψη αυτή μπορεί να περιγράφει με λίγα λόγια το έγγραφο και να επικεντρώνεται στα σημεία που είναι ενδιαφέροντα για αυτόν που κάνει την σημείωση. Βέβαια για κάθε έγγραφο υπάρχει συνήθως μια περίληψη από τον συγγραφέα του, αλλά όταν ένας αναγνώστης γράφει την δική του περίληψη μπορεί να τονίσει τα σημεία που είναι πιο ενδιαφέροντα για αυτόν.

Οι σημειώσεις που αναφέρονται στα κύρια σημεία των εγγράφων είναι πολύ χρήσιμες για την κατανόηση του εγγράφου καθώς και για να ξαναθυμηθούμε γρήγορα το περιεχόμενό του. Δηλαδή μπορούμε να αναζητήσουμε τους στόχους, τα συμπεράσματα και τα αποτελέσματα ενός άρθρου καθώς και τα προβλήματα με τα οποία ασχολείται ή επιλύει. Επίσης μπορούμε να δούμε πόσο αξιόλογα είναι ορισμένα έγγραφα ή τμήματα τους που μας ενδιαφέρουν, αλλά και να διαβάσουμε τυχόν περιλήψεις που έχουμε γράψει για κάποια τμήματα του άρθρου. Ακόμη μπορούμε να αντιληφθούμε πόσο σημαντικό είναι ένα άρθρο από τις σημειώσεις που έχουν κάνει σε αυτό κάποια άλλα άτομα που το έχουν μελετήσει.

5.3.4 Σημειώσεις που συσχετίζουν έγγραφα μεταξύ τους

Οι κατηγορίες σημειώσεων που συσχετίζουν έγγραφα μεταξύ τους είναι περιπτώσεις της μετακατηγορίας *is_annotated_by_existing_document* (βλ. ενότητα 5.2) και χρησιμοποιούνται για να εκφράσουμε τις σχέσεις μεταξύ των εγγράφων που έχουμε μελετήσει. Συνήθως μας ενδιαφέρει να σημειώσουμε ποιά έγγραφα ασχολούνται με όμοιο ή σχετικό θέμα καθώς και ποιά έγγραφα περιγράφουν εξέλιξη μιας εργασίας που περιγράφεται σε ένα άλλο έγγραφο. Επίσης πολύ συχνά θέλουμε να συσχετίσουμε μία φράση ή τμήμα ενός εγγράφου με μία επεξήγηση του που γίνεται μέσω ενός άλλου εγγράφου.

Οι κατηγορίες σημειώσεων που συσχετίζουν έγγραφα απεικονίζονται στο σχήμα 5.7 και είναι οι εξής:

- Η κατηγορία *has_same_theme* δηλώνει ότι το έγγραφο ή το τμήμα του εγγράφου που σημειώθηκε διαπραγματεύεται όμοιο θέμα με το έγγραφο που αποτελεί το περιεχόμενο της σημείωσης. Σε αυτή τη κατηγορία σημείωσης δεν παίζει ρόλο η κατεύθυνση της σύνδεσης, δηλαδή αν το *Article'Clio95* έχει όμοιο θέμα με το *Article'Cultur94* τότε η και το *Article'Cultur94* έχει όμοιο θέμα με το *Article'Clio95*. Βέβαια η έννοια της ομοιότητας είναι πολύ σύνθετη και υποκειμενική, δηλαδή ένας μελετητής μπορεί να θεωρήσει ότι δύο έγγραφα είναι όμοια ενώ ένας άλλος να μην το αντιληφθεί ή να πιστεύει ότι δεν είναι όμοια.

- Η κατηγορία *has_related_theme* δηλώνει ότι το έγγραφο ή το τμήμα του εγγράφου που σημειώθηκε διαπραγματεύεται ένα σχετικό θέμα με αυτό του εγγράφου που αποτελεί το περιεχόμενο της σημείωσης. Και σε αυτήν την περίπτωση η έννοια *σχετικό θέμα είναι υποκειμενική.*

- Η κατηγορία *expresses_further_work* δηλώνει ότι το έγγραφο ή το τμήμα του εγγράφου που σημειώθηκε παρουσιάζει μία εξέλιξη εργασίας που παρουσιάστηκε στο έγγραφο που αποτελεί το περιεχόμενο της σημείωσης. Για παράδειγμα το άρθρο *Article'sis96* εκφράζει εξέλιξη εργασίας που παρουσιάστηκε στο άρθρο *Article'sis94*.

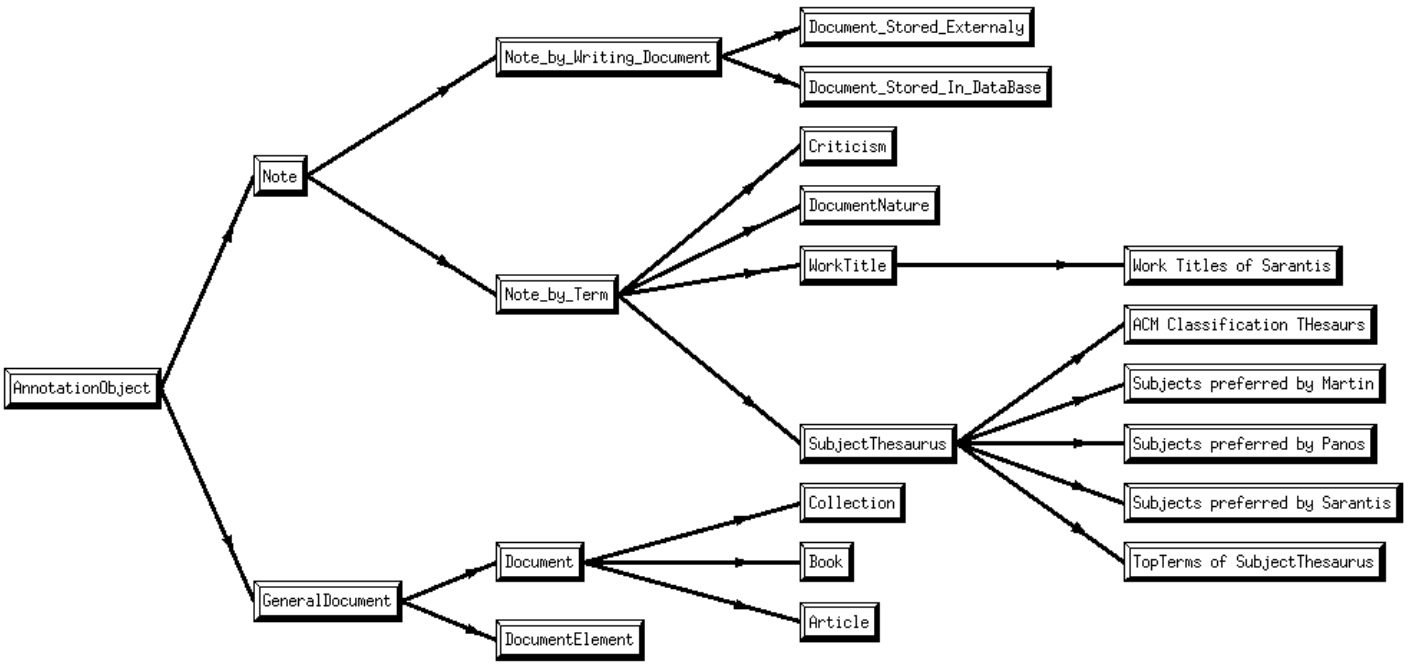
- Η κατηγορία *has_different_opinion_of* δηλώνει ότι το έγγραφο ή το τμήμα του εγγράφου που σημειώθηκε περιγράφει διαφορετικές απόψεις από αυτές που παρουσιάζονται στο έγγραφο που αποτελεί το περιεχόμενο της σημείωσης.

- Η κατηγορία *explained_with* δηλώνει ότι το έγγραφο ή το τμήμα του εγγράφου που σημειώθηκε επεξηγείται από το έγγραφο που αποτελεί το περιεχόμενο της σημείωσης. Αυτή είναι η κατηγορία σημείωσης που συναντάμε στα hypertext συστήματα, δηλαδή μία φράση ή τμήμα εγγράφου μας παραπέμπει για την ερμηνεία του (ή για περισσότερες πληροφορίες) σε ένα άλλο έγγραφο.

Οι σημειώσεις που συσχετίζουν έγγραφα μεταξύ τους βοηθούν τους μελετητές να διαπιστώσουν άμεσα τις συσχετίσεις των εγγράφων που έχουν μελετήσει. Βέβαια τις σημειώσεις που ανήκουν στις κατηγορίες *has_same_theme* και *has_related_theme* ίσως να μπορούσαμε να τις είχαμε διαπιστώσει έμμεσα (π.χ. αν δύο έγγραφα ταξινομούνται στον ίδιο ή σε σχετικούς θεματικούς όρους ή εργασίες), αλλά μερικές φορές θέλουμε να υπάρχει άμεσα αυτή η σύνδεση.

5.4 Οι τύποι των περιεχομένων της σημείωσης

Η σημείωση μπορεί να είναι είτε ένα έγγραφο (πολύμορφο ή μη) είτε ένας όρος ενός ελεγχόμενου λεξιλόγιου. Συγκεκριμένα οι τύποι των περιεχομένων της σημείωσης που υποστηρίζουμε καθώς και οι ιεραρχίες μεταξύ τους απεικονίζονται στην εικόνα 5.8. Η κλάση *AnnotationObject* έχει ως περιπτώσεις όλες τις σημειώσεις και έχει ως εξειδικεύσεις τις κλάσεις *GeneralDocument* και *Note*. Δηλαδή η σημείωση είναι είτε ένα υπάρχον έγγραφο ή τμήμα του είτε ένα σύντομο σχόλιο (*Note*). Το σύντομο σχόλιο *Note* μπορεί να είναι είτε ένα έγγραφο, που αποθηκεύεται στην βάση δεδομένων ή εξωτερικά από αυτή, είτε ένας όρος ενός ελεγχόμενου λεξιλόγιου. Οι τύποι ελεγχόμενων λεξιλόγιων που ορίζονται στο μοντέλο περιγράφονται στις παρακάτω ενότητες.



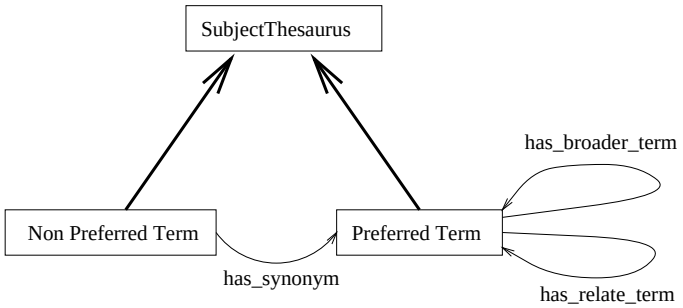
Σχήμα 5.8: Οι τύποι των περιεχομένων της σημείωσης.

Στην εικόνα φαίνονται οι τύποι των περιεχομένων της σημείωσης και οι μεταξύ τους συσχετίσεις. Το βέλος δηλώνει την σχέση εξειδίκευσης.

5.4.1 Ο θεματικός θησαυρός όρων (SubjectThesaurus)

Ο Θεματικός θησαυρός όρων είναι ένα σύνολο λέξεων ή φράσεων που έχουν επιλεγεί για να εκφράζουν τα θέματα ενός τομέα ενασχόλησης (π.χ. επιστήμη υπολογιστών) και συσχετίζονται μεταξύ τους. Οι βασικές σχέσεις μεταξύ των όρων ενός θησαυρού φαίνονται στο σχήμα 5.9.

Οι όροι ενός θησαυρού χωρίζονται σε δόκιμους (preferred) και μη δόκιμους (non preferred) και κάθε μη δόκιμος έχει ένα συνώνυμο (synonym) δόκιμο όρο, π.χ. το *AI* είναι συνώνυμο του *Artificial Intelligence*. Ενώ κάθε δόκιμος όρος έχει ευρύτερους, στενότερους και σχετικούς όρους [ISO86]. Για παράδειγμα στην εικόνα 5.10 βλέπουμε τις συσχετίσεις του όρου *Database Management* με τους άλλους όρους του θησαυρού. Ο όρος *Information Systems* εκφράζει μια γενικότερη έννοια που συμπεριλαμβάνει και αυτό που εκφράζει ο όρος *Database Management*, ενώ οι όροι *Logical Design*, *Physical Design*, *Database Languages* εκφράζουν ειδικότερα θέματα με τα οποία ασχολείται η *Διαχείριση βάσεων δεδομένων*



Σχήμα 5.9: Η δομή του θεματικού θησαυρού

Οι όροι του θησαυρού χωρίζονται σε δόκιμους και μη δόκιμους. Οι μή−δόκιμοι όροι έχουν ένα συνώνυμο δόκιμο όρο. Ενώ οι δόκιμοι όροι μπορουν να συνδέονται με ευρήτερους όρους και σχετικούς όρους.

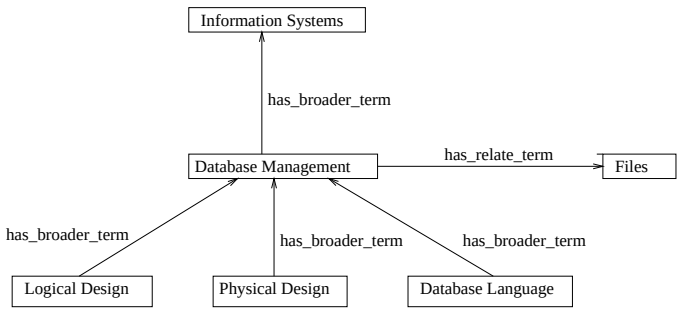
(Database Management). Επίσης τα *Αρχεία (Files)* είναι πολύ σχετικά με τη *Διαχείριση βάσεων δεδομένων (Database Management)* αλλά δεν υπάρχει μεταξύ τους μια ξεκάθαρη σχέση ευρύτερου ή στενότερου θέματος.

Στο μοντέλο που δημιουργήσαμε υπάρχει η δυνατότητα να ορίσουμε συλλογές από όρους. Δηλαδή μπορούμε να ορίσουμε μία κλάση, στο S_Class επίπεδο της γλώσσας Telos, με το όνομα *"Όροι που προτιμούνται από τον Σαράντη"* και έπειτα να ταξινομήσουμε τους όρους που θέλουμε σε αυτήν την κλάση. Ο μηχανισμός αυτός δημιουργήθηκε από την ανάγκη να βρίσκουμε εύκολα όσους όρους χρησιμοποιούμε πολύ συχνά (π.χ. για να εκφράσουμε με αυτούς μια σημείωση).

Ο θεματικός θησαυρός όρων είναι πολύ χρήσιμος για την ταξινόμηση των εγγράφων στην θεματική περιοχή με την οποία ασχολούνται. Διότι με την βοήθεια των συσχετίσεων των όρων του θησαυρού μπορούμε να βρούμε έναν ή περισσότερους όρους που εκφράζουν το περιεχόμενο ενός εγγράφου. Επίσης με την βοήθεια του θησαυρού όρων μπορούμε να αυξήσουμε την ακρίβεια (precision) ή την ανάκληση (recall) δεδομένων μιας αναζήτησης. Για παράδειγμα αν θέλαμε να βρούμε τα έγγραφα που ασχολούνται με *Database Management* μπορούμε να βρούμε τα έγγραφα που έχουν ταξινομηθεί σε αυτόν τον όρο αλλά και αυτά που έχουν ταξινομηθεί στους στενότερους όρους του, στους στενότερους όρους των στενότερων όρων του, κ.ο.κ. .

Στην συγκεκριμένη εφαρμογή που υλοποιήσαμε, έχουμε εισάγει στην βάση δεδομένων το σύστημα ταξινόμησης του ACM[4] [Cou94]. Η επιλογή αυτή έγινε γιατί ο θησαυρός αυτός είναι αρκετά καλός για την επιστήμη υπολογιστών και επίσης τα περισσότερα άρθρα

[4]Association for Computing Machinery

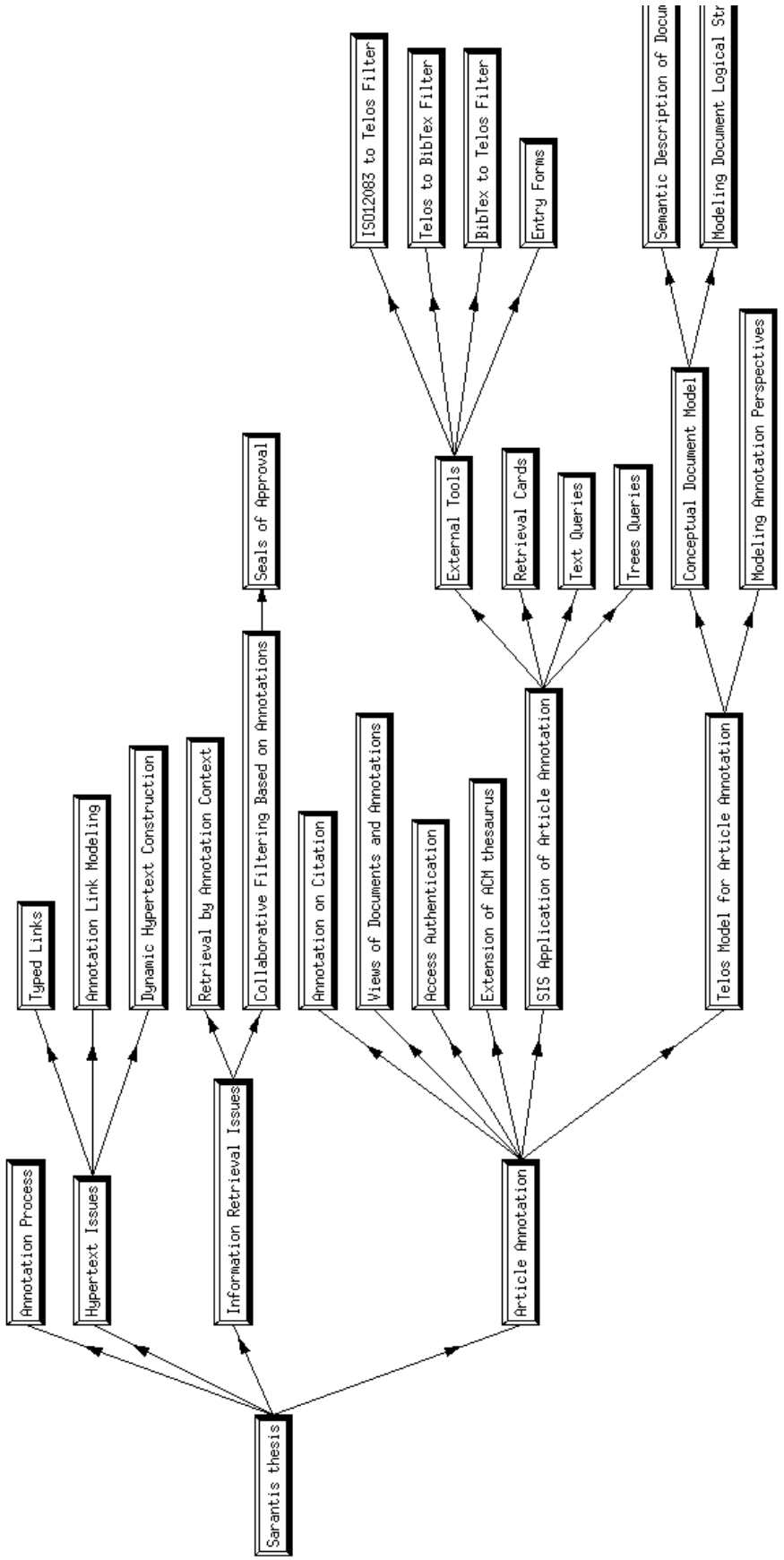


Σχήμα 5.10: Όροι θεματικού θησαυρού και οι συσχετίσεις τους

είναι ήδη ταξινομημένα σύμφωνα με αυτόν τον θησαυρό όρων. Βέβαια μπορούμε να αυξήσουμε τον θησαυρό για να εκφράσουμε ορισμένα θέματα που δεν τα καλύπτει και μας ενδιαφέρουν.

5.4.2 Οι τίτλοι εργασίας (WorkTitle)

Μια συγκεκριμένη εργασία εκφράζεται με ένα *τίτλο εργασίας* και αντίστροφα ένας *τίτλος εργασίας* εκφράζει μια μόνο συγκεκριμένη εργασία. Μια εργασία μπορεί να διασπάται σε δύο ή περισσότερες συγκεκριμένες υποεργασίες, οι οποίες μπορούν και αυτές με την σειρά τους να διασπώνται σε υποεργασίες, κ.ο.κ. . Επομένως δημιουργείται ένα δέντρο εργασιών (ένα παράδειγμα φαίνεται στο σχήμα 5.11). Ορισμένοι τίτλοι εργασιών έχουν νόημα μόνο στο συγκεκριμένο περιβάλλον (context). Για παράδειγμα στο σχήμα 5.11 φαίνεται ο τίτλος *''SIS Application of Article Annotation''* και έχει ως υποεργασία την εργασία *''Text Queries''*, αλλά σε κάθε εφαρμογή του SIS θα υπάρχουν Text Queries και για αυτό το όνομα τέτοιων εργασιών εμπεριέχει και το όνομα της γενικότερης εργασίας που ανήκει (δηλαδή στην περίπτωση μας ο τίτλος της εργασίας *''Text Queries''* είναι *''SIS Application of Article Annotation'Text Queries''*). Επίσης ορισμένοι τίτλοι εργασιών έχουν απλά οργανωτικό χαρακτήρα, όπως π.χ. ο τίτλος εργασίας *''Information Retrieval Issues''* της εικόνας 5.11. Ενώ μια άλλη δυνατότητα που προσφέρεται είναι η συσχέτιση, μέσω της κατηγορίας *related_theme*, ενός τίτλου εργασίας με την θεματική περιοχή που ανήκει. Δηλαδή συσχετίζουμε μια περίπτωση του *WorkTitle* με μια περίπτωση του *SubjectThesaurus* (π.χ. οι εργασίες *''Retrieval by Annotation Context''* και *''Collaborative Filtering based on Annotations''* ασχολούνται με το θέμα του *''Information Retrieval''* το οποίο αποτελεί όρο του θησαυρού). Έτσι μπορούμε να βρούμε τις εργασίες που είναι σχετικές με μια θεματική περιοχή.



Σχήμα 5.11: Ένα παράδειγμα τίτλων εργασίων

Στο σχήμα φαίνονται οι τίτλοι εργασιών που είναι υποεργασίες της εργασίας *Sarantis thesis* καθώς και πώς συσχετίζονται μεταξύ τους.

Όμως ένα άτομο είναι δυνατόν να ενδιαφέρεται για ορισμένες εργασίες και ίσως όχι για όλες τις υποεργασίες τους. Άρα παρουσιάζεται η ανάγκη να ορίσουμε συλλογές από εργασίες ώστε να μπορεί κάποιος να ομαδοποιεί τις εργασίες που τον ενδιαφέρουν. Αλλά καθώς η κατάσταση των εργασιών αλλάζει με την πάροδο του χρόνου (δηλαδή άλλες εργασίες ολοκληρώνονται και νέες δημιουργούνται) παρουσιάζεται και η ανάγκη για δημιουργία συλλογών σε σχέση με τον χρόνο. Για παράδειγμα θα μπορούσαμε να χωρίσουμε τις εργασίες σε *''Ενεργές''* και *''Ολοκληρωμένες''*. Επίσης τις εργασίες που έχουν τελειώσει (δηλαδή τις *''Ολοκληρωμένες''*) θα μπορούσαμε να τις κατατάξουμε σε συλλογές ανάλογα με την χρονική περίοδο (π.χ. George's WorkTitles of 1995) ή ανάλογα με την γενικότερη εργασία που ανήκουν (π.χ. George's WorkTitles of AQUARELL project).

Συνήθως οι τίτλοι εργασιών δημιουργούνται μια φορά και έπειτα, με την πάροδο του χρόνου, προσθέτουμε νέους και τους συσχετίζουμε με αυτούς που υπάρχουν. Επομένως αν έχουμε καλά ορισμένες τις εργασίες και έχουμε οργανώσει προσεκτικά τις μεταξύ τους συσχετίσεις τότε μπορούμε να κάνουμε γρήγορα μία εύστοχη σημείωση. Αφού απλά θα βάλουμε μια σύνδεση από το έγγραφο στην εργασία. Μερικές φορές όμως, κατά την διάρκεια της δημιουργίας σημειώσεων, εμφανίζεται η ανάγκη να δημιουργήσουμε νέους τίτλους εργασιών ή να διασπάσουμε τους υπάρχοντες σε υποεργασίες με αποτέλεσμα να καθυστερεί η δημιουργία σημειώσεων.

5.4.3 Η φύση του εγγράφου (DocumentNature)

Η κλάση *DocumentNature* έχει ως περιπτώσεις τις φράσεις που εκφράζουν την φύση του εγγράφου Στο συγκεκριμένο μοντέλο που υλοποιήσαμε δίνεται η δυνατότητα να χαρακτηρίσουμε ένα έγγραφο ως θεωρητικό, αναλυτικό, σχεδιαστικό, περιγραφικό, εμπορικό, σύνθετο και ως ανασκόπηση. Οι φράσεις (ή λέξεις) αυτού του λεξιλόγιου μπορούν να σχετίζονται μεταξύ τους, δηλαδή αν ένα έγγραφο το χαρακτηρίσουμε θεωρητικό τότε θα πρέπει εν μέρη να είναι και αναλυτικό. Βέβαια δεν περιοριζόμαστε σε αυτούς μόνο τους χαρακτηρισμούς της φύσης του εγγράφου, αλλά μπορεί να ορίσουμε και νέους ανάλογα με τις ανάγκες των χρηστών του συστήματος σημειώσεων.

5.4.4 Η κριτική του εγγράφου (Criticism)

Η κλάση *Criticism* έχει ως περιπτώσεις τις φράσεις που εκφράζουν μια κριτική για το έγγραφο. Στο συγκεκριμένο μοντέλο που υλοποιήσαμε έχουμε ορίσει της εξής φράσεις: *Accept* (Αποδοχή), *Reject* (Απόρριψη), *Partial_Accept* (μερική αποδοχή), *Very_Interesting* (Πολύ ενδιαφέρον), *Interesting* (Ενδιαφέρον), *Uninteresting* (Δεν έχει ενδιαφέρον), *Good* (Καλό),

Incomprehensible (Ακατανόητο), *Need_Evolution* (Απαιτείται ανάπτυξη), *Need_Correction* (Απαιτείται διόρθωση). Οι παραπάνω φράσεις σχετίζονται μεταξύ τους, π.χ. αν για ένα έγγραφο απαιτείται περισσότερη ανάπτυξη τότε είναι και μερικώς αποδεκτό.

5.5 Η ενημέρωση σημειώσεων

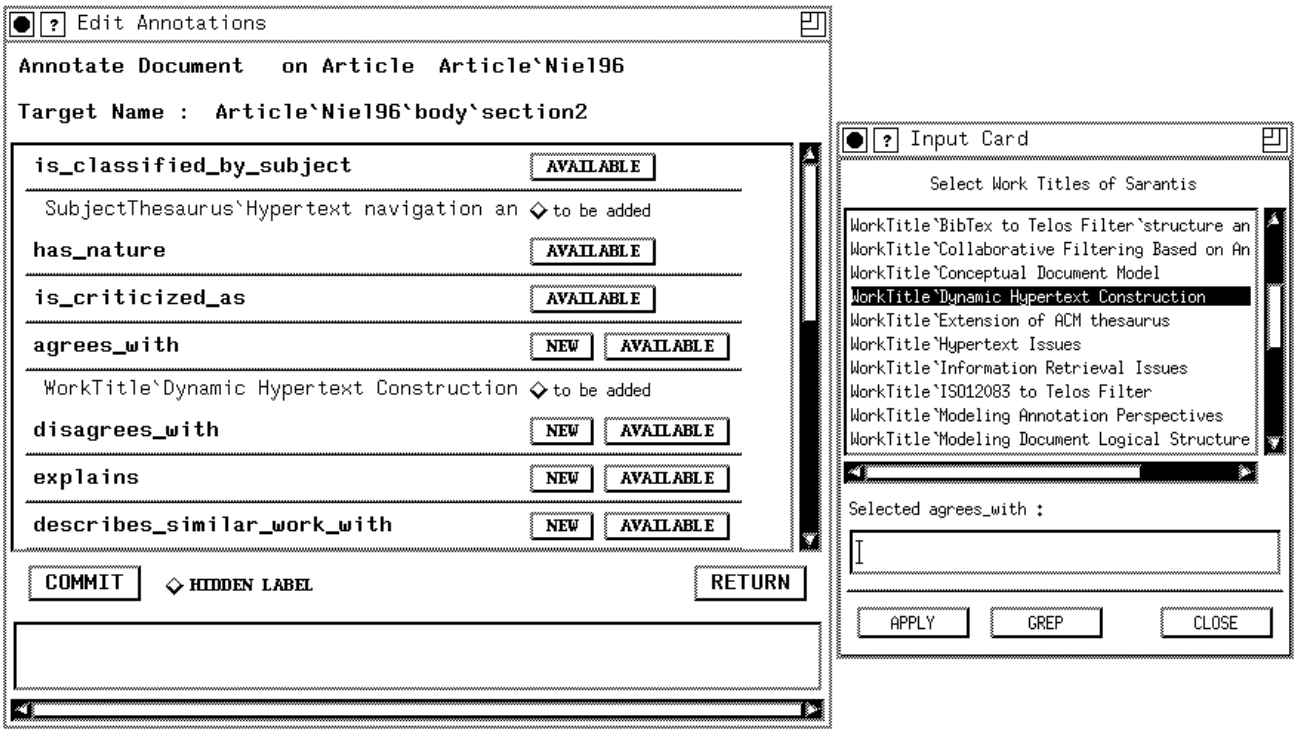
Η ενημέρωση σημειώσεων γίνεται μέσω των Δελτίων Ενημέρωσης (Entry Forms) [Δασ96a] που προσαρμόστηκαν στο μοντέλο σημειώσεων. Η προσαρμογή αυτή έγινε μέσω ενός μοντέλου που ορίζει: τις διαδικασίες ενημέρωσης, προσδιορίζει για κάθε διαδικασία τα αντικείμενα που μπορεί να ενημερώσει και ορίζει τις λειτουργίες (π.χ. εισαγωγή, διαγραφή, αλλαγή γνωρισμάτων, ταξινόμηση, κ.α.) που θέλουμε να εκτελούνται πάνω σε αυτά τα αντικείμενα.

Υλοποιήσαμε δύο διαδικασίες ενημέρωσης σημειώσεων που διαφέρουν στον τρόπο με τον οποίο βρίσκουμε τα τμήματα του εγγράφου που θέλουμε να σημειώσουμε. Επίσης υλοποιήσαμε τις διαδικασίες για την ενημέρωση των όρων του θεματικού θησαυρού (SubjectThesaurus), των τίτλων εργασιών (WorkTitle), των συλλογών από όρους ή τίτλους εργασίας καθώς και για την ενημέρωση των άλλων ελεγχόμενων λεξιλόγιων (Document-Nature, Criticism).

Οι διαδικασίες ενημέρωσης των σημειώσεων είναι οι εξής:

- Επιλέγουμε το έγγραφο που θέλουμε να σημειώσουμε και τότε παρουσιάζεται μια λίστα με όλα τα λογικά του τμήματα. Έπειτα μπορούμε να επιλέξουμε ένα λογικό τμήμα και να ενημερώσουμε τις σημειώσεις σε αυτό. Ένα παράδειγμα είναι το δελτίο ενημέρωσης των σημειώσεων της δεύτερης ενότητας του άρθρου *Article'Niel96* που φαίνεται στην εικόνα 5.12. Αυτή η διαδικασία ενημέρωσης προτιμάται όταν υπάρχει ήδη η λογική δομή του εγγράφου στην βάση δεδομένων και γνωρίζουμε σε ποιό τμήμα του εγγράφου θέλουμε να κάνουμε την σημείωση.

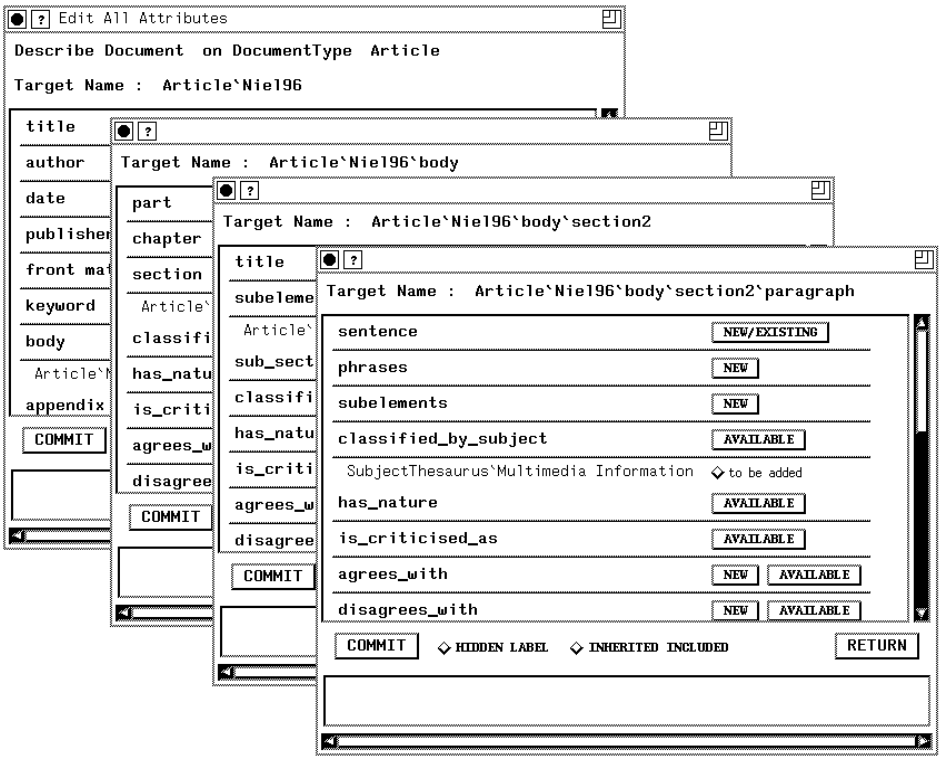
- Επιλέγουμε τον τύπο του εγγράφου που θέλουμε να σημειώσουμε (π.χ. άρθρο ή βιβλίο ή συλλογή εγγράφων) και τότε παρουσιάζεται μια λίστα με όλα τα έγγραφα του τύπου που επιλέξαμε. Έπειτα μπορούμε είτε να δημιουργήσουμε ένα νέο έγγραφο είτε να διαγράψουμε ή να μετονομάσουμε ένα υπάρχον είτε να ενημερώσουμε τα γνωρίσματα του. Η ενημέρωση των γνωρισμάτων του επικεντρώνεται στην ενημέρωση της λογικής δομής του και των σημειώσεων στο έγγραφο. Έπειτα από το δελτίο ενημέρωσης του εγγράφου μπορούμε να ανοίξουμε τα δελτία ενημέρωσης των λογικών τμημάτων του, έπειτα τα δελτία από τα λογικά τμήματα των λογικών



Σχήμα 5.12: Ενημέρωση σημειώσεων σε ένα τμήμα του εγγράφου

τμημάτων του εγγράφου, κ.ο.κ. . Με αυτό τον τρόπο μπορούμε, για να κάθε υπάρχον ή νέο λογικό τμήμα του εγγράφου, να ανοίξουμε το δελτίο ενημέρωσης του. Ένα παράδειγμα φαίνεται στην εικόνα 5.13, όπου έχουμε επιλέξει το άρθρο *Article'Niel96* και ενημερώνουμε τις σημειώσεις στη πρώτη παράγραφο της τρίτης ενότητας του άρθρου. Αυτή η διαδικασία ενημέρωσης προτιμάται όταν θέλουμε να εισάγουμε την λογική δομή του εγγράφου και να κάνουμε ταυτόχρονα και σημειώσεις στα διάφορα τμήματά του.

Σε κάθε δελτίο ενημέρωσης των λογικών τμημάτων του εγγράφου υπάρχουν οι κατηγορίες σημειώσεων που μπορούμε να κάνουμε σε αυτά. Επίσης σε κάθε στάδιο ενημέρωσης παρέχεται βοήθεια για τον προσδιορισμό του τύπου του περιεχομένου της σημείωσης καθώς και μια λίστα με υπάρχουσες σημειώσεις που μπορούμε να χρησιμοποιήσουμε. Για παράδειγμα όταν θέλουμε να σημειώσουμε ένα έγγραφο σε σχέση με μια εργασία τότε μπορούμε να επιλέξουμε από μια λίστα εργασιών (οι εργασίες αυτές μπορεί είτε να είναι όλες οι εργασίες που υπάρχουν στη βάση είτε οι εργασίες μια συλλογής, π.χ. ''WorkTitle of AQUARELL project'').



Σχήμα 5.13: Ενημέρωση της λογικής δομή και των σημειώσεων σε ένα τμήμα του εγγράφου

5.6 Παρατηρήσεις

Το ενδιαφέρον του μοντέλου σημειώσεων επικεντρώνεται στη προσπάθεια να εκφράσουμε, μέσω των κατηγοριών των σημειώσεων, τις οπτικές γωνίες δημιουργίας σημειώσεων. Μάλιστα τα ονόματα των κατηγοριών είναι κατάλληλα επιλεγμένα ώστε να σχηματίζεται μια πρόταση, π.χ. όταν μια παράγραφος παρουσιάζει επιχειρήματα για την αξία μιας εργασίας (π.χ. ‘‘Document Annotation by SIS’’) τότε σχηματίζεται η εξής πρόταση:

Article‘saran95‘section2‘paragraph3 agrees_with Document Annotation by SIS.

Επίσης με τον ορισμό των τύπων των περιεχομένων της σημείωσης προσπαθούμε να προσδιορίσουμε τις σημειώσεις που θέλουμε να κάνουμε καθώς και να τις συσχετίσουμε μεταξύ τους (π.χ. όταν ανήκουν σε ένα ελεγχόμενο λεξιλόγιο). Επομένως μπορούμε να κάνουμε εκφραστικές σημειώσεις, αφού προσδιορίζουμε την οπτική γωνία και το περιεχόμενο της σημείωσης, και μάλιστα με γρήγορο τρόπο, αφού συνήθως απλά δημιουργούμε μια σύνδεση (π.χ. συνδέουμε ένα τμήμα του εγγράφου με ένα όρο ενός

ελεγχόμενου λεξιλόγιου).

Βέβαια ίσως κάποιοι χρήστες του συστήματος να μην καλύπτονται πλήρως από τις κατηγορίες σημειώσεων που ορίσαμε. Για τον λόγο αυτό δημιουργήσαμε ένα πρόγραμμα εισαγωγής νέων κατηγοριών σημειώσεων ενώ παράλληλα δημιουργούμε αυτόματα, στο μέσο επαφής (interface), τις αντίστοιχες ερωτήσεις για την αναζήτηση σημειώσεων της νέας κατηγορίας.

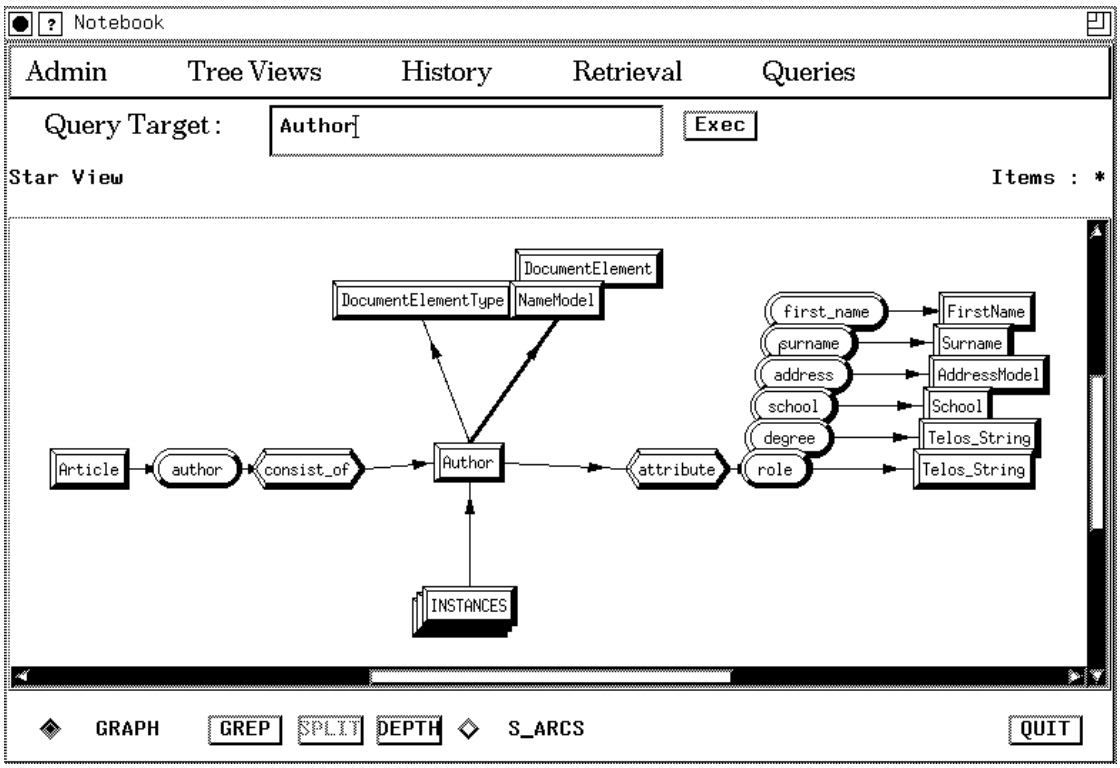
Κεφάλαιο 6

Η Χρήση του συστήματος

Στο κεφάλαιο αυτό περιγράφεται η χρήση προκαθορισμένων ερωτήσεων για την άντληση της πληροφορίας που υπάρχει στην βάση των έγγραφων και των σημειώσεων σε αυτά. Οι προκαθορισμένες ερωτήσεις γίνονται μέσω του Σημασιολογικού Συστήματος Ευρετηριασμού (ΣΣΕ), το οποίο περιγράφεται στην ενότητα 6.1. Υπάρχουν δύο κατηγορίες ερωτήσεων , οι ερωτήσεις 1ης τάξεως και οι αναδρομικές ερωτήσεις. Η σημασία και η χρήση τους περιγράφονται στις αντίστοιχες ενότητες του κεφαλαίου αυτού.

Η κατασκευή προκαθορισμένων ερωτήσεων σκοπό έχει να διευκολύνει το χρήστη παρέχοντάς του τη δυνατότητα να έχει με μορφή καταλόγου τις ερωτήσεις που γίνονται συχνότερα στο σύστημα και έχουν τη μεγαλύτερη σημασία. Επιπλέον μπορεί να αντλήσει την πληροφορία χωρίς να είναι υποχρεωμένος να ξέρει μια ερωτηματική γλώσσα. Γι' αυτό το λόγο πρέπει οι ερωτήσεις που παρέχονται από το σύστημα να είναι αρκετά γενικές έτσι ώστε να μπορούν να εκφράσουν την αναζήτηση της πληροφορίας που θέλουμε. Από την άλλη πλευρά όμως πρέπει να είναι και αρκετά εξειδικευμένες, έτσι ώστε να μην έχουν ως αποτέλεσμα μεγάλα σύνολα απαντήσεων που κουράζουν το χρήστη ή κάνουν την πληροφορία μη χρησιμοποιήσιμη.

Οι προκαθορισμένες ερωτήσεις γίνονται σε σχέση με τις βασικές έννοιες του μοντέλου που περιγράφηκε στα κεφάλαια 4 και 5. Τέτοιες είναι τα έγγραφα, οι σημειώσεις, τα πρόσωπα, κ.λ.π., καθώς επίσης και οι βασικές σχέσεις μεταξύ των εννοιών αυτών, όπως οι κατηγορίες σημειώσεων, η σχέση μέρους/όλου (π.χ. το έγγραφο αποτελείται από τμήματα), κ.λ.π. . Ο διαχωρισμός των ερωτήσεων σε ''θέματα'' και η χρήση παραμέτρων, όπως θα περιγραφεί παρακάτω, δίνει σχετικά μικρά και ευανάγνωστα σύνολα απαντήσεων.



Σχήμα 6.1: Η επαφή χρήσεως του ΣΣΕ.

Στην εικόνα βλέπουμε τις βασικές λειτουργίες που παρέχονται από την επαφή χρήσεως του ΣΣΕ, καθώς και τον αστεροειδή γράφο για μια οντότητα του μοντέλου. Η τονισμένη γραμμή παριστάνει τη σχέση γενίκευσης, η μη τονισμένη γραμμή που έχει φορά προς το πάνω μέρος παριστάνει τη σχέση ταξινόμησης ενώ οι υπόλοιπες τη σχέση απόδοσης γνωρίσματος.

6.1 Σύντομη περιγραφή της επαφής χρήσης του ΣΣΕ και του μοντέλου των προκαθορισμένων ερωτήσεων

Στο σχήμα 6.1 φαίνεται το αρχικό παράθυρο της επαφής χρήσεως. Υπάρχει ένα πεδίο όπου τίθεται ο στόχος της ερώτησης (Query Target) αν αυτός υπάρχει. Μπορεί να είναι μια οποιαδήποτε οντότητα του μοντέλου.

Οι επιλογές που προσφέρονται είναι οι εξής :

- *Admin* χρησιμοποιείται για την επικοινωνία με άλλα εργαλεία (επεξεργαστές κειμένου, εργαλεία ενημέρωσης δεδομένων, κλπ).

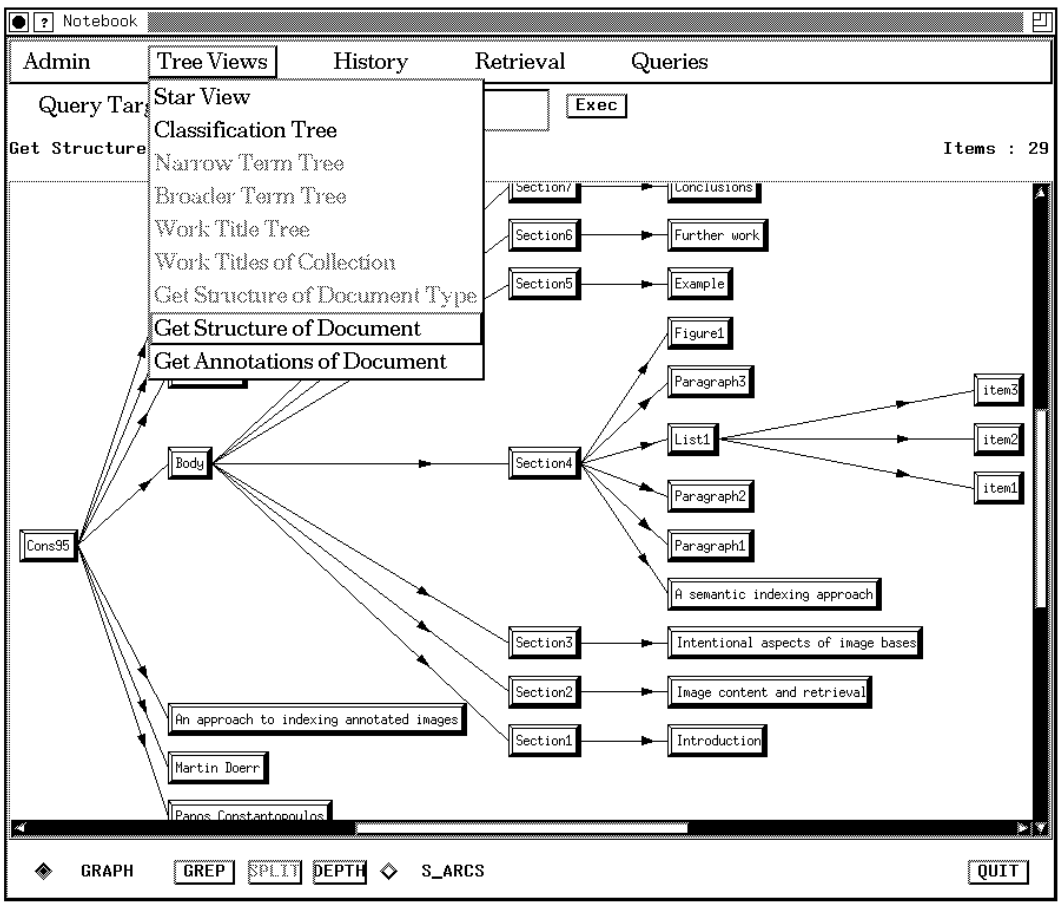
- *Views* χρησιμοποιείται για την εκτέλεση προκαθορισμένων αναδρομικών ερωτήσεων, των οποίων τα αποτελέσματα δίνονται σε μορφή γράφου. Οι ερωτήσεις αυτές

καθορίζονται από το δημιουργό του μοντέλου ή το χρήστη αλλά υπάρχουν δύο οι οποίες δίνονται έτοιμες. Η μια είναι η παράσταση του αστεροειδούς γράφου (star view). Ένας τέτοιος γράφος φαίνεται στο σχήμα 6.1. Ο στόχος της ερώτησης βρίσκεται στο κέντρο του γράφου ενώ ταυτόχρονα φαίνεται η σχέση του με τις άλλες οντότητες του μοντέλου. Η άλλη είναι η παράσταση της ιεραρχίας γενίκευσης , ένα παράδειγμα φαίνεται στην εικόνα 5.8. Στο σχήμα 6.2 φαίνεται ο κατάλογος επιλογών της *Views* για το μοντέλο που υλοποιήσαμε καθώς και το αποτέλεσμα (σε μορφή γράφου) μιας ερώτησης.

- *History* χρησιμοποιείται για την επανεκτέλεση ερωτήσεων που έχουν γίνει στο παρελθόν.
- *Retrieval* χρησιμοποιείται για ερωτήσεις με παραπάνω από μια παραμέτρους.
- *Queries* χρησιμοποιείται για την εκτέλεση προκαθορισμένων ερωτήσεων, των οποίων τα αποτελέσματα δίνονται σε μορφή κειμένου. Στο σχήμα 6.3 φαίνεται το αποτέλεσμα της εκτέλεσης μιας τέτοιας ερώτησης, ο κατάλογος επιλογών της *Queries* με τις προκαθορισμένες ερωτήσεις για για το μοντέλο που υλοποιήσαμε .
- *Help* παρέχει πληροφορία στο χρήστη για τις λειτουργίες του συστήματος.

Η παράσταση των προκαθορισμένων ερωτήσεων σε TELOS αποτελείται από μια σειρά κλάσεων που περιγράφονται περιληπτικά παρακάτω. Η κλάση **Model** είναι μετά-μετακλάση και έχει το γνώρισμα *uiMenus* που παίρνει τιμές στην κλάση **MenuDescription**. Η **MenuDescription** έχει τα γνωρίσματα *queryMenu, retrievalMenu, viewMenu* που παίρνουν τιμή στην κλάση **SubMenu**. Το πρώτο γνώρισμα αντιστοιχεί στις ερωτήσεις της επιλογής *Queries*, το δεύτερο της *Retrieval* και το τρίτο της *Views*. Η κλάση **SubMenu** έχει το γνώρισμα *commands* που παίρνει τιμές στην κλάση **QueryMacro**. Το γνώρισμα αυτό αντιστοιχεί στις επιλογές των καταλόγων που φαίνονται στα σχήματα 6.2, 6.3. Η **QueryMacro** έχει τα γνωρίσματα *code*, *inputType*, *iterator*, και *outputHeader*. Τα τρία πρώτα παίρνουν τιμές στην **String**, οι οποίες είναι συντετμημένα ονόματα συναρτήσεων της ΒΣΕΣ και τις οποίες εκτελεί ο διερμηνευτής για να αντλήσει πληροφορία από τη βάση. Το δεύτερο γνώρισμα έχει σχέση με το πώς παρουσιάζονται τα αποτελέσματα των ερωτήσεων σε μορφή κειμένου, αφού υπάρχει η δυνατότητα εκτός από τα ονόματα των οντοτήτων να παρουσιάζονται και ορισμένα από τα γνωρίσματά τους.

Περίπτωση της **Model** είναι η κλάση **NOTEBOOK** που εκφράζει την παρούσα εφαρμογή. Το γνώρισμα *uiMenus* παίρνει την τιμή **NOTEBOOKMenus** που είναι περίπτωση



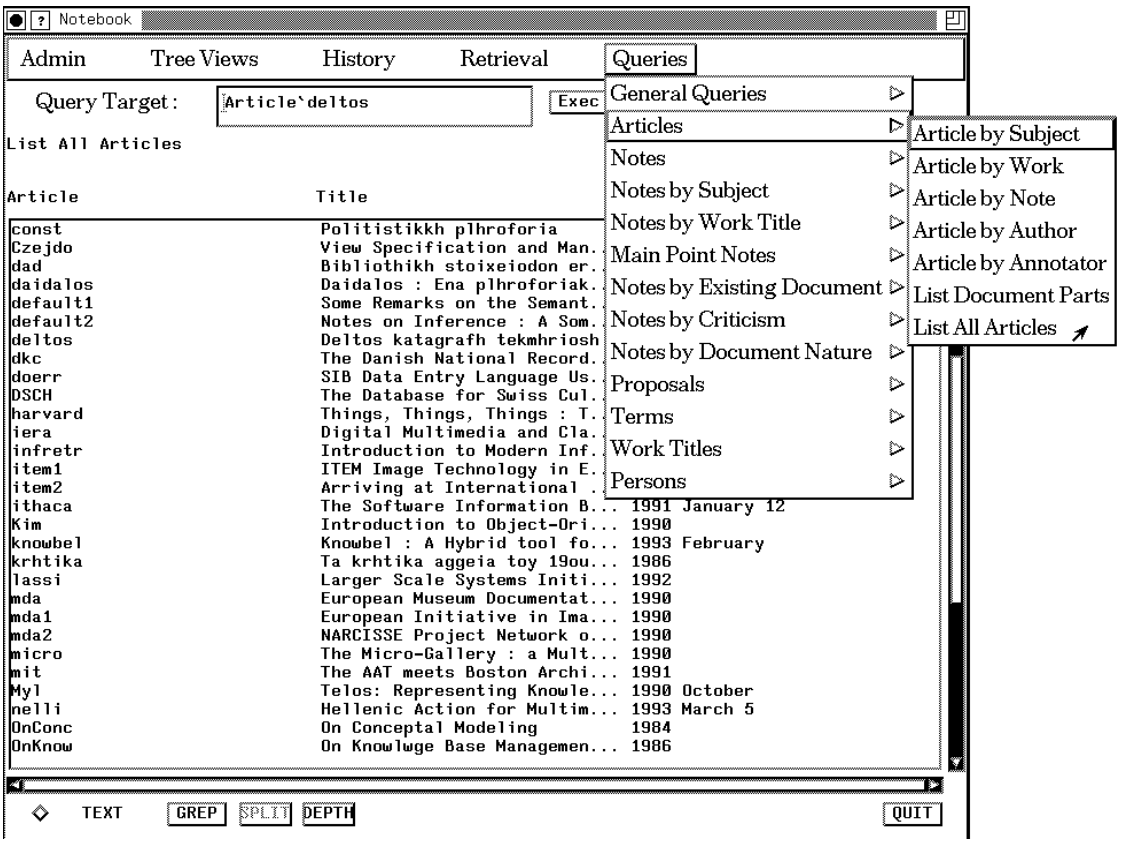
Σχήμα 6.2: Προκαθορισμένες ερωτήσεις σε μορφή γράφου

Στην εικόνα φαίνεται ο κατάλογος των προκαθορισμένων ερωτήσεων σε μορφή γράφου καθώς και ένα παράδειγμα εκτέλεσης της ερώτησης *Get Structure of Document*. Ο γράφος παριστάνει την λογική δομή του άρθρου *Article'Cons95*.

της **MenuDescription** με τιμές στα γνωρίσματα *queryMenu*, *retrievalMenu*, *viewMenu* τους αντίστοιχους καταλόγους επιλογών με προκαθορισμένες ερωτήσεις.

6.2 Ερωτήσεις 1ης τάξεως

Με τον όρο ερωτήσεις πρώτης τάξεως περιγράφονται εκείνες οι ερωτήσεις που α) δεν είναι αναδρομικές και β) έχουν συνήθως αποτέλεσμα σε μορφή κειμένου. Όπως αναφέρθηκε στην προηγούμενη παράγραφο οι ερωτήσεις αυτού του τύπου υπάρχουν στο ΣΣΕ κάτω από την επιλογή *Queries*. Όπως φαίνεται στο σχήμα 6.3 κατά την επιλογή αυτή παρουσιάζεται ένας κατάλογος από επιλογές κάθε μία από τις οποίες αποτελείται

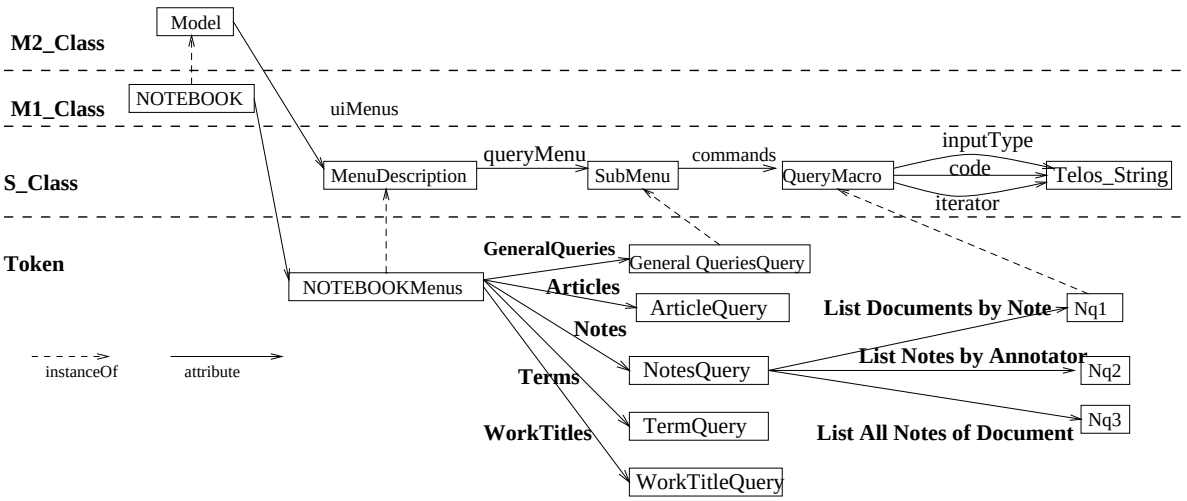


Σχήμα 6.3: Αποτέλεσμα προκαθορισμένης ερώτησης σε μορφή κειμένου

Στην εικόνα φαίνεται ο κατάλογος των προκαθορισμένων ερωτήσεων σε μορφή κειμένου καθώς και ένα παράδειγμα εκτέλεσης της ερώτησης *List All Articles*.

από έναν άλλο κατάλογο από εκτελέσιμες πλέον εντολές. Σύμφωνα με το μοντέλο που παρουσιάστηκε παραπάνω, ο πρώτος κατάλογος επιλογών είναι οι τιμές του γνωρίσματος *queryMenu* της οντότητας **NOTEBOOKMenus**. Κάθε τιμή είναι περίπτωση της κλάσης **SubMenu**, δηλαδή κάθε τιμή αποτελεί έναν υποκατάλογο από επιλογές. Οι επιλογές αυτές είναι εκτελέσιμες. Οι εκτελέσιμες επιλογές είναι περιπτώσεις της κλάσης **QueryMacro** και αποτελούν τιμές του γνωρίσματος *commands* από τη **SubMenu**. Στο σχήμα 6.4 φαίνεται ο τρόπος παράστασης των ερωτήσεων. Τα ονόματα των επιλογών είναι τα ίδια με τα ονόματα των γνωρισμάτων που αποτελούν περιπτώσεις του γνωρίσματος *commands* έτσι ο αναγνώστης μπορεί να συγκρίνει το μοντέλο με την τελική παρουσίαση που φαίνεται στο σχήμα 6.3.

Οι επιλογές του αρχικού καταλόγου αφορούν γενικές έννοιες του μοντέλου. Οι έννοιες αυτές είναι τα έγγραφα, οι σημειώσεις στα έγγραφα, οι σημειώσεις που εκφράζουν



Σχήμα 6.4: Ένα υποσύνολο του μοντέλου των προκαθορισμένων ερωτήσεων

Στην εικόνα φαίνεται ο τρόπος παράστασης των προκαθορισμένων ερωτήσεων στη γλώσσα TELOS. Με τονισμένα γράμματα δηλώνονται τα ονόματα των γνωρισμάτων που αποτελούν ταυτόχρονα τους τίτλους των επιλογών στους καταλόγους όπως φαίνεται στο σχήμα 6.3

ταξινόμηση σε θεματικό θησαυρό, οι σημειώσεις σχετικά με τις εργασίες, οι σημειώσεις σχετικά με τα κύρια σημεία του εγγράφου, οι σημειώσεις που συσχετίζουν τα έγγραφα μεταξύ τους, οι σημειώσεις που εκφράζουν ταξινόμηση σε σχέση με την φύση του εγγράφου, οι σημειώσεις που εκφράζουν ταξινόμηση σε σχέση με την κριτική που κάνουμε στα έγγραφα, οι σημειώσεις που εκφράζουν πρόταση για νέες εργασίες, οι όροι του θεματικού θησαυρού, οι τίτλοι των εργασιών και τα πρόσωπα. Εξαίρεση αποτελεί η πρώτη επιλογή του καταλόγου που επιγράφεται *General Queries* και μέσω των επιλογών της μπορεί να πάρει κανείς ως αποτέλεσμα τις περιπτώσεις μιας οντότητας (*List Objects Of Kind*), τις οντότητες των οποίων μία οντότητα είναι περίπτωση (*List Kinds Of Object*), τις υποκλάσεις (*List Special Kinds Of*) και τις υπερκλάσεις (*List General Kinds Of*) μιας οντότητας [1]. Ο λόγος που υπάρχουν οι ερωτήσεις αυτές ξεχωριστά είναι ότι α) ο χρήστης μπορεί να τις χρησιμοποιεί συχνά και γι'αυτό είναι καλύτερα να του δίνονται ως βασική επιλογή, και β) ανακαλούν πληροφορία με τις βασικότερες σχέσεις που δίνει η TELOS, δηλαδή τη σχέση γενίκευσης και ταξινόμησης. Παρ'όλα αυτά, όπως αναφέρεται και παρακάτω, πολλές φορές ερωτήσεις των επόμενων επιλογών δίνουν τα ίδια αποτελέσματα.

Οι υποκατάλογοι που υπάρχουν εκτελούν συγκεκριμένες ερωτήσεις με βάση μία ή και

[1] Κατά σύμβαση με τον όρο *Object* εννοείται η περίπτωση, συνήθως σε επίπεδο **Token**, ενώ με τον όρο **Kind** κάποιο είδος κλάσης.

καμία παράμετρο. Οι ερωτήσεις με καμία παράμετρο δίνουν συνήθως ως αποτέλεσμα όλες τις περιπτώσεις της κλάσης με την οποία περιγράφεται η έννοια αυτή στο μοντέλο μας. Κατά σύμβαση αρχίζουν με την έκφραση *List All*. Όλες οι επιλογές έχουν μια ερώτηση του τύπου αυτού. Έτσι υπάρχουν οι ερωτήσεις *List All Articles, List All Collections, List All Collections of Work Titles, List All Work Titles, List All Collections of Terms, List All Terms, List All Proposals, List All Authors, List All Annotators* κλπ στις αντίστοιχες επιλογές του καταλόγου. Οι ερωτήσεις αυτές είναι απαραίτητες γιατί κάποιος που δεν έχει γνώση των οντοτήτων του μοντέλου μπορεί έτσι να δει το συνολικό περιεχόμενο της βάσης γύρω από ένα θέμα. Τα ίδια αποτελέσματα με την εκτέλεση μιας ερώτησης τύπου *List All* θα μπορούσε να πάρει κανείς π.χ. στην περίπτωση των άρθρων θέτοντας την κλάση **Article** ως στόχο της ερώτησης και εκτελώντας την ερώτηση *List Objects Of* της επιλογής *General Queries*.

Οι ερωτήσεις με μία παράμετρο εκτελούνται με όρισμα τον στόχο της ερώτησης που τίθεται στη θέση που υποδεικνύεται από την επιγραφή *Query Target* (βλ. σχήματα 6.3, 6.2). Τέτοιες ερωτήσεις έχουν όλες οι επιλογές (οι κατάλογοι). Στις επιλογές αυτές το είδος της παραμέτρου δηλώνεται είτε με τη λέξη *by* είτε με τη λέξη *Of*. Έτσι μπορεί κάποιος να εκτελέσει πρώτα μία ερώτηση του τύπου *List All* και να πάρει όλες τις δυνατές τιμές της παραμέτρου και στη συνέχεια να διαλέξει μία με βάση την οποία να κάνει μια ερώτηση. Για παράδειγμα εκτελώντας την ερώτηση *List All Articles* παίρνει κανείς ως αποτέλεσματα όλα τα άρθρα που υπάρχουν στο σύστημα και στη συνέχεια διαλέγοντας ένα από αυτά και εκτελώντας την ερώτηση *List All Notes of Document* παίρνει ως αποτέλεσμα όλες τις σημειώσεις που έχουν γίνει στο δεδομένο άρθρο ή στα τμήματα του. Ερωτήσεις με παραπάνω από μία παράμετρο ή σε μια όψη (view) της βάσης δεν υποστηρίζει, προς το παρόν, το ΣΣΕ . Παρ'όλα αυτά στο μέλλον θα είναι δυνατόν το αποτέλεσμα μιας ερώτησης να ''φιλτράρεται'' μέσω μιας άλλης και έτσι το αποτέλεσμα των διαδοχικών ερωτήσεων σε ένα σύνολο οντοτήτων θα είναι το ίδιο με το να είχε γίνει η ερώτηση με ίσο αριθμό παραμέτρων. Βέβαια υπάρχει η δυνατότητα να ορίσουμε κάρτες αναζήτησης που αποτελούνται από πολλές ερωτήσεις, όπως θα περιγράψουμε στην ενότητα 6.3.

Αναλυτικά, οι εκτελέσιμες ερωτήσεις που υπάρχουν στις επιλογές (menus) περιγράφονται στις παρακάτω υποενότητες.

6.2.1 Η επιλογή Articles

Στην επιλογή *Articles* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *Article by Subject* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλα τα άρθρα που έχουν ταξινομηθεί σε αυτό τον όρο.

- *Article by Immediately Narrow Subjects* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλα τα άρθρα που έχουν ταξινομηθεί σε αυτό τον όρο ή σε κάποιον από τους άμεσους (δηλ. συνδέονται με μία σύνδεση) στενότερους όρους του.

- *Article by All Narrow Subjects* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλα τα άρθρα που έχουν ταξινομηθεί σε αυτό τον όρο ή σε κάποιον από τους στενότερους όρου του.

- *Article by Work* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλα τα άρθρα που έχουν σημειώσεις σχετικά με αυτό το τίτλο εργασίας.

- *Article by Immediately Parts of Work* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλα τα άρθρα που έχουν σημειώσεις σχετικά με αυτή την εργασία ή με κάποια από τις άμεσες υποεργασίες της.

- *Article by All Parts of Work* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλα τα άρθρα που έχουν σημειώσεις σχετικά με αυτή την εργασία ή με κάποια από τις υποεργασίες της ή με κάποια από τις υποεργασίες των υποεργασιών της κ.ο.κ.

- *Article by Note* : Η ερώτηση αυτή παίρνει ως παράμετρο μία σημείωση (δηλ. μία περίπτωση της κλάσης *AnnotationObject* ,βλ. ενότητα 5.3) και έχει ως αποτέλεσμα όλα τα άρθρα που έχουν σημειωθεί με αυτή τη σημείωση.

- *Article by Annotator* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όνομα (το μικρό όνομα ή το επώνυμο ή το πλήρες όνομα) και έχει ως αποτέλεσμα όλα τα άρθρα στα οποία έχει κάνει σημειώσεις το πρόσωπο με αυτό το όνομα.

- *Article by Author* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όνομα (το μικρό όνομα ή το επώνυμο ή το πλήρες όνομα) και έχει ως αποτέλεσμα όλα τα άρθρα που έχουν συγγραφέα με αυτό το όνομα.

- *List Parts of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλα τα λογικά τμήματα από τα οποία αποτελείται αυτό.

- *List All Articles* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλα τα άρθρα που υπάρχουν στην βάση δεδομένων.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση άρθρων δίνοντας κάποιο χαρακτηριστικό τους, εκτός από την ερώτηση List Parts of Document που δίνει τα τμήματα του εγγράφου. Τα αποτελέσματα των ερωτήσεων περιλαμβάνουν το κωδικό όνομα του άρθρου, τον τίτλο του, την ημερομηνία δημοσίευσης (βλ σχήμα 6.3). Επίσης στις επτά πρώτες προκαθορισμένες ερωτήσεις παρουσιάζεται και η σημείωση με την οποία έγινε η εύρεση του άρθρου.

6.2.2 Η επιλογή Notes

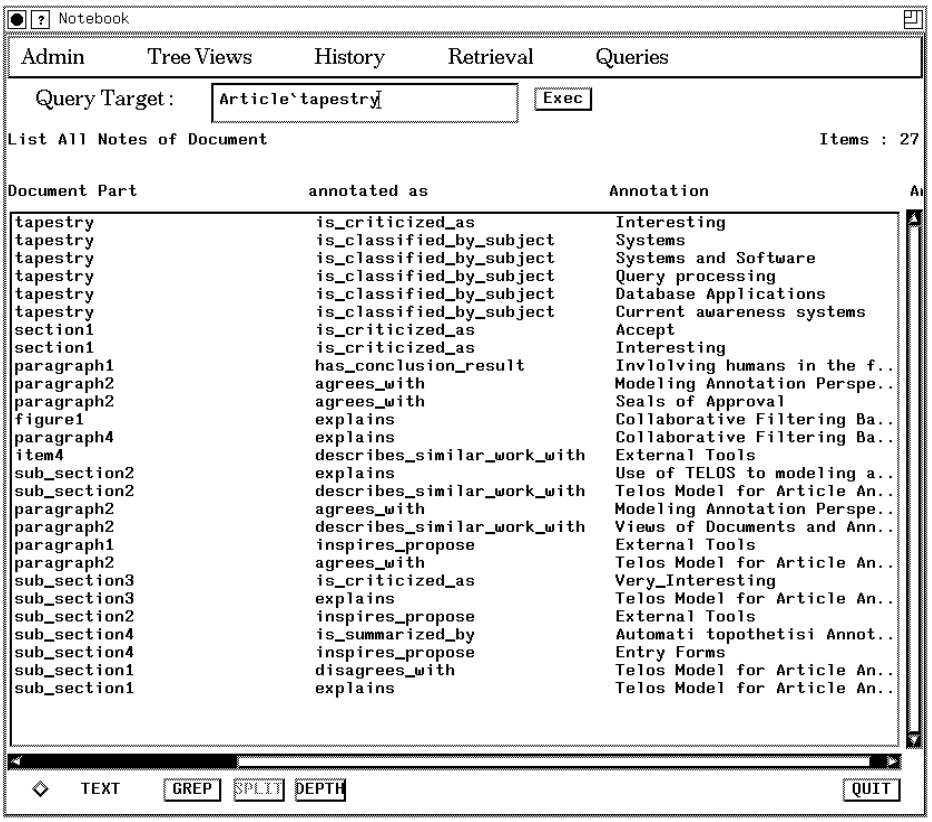
Στην επιλογή *Notes* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List Documents by Note* : Η ερώτηση αυτή παίρνει ως παράμετρο μία σημείωση (δηλ. μία περίπτωση της κλάσης *AnnotationObject* ,βλ. ενότητα 5.3) και έχει ως αποτέλεσμα όλα τα έγγραφα ή τμήματά τους που έχουν σημειωθεί με αυτή τη σημείωση. Η ερώτηση αυτή μοιάζει με την ερώτηση *Article by Note* (βλ. ενότητα 6.2.1). Αλλά η διαφορά τους είναι ότι η ερώτηση *Article by Note* έχει ως αποτέλεσμα τα άρθρα που σε κάποιο τμήμα τους έχουν σημειωθεί με την σημείωση, που δίνουμε ως παράμετρο, ενώ η ερώτηση *List Documents by Note* έχει ως αποτέλεσμα τα τμήματα του άρθρου στα οποία έγινε η σημείωση.

- *List Notes by Annotator* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όνομα (το μικρό όνομα ή το επώνυμο ή το πλήρες όνομα) και έχει ως αποτέλεσμα όλες τις σημειώσεις που έχει κάνει το πρόσωπο με αυτό το όνομα. Βέβαια μαζί με την σημείωση παρουσιάζει και το τμήμα στο οποίο έγινε καθώς και την κατηγορία της σημείωσης.

- *List All Notes of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις που έχουν γίνει σε αυτό το έγγραφο.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση πληροφοριών σχετικά με τις σημειώσεις. Τα αποτελέσματα των παραπάνω ερωτήσεων περιλαμβάνουν τα εξής:



Σχήμα 6.5: Ένα παράδειγμα παρουσίασης αποτελεσμάτων

Στην εικόνα φαίνονται τα αποτελέσματα της εκτέλεσης της ερώτησης *List All Notes of Document* με παράμετρο το άρθρο *Article'tapestry*. Η κάθε γραμμή αποτελέσματος μοιάζει με μια πρόταση.

το τμήμα του εγγράφου στο οποίο έγινε η σημείωση, την κατηγορία της σημείωσης, το περιεχόμενο της σημείωσης, το όνομα αυτού που έκανε την σημείωση και την ημερομηνία της σημείωσης. Η κάθε γραμμή αποτελέσματος μοιάζει με μια πρόταση, δηλ. έχει το υποκείμενο (το έγγραφο) το ρήμα (η κατηγορία σημείωσης) και το αντικείμενο ή τον προσδιορισμό (το περιεχόμενο της σημείωσης). Ένα παράδειγμα φαίνεται στην εικόνα 6.5.

6.2.3 Η επιλογή ''Notes by Subject''

Στην επιλογή *''Notes by Subject''* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *Notes by Term* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλες τις σημειώσεις που έχουν γίνει με αυτό τον

όρο.

- *Notes by Immediately Narrow Terms* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλες τις σημειώσεις που έχουν γίνει με αυτό τον όρο ή με τους άμεσους στενότερους όρους του (δηλ. συνδέονται με μία σύνδεση).

- *Notes by All Narrow Terms* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλες τις σημειώσεις που έχουν γίνει με αυτό τον όρο ή με τους στενότερους όρους του.

- *List Notes of Document by All Subjects* : Η αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις με όρους του θεματικού θησαυρου, που έχουν γίνει με αυτό.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση των εγγράφων ή τμημάτων τους που έχουν ταξινομηθεί σε κάποιο θεματικό όρο. Η μορφή παρουσίασης των αποτελεσμάτων τους είναι η ίδια όπως και στην υποενότητα 6.2.2.

6.2.4 Η επιλογή ‘‘Notes by Work Title”

Στην επιλογή *‘‘Notes by Work Title’’* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List Documents that Agrees by Work* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλα τα έγγραφα που σημειώθηκαν ότι συμφωνούν με αυτή την εργασία (δηλ. η σύνδεση της σημείωσης είναι περίπτωση της κατηγορίας agrees_with).

- *List Documents that Disagrees by Work* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλα τα έγγραφα που σημειώθηκαν ότι διαφωνούν με αυτή την εργασία (δηλ. η σύνδεση της σημείωσης είναι περίπτωση της κατηγορίας disagrees_with).

- *List Documents that Explain Work* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλα τα έγγραφα που σημειώθηκαν ότι επεξηγούν αυτή την εργασία (δηλ. η σύνδεση της σημείωσης είναι περίπτωση της κατηγορίας explain_of).

- *List Documents that Describes Similar Work* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλα τα έγγραφα που σημειώθηκαν ότι

περιγράφουν παρόμοια εργασία με αυτή την εργασία (δηλ. η σύνδεση της σημείωσης είναι περίπτωση της κατηγορίας describes_similar_work_with).

- *List Notes of Document by All Work Titles* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις σχετικά με τις εργασίες που έχουν γίνει σε αυτό το έγγραφο.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση πληροφοριών σχετικά με τις εργασίες. Τα αποτελέσματα των παραπάνω ερωτήσεων περιλαμβάνουν τα εξής : το έγγραφο στο οποίο έγινε η σημείωση, τον σχολιαστή, την ημερομηνία δημιουργίας της σημείωσης, το βάρος της σημείωσης και το σχόλιο για την σημείωση.

6.2.5 Η επιλογή ‘‘Main Point Notes”

Στην επιλογή *‘‘Main Point Notes’’* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List Conclusion or Result of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις σχετικά με τα συμπεράσματα ή τα αποτελέσματα του εγγράφου.

- *List Problems Discovered of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις σχετικά με τα προβλήματα που ανακαλύπτει το έγγραφο.

- *List Problems Solved of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις σχετικά με τα προβλήματα που επιλύονται στο έγγραφο.

- *List Aim of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις σχετικά με τους σκοπούς του εγγράφου.

- *List All Main Point Notes of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις σχετικά με όλα τα κύρια σημεία του εγγράφου.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση πληροφοριών σχετικά με τις σημειώσεις στα κύρια σημεία του εγγράφου.Τα αποτελέσματα των παραπάνω ερωτήσεων περιλαμβάνουν τα εξής: το έγγραφο στο οποίο έγινε η σημείωση, την σημείωση,

τον σχολιαστή, την ημερομηνία δημιουργίας της σημείωσης, το βάρος της σημείωσης και το σχόλιο για την σημείωση.

6.2.6 Η επιλογή ''Notes by Existing Document''

Στην επιλογή *''Notes by Existing Document''* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List Similar theme of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλα τα έγγραφα που σημειώθηκαν ότι έχουν παρόμοιο θέμα.
- *List Relate theme of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλα τα έγγραφα που σημειώθηκαν ότι έχουν σχετικό θέμα.
- *List Further work of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλα τα έγγραφα που σημειώθηκαν ότι περιγράφουν εξέλιξη της εργασίας που περιγράφεται σε αυτό.
- *List Explanation of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλα τα έγγραφα που σημειώθηκαν ότι είναι επεξηγήσεις σε αυτό.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την παρουσίαση των συσχετίσεων μεταξύ των εγγράφων. Τα αποτελέσματα των παραπάνω ερωτήσεων περιλαμβάνουν τα εξής: το έγγραφο στο οποίο έγινε η σημείωση, το έγγραφο που αποτελεί την σημείωση, τον σχολιαστή, την ημερομηνία δημιουργίας της σημείωσης, το βάρος της σημείωσης και το σχόλιο για την σημείωση (αν υπάρχουν).

6.2.7 Η επιλογή ''Notes by Criticism''

Στην επιλογή *Notes by by Criticism* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List All Criticism Notes of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις που εκφράζουν κριτική για το έγγραφο αυτό.
- *List All Documents Criticised by Term* : Η ερώτηση αυτή παίρνει ως παράμετρο έναν όρο που εκφράζει κριτική (δηλ. περίπτωση της κλάσης Criticism) και έχει ως αποτέλεσμα όλα τα έγγραφα στα οποία έχει γίνει κριτική με αυτό τον όρο.

- *List All Criticism Terms* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλους τους όρους που εκφράζουν κριτική (δηλ. όλες τις περιπτώσεις της κλάσης DocumentNature).

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση πληροφοριών σχετικά με την κριτική που έχει γίνει στα έγγραφα.

6.2.8 Η επιλογή ‘‘Notes by Document Nature”

Στην επιλογή *Notes by by Criticism* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List All Notes on Nature of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις σημειώσεις σχετικά με την φύση του εγγράφου αυτού.

- *List All Documents by Nature* : Η ερώτηση αυτή παίρνει ως παράμετρο έναν όρο που εκφράζει την φύση του εγγράφου (δηλ. περίπτωση της κλάσης DocumentNature) και έχει ως αποτέλεσμα όλα τα έγγραφα τα οποία έχουν αυτή τη φύση.

- *List All Nature Terms* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλους τους όρους που εκφράζουν την φύση του εγγράφου (δηλ. όλες τις περιπτώσεις της κλάσης DocumentNature).

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση πληροφοριών σχετικά με την φύση των εγγράφων.

6.2.9 Η επιλογή Proposals

Στην επιλογή *Proposals* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List Proposals Inspired of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο ή τμήμα του και έχει ως αποτέλεσμα όλες τις προτάσεις που έχουν γίνει και σχετίζονται με αυτό το έγγραφο.

- *List Proposals by Subject Term* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλες τις προτάσεις που έχουν γίνει και σχετίζονται με αυτό το θέμα.

- *List Proposals by Work* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλες τις προτάσεις που έχουν γίνει και σχετίζονται με αυτή την εργασία.

- *List Proposals by Person* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όνομα (το μικρό όνομα ή το επώνυμο ή το πλήρες όνομα) και έχει ως αποτέλεσμα όλες τις προτάσεις που έχουν γίνει από το πρόσωπο με αυτό το όνομα.

- *List All Proposals* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλες τις προτάσεις που υπάρχουν στην βάση δεδομένων.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση πληροφοριών σχετικά με τις προτάσεις που υπάρχουν στη βάση δεδομένων. Τα αποτελέσματα των παραπάνω ερωτήσεων περιλαμβάνουν τα εξής: το έγγραφο με το οποίο σχετίζεται η πρόταση, την σημείωση που εκφράζει την πρόταση, τον σχολιαστή, την ημερομηνία δημιουργίας της σημείωσης, το βάρος της σημείωσης και το σχόλιο για την σημείωση (αν υπάρχουν).

6.2.10 Η επιλογή Subject Terms

Στην επιλογή *Subject Terms* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List Synonym Term* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλους τους συνώνυμους όρους του.

- *List Relate Term* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλους τους σχετικούς προς αυτόν όρους.

- *List Terms of Collection* : Η ερώτηση αυτή παίρνει ως παράμετρο το όνομα μιας συλλογής όρων και έχει ως αποτέλεσμα όλους τους όρους της συλλογής αυτής.

- *List All Collections of Terms* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλες τις συλλογές θεματικών όρων που έχουν οριστεί στη βάση.

- *List All Top Terms* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλους τους όρους που δεν έχουν ευρύτερους όρους, δηλαδή είναι οι γενικότεροι όροι του θησαυρού.

- *List All Terms* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλους τους όρους του θησαυρού.

Επίσης στην επιλογή *Subject Terms* υπάρχουν και οι εξής επιλογές:

- *List Narrow Terms* : Αυτή είναι μια επιλογή που έχει δύο ερωτήσεις. Η ερώτηση *List Immediately Narrow Terms* παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού

και έχει ως αποτέλεσμα όλους τους άμεσους (δηλ. συνδέονται με μία σύνδεση) στενότερους όρους του. Ενώ η ερώτηση *List All Narrow Terms* παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλους τους στενότερους όρους του.

- *List Broader Terms* : Αυτή είναι μια επιλογή που έχει δύο ερωτήσεις. Η ερώτηση *List Immediately Broader Terms* παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλους τους άμεσους (δηλ. συνδέονται με μία σύνδεση) ευρύτερους όρους του. Ενώ η ερώτηση *List All Broader Terms* παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλους τους ευρύτερους όρους του.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση πληροφοριών σχετικά με τους όρους του θεματικού θησαυρού.

6.2.11 Η επιλογή Work Titles

Στην επιλογή *Work Titles* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List Immediately Work Parts* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλες τις υποεργασίες στις οποίες διασπάται.
- *List All Work Parts* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα όλες τις υποεργασίες στις οποίες διασπάται, τις υποεργασίες των υποεργασιών της, κ.ο.κ..
- *List Work that Target Belongs* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα την εργασία στην οποία ανήκει (δηλαδή η παράμετρος της ερώτησης είναι υποεργασία του αποτελέσματος).
- *List Relate Subject Terms* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα τους όρους του θεματικού θησαυρού που με τους οποίους είναι σχετική αυτή η εργασία.
- *List Relate Work* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα τίτλο εργασίας και έχει ως αποτέλεσμα τους τίτλους των εργασιών που είναι σχετικές με αυτή την εργασία.
- *List Work Titles of Collection* : Η ερώτηση αυτή παίρνει ως παράμετρο το όνομα μιας συλλογής τίτλων εργασιών και έχει ως αποτέλεσμα όλους τους τίτλους εργασιών που ανήκουν σε αυτή.

- *List All Collections of Work Titles* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλες τις συλλογές τίτλων εργασιών που έχουν ορισθεί στη βάση.

- *List All Work Titles* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλους τους τίτλων εργασιών που υπάρχουν στη βάση.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση πληροφοριών σχετικά με τις εργασίες.

6.2.12 Η επιλογή Persons

Στην επιλογή *Persons* υπάρχουν οι εξής εκτελέσιμες ερωτήσεις:

- *List Annotators by Subject Term* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα όρο του θεματικού θησαυρού και έχει ως αποτέλεσμα όλα τα πρόσωπα που έχουν κάνει κάποια σημείωση με αυτό τον όρο.

- *List All Annotators of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο και έχει ως αποτέλεσμα όλα τα πρόσωπα που έχουν κάνει κάποια σημείωση σε αυτό το έγγραφο.

- *List All Authors of Document* : Η ερώτηση αυτή παίρνει ως παράμετρο ένα έγγραφο και έχει ως αποτέλεσμα όλους τους συγγραφείς του εγγράφου.

- *List All Authors* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλους τους συγγραφείς που υπάρχουν στη βάση.

- *List All Annotators* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλα τα πρόσωπα που έχουν κάνει κάποια σημείωση στα έγγραφα της βάσης δεδομένων.

- *List All Persons* : Η ερώτηση αυτή δεν παίρνει παράμετρο και έχει ως αποτέλεσμα όλους τα πρόσωπα που υπάρχουν στη βάση.

Οι ερωτήσεις αυτής της επιλογής έχουν ώς στόχο την εύρεση πληροφοριών σχετικά με τα πρόσωπα που υπάρχουν στη βάση.

6.3 Αναζήτηση με πολλές παραμέτρους

Ερωτήσεις πρώτης τάξεως μπορεί να γίνουν επίσης και μέσω της επιλογής *Retrieval*. Η επιλογή αυτή παρέχει τη δυνατότητα ερωτήσεων με παραπάνω από μια παραμέτρους. Ουσιαστικά πρόκειται για μία κάρτα που αποτελείται από ένα σύνολο ερωτήσεων, όπου η κάθε ερώτηση παίρνει ώς παράμετρο μια λογική έκφραση (με χρήση του λογικού "και" και του λογικού "ή") από αντικείμενα ή κλάσεις του ιδίου τύπου. Κάθε ερώτηση εκτελείται με βάση τις δικές της παραμέτρους και δίνει ως αποτέλεσμα ένα σύνολο από αντικείμενα. Έπειτα από τα σύνολα των αποτελεσμάτων παίρνουμε την τομή τους και προκύπτει ένα σύνολο στο οποίο ικανοποιούνται όλες οι ερωτήσεις. Ένα παράδειγμα κάρτας αναζήτησης βλέπουμε στην εικόνα 6.6. Δίπλα σε κάθε πεδίο υπάρχει ο τύπος της παραμέτρου ή ο τρόπος που χρησιμοποιούμε την παράμετρο (π.χ. δημοσιεύτηκε μετά την ημερομηνία). Επίσης , όπως φαίνεται και στο σχήμα, δεν είναι απαραίτητο να δώσουμε παραμέτρους σε όλες τις ερωτήσεις, δηλαδή όταν δεν δίνουμε παράμετρο σε μια ερώτηση τότε ουσιαστικά δεν παίρνουμε υπόψην αυτή την ερώτηση.

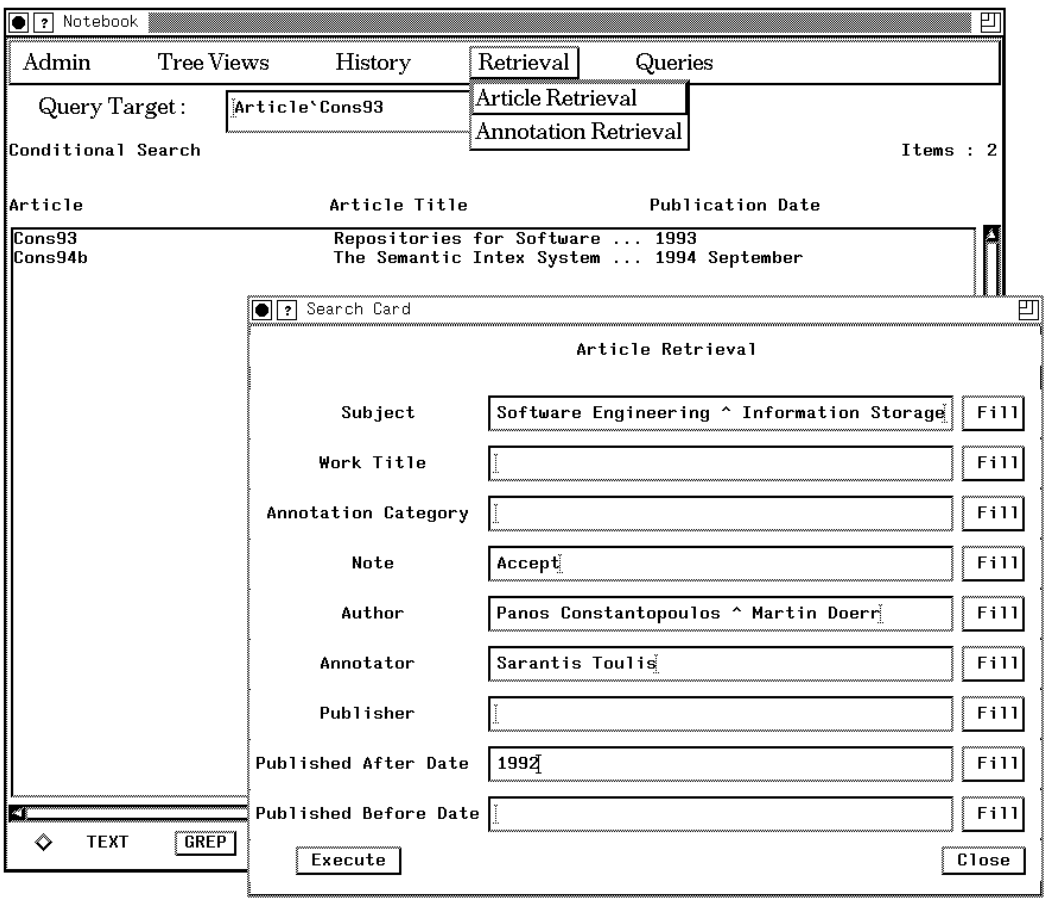
Η κάρτα αναζήτησης χρησιμοποιείται όταν θέλουμε να κάνουμε αναζήτηση αντικειμένων με βάση περισσότερα από ένα χαρακτηριστικά τους και με περισσότερες από μία παραμέτρους αυτά. Στην εφαρμογή που υλοποιήσαμε κρίναμε ότι είναι απαραίτητο να υπάρχει μία κάρτα αναζήτησης εγγράφων και μία κάρτα αναζήτησης σημειώσεων. Η κάρτα αναζήτησης εγγράφων, που φαίνεται στο σχήμα 6.6, περιλαμβάνει της εξής ερωτήσεις :

- *Subject* : παίρνει ως παράμετρο ένα λογικό συνδυασμό αντικειμένων τύπου Subject-Thesaurus (δηλ. όροι του θεματικού θησαυρού) και έχει ως αποτέλεσμα τα έγγραφα που έχουν ταξινομηθεί σε αυτό που εκφράζει ο συνδυασμός των όρων.

- *WorkTitle* : παίρνει ως παράμετρο ένα λογικό συνδυασμό αντικειμένων τύπου WorkTitle (δηλ. τίτλοι εργασιών) και έχει ως αποτέλεσμα τα έγγραφα που σχετίζονται με αυτό που εκφράζει ο συνδυασμός των τίτλων εργασιών.

- *Annotation Category* : παίρνει ως παράμετρο ένα λογικό συνδυασμό αντικειμένων τύπου AnnotationCategory (βλ. ενότητα 5.2) και έχει ως αποτέλεσμα τα έγγραφα που έχουν σημειωθεί κάτω από τις κατηγορίες που ορίζει αυτός ο λογικός συνδυασμός.

- *Note* : παίρνει ως παράμετρο ένα λογικό συνδυασμός σημειώσεων και έχει ως αποτέλεσμα τα έγγραφα που έχουν δεχθεί τις σημειώσεις που ορίζει αυτός ο



Σχήμα 6.6: Η κάρτα αναζήτησης εγγράφων

συνδυασμός.

- *Author* : παίρνει ως παράμετρο ένα λογικό συνδυασμός συγγραφέων και έχει ως αποτέλεσμα τα έγγραφα που έχουν ως συγγραφέα αυτό που ορίζει ο συνδυασμός.

- *Annotator* : παίρνει ως παράμετρο ένα λογικό συνδυασμός ονομάτων και έχει ως αποτέλεσμα τα έγγραφα που έχουν δεχθεί σημειώσεις από τα πρόσωπα που ορίζει ο συνδυασμός αυτός.

- *Publisher* : παίρνει ως παράμετρο ένα λογικό συνδυασμός από περιοδικά ή συνέδρια παρουσίασης εγγράφων και έχει ως αποτέλεσμα τα έγγραφα που έχουν παρουσιαστεί εκεί που ορίζει ο συνδυασμός αυτός.

- *Published After Date* : παίρνει ως παράμετρο μία ημερομηνία και έχει ως αποτέλεσμα

τα έγγραφα που έχουν δημοσιευτεί πριν από αυτή την ημερομηνία ή κατά την διάρκεια της (αν δεν είναι ακριβής ημερομηνία π.χ. 1995).

- *Published Before Date*: παίρνει ως παράμετρο μία ημερομηνία και έχει ως αποτέλεσμα τα έγγραφα που έχουν δημοσιευτεί μετά από αυτή την ημερομηνία ή κατά την διάρκεια της.

Ενώ κάρτα αναζήτησης σημειώσεων περιλαμβάνει της εξής ερωτήσεις :

- *Annotation Category* : παίρνει ως παράμετρο ένα λογικό συνδυασμό αντικειμένων τύπου AnnotationCategory (βλ. ενότητα 5.2) και έχει ως αποτέλεσμα όλους τους συνδέσμους σημειώσεων που εντάσσονται κάτω από τις κατηγορίες που ορίζει αυτός ο λογικός συνδυασμός.

- *Note* : παίρνει ως παράμετρο ένα λογικό συνδυασμός σημειώσεων και έχει ως αποτέλεσμα τους συνδέσμους των σημειώσεων που ορίζει αυτός ο συνδυασμός.

- *Annotator* : παίρνει ως παράμετρο ένα λογικό συνδυασμός ονομάτων και έχει ως αποτέλεσμα τους συνδέσμους (links) των σημειώσεων που έχουν κάνει τα πρόσωπα που ορίζει ο συνδυασμός αυτός.

- *Author* : παίρνει ως παράμετρο ένα λογικό συνδυασμός συγγραφέων και έχει ως αποτέλεσμα τους συνδέσμους των σημειώσεων που έχουν γίνει σε έγγραφα που έχουν ως συγγραφέα αυτό που ορίζει ο συνδυασμός.

- *Weight >= from* : παίρνει ως παράμετρο ένα βάρος (από 1 έως 5) και έχει ως αποτέλεσμα τους συνδέσμους των σημειώσεων που χαρακτηρίζονται με βάρος μεγαλύτερο ή ίσο με αυτό της παραμέτρου.

- *Weight <= from* : παίρνει ως παράμετρο ένα βάρος (από 1 έως 5) και έχει ως αποτέλεσμα τους συνδέσμους των σημειώσεων που χαρακτηρίζονται με βάρος μικρότερο ή ίσο με αυτό της παραμέτρου.

- *Published After Date*: παίρνει ως παράμετρο μία ημερομηνία και έχει ως αποτέλεσμα τους συνδέσμους των σημειώσεων που έγιναν πριν από αυτή την ημερομηνία ή κατά την διάρκεια της (αν δεν είναι ακριβής ημερομηνία π.χ. 1995).

- *Published Before Date*: παίρνει ως παράμετρο μία ημερομηνία και έχει ως αποτέλεσμα τους συνδέσμους των σημειώσεων που έγιναν μετά από αυτή την ημερομηνία ή κατά την διάρκεια της (αν δεν είναι ακριβής ημερομηνία π.χ. 1995).

Η μορφή παρουσίασης των σημειώσεων είναι ίδια με αυτή που φαίνεται στο σχήμα 6.5.

6.4 Αναδρομικές προκαθορισμένες ερωτήσεις

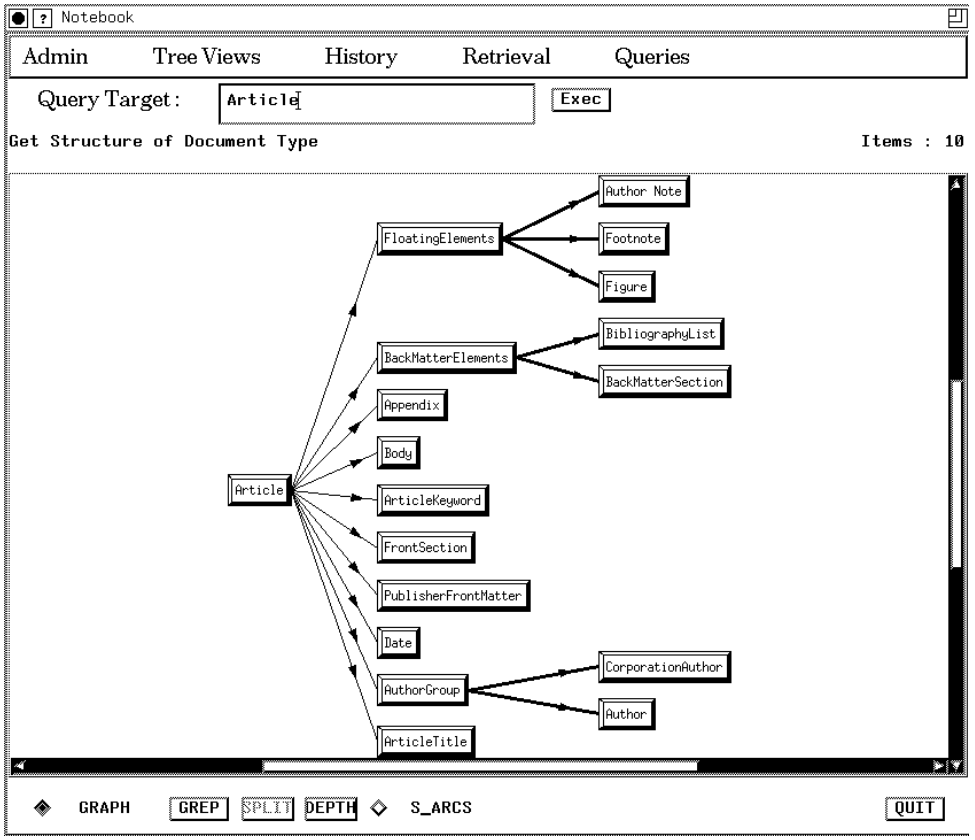
Οι αναδρομικές ερωτήσεις χρησιμοποιούν τις συναρτήσεις της ΒΣΕΣ οι οποίες δίνουν το μεταβατικό περίβλημα μιας σχέσης. Έτσι δίνοντας έναν κόμβο του σημασιολογικού δικτύου ως αρχικό, καθώς και μία ή περισσότερες κατηγορίες ή μετακατηγορίες γνωρισμάτων, αρχικά λαμβάνεται το σύνολο των κόμβων που είναι συνδεδεμένοι με τον αρχικό μέσω της κατηγορίας (ή της μετακατηγορίας) γνωρισμάτων, και μετά για κάθε στοιχείο του συνόλου αυτού επαναλαμβάνεται η ίδια διαδικασία, ώσπου το σύνολο των κόμβων κάποιας επανάληψης να είναι κενό. Για παράδειγμα, ο γράφος που φαίνεται στο σχήμα 4.8 είναι αποτέλεσμα μιας αναδρομικής ερώτησης για τα γνωρίσματα που είναι περιπτώσεις της μετακατηγορίας *consist_of* με αρχικό κόμβο το αντικείμενο **Article'Cons95**. Αναδρομικές ερωτήσεις μπορούν να γίνουν και κατά την αντίστροφη κατεύθυνση, δηλαδή θέτοντας ως αρχικό τον κόμβο **Article'Cons95'Body'Chapter2** στο σχήμα 4.8, το αποτέλεσμα είναι οι κόμβοι **Article'Cons95'Body** και **Article'Cons95**. Επίσης θα μπορούσε η αναδρομή να γίνει και ως προς τις δυο κατευθύνσεις, έτσι θέτοντας ως αρχικό τον κόμβο **Article'Cons95'Body'Chapter1** το αποτέλεσμα είναι οι κόμβοι **Article'Cons95, Article'Cons95'Body** και οι κόμβοι **Article'Cons95'Body'Chapter1'section1, Article'Cons95'Body'Chapter1'section2, Article'Cons95'Body'Chapter1'section3**. Η διάσχιση του δικτύου προς οποιαδήποτε κατεύθυνση έχει το ίδιο κόστος αφού στην TELOS όλοι οι σύνδεσμοι είναι διπλής κατεύθυνσης.

Όπως αναφέρθηκε παραπάνω οι αναδρομικές ερωτήσεις ομαδοποιούνται κάτω από την επιλογή *Views* και το αποτέλεσμά τους παρουσιάζεται σε γραφική μορφή, όπως φαίνεται στο σχήμα 6.2. Στο μοντέλο οι αναδρομικές ερωτήσεις είναι τιμές στο γνώρισμα *viewMenu* της **MenuDescription**.

Οι αναδρομικές προκαθορισμένες ερωτήσεις, όπως φαίνονται και στο σχήμα 6.2, είναι οι εξής:

- *Narrow Term Tree* : Η ερώτηση αυτή παίρνει ως όρισμα ένα όρο του θεματικού θησαυρού και επιστρέφει την ιεραρχία που αποτελείται από τους στενότερους όρους του. Δηλαδή διασχίζει προς την αντίθετη κατεύθυνση την κατηγορία *has_broader_term*.

- *Broader Term Tree* : Η ερώτηση αυτή παίρνει ως όρισμα ένα όρο του θεματικού θησαυρού και επιστρέφει την ιεραρχία που αποτελείται από τους ευρήτερους όρους του. Δηλαδή διασχίζει προς την ορθή φορά την κατηγορία *has_broader_term*.

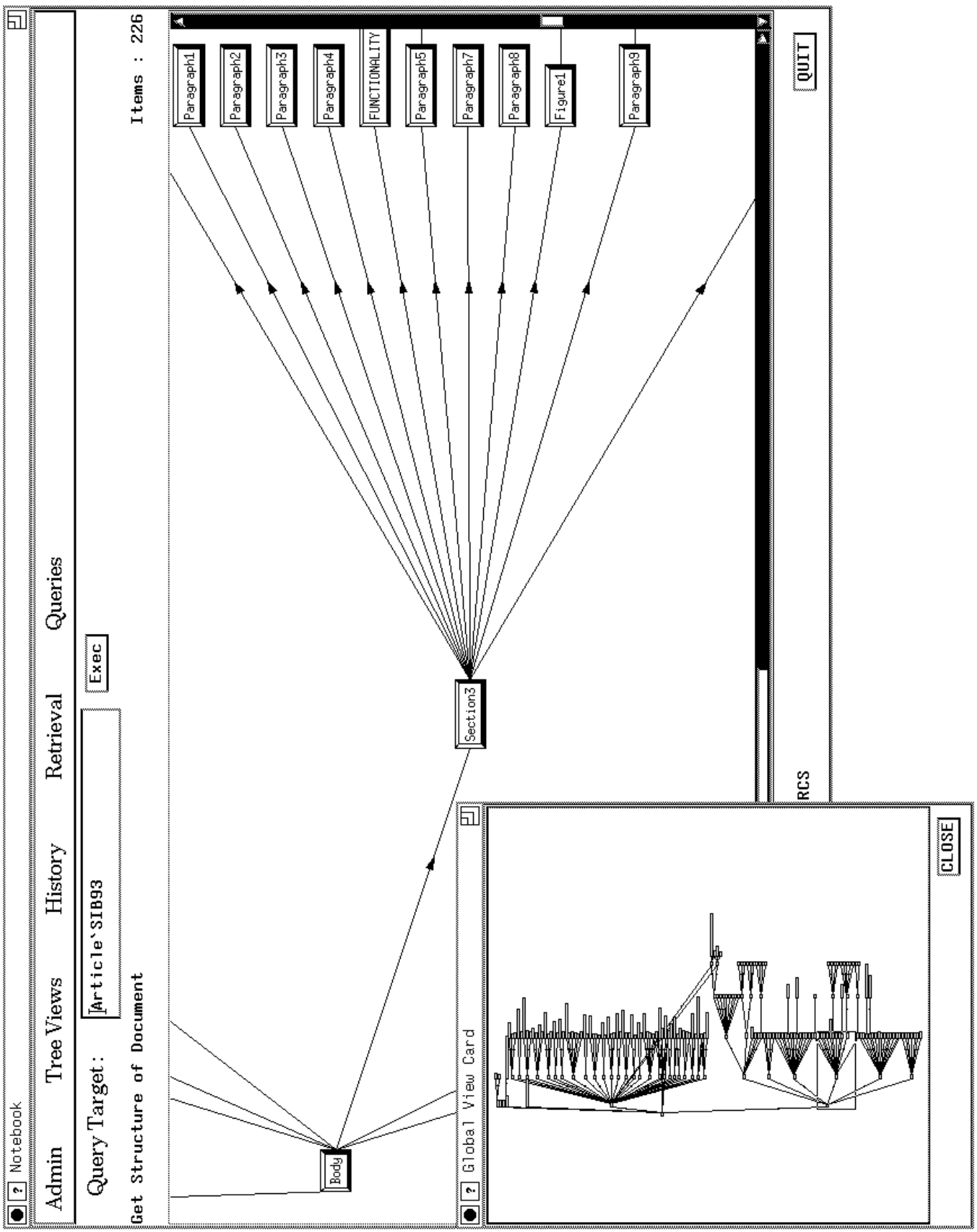


Σχήμα 6.7: Τα λογικά τμήματα του *Article*

- *Work Title Tree* : Η ερώτηση αυτή παίρνει ως όρισμα ένα τίτλο εργασίας και επιστρέφει την ιεραρχία που αποτελείται από τις υποεργασίες της (ένα παράδειγμα φαίνεται στην εικόνα 5.11). Δηλαδή διασχίζει την κατηγορία *divided_into* της κλάσης *WorkTitle*.

- *Work Titles of Collection* : Η ερώτηση αυτή παίρνει ως όρισμα μία συλλογή εργασιών και επιστρέφει τον γράφο που δείχνει τις μεταξύ τους συσχετίσεις. Η ερώτηση διασχίζει τις κατηγορίες *divided_into* και *relate_work* της κλάσης *WorkTitle*. Το μειονέκτημα του γράφου αυτού είναι ότι δεν παρουσιάζει την κατηγορία του συνδέσμου για να γνωρίζουμε ποια είναι η σχέση μεταξύ των εργασιών.

- *Get Structure of Document Type* : Η ερώτηση αυτή παίρνει ως όρισμα ένα τύπο εγγράφου ή τμήματος του και και επιστρέφει τα τμήματα από τα οποία αποτελείται καθώς και την ιεραρχία εξειδίκευσης των τμημάτων αυτών. Στην εικόνα 6.7 φαίνεται η εκτέλεση αυτής της ερώτησης για τη κλάση *Article*.



Σχήμα 6.8: Η λογική δομή του άρθρου SIB

Στην εικόνα φαίνεται η λογική δομή του άρθρου SIB. Με την βοήθεια της *Γενικής Όψης (Global View)* μπορούμε να κινούμαστε στο γράφο που παρουσιάζεται στην επαφή χρήσης (interface).

- *Get Structure of Document* : Η ερώτηση αυτή παίρνει ως όρισμα μία περίπτωση εγγράφου ή τμήματος του και επιστρέφει όλα τα τμήματα από τα οποία αυτό αποτελείται καθώς και τα τμήματα στα οποία αυτά αναφέρονται (δηλαδή συνδέονται με συνδέσεις της κατηγορίας *refer* του *GeneralDocument*). Συγκεκριμένα διασχίζει προς την ορθή φορά την μετακατηγορία *consist_of*. Με την βοήθεια αυτής της ερώτησης μπορούμε να περιπλανηθούμε στα λογικά τμήματα του εγγράφου. Ένα παράδειγμα εκτέλεσης της ερώτησης φαίνεται στο σχήμα 6.8.

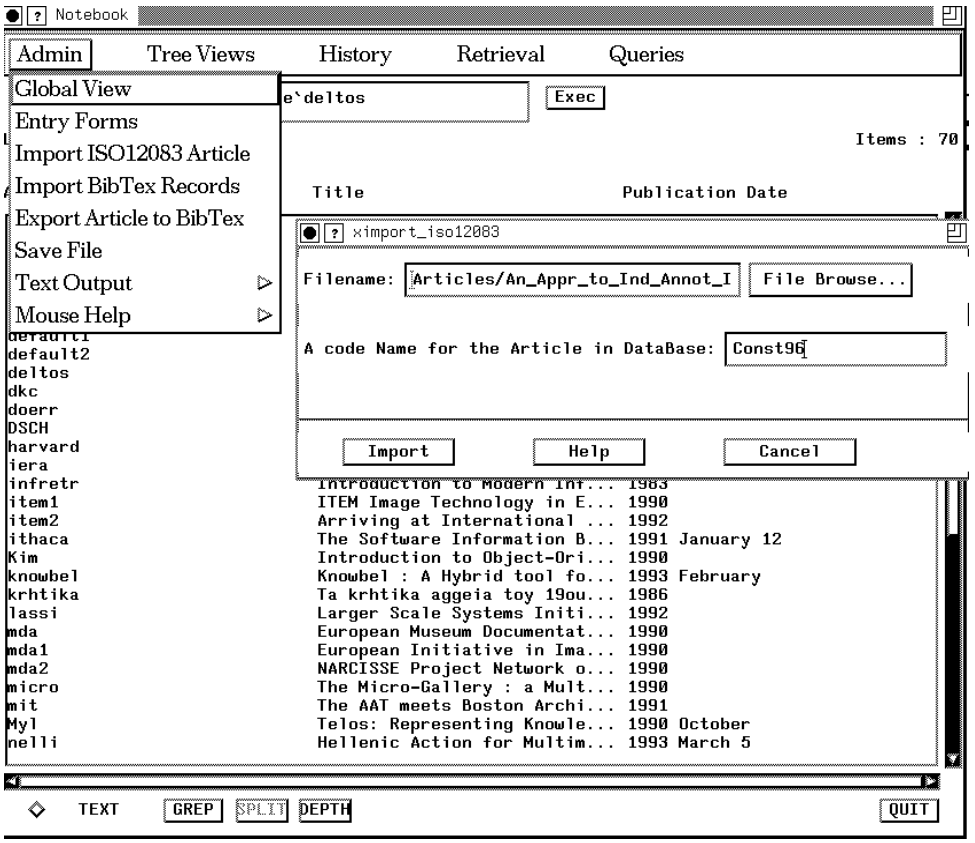
- *Get Annotations of Document* : Η ερώτηση αυτή παίρνει ως όρισμα μία περίπτωση εγγράφου ή τμήματος του και επιστρέφει όλες τις σημειώσεις που έχουν γίνει σε αυτό ή σε κάποιο τμήμα του. Δηλαδή διασχίζει τις κατηγορίες σημειώσεων (βλ. ενότητα 5.3) και την μετακατηγορία *consist_of*. Το μειονέκτημα του γράφου αυτού είναι ότι δεν παρουσιάζει την κατηγορία του συνδέσμου για να γνωρίζουμε ποια είναι η η οπτική γωνία κάτω από την οποία έγινε η σημείωσης.

Οι αναδρομικές ερωτήσεις γίνονται με βάση τις κύριες κατηγορίες αφηρημένων γνωρισμάτων που περιγράφηκαν στο κεφάλαιο 4 και 5. Οι κατηγορίες που χρησιμοποιούνται στις ερωτήσεις αυτού του είδους πρέπει να έχουν τη μεταβατική ιδιότητα για να έχουν και νόημα οι γράφοι οι οποίοι παρουσιάζουν τα αποτελέσματα. Συνήθως τις χρησιμοποιούμε για να διευκολύνουμε την περιπλάνηση στο χώρο της πληροφορίας (π.χ. για την περιπλάνηση στα τμήματα του εγγράφου και στις σημειώσεις που έχουν γίνει σε αυτά). Ένα σημαντικό μειονέκτημα των γράφων αυτών είναι ότι ορισμένες φορές απαιτείται για την κατανόηση του να υπάρχουν και τα ονόματα των συνδέσμων ή των κατηγοριών τους (π.χ. στην ερώτηση *Get Annotations of Document*). Στο μέλλον προβλέπεται να υπάρχει ένας μηχανισμός για να προσδιορίζει, ο δημιουργός της ερώτησης, ποιοί σύνδεσμοι να παρουσιάζονται με το όνομα τους ή το όνομα της κατηγορίας τους.

6.5 Τα εργαλεία υποστήριξης του συστήματος

Τα εργαλεία υποστήριξης του συστήματος (βλ. εικόνα 6.9) χρησιμοποιούνται για την εισαγωγή και εξαγωγή δεδομένων στο σύστημα καθώς και για την διαχείριση της επαφής χρήσης. Συγκεκριμένα υπάρχουν τα εξής εργαλεία :

- *Global View* : παρουσιάζει σε σμίκρυνση τον γράφο που υπάρχει στην περιοχή εμφάνισης αποτελεσμάτων της επαφής χρήσης (interface) και βοηθάει στην μετακίνηση στο κομμάτι του γράφου που φαίνεται στην επαφή χρήσης (βλ. εικόνα 6.8).



Σχήμα 6.9: Τα εργαλεία υποστήριξης του συστήματος

Στην εικόνα φαίνονται τα εργαλεία που χρησιμοποιούνται για την εισαγωγή και εξαγωγή δεδομένων στο σύστημα καθώς και για την διαχείριση της επαφής χρήσης.

- *Entry Forms* : Η επιλογή αυτή εκτελεί την κλήση του εργαλείου EntryForms (Δελτία Ενημέρωσης) για την ενημέρωση των δεδομένων της βάσης, όπως παρουσιάστηκε στις ενότητες 4.4.2 και 5.5.

- *Import ISO12083 Article* : Η επιλογή αυτή εκτελεί την κλήση του εργαλείου για την αυτόματη εισαγωγή ενός άρθρου, που ακολουθεί το πρότυπο ISO 12083, στη βάση δεδομένων. Το εργαλείο περιγράφεται στην ενότητα 4.4.1 και η επαφή χρήσης του φαίνεται στην εικόνα 6.9.

- *Import BibTex Records* : Η επιλογή αυτή εκτελεί την κλήση του προγράμματος για την αυτόματη εισαγωγή βιβλιογραφικών δεδομένων από την μορφή BIBTEX στη μορφή παράστασης άρθρων που υλοποιήσαμε στο μοντέλο μας, όπως παρουσιάστηκε στην ενότητα 4.4.1.

- *Export Article to BibTex* : Η επιλογή αυτή εκτελεί την κλήση του προγράμματος για την αυτόματη δημιουργία της BIBTEX εγγραφής (record) ενός άρθρου που υπάρχει στην βάση.

- *Save File* : Η επιλογή αυτή αποθηκεύει σε ένα αρχείο, που επιλέγουμε, τα αποτελέσματα της ερώτησης που εκτελέσαμε. Τα αρχεία αυτά μπορούμε να τα χρησιμοποιήσουμε για την συγγραφή αναφορών.

Κεφάλαιο 7

Συμπεράσματα και μελλοντικές κατευθύνσεις

Στα πλαίσια της παρούσας εργασίας σχεδιάσαμε και υλοποιήσαμε ένα πληροφοριακό σύστημα που υποστηρίζει την διαχείριση εγγράφων και την δημιουργία σημειώσεων σε αυτά. Το μοντέλο παράστασης των πληροφοριών υλοποιήθηκε στη γλώσσα παράστασης γνώσης Telos, η αναζήτηση των πληροφοριών γίνεται μέσω του Σημασιολογικού Συστήματος Ευρετηριασμού (ΣΣΕ) και η ενημέρωση της βάσης γίνεται είτε με αυτόματο τρόπο είτε με την χρήση των Δελτίων Ενημέρωσης (Entry Forms). Στο κεφάλαιο αυτό θα αξιολογήσουμε τη λειτουργία του συστήματος και θα αναφερθούμε στις μελλοντικές του προεκτάσεις.

7.1 Η λειτουργία του συστήματος

Για να γίνει η αξιολόγηση του συστήματος θα πρέπει αρχικά να προσδιορίσουμε τις εργασίες που γίνονται με το σύστημα και την ροή εκτέλεσης των εργασιών αυτών. Σε αυτήν την ενότητα προσδιορίζουμε τις εργασίες του συστήματος που θα αποτελέσουν το σημείο αναφοράς για την αξιολόγηση. Συγκεκριμένα οι κύριες εργασίες είναι: α) η εισαγωγή εγγράφων, β) η δημιουργία σημειώσεων και γ) η αναζήτηση εγγράφων και σημειώσεων που βρίσκονται στη βάση δεδομένων.

Η εισαγωγή άρθρων, που περιγράφεται στο 4.4.2, γίνεται είτε με αυτόματο τρόπο είτε με την χρήση των Δελτίων Ενημέρωσης (Entry Forms). Αν το άρθρο είναι διαθέσιμο σε ηλεκτρονική μορφή σύμφωνα με το πρότυπο ISO12083 ή αν τα βιβλιογραφικά του στοιχεία είναι σε BIBTEX (ή σε κάποια άλλη μορφή για την οποία υπάρχει φίλτρο σε BIBTEX), τότε η εισαγωγή των πληροφοριών για το άρθρο γίνεται αυτόματα. Επίσης στο μέλλον

προβλέπεται η αυτόματη εισαγωγή εγγράφων που περιγράφονται με ένα SGML DTD[1]. Ενώ στη περίπτωση που η εισαγωγή άρθρων δεν μπορεί να γίνει αυτόματα τότε χρησιμοποιούμε τα Δελτία Ενημέρωσης. Με τον τρόπο αυτό μπορούμε να ανοίξουμε ένα δελτίο για να ενημερώσουμε τα βιβλιογραφικά στοιχεία του άρθρου και έπειτα από αυτό το δελτίο να ανοίξουμε άλλα δελτία για να ενημερώσουμε την λογική δομή του. Επομένως η εισαγωγή των εγγράφων γίνεται με κατευθυνόμενο τρόπο (δηλ. παρέχεται βοήθεια) και υπάρχει έλεγχος για την τήρηση των περιορισμών ακεραιότητας. Επίσης αξίζει να διαθέσουμε χρόνο για την εισαγωγή ενός εγγράφου, μέσω των Δελτίων Ενημέρωσης, διότι η εισαγωγή θα γίνει μία μόνο φορά και από ένα μόνο άτομο, ενώ έπειτα θα το έχουν στη διάθεσή τους όλα τα άτομα που χρησιμοποιούν το σύστημα.

Αφού γίνει η εισαγωγή εγγράφων στη βάση τότε η επόμενη εργασία είναι η δημιουργία σημειώσεων σε αυτά. Η εισαγωγή σημειώσεων, όπως παρουσιάζεται στην ενότητα 5.5, γίνεται μέσω των Δελτίων Ενημέρωσης με μια από τις δυο διαδικασίες που παρέχουμε. Δηλαδή είτε διαλέγουμε ένα λογικό τμήμα ενός εγγράφου και κάνουμε σημειώσεις σε αυτό είτε διαλέγουμε ένα έγγραφο και καθώς γίνεται η πλοήγηση (navigate) στα λογικά του τμήματα κάνουμε σημειώσεις σε αυτά (βλ. εικόνες 5.12, 5.13). Τα βασικά σημεία που πρέπει να προσέξουμε για την γρήγορη και εύστοχη δημιουργία σημειώσεων είναι τα εξής:

- Οι κατηγορίες σημειώσεων. Οι νέοι χρήστες θα πρέπει να μελετήσουν τις κατηγορίες κατάταξης των σημειώσεων[2], όπως αυτές παρουσιάζονται στην ενότητα 5.3, ώστε να αποφευχθούν παρερμηνείες. Ενώ οι έμπειροι χρήστες του συστήματος έχουν την δυνατότητα να προτείνουν νέες κατηγορίες σημειώσεων, αφού το Telos μοντέλο είναι επεκτάσιμο, που να καλύπτουν τις ανάγκες για καλύτερη ταξινόμηση των σημειώσεων.

- Τα ελεγχόμενα λεξιλόγια όρων. Οι όροι των λεξιλογίων αυτών εκφράζουν, πολύ συχνά, το περιεχόμενο της σημείωσης που θέλουμε να κάνουμε (βλ. ενότητα 5.4). Για τον λόγο αυτό θα πρέπει να είναι καλά οργανωμένα, δηλαδή να υπάρχουν μεταξύ των όρων συσχετίσεις, να προσθέτουμε νέους όρους που κάνουν πιο εκφραστικό το λεξιλόγιο και να δημιουργούμε συλλογές από όρους (βλ. ενότητα 5.4.2) για να βρίσκουμε πιο εύκολα αυτούς που μας ενδιαφέρουν.

Τα δυο αυτά βασικά σημεία είναι ουσιαστικά και η διαδικασία δημιουργίας της σημείωσης, δηλαδή πρώτα επιλέγουμε μια κατηγορία σημείωσης και έπειτα προσδιο-

[1] Εργασία που βρίσκεται σε εξέλιξη στην Ομάδα Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του Ινστιτούτου Πληροφορικής του ΙΤΕ

[2] Τα ονόματα των κατηγοριών είναι πολύ εκφραστικά, αλλά είναι προτιμότερο να μελετήσουμε μία τουλάχιστον φορά την σημασία των κατηγοριών για να αποφύγουμε παρερμηνείες.

ρίζουμε το περιεχόμενο της σημείωσης (ή ένας όρος του ελεγχόμενου λεξιλογίου ή ένα τμήμα υπαρκτού εγγράφου ή ένα νέο έγγραφο που γράφουμε για να εκφράσουμε την σημείωση).

Οι σημειώσεις που γίνονται στα έγγραφα αποτελούν μια πολύτιμη πληροφορία την οποία θα αναζητήσουμε όταν θα την χρειαστούμε. Η αναζήτηση της πληροφορίας, που περιγράψαμε στο κεφάλαιο 6, γίνεται είτε με ερωτήσεις που παίρνουν μια παράμετρο είτε με μια κάρτα αναζήτησης που αποτελείται από πολλές ερωτήσεις οι οποίες παίρνουν ένα λογικό συνδυασμό παραμέτρων (λογικό ''και'' και ''ή''). Ενώ τα αποτελέσματα παρουσιάζονται με την μορφή κειμένου ή γράφου. Συγκεκριμένα τα αποτελέσματα των ερωτήσεων, για την αναζήτηση σημειώσεων, παρουσιάζονται με την μορφή προτάσεων (βλ. ενότητα 5.6), ώστε να είναι πιο κατανοητά.

Όμως οι εργασίες που περιγράψαμε γίνονται σε μια βάση δεδομένων που ίσως να μην ανήκει σε ένα μόνο χρήστη. Η βάση δεδομένων μπορεί να είναι ατομική ή να ανήκει σε μια (μικρή ή μεγάλη) ομάδα ατόμων. Έτσι δημιουργούνται προβλήματα δικαιοδοσίας στα δεδομένα (δηλαδή τί είδους πρόσβαση έχει κάποιος στα δεδομένα). Επομένως για την αλλαγή ή εισαγωγή δεδομένων θα πρέπει να γίνεται αναγνώριση της ταυτότητας του χρήστη ώστε να ελέγχεται η πρόσβαση στα δεδομένα και να γίνονται κάποιοι αυτοματισμοί (π.χ. όταν εισάγεται μια σημείωση τότε αυτόματα να εισάγεται το όνομα του σχολιαστή ως γνώρισμα της). Επίσης κατά τη αναζήτηση, σε μια βάση όπου κάνουν σημειώσεις περισσότερα από ένα άτομα, θα πρέπει να δηλώνουμε και το όνομα του σχολιαστή που μας ενδιαφέρει (δηλ. αναγκαζόμαστε να χρησιμοποιήσουμε τις κάρτες αναζήτησης γιατί οι άλλες ερωτήσεις παίρνουν μία μόνο παράμετρο).

Στην επόμενη ενότητα γίνεται η αξιολόγηση του συστήματος για μια βάση στην οποία έχουν πρόσβαση ανάγνωσης και εγγραφής τα 5 μέλη μιας ομάδας συνεργατών.

7.2 Αξιολόγηση του συστήματος

Κατά την διάρκεια σχεδίασης και υλοποίησης του συστήματος δεχθήκαμε την κριτική συνεργατών που το χρησιμοποίησαν και η κριτική αυτή οδήγησε στην επανασχεδίαση του. Έτσι η ανάπτυξη του συστήματος ακολούθησε μια επαναληπτική διαδικασία αξιολόγησης−επανασχεδίασης. Χρησιμοποιήσαμε 10 χρήστες από τους οποίους οι 5 γνώριζαν το γενικό περιβάλλον του ΣΣΕ ενώ οι άλλοι 5 δεν γνώριζαν αυτό το περιβάλλον. Για να γίνει η αξιολόγηση του συστήματος ο κάθε χρήστης ακολουθεί ένα πλάνο εργασίων και συγχρόνως κάνει ορισμένες παρατηρήσεις σχετικά με αυτές. Το πλάνο αυτό

περιλαμβάνει την εισαγωγή άρθρων και την ταυτόχρονη δημιουργία σημειώσεων σε αυτά, τη δημιουργία σημειώσεων σε άρθρα που υπάρχουν ήδη στη βάση και την αναζήτηση πληροφοριών σε μια βάση όπου υπάρχουν σημειώσεις από τα μέλη μιας ομάδας. Ο κάθε χρήστης έκανε εισαγωγή 3 εως 6 άρθρων, που είχε μελετήσει, καθώς και των σημειώσεων που είχε κάνει σε αυτά. Έπειτα αναζήτησε άρθρα μέσω των προκαθορισμένων ερωτήσεων και είδε τις σημειώσεις των άλλων χρηστών. Κατά την διάρκεια των εργασιών αυτών, οι χρήστες έκαναν παρατηρήσεις για την λειτουργία του συστήματος και τις δυνατότητές του. Ενώ στο τέλος χρησιμοποιήσαμε και τα ερωτηματολόγια αξιολόγησης της ευχρηστίας της επαφής χρήσης, που περιγράφονται στο [RJ89]. Οι ερωτήσεις που χρησιμοποιήσαμε παρουσιάζονται στο παράρτημα Α και τα τελικά αποτελέσματα για τα δύο διαφορετικά είδη χρηστών στους πίνακες 7.1 και 7.2.

Επομένως η ευχρηστία και λειτουργικότητα του συστήματος κυμαίνεται από μέτρια έως πολύ ικανοποιητική, για τους χρήστες που δεν γνωρίζουν το Σημασιολογικό Σύστημα Ευρετηριασμού (ΣΣΕ), και είναι πολύ ικανοποιητική για αυτούς που το γνωρίζουν. Δηλαδή η ευχρηστία του συστήματος γίνεται πολύ ικανοποιητική μετά την απόκτηση εμπειρίας για την χρήση του.

Επίσης θα πρέπει να τονίσουμε την ιδιαίτερη ικανοποίηση των χρηστών από τα εξής:

- Η αυτόματη εισαγωγή της βιβλιογραφίας, της λογικής δομής εγγράφων και η διασύνδεση με το περιεχόμενο τους.
- Η καθοδήγηση και πρόληψη λαθών κατά την εισαγωγή δεδομένων μέσω των Δελτίων Ενημέρωσης (entry forms). Ο πειθαρχημένος τρόπος δημιουργίας σημειώσεων βοηθάει στην δημιουργία κατανοητών σημειώσεων καθώς και στην αναζήτηση τους.
- Η διάθεση προκαθορισμένων ερωτήσεων που καλύπτουν τις απαιτήσεις για την αναζήτηση πληροφοριών αντί για την παροχή μόνο μιας ερωτηματικής γλώσσας.
- Η παρουσίαση των αποτελεσμάτων με την μορφή γράφων, που διευκολύνουν την κατανόηση και την περιπλάνηση (navigate) στο χώρο της πληροφορίας.

Αλλά από τις εργασίες των χρηστών και τις συζητήσεις μαζί τους προσδιορίστηκαν και οι εξής αδυναμίες του συστήματος:

- Στην εισαγωγή σημειώσεων δεν εισάγεται αυτόματα το όνομα του σχολιαστή ούτε η ημερομηνία σχολιασμού. Η αδυναμία αυτή υπάρχει γιατί το σύστημα των Δελτιών Ενημέρωσης δεν δίνει την δυνατότητα έμμεσων ενημερώσεων της βάσης, δηλαδή

	Πολύ ικανοποιητική	Μέτρια ικανοποιητική	Καθόλου ικανοποιητική
Σαφήνεια παρουσίασης	60%	40%	
Συνέπεια	80%	20%	
Ενδείξεις για την πορεία του συστήματος	80%	20%	
Κατανοητή λειτουργία του συστήματος	80%	20%	
Πρόληψη και αποφυγή λαθών	60%	40%	

Πίνακας 7.1: Αξιολόγηση από χρήστες που **δεν** γνώριζαν το περιβάλλον του ΣΣΕ.
Στο πίνακα φαίνεται το ποσοστό των χρηστών ανάλογα με τον βαθμό ικανοποίησης τους για τις διάφορες οπτικές γωνίες αξιολόγησης του συστήματος. Η αξιολόγηση έγινε από 5 χρήστες.

	Πολύ ικανοποιητική	Μέτρια ικανοποιητική	Καθόλου ικανοποιητική
Σαφήνεια παρουσίασης	100%		
Συνέπεια	100%		
Ενδείξεις για την πορεία του συστήματος	100%		
Κατανοητή λειτουργία του συστήματος	100%		
Πρόληψη και αποφυγή λαθών	80%	20%	

Πίνακας 7.2: Αξιολόγηση από χρήστες που γνώριζαν το περιβάλλον του ΣΣΕ.
Στο πίνακα φαίνεται το ποσοστό των χρηστών ανάλογα με τον βαθμό ικανοποίησης τους για τις διάφορες οπτικές γωνίες αξιολόγησης του συστήματος. Η αξιολόγηση έγινε από 5 χρήστες.

μια αλλαγή στα δεδομένα δεν ενεργοποιεί κάποιο πρόγραμμα που θα κάνει κάποια ενημέρωση της βάσης.

- Σε ένα έγγραφο ή τμήμα εγγράφου που εντοπίζουμε στο σύστημα παρουσίασης (δηλ. η επαφή χρήσης του ΣΣΕ) δεν μπορούμε να κάνουμε απευθείας σημειώσεις (δηλ. να το επιλέξουμε και να βγάλει την κάρτα εισαγωγής σημειώσεων), αλλά θα πρέπει να πάμε μέσω των Δελτίων Ενημέρωσης. Αυτό είναι μέρος ενός γενικότερου προβλήματος εισαγωγής δεδομένων στο ΣΣΕ, για το οποίο γίνονται μελέτες αντιμετώπισης του.

- Οι αναζητήσεις πληροφοριών (ερωτήσεις 1ης τάξεως και αναδρομικές) γίνονται σε όλα τα δεδομένα της βάσης και δεν υπάρχει δυνατότητα να γίνουν πάνω σε όψεις (views). Αυτή η αδυναμία οφείλεται στην έλλειψη μηχανισμών διαχείρισης όψεων στη Telos. Ενώ αντιμετωπίζεται εν μέρη με την χρήση των Καρτών Αναζήτησης (Retrieval Card).

Επίσης συγκρίναμε το σύστημα υπομνηματισμού που υλοποιήσαμε με άλλα παρόμοια συστήματα που μελετήσαμε. Η σύγκριση έγινε ως προς ορισμένα βασικά χαρακτηριστικά τους, όπως φαίνεται στον πίνακα 7.3. Το σύμβολο - σημαίνει ότι το σύστημα δεν παρέχει την αντίστοιχη λειτουργία, το σύμβολο $\surd$ σημαίνει ότι το σύστημα παρέχει την αντίστοιχη λειτουργία και το σύμβολο + σημαίνει ότι το σύστημα παρέχει πολύ ικανοποιητικά την αντίστοιχη λειτουργία. Όπως φαίνεται από τον πίνακα το σύστημα που υλοποιήσαμε με το SIS υπερέχει στον προσδιορισμό κατηγοριών σημειώσεων και άρα οργανώνει και αναζητεί αποδοτικότερα τις σημειώσεις στα έγγραφα. Ενώ υστερεί στο ότι δεν έχει υλοποιηθεί η επαφή χρήσης του για το World Wide Web (WWW). Επίσης θα πρέπει να τονίσουμε ότι όλα τα συστήματα που χρησιμοποιούνται μέσω του WWW έχουν ως κύριο στόχο τους την δημιουργία σημειώσεων σε ηλεκτρονικά έγγραφα, ενώ το δικό μας σύστημα ικανοποιεί και την ανάγκη δημιουργίας σημειώσεων σε μη ηλεκτρονικά έγγραφα, αρκεί να έχουμε την πληροφορία για την λογική δομή τους.

Τέλος, δεν κάναμε μετρήσεις για την ταχύτητα απόκρισης[3] διότι αυτές στην ουσία θα αξιολογούσαν όχι την παρούσα εφαρμογή, αλλά τις επιδόσεις του ΣΣΕ.

[3]Χρόνος απόκρισης είναι το χρονικό διάστημα που μεσολαβεί μεταξύ της ενεργοποίησης της ερώτησης και της παρουσίασης των αποτελεσμάτων στην οθόνη

	Συστήματα							
Χαρακτηριστικά	SIS	ComMentor	CoNote	P.C.	HyperNews	NetQ	BiblioText	Tapestry
Κατηγορίες σημείωσης	++	-	-	+	+	-	-	-
Αναζήτηση Σημειώσεων	+	√	√	-	-	-	√	+
Αναζήτηση Εγγράφων	+	√	-	-	-	-	√	+
Ομάδες συνεργατών	√	+	+	-	-	-	√	-
Ομάδες εγγράφων	√	√	+	+	-	√	√	-
Εξουσιοδοτήσεις χρηστών	√	+	+	+	+	-	√	√
Σχολιασμός σημείωσης	+	√	+	+	+	-	+	+
Θεματική ταξινόμηση	+	-	-	-	-	-	√	√
Κριτική εγγράφων	+	√	√	+	+	+	-	√
Χρήση μέσω του WWW	-	+	+	+	+	+	-	-

Πίνακας 7.3: Σύγκριση συστημάτων υπομνηματισμού.

Η σύγκριση γίνεται ως προς ορισμένα βασικά χαρακτηριστικά που προσδιορίζουν τις δυνατότητες των συστημάτων υπομνηματισμού. Το P.C. είναι η συντομογραφία του Principia Cybernetica.

- : Δεν παρέχει την λειτουργία.

√ : Παρέχει την λειτουργία.

+ : Παρέχει πολύ ικανοποιητικά την λειτουργία.

7.3 Μελλοντικές επεκτάσεις

Μια άμεση επέκταση του συστήματος είναι να εισάγουμε στο σύστημα νέους τύπους εγγράφων (π.χ. HTML, Postscript, κ.α.). Δηλαδή να ορίσουμε την λογική δομή τους και να δημιουργήσουμε μεταφραστές (parsers) για την αυτόματη εξαγωγή της λογικής δομής και την διασύνδεση της με το περιεχόμενο του εγγράφου (π.χ με εξωτερικά εργαλεία παρουσιάζεται το περιεχόμενό του). Όσον αφορά τα SGML έγγραφα, βρίσκεται σε εξέλιξη μια εργασία[4] [Τρα96b] για την δημιουργία ενός μοντέλου περιγραφής του, το οποίο θα δώσει την δυνατότητα διαχείρισης των εγγράφων με το Σημασιολογικό Σύστημα Ευρετηριασμού (ΣΣΕ). Ενώ για έγγραφα που περιγράφονται με κάποια γλώσσα περιγραφής σελίδας (page description language), π.χ. Postscript, θα μπορούσαν να δημιουργηθούν αλγόριθμοί για την εξαγωγή της λογικής δομής του εγγράφου όπως φαίνεται από μελέτες που έχουν γίνει [HI93], [FNK92].

Επίσης θα μπορούσαμε να βελτιώσουμε την διαχείριση των δεδομένων δίνοντας την δυνατότητα ορισμού όψεων (views) της βάσης δεδομένων. Με τον τρόπο αυτό θα μπορούσαμε να κατευθύνουμε τις αναζητήσεις πληροφορίας σε μία ή περισσότερες όψεις της βάσης, π.χ. αν η βάση ανήκει σε μια ομάδα τότε ορίζουμε για κάθε μέλος της μια όψη που περιλαμβάνει μόνο τις δικές του σημειώσεις και τα έγγραφα στα οποία έχουν γίνει. Η επιλογή μιας ή περισσότερων όψεων θα γίνεται από την επαφή χρήσης (interface) του ΣΣΕ. Ακόμη μια βελτίωση προς την ίδια κατεύθυνση είναι να μπορούμε να εκτελέσουμε τις αναζητήσεις σε δύο ή περισσότερες βάσεις (ή όψεις διαφορετικών βάσεων) συγχρόνως. Με τον τρόπο αυτό θα μπορούν διαφορετικές ομάδες να ανταλλάξουν και να συγκρίνουν τις σημειώσεις που έχουν κάνει, αλλά και να αναζητούν έγγραφα μέσω των σημειώσεων αυτών.

Μια άλλη επέκταση του συστήματος είναι η δημιουργία ενός μοντέλου περιγραφής προσώπων. Η περιγραφή προσώπων θα περιλαμβάνει πληροφορίες για τις ικανότητες τους, τις γνώσεις τους (τυπικές και ουσιαστικές) και τις εργασίες που έχουν κάνει καθώς και τις κριτικές άλλων ατόμων για αυτούς. Με τον τρόπο αυτό θα μπορούμε να προσδιορίζουμε την ποιότητα ενός σχολιαστή χωρίς να τον γνωρίζουμε προσωπικά, π.χ. όταν έχει φύγει από μια ομάδα αλλά οι σημειώσεις του έχουν παραμείνει ή αν πρόκειται για άτομο μιας άλλης ομάδας. Έτσι θα μπορούμε να αξιολογήσουμε και να αποφασίσουμε αν μας ενδιαφέρουν οι σημειώσεις που έχει κάνει στα έγγραφα.

Επίσης μια άλλη βελτίωση είναι η δημιουργία ενός μηχανισμού για την καλύτερη παρουσίαση του περιεχομένου του εγγράφου. Στην παρούσα κατάσταση χρησιμοποιούμε

[4] Η μεταπτυχιακή εργασία της Αθηνάς Τραψιώτη.

τον *vi* επεξεργαστή κειμένου (ή κάποιον άλλο επεξεργαστή που ορίζει ο χρήστης) για την παρουσίαση του περιεχομένου ενός τμήματος του εγγράφου. Αλλά θα μπορούσε να υλοποιηθεί μια αρχιτεκτονική πελάτη−εξυπηρετητή (client−server), όπου ο πελάτης (client) θα κάνει αιτήσεις για την παρουσίαση ενός τμήματος του εγγράφου με έναν επεξεργαστή και από την επιλογή *Admin* της επαφής χρήσης θα ορίζονται τα χαρακτηριστικά της παρουσίασης. Τα χαρακτηριστικά αυτά δηλώνουν αν θα εμφανίζονται όλα τα έγγραφα σε ένα μόνο επεξεργαστή ή αν για κάθε έγγραφο θα εμφανίζονται ένας ή περισσότεροι επεξεργαστές κειμένου. Με το τρόπο αυτό θα γίνεται η περιπλάνηση στο περιεχόμενο του εγγράφου σύμφωνα με τις απαιτήσεις του κάθε χρήστη.

Τέλος, η δημιουργία του μοντέλου υπομνηματισμού μας οδήγησε σε σκέψεις για τον ορισμό ενός πρότυπου μοντέλου υπομνηματισμού. Δηλαδή να προσδιοριστούν οι διαδικασίες δημιουργίας σημειώσεων, οι κατηγορίες σημειώσεων και οι τύποι του περιεχομένου των σημειώσεων. Με τον τρόπο αυτό θα μπορούμε να κατανοούμε τις σημειώσεις που έχουν κάνει άλλοι μελετητές και άρα να υποστηριχτεί η ανταλλαγή σημειώσεων και έξω από τα πλαίσια μιας ομάδας συνεργατών.

7.4 Επίλογος

Το σύστημα που περιγράψαμε χρησιμοποιείται από τον συγγραφέα καθώς και από δύο μεταπτυχιακούς φοιτητές[5], που εκπονούν την διδακτορική του διατριβή, για την διαχείριση της βιβλιογραφίας τους και των σημειώσεων σε αυτή. Επίσης προβλέπεται να χρησιμοποιηθεί και από τα μέλη της Ομάδας Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του Ινστιτούτου Πληροφορικής του ΙΤΕ. Ενώ, μετά από ένα διάστημα εσωτερικής χρήσης στην ομάδα, θα γίνει προσπάθεια προώθησης του συστήματος στην αγορά.

[5]Οι φοιτητές αυτοί είναι ο Μάνος Θεοδωράκης και ο Γιάννης Τζίτζικας.

Παράρτημα Α

Ερωτήσεις για την αξιολόγηση του συστήματος

Οι ερωτήσεις χωρίζονται ανάλογα με την οπτική γωνία αξιολόγησης του συστήματος. Ο απαντήσεις είναι μια από τις εξής: *Πάντα*, *Τις περισσότερες φορές*, *Μερικές φορές*, *Ποτέ*.

Σαφήνεια παρουσίασης

1. Έχει κάθε οθόνη ένα τίτλο ή μια περιγραφή;

2. Είναι εμφανές που θα δώσει τα δεδομένα ο χρήστης και σε τί μορφή;

3. Οι διαφορετικοί τύποι πληροφορίας διαχωρίζονται μεταξύ τους;

4. Η παρουσίαση της πληροφορίας χρησιμοποιεί χρώματα;

5. Αν χρησιμοποιεί χρώματα τότε είναι εύκολο να δούμε την πληροφορία και σε μονόχρωμη οθόνη;

6. Είναι εύκολο να βρούμε την πληροφορία που θέλουμε στην οθόνη;

Συνέπεια

1. Τα χρώματα χρησιμοποιούνται με την ίδια σημασία σε όλη τη λειτουργία του συστήματος;

2. Οι κωδικοί και τα αρχικά έχουν την ίδια σημασία σε όλη τη λειτουργία του συστήματος;

3. Τα εικονίδια και τα σύμβολα έχουν την ίδια σημασία σε όλη τη λειτουργία του συστήματος;

4. Ο ίδιος τύπος πληροφορίας παρουσιάζεται στην ίδια θέση και με τον ίδιο τρόπο;

5. Ο τρόπος εισαγωγής της πληροφορίας είναι ο ίδιος σε όλη τη λειτουργία του συστήματος;

6. Ο τρόπος που διαλέγουμε τις επιλογές (π.χ. μενού) είναι ο ίδιος σε όλη τη λειτουργία του συστήματος;

7. Ο τρόπος που το σύστημα απαντάει στις ενέργειες του χρήστη είναι συνεπής σε όλη τη λειτουργία του συστήματος;

Ενδείξεις για την πορεία του συστήματος

1. Τα μηνύματα που παρουσιάζει το σύστημα είναι κατάλληλα;

2. Είναι φανερό τί ενέργειες μπορεί να κάνει ο χρήστης κάθε στιγμή;

3. Είναι φανερό τί τύπου πληροφορία θα εισάγει ο χρήστης;

4. Είναι φανερό τί αλλαγές συμβαίνουν στην οθόνη ως αποτέλεσμα των ενεργειών του χρήστη;

5. Υπάρχουν μηνύματα προσδιορισμού της τρέχουσας κατάστασης;

6. Πληροφορεί το σύστημα για την αποτυχία ή επιτυχία της ενέργειας του χρήστη;

7. Πληροφορεί το σύστημα για την καθυστέρηση του αποτελέσματος που περιμένει ο χρήστης;

8. Είναι κατανοητά τα μηνύματα λάθους;

Κατανοητή λειτουργία του συστήματος

1. Είναι κατανοητή η κατάσταση του συστήματος κατά την διάρκεια μιας εργασίας;

2. Είναι κατανοητό τί πρέπει να κάνει ο χρήστης ώστε να ολοκληρώσει μια εργασία;

3. Όταν ο χρήστης έχει να διαλέξει από ένα σύνολο επιλογών, είναι κατανοητή η σημασία της κάθε μιας;

4. Είναι φανερό σε ποιό σημείο του συστήματος είναι ο χρήστης;
5. Είναι κατανοητή η οργάνωση και δομή του συστήματος;
6. Είναι κατανοητές οι κατηγορίες σημειώσεων που υποστηρίζει το σύστημα;
7. Είναι κατανοητή η παρουσίαση της πληροφορίας που ψάχνει ο χρήστης;
8. Είναι κατανοητός το τρόπος αναζήτησης της πληροφορίας με το σύστημα;
9. Είναι κατανοητός το τρόπος εισαγωγής της πληροφορίας;

Πρόληψη και αποφυγή λαθών

1. Το σύστημα επικυρώνει τα δεδομένα που εισάγει ο χρήστης πριν τα δώσει για επεξεργασία;
2. Το σύστημα πληροφορεί τον χρήστη όταν εντοπίζει ένα λάθος;
3. Σε κάθε σημείο της εργασίας υπάρχει τρόπος διαφυγής (cancel) ώστε να αποφευχθούν λάθος καταστάσεις;
4. Υπάρχει δυνατότητα επαναφοράς στην προηγούμενη κατάσταση (undo) του συστήματος;
5. Προστατεύεται το σύστημα από τα συνηθισμένα λάθη;
6. Για μια καταστροφική ενέργεια υπάρχει και δεύτερη επιβεβαίωση από τον χρήστη;
7. Το σύστημα αποτρέπει τους χρήστες από ενέργειες που δεν έχουν το δικαίωμα να κάνουν;
8. Για κάθε λάθος που συμβαίνει δίνει το σύστημα τις κατάλληλες πληροφορίες για τον εντοπισμό και την λύση του προβλήματος;

Παράρτημα Β

Το Document Type Definition (DTD) του ISO12083 για τα άρθρα

```
<!-- This is the ISO12083:1993 document type definition for an Article.
     It can be embedded in a Serial.                                    -->
<!-- Copyright: (C) International Organization for Standardization 1993.
Permission to copy in any form is granted for use with conforming SGML
systems and applications as defined in ISO 8879:1986, provided this
notice is included in all copies.                                       -->

<!-- ================================================================== -->
<!--    PUBLIC DOCUMENT TYPE DEFINITION SUBSET                          -->
<!-- ================================================================== -->

<!-- ================================================================== -->
<!--    Entity Naming Conventions                                       -->
<!-- ================================================================== -->

<!-- Prefix = where used:
     p.  = in paragraphs (also in phrases if .ph suffix)
     s.  = in sections (i.e., among paragraphs)
     i.  = where allowed by inclusion exceptions
     m.  = content model or declared content
     a.  = attribute definition
     NONE= specific use defined in models
```

```
    Suffix = allowed content:
     .ph = elements whose content is %m.ph;
     .d  = elements whose content has same model as defaults
     .zz = for subelements
     NONE= individually defined elements                                    -->

<!ENTITY % doctype   "article" -- default document type generic identifier -->


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Specialized Elements                                                -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ENTITY % ade.ph  "street|city|state|country|postcode|san|email|postbox|
                    phone|fax"  -- address elements                        -->
<!ENTITY % bib     "author|corpauth|msn|sertitle|location|date|pages|subject|
                    othinfo" -- bibliographic, date is the publication date -->

<!-- The following 4 declarations are specific to Articles                 -->
<!ENTITY % bmsec.d "ack|vita"             -- back matter cps elements      -->
<!ENTITY % fmsec.d "ded|abstract|supmatl" -- front matter sections         -->
<!ENTITY % pub     "sponsor|contract|reprint|cpyrt|date|pubname|location|
                    confgrp|avail|history"
                    -- pubfront, date is the publication date              -->
<!ENTITY % pub.ph  "fpage|lpage|coden|acqno|issn|acidfree|price|extent|artid|
                    category|pubid"
                    -- publication related front matter material          -->


<!-- ====================================================================== -->
<!--    Basic Document Elements                                             -->
<!-- ====================================================================== -->

<!ENTITY % i.float "figgrp|footnote|note" -- floating elements             -->
<!ENTITY % p.el    "deflist|orgaddr|indaddr|artwork|bq|lit|date|biblist|author
                    |corpauth|keyword|keyphras|poem|nameloc|indxflag"
                    -- general                                             -->
<!ENTITY % p.em.ph "emph" -- emphasis                                      -->
<!ENTITY % p.lst.d "list" -- list                                          -->
```

```
<!-- The following declaration is specific to Articles                    -->
<!ENTITY % p.rf.ph "noteref|fnoteref|figref|tableref|artref|appref|citeref|
                    secref|formref" -- references                         -->

<!ENTITY % p.tbl   "table" -- table matter                                 -->
<!ENTITY % p.form  "formula|dformula|dformgrp" -- mathematical formulas   -->
<!ENTITY % p.zz    "(%p.el;)|(%p.tbl;)|(%p.lst.d;)|(%p.form;)"
                    -- paragraph subelements                               -->

<!ENTITY % p.zz.ph "q|pages|(%p.em.ph;)|(%p.rf.ph;)" -- phrases           -->
<!ENTITY % s.zz    "p|(%p.zz;)" -- section subelements                    -->


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Models                                                            -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ENTITY % m.addr  "(%ade.ph;)*" -- address (no name)                     -->
<!ENTITY % m.bib   "(no?, title, (%bib;)*)" -- bibliographic entry        -->
<!ENTITY % m.copy  "(date|cpyrtnme|cpyrtclr)+" -- copyright notice data   -->
<!ENTITY % m.date  "(#PCDATA)" -- date                                    -->
<!ENTITY % m.fig   "EMPTY" -- default FIG content                         -->
<!ENTITY % m.sec   "(title?, (%s.zz;)*, section*)" -- section             -->
<!ENTITY % m.name  "((fname? & surname), (degree|school)*, role*, (%ade.ph;)*,
                    aff?)" -- name components                             -->
<!ENTITY % m.org   "(orgname, orgdiv*, %m.addr;)" -- organization name    -->
<!ENTITY % m.ph    "(#PCDATA|(%p.zz.ph;)|(%p.form;))*" -- phrase model    -->
<!ENTITY % m.pseq  "(p, (p|(%p.zz;))*)" -- P with sequence                -->
<!ENTITY % m.poem  "(stanza+|poemline+)" -- poetry sub-elements           -->

<!-- ================================================================== -->
<!--    THE DOCUMENT STRUCTURE                                          -->
<!-- ================================================================== -->

<!--       ELEMENT             MIN  CONTENT               (EXCEPTIONS)  -->
<!ELEMENT (%doctype;)          - -  (front, body, appmat?, back?)
                                                        +(%i.float;)      >
```

```
<!-- ================================================================== -->
<!--    FRONT MATTER ELEMENTS                                           -->
<!-- ================================================================== -->

<!-- The following declaration is specific to Articles                  -->
<!ELEMENT front               O O  (titlegrp, authgrp, date?, pubfront?,
                                    ((%fmsec.d;)|keyword|keyphras)*)      >

<!ELEMENT (%fmsec.d;)         - O  %m.sec;                                >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--   Title Group                                                      -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT titlegrp            O O  (title, subtitle*)                     >
<!ELEMENT (title|subtitle)    - O  %m.ph;                                 >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Author Group                                                    -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT authgrp             O O  (author|corpauth|aff)*                 >
<!ELEMENT author              - O  %m.name;                               >
<!ELEMENT (fname|surname|role|degree|orgname|orgdiv)
                              - O  (#PCDATA)                              >

<!ELEMENT (aff|corpauth|school)
                              - O  %m.org;                                >

<!ELEMENT (%ade.ph;)          - O  (#PCDATA)                              >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Publisher's Front Matter                                        -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT pubfront            - O  ((%pub;) | (%pub.ph;))*                >
```

```
<!ELEMENT (%pub.ph;|contract)  - O  (#PCDATA)                              >
<!ELEMENT (pubname|avail|sponsor)
                               - O  %m.org;                                >
<!ELEMENT reprint              - O  (%m.org;|%m.name;)                     >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--            Article's History                                         -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!-- The following 2 declarations are specific to Articles               -->
<!ELEMENT history              - O  (received|accepted|revised|misc)*      >
<!ELEMENT (received|accepted|revised)
                               - O  (date)                                 >
<!ELEMENT misc                 - O  (p*, date)                             >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Copyright                                                          -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT cpyrt                - -  %m.copy;                               >
<!ELEMENT cpyrtclr             - -  %m.org;                                >
<!ELEMENT cpyrtnme             - -  (%m.org;|%m.name;)                     >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Conference Group                                                   -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT confgrp              - -  (no?, confname, date?, location?,
                                    sponsor?)                              >
<!ELEMENT confname             - O  (#PCDATA)                              >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Date                                                               -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
```

```
<!ELEMENT date                - O  %m.date;                                  >


<!-- ================================================================== -->
<!--    BODY ELEMENTS                                                   -->
<!-- ================================================================== -->

<!-- +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Body Structure                                                  -->
<!-- +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT body                O O  (part+|chapter+)                        >
<!ELEMENT part                - O  (no?, title?, (%s.zz;)*, chapter+)      >
<!ELEMENT chapter             - O  (no?, %m.sec;)                          >
<!ELEMENT section             - O  (no?, title?, (%s.zz;)*, subsect1*)     >
<!ELEMENT subsect1            - O  (no?, title?, (%s.zz;)*, subsect2*)     >
<!ELEMENT subsect2            - O  (no?, title?, (%s.zz;)*, subsect3*)     >
<!ELEMENT subsect3            - O  (no?, title?, (%s.zz;)*, subsect4*)     >
<!ELEMENT subsect4            - O  (no?, title?, (%s.zz;)*, subsect5*)     >
<!ELEMENT subsect5            - O  (no?, title?, (%s.zz;)*, subsect6*)     >
<!ELEMENT subsect6            - O  (no?, title?, (%s.zz;)*)                >
<!ELEMENT no                  - O  (#PCDATA)                               >


<!-- +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Section Subelements                                             -->
<!-- +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT p                   - O  (#PCDATA|(%p.zz.ph;)|(%p.zz;))*         >


<!-- +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Paragraph Subelements                                           -->
<!-- +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT bq                  - -  %m.pseq;                                >
<!ELEMENT indaddr             - O  %m.name; -- individual address        -->
<!ELEMENT orgaddr             - O  %m.org;  -- organization address      -->
<!ELEMENT artwork             - O  EMPTY                                   >
```

```
<!ELEMENT lit                  - -  CDATA                                  >

<!ELEMENT (%p.lst.d;)          - -  (head?, item)*                         >
<!ELEMENT item                 - O  %m.pseq;                               >

<!ELEMENT deflist              - -  ((head, ddhd)?, term, dd)*             >
<!ELEMENT (term|head|ddhd)     - O  %m.ph;                                 >
<!ELEMENT dd                   - O  %m.pseq;                               >

<!ELEMENT biblist              - O  (head?, citation)*                     >
<!ELEMENT citation             - O  %m.bib;                                >
<!ELEMENT (othinfo|subject|sertitle)
                               - O  %m.ph;                                 >

<!ELEMENT location             - O  %m.addr;                               >
<!ELEMENT (msn|pages)          - O  (#PCDATA)                              >
<!ELEMENT keyword              - O  (#PCDATA)                              >
<!ELEMENT keyphras             - O  (#PCDATA)                              >
<!ELEMENT indxflag             - O  EMPTY                                  >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Poetry                                                            -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT poem                 - O  %m.poem;                               >
<!ELEMENT stanza               - O  (poemline)+                            >
<!ELEMENT poemline             - O  (#PCDATA|cline|%p.em.ph;)*             >
<!ELEMENT cline                - O  (#PCDATA|%p.em.ph;)*                   >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Phrases                                                           -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT q                    - -  %m.ph;                                 >
<!ELEMENT (%p.em.ph;)          - -  %m.ph;                                 >
<!ELEMENT (%p.rf.ph;)          - -  (#PCDATA)                              >
```

```
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    For HyTime Links                                                     -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT nameloc              - O (nmlist*) -- assigns a local ID to
                                   named objects                            -->
<!ELEMENT nmlist               - O (#PCDATA) -- list of local ID or entity
                                   names                                    -->


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Floating Elements                                                    -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT figgrp               - -  (title? & fig*)                          >
<!ELEMENT fig                  - O  %m.fig;                                  >
<!ELEMENT footnote             - -  (no?, %m.pseq;)          -(%i.float;)    >
<!ELEMENT note                 - -  (no?, %m.pseq;)                          >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Tables                                                               -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ELEMENT table                - -  (no?, title?, tbody)      -(%i.float;)   >
<!ELEMENT tbody                - O  (head*, tsubhead*, row*)                 >
<!ELEMENT row                  - O  (tstub?, cell*)                          >
<!ELEMENT tsubhead             - O  %m.ph;                                   >
<!ELEMENT (tstub|cell)         - O  %m.pseq;                                 >


<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->
<!--    Mathematics                                                          -->
<!-- ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++ -->

<!ENTITY % maths PUBLIC "ISO 12083:1993//DTD Mathematics//EN"                >
%maths;
```

```
<!-- Remove these comments if the formulas follow a NOTATION rather than SGML.
<!ELEMENT dformgrp             - O  (dformula)+                              >
<!ELEMENT (formula|dformula)   - -  CDATA                                     >
-->


<!-- ========================================================================= -->
<!--    APPENDIX ELEMENTS                                                      -->
<!-- ========================================================================= -->

<!ELEMENT appmat               - O  (appendix+)                              >
<!ELEMENT appendix             - O  (no?, %m.sec;)                           >


<!-- ========================================================================= -->
<!--    BACK MATTER ELEMENTS                                                   -->
<!-- ========================================================================= -->

<!-- The following declaration is specific to Articles                        -->
<!ELEMENT back                 - O  (%bmsec.d;|biblist)*                     >

<!ELEMENT (%bmsec.d;)          - O  %m.sec;                                  >
```

Βιβλιογραφία

[Ντα93] Κώστας Νταντουρής. Βιβλιοθήκη Στοιχειωδών Ερωτηματικών Συναρτήσεων και Επεξεργασία Ερωτήσεων για την Γλώσσα TELOS. Μεταπτυχιακή εργασία, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης, Ιούλιος 1993.

[Γεω94a] Γιώργος Γεωργιανάκης. Ο Μηχανισμός Αποθήκευσης και Διαχείρησης Οντοτήτων για την Γλώσσα Παράστασης Γνώσης TELOS. Μεταπτυχιακή εργασία, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης, Φεβρουάριος 1994.

[Χρι94b] Μαρία Χριστοφοράκη. Τεκμηρίωση πολιτιστικών αγαθών με το σύστημα ΚΛΕΙΩ (CLIO). Μεταπτυχιακή εργασία, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης, Φεβρουάριος 1994.

[Θεο95] Μάνος Θεοδωράκης. Εμβέλεια ονόματος σε σημασιολογικά μοντέλα δεδομένων. Μεταπτυχιακή εργασία, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης, Σεπτέβριος 1995.

[Δασ96a] Δημήτρης Δασκαλάκης. Διαλογική ενημέρωση οντοκεντρικών βάσεων δεδομένων. Μεταπτυχιακή εργασία, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης, Φεβρουάριος 1996.

[Τρα96b] Αθηνά Τραψιώτη. Σύστημα διαχείρισης SGML εγγράφων σε βάσεις γνώσεων. Μεταπτυχιακή εργασία, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης, 1996.

[Bai92] Henry S. Baird. Anatomy of a versatile page reader. *Proceedings of the IEEE*, pages 1059--1065, 1992.

[CBY89] Timothy Caltin, Paulette Bush, and Nicole Yankelovich. Internote: extending a hypermedia framework to support annotative colladoration. In *Hypertext '89 Proceedings*, pages 365--378, 1989.

[CD94] Panos Constantopoulos and Martin Doerr. The semantic index system: A brief presentation. Institute of Computer Science Foundation ofResearch and Technology-Hellas, 1994. URL:http://www.ics.forth.gr/proj/isst/Systems/SIS.

[CFP95] Augusto Celentano, Maria Grazia Fugini, and Silvano Pozzi. Knowledge-based document retrieval in office environments: The kabiria system. *ACM Transactions on Information Systems*, 13(4):237--268, July 1995.

[Che86] Pehong Chen. Gnu emacs BibTEX-Mod. Technical Report 87 317, Computer science division, University of California, Berkeley, October 1986.

[Cou94] Neal Coulter, editor. *ACM guide to Computing Literature*. ACM press, 1994.

[CR94] V. Christophides and A Rizk. Querying structured documents with hypertext links using OODBMS. In *Proceedings of the ECHT'94 European Conference on Hypermedia Technologies*, pages 186--197, 1994.

[DD92] C. Dadouris and M. Doerr. The programmatic query interface for the sib user's manual. Technical Report ITHACA.FORT.92.E2#8, Institute of Computer Science Foundation of Research and Technology-Hellas, December 1992.

[DH94] Jim Davis and Dan Huttenlocher. CoNote: small group annotation experiment. Technical report, Computer Science, Cornell University, 1994.

[FNK92] Hiromichi Fujisawa, Yasuaki Nakano, and Kiyomichi Kurino. Segmentation methods for character recognition: From segmentation to document structure analysis. *Proceedings of the IEEE*, 80:1079--1092, July 1992.

[Fur89] Richard Furuta. An object-based taxonomy for abstract structure in document models. *Computer*, 32(6):494--504, 1989.

[GNOT92] David Goldberg, David Nichols, Brian Oki, and Douglas Terry. Using collaborative filtering to wave an information tapesty. *Communications of the ACM*, 35(12), December 1992.

[Gor88] M. Gordon. Probabilistic and genetic algorithms in document retrieval. *Communications ACM*, 31(10), 1988.

[Gra95] Wayne Gramlich. Tree annotation set proposal, November 1995. SUN laboratories.

[GRSW95] R. Greenwood, I. Robertson, R. Snowdon, and B. Warboys. Active models in business. Technical report, Informatics Process Group, Department of Computer Science, University of Manchester, 1995.

[HI93] Tao Hu and Rolf Ingold. A mixed approach toward an efficient logical structure recognition from document images. *Electronic Publishing*, 6(4):457--468, 1993.

[HM93] Michael A. Harrison and Ethan V. Munson. On integrated bibliography processing, April 1993. Technical Report.

[ISO86] Documentation - guidelines for the establishment and development of monolingual thesauri. International Standard Organization (ISO), reference number 2788, 1986.

[KLK89] Won Kim, F.H. Lochovsky, and Won Kim, editors. *Object-Oriented Concepts, Databases and Applications.* ACM Press, 1989.

[LaL94] Daniel LaLiberte. Hypernews, 1994. URL:http://union.ncsa.uiuc.edu/HyperNews/get/hypernews.html.

[Lam86] Leslie Lamport, editor. LaTeX*: A Document Preparation System. User's Guide and Reference Manual.* Addison-Wesley Publishing Company, Massachusetts, 1986.

[Mar] Mark Antre Schenk. Wit. http://eliot.unil.ch:8085/docs/wda-article.html.

[MB93] Jessics L. Milstead and Mary C. Berger. The engineering information thesaurus development project. *Information Services & Use*, 13:71--80, 1993.

[MBJK90] J. Mylopoulos, A. Borgida, M. Jarke, and M. Koubarakis. Telos: Representing knowledge about information systems. *ACM Transactions on Information Systems*, 8(4):352--362, October 1990.

[MR88] H. Mili and R. Rada. Merging thesauri: Principles and evaluation. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 10(2), 1988.

[MTT+91] Masaaki Mizuno, Yoshitake Tsuji, Toshiyuki Tanaka, Haruhiko Tanaka, Masao Iwashita, and Tsutomu Temma. Document recognition system with layout structure generator. *NEC Research and Development*, pages 430--437, July 1991.

[Mur89] Makoto Murata. An object-oriented interpretation of oda. *Woodman '89: Workshop on Object-Oriented Document Manipulation*, May 1989.

[Nie84] Jakob Nielsen. How readers annotate textbooks and manuals. Technical Report DAIMI PB-182, Compute Science Department, Aarhus University, October 1984.

[Pat88] Oren Patashnik. BibTEXing. Computer Science Department, Stanford University, California, January 1988.

[PR94] Gilbert B. Porter and Emil V. Rainero. Document reconstruction: A system for recovering document structure from layout. *Proceedings of the Conference on Electronic Publishing*, 1994.

[RJ89] Susannah Ravden and Graham Johnson, editors. *Evaluating Usability of Human-Computer Interfaces: A Practical Method.* Ellis Horwood Limited, 1989.

[RM87] Roy Rada and Brian Martin. Augmenting thesauri for information systems. *ACM Transactions on Office Information Systems*, 5(4):378--392, October 1987.

[RMW95] Martin Roscheisen, Christian Mogensen, and Terry Winograd. A platform for third-party value-added informtion providers: Architecture, protocols, and usage examples, 1995. Technical Report CSDTR/DLTR.

[SK93] June P. Silvester and Paul H. Klingbiel. An operational system for subject switching between controled vocabularies. *Information processing & Management*, 2(1):47--59, 1993.

[SMC+96] Bruce Schatz, William Mischo, Timothy Cole, Joseph Hardin, Ann Bishop, and Hsinchun Chen. Federating diverse collections of scientific literature. *IEEE Computer*, 20(5), May 1996. Digital Library Initiative: University of Illinois.

[Sve89] Elaine Svenonius. Design of controlled vocabularies. In *Encyclopedia of Library and Information Science, Marcel Dokker*, 1989.

[Van90] Michael Vanter. Bibliotext: a hypertext browser for bibliographic data and notes, 1990. Technical Report No. UCB/CSD 90/603.