

Σύστημα Παροχής Πληροφοριών κατ' Απαίτηση από Ετερογενείς Πηγές Πληροφορίας στο Διαδίκτυο

Διατριβή που εκπονήθηκε στο Πολυτεχνείο Κρήτης για την εκπλήρωση των
υποχρεώσεων για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης

ΧΡΥΣΑ ΤΣΙΝΑΡΑΚΗ

Ιανουάριος 2000

Περίληψη

Παρ' όλο που ο *Παγκόσμιος Ιστός* (*World Wide Web – WWW*) έχει κάνει ευκολότερη την πρόσβαση των χρηστών, ειδικών και μη, σε πληροφορία που τους ενδιαφέρει, υποφέρει σε ότι αφορά την παροχή υποστήριξης για τον εύκολο και γρήγορο εντοπισμό και την ανάκτηση της πληροφορίας αυτής. Οι χρήστες των μηχανών αναζήτησης στον Παγκόσμιο Ιστό συχνά παίρνουν πίσω ένα πολύ μεγάλο αριθμό από μη σχετικά έγγραφα σαν απάντηση των ερωτήσεών τους. Το χαρακτηριστικό που λείπει από τον Παγκόσμιο Ιστό είναι η παροχή λειτουργικότητας αντίστοιχης με αυτή που παρέχουν οι βάσεις δεδομένων, που διαθέτουν μηχανισμούς που εγγυώνται τη γρήγορη, ακριβή και αποδοτική πρόσβαση για την ανάκτηση της πληροφορίας που αποθηκεύεται σ' αυτές. Μερική ευθύνη για την κατάσταση αυτή φέρει η HTML, πάνω στην οποία αρχικά βασίστηκε ο Παγκόσμιος Ιστός: Η HTML είναι πολύ πιο κοντά στη νοοτροπία των γλωσσών περιγραφής διάταξης σελίδων απ' ότι των γλωσσών περιγραφής περιεχομένου. Προς την κατεύθυνση της λύσης των προβλημάτων αυτών, τα τελευταία χρόνια ορίζεται μια «νέα γενιά» προτύπων για τον Παγκόσμιο Ιστό (XML για τη σύνταξη εγγράφων, RDF για την περιγραφή μεταδεδομένων, XQL για την υποβολή ερωτήσεων).

Η παρούσα διατριβή παρουσιάζει μια μεθοδολογία και ένα σύστημα που ενσωματώνει τα καινούρια πρότυπα του Παγκοσμίου Ιστού και την τεχνολογία των βάσεων δεδομένων για να αυξήσει την αποδοτικότητα της ανάκτησης πληροφορίας από τους χρήστες. Το σύστημα «*Παροχής Πληροφοριών κατ' Απαίτηση*» («*Information Provision on Demand*») παρέχει στους χρήστες του, μέσω του Παγκοσμίου Ιστού, τόσο τη δυνατότητα εντοπισμού και ανάκτησης πληροφορίας που σχετίζεται με τα ενδιαφέροντά τους όσο και ενημέρωση για πληροφορία που γίνεται διαθέσιμη και ταιριάζει με αυτά.

Το σύστημα «*Παροχής Πληροφοριών κατ' Απαίτηση*» μπορεί να θεωρηθεί παροχέας θεματικής πληροφορίας στον Παγκόσμιο Ιστό, καθώς παρέχει υπηρεσίες πληροφόρησης και ανάκτησης πληροφορίας μέσα από το συνδυασμό χαρακτηριστικών των Ψηφιακών Βιβλιοθηκών, των Συστημάτων Νέων κατ' Απαίτηση, των Διαμεσολαβητών και των Πρακτόρων.

Η πληροφορία στην οποία παρέχει πρόσβαση με ενιαίο τρόπο το σύστημα «*Παροχής Πληροφοριών κατ' Απαίτηση*» αποθηκεύεται σε ένα σύνολο ανεξάρτητων ετερογενών *Πηγών Πληροφορίας* (*Information Sources*). Ενιαία πρόσβαση στα περιεχόμενα των πηγών πληροφορίας παρέχεται από ένα *Διαμεσολαβητή* (*Mediator*), που αναλαμβάνει να κάνει διαφανή την πρόσβαση των χρηστών στην πληροφορία που τους ενδιαφέρει, ανεξάρτητα από την πηγή πληροφορίας στην οποία η πληροφορία αυτή αποθηκεύεται.

Οι πηγές πληροφορίας είναι ανεξάρτητα πληροφοριακά συστήματα που μπορεί να διαφέρουν μεταξύ τους τόσο στη μορφή της πληροφορίας που περιέχουν (π.χ. δομημένα έγγραφα, παραδοσιακά δεδομένα, αντικείμενα πολυμέσων κλπ.) όσο και στα τεχνικά χαρακτηριστικά τους όπως είναι το είδος της βάσης δεδομένων τους (π.χ. σχεσιακή, αντικειμενοστραφής κλπ.), η οργάνωση της βάσης δεδομένων τους σε επίπεδο σχήματος και η γλώσσα μέσω της οποίας εξάγουν τη λειτουργικότητά τους. Επιπλέον, οι πηγές πληροφορίας που αποτελούν μέρος του συστήματος «*Παροχής Πληροφοριών κατ'*

Απαίτηση» μπορούν να χαρακτηριστούν ως συγχρονιζόμενες, καθώς το περιεχόμενό τους ενημερώνεται από *Πράκτορες (Agents)* για αντικείμενα που οι τελευταίοι εντοπίζουν στο *Διαδίκτυο (Internet)* και το περιεχόμενό τους ταιριάζει με αυτό των πηγών πληροφορίας.

Ο σχεδιασμός τόσο της συνολικής αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» όσο και των τμημάτων από τα οποία αυτή απαρτίζεται, βασίστηκε στο συνδυασμό των ώριμων τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό (XML, RDF, XQL κλπ.). Επιπλέον, ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» βασίστηκε στη χρήση της UML για την περιγραφή της αρχιτεκτονικής και της συμπεριφοράς του συστήματος.

Αφιέρωση

Στη μαμά και τον αδερφό μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Σταύρο Χριστοδουλάκη για την επίβλεψη του στην ολοκλήρωση αυτής της διατριβής και για τις εμπειρίες που μου πρόσφερε στα πλαίσια της εργασίας μου στο Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων και Εφαρμογών του Πολυτεχνείου Κρήτης.

Θα ήθελα επίσης να ευχαριστήσω προκαταβολικά τους αναγνώστες της διατριβής, τον καθηγητή κ. Μιχάλη Πατεράκη και τον επίκουρο καθηγητή κ. Μανόλη Κουμπαρακή για το χρόνο που θα διαθέσουν για την ανάγνωση της διατριβής και τις τυχόν παρατηρήσεις τους πάνω σε αυτήν.

Θα ήθελα ακόμα να ευχαριστήσω όλους τους συνεργάτες μου στο Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων και Εφαρμογών του Πολυτεχνείου Κρήτης για τη συνεργασία τους.

Τέλος, θα ήθελα να ευχαριστήσω όλους όσους αγωνίστηκαν τα τελευταία δύο χρόνια, παρά τις αντίξοες συνθήκες, για τη διασφάλιση του επιπέδου του Πολυτεχνείου Κρήτης και της αξίας των πτυχίων που προέρχονται απ' αυτό.

Χρύσα Τσιναράκη

Πολυτεχνείο Κρήτης
Ιανουάριος 2000

Πίνακας Περιεχομένων

1. Εισαγωγή	1
1.1. Αναγκαιότητα – Συνεισφορά.....	4
1.2. Περιγραφή του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»	6
1.3. Στόχοι της Διατριβής	8
1.4. Δομή της Διατριβής	10
2. Σχετική Έρευνα	13
2.1. Ψηφιακές Βιβλιοθήκες.....	14
2.2. Διαμεσολαβητές	15
2.3. Συστήματα Νέων κατ' Απαίτηση.....	16
2.4. Μηχανές Αναζήτησης.....	18
2.5. Πράκτορες	20
2.6. Αναπαράσταση Δομημένων Εγγράφων σε Βάσεις Δεδομένων	21
2.7. Το πρόγραμμα VENIVA.....	23
2.8. Ανακεφαλαίωση	25
3. Συσχετιζόμενη Εργασία	26
3.1. UML (Unified Modeling Language).....	26
3.1.1. Διαγράμματα Παράταξης (Deployment Diagrams)	27
3.1.2. Διαγράμματα Περιπτώσεων Χρήσης (Use Case Diagrams)	28
3.1.3. Διαγράμματα Κλάσεων (Class Diagrams)	29
3.1.4. Διαγράμματα Αλληλεπίδρασης (Interaction Diagrams).....	30
3.1.4.1. Σειριακά Διαγράμματα (Sequence Diagrams)	30
3.1.4.2. Διαγράμματα Συνεργασίας (Collaboration Diagrams)	31
3.1.5. Διαγράμματα Δραστηριοτήτων (Activity Diagrams).....	31
3.2. Πρότυπα Αναπαράστασης Εγγράφων.....	32
3.2.1. eXtensible Markup Language (XML).....	37
3.2.1.1. Η Δομή και το Περιεχόμενο των XML Εγγράφων	38
Document Type Declaration (DTD).....	40
Γλώσσες Περιγραφής XML Σχημάτων.....	41
XML Schema Language	47
3.2.1.2. Πληροφορία που περιέχεται σε XML Έγγραφα	51
3.2.1.3. XML Namespaces.....	53
3.2.1.4. Υποστήριξη Συνδέσμων	54
XLL ή XLink (eXtensible Link Language).....	54
XPointer (XML Pointer Language)	56
XPath (XML Path Language).....	58
3.2.1.5. Σύνδεση Στυλ με XML Έγγραφα.....	59
3.2.1.6. Κωδικοποίηση XML Εγγράφων	61
3.2.1.7. Ανταλλαγή Τμημάτων XML Εγγράφων.....	61
3.2.1.8. Εφαρμογές της XML	62
3.3. Πρότυπα Αναπαράστασης Μεταδεδομένων	68
3.3.1. Το Πλαίσιο Warwick	69
3.3.1.1. Digital Object Architecture (DOA)	72
3.3.1.2. Meta-Object Facility/XML-based Metadata Interchange (MOF/XMI).....	73
3.3.1.3. Resource Description Framework (RDF).....	75
Το μοντέλο της RDF.....	77
Η σύνταξη της RDF	78
RDF σχήματα	79
3.3.2. Λεξικά Μεταδεδομένων	80
3.3.2.1. Dublin Core	84

3.4. Document Object Model (DOM)	88
3.5. Γλώσσες Διαχείρισης & Αναζήτησης Εγγράφων	89
3.5.1. XML-QL	90
3.5.2. XQL (XML Query Language).....	92
3.6. Ανακεφαλαίωση	98
4. Αρχιτεκτονική του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»	99
4.1. Γενική Περιγραφή της Αρχιτεκτονικής του Συστήματος	99
4.2. Τμήματα της Αρχιτεκτονικής του Συστήματος.....	101
4.3. Λειτουργικότητα του Συστήματος	103
4.4. Ανακεφαλαίωση	105
5. Διαμεσολαβητής	106
5.1. Αρχιτεκτονική του Διαμεσολαβητή.....	108
5.2. Τμήματα της Αρχιτεκτονικής του Διαμεσολαβητή	109
5.2.1. Επίπεδο Διεπαφής του Διαμεσολαβητή.....	110
5.2.2. Παρουσιαστές.....	110
5.2.3. Ο Πυρήνας του Διαμεσολαβητή	111
5.2.4. Περιτυλίγματα	112
5.3. Λειτουργικότητα του Διαμεσολαβητή.....	112
5.4. Γλώσσα του Διαμεσολαβητή.....	117
5.4.1. Υποστήριξη Ερωτήσεων και Αιτήσεων	120
5.4.2. Επιστρεφόμενη Πληροφορία	126
5.5. Βάση Δεδομένων του Διαμεσολαβητή	128
5.5.1. Μεταπληροφορία Σχετική με τις Πηγές Πληροφορίας.....	131
5.5.2. Λειτουργικότητα της Βάσης Δεδομένων του Διαμεσολαβητή	135
5.6. Ανακεφαλαίωση	141
6. Πηγές Πληροφορίας	142
6.1. Αρχιτεκτονική των Πηγών Πληροφορίας.....	143
6.2. Λειτουργικότητα των Πηγών Πληροφορίας.....	145
6.3. Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα	147
6.3.1. Αρχιτεκτονική των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα	148
6.3.2. Γλώσσα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα	150
6.3.2.1. Υποστήριξη Ερωτήσεων και Αιτήσεων	151
6.3.2.2. Υποστήριξη Εργαλείων Διαχείρισης των Πηγών Πληροφορίας.....	152
6.3.2.3. Επιστρεφόμενη Πληροφορία	154
6.3.2.4. Περιτυλίγματα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα	154
6.3.3. Μέθοδοι Πρόσβασης.....	159
6.3.4. Βάση Δεδομένων των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα.....	159
6.3.5. Λειτουργικότητα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα	162
6.3.6. Αναπαράσταση Δομημένων Εγγράφων	164
6.3.6.1. Μοντέλο Αναπαράστασης Δομημένων Εγγράφων	164
6.3.6.2. XML Έγγραφα.....	174
6.3.6.3. Σχήμα Βάσης Δεδομένων	176
6.3.6.4. Επίπεδο Διεπαφής με XML Έγγραφα.....	179
6.4. Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα.....	181
6.4.1. Γλώσσα των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα	182
6.4.2. Βάση Δεδομένων των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα.....	184
6.4.3. Λειτουργικότητα των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα	192
6.5. Ανακεφαλαίωση	193
7. Διαχείριση των Πηγών Πληροφορίας.....	194

7.1. Δημιουργία Πηγών Πληροφορίας.....	195
7.2. Υποστήριξη Νέων Κλάσεων από τις Πηγές Πληροφορίας.....	199
7.2.1. Λειτουργία του Εργαλείου SupportDTD	200
7.2.2. Η Γλώσσα ΚΑΔΔΑ (Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Αντικειμένων).....	208
7.2.2.1. Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Δομημένων Εγγράφων	211
7.2.2.2. Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Παραδοσιακών Δεδομένων	216
7.2.2.3. Σύνταξη της Γλώσσας ΚΑΔΔΑ.....	220
7.2.3. Δομή των Κλάσεων	224
7.3. Εισαγωγή Αντικειμένων	227
7.3.1. Λειτουργία των Μηχανισμών Εισαγωγής Αντικειμένων	228
7.3.2. Εφαρμογές για την Εισαγωγή Αντικειμένων	232
7.4. Αναζήτηση & Ανάκτηση Αντικειμένων	233
7.4.1. Αναζήτηση & Ανάκτηση Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα	233
7.4.2. Αναζήτηση & Ανάκτηση Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα.....	234
7.5. Δεικτοδότηση Απομακρυσμένων Αντικειμένων	235
7.5.1. Δεικτοδότηση Απομακρυσμένων Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα	235
7.5.2. Δεικτοδότηση Απομακρυσμένων Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα.....	239
7.6. Ενημέρωση των Πηγών Πληροφορίας.....	241
7.6.1. Διαχείριση Απομακρυσμένων Αντικειμένων	242
7.6.2. Λειτουργία του Εργαλείου ISAgent.....	243
7.7. Ανακεφαλαίωση	246
8. Εφαρμογή του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».....	247
8.1. Περιβάλλον της Εφαρμογής	247
8.2. Αρχιτεκτονική της Εφαρμογής	248
8.3. Λειτουργία της Εφαρμογής	250
8.3.1. Αρχικοποίηση & Εξέλιξη των Βάσεων Δεδομένων των Πηγών Πληροφορίας.....	251
8.3.2. Ενημέρωση των Πηγών Πληροφορίας	266
8.3.2.1. Εισαγωγή Εγγράφου σε Πηγή Πληροφορίας που περιέχει Δομημένα Έγγραφα	267
8.3.2.2. Δεικτοδότηση Εγγράφου σε Πηγή Πληροφορίας που περιέχει Δομημένα Έγγραφα	268
8.3.2.3. Εισαγωγή Αντικειμένου σε Πηγή Πληροφορίας που περιέχει Παραδοσιακά Δεδομένα.....	269
8.3.2.4. Δεικτοδότηση Αντικειμένου σε Πηγή Πληροφορίας που περιέχει Παραδοσιακά Δεδομένα.....	270
8.3.3. Αλληλεπίδραση με τους Χρήστες	271
8.4. Ανακεφαλαίωση	277
9. Ανακεφαλαίωση – Συνεισφορά – Μελλοντικές Επεκτάσεις.....	278
9.1. Ανακεφαλαίωση	278
9.2. Συμπεράσματα – Συνεισφορά	281
9.3. Μελλοντικές Επεκτάσεις	284
Αναφορές	285

Πίνακας Σχημάτων

Σχήμα 1-1: Αρχιτεκτονική του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»	7
Σχήμα 2-1: Η Αρχιτεκτονική του Συστήματος που αναπτύχθηκε στα πλαίσια του προγράμματος VENIVA	23
Σχήμα 3-1: XML και XML Εφαρμογές	63
Σχήμα 3-2: Δοχείο του πλαισίου Warwick	71
Σχήμα 3-3: Πληροφορία, Μεταπληροφορία και Μετα-Μεταπληροφορία	74
Σχήμα 3-4: Γενικευμένη RDF περιγραφή πόρου:	77
Σχήμα 7-1: Η Δομή ενός ΚΑΔΔΑ Συνόλου Κανόνων	208
Σχήμα 7-2: Το RDF Μοντέλο των Συνόλων Κανόνων που αντιστοιχούν σε DTDs που περιγράφουν τη Δομή Εγγράφων	210
Σχήμα 7-3: Το RDF Μοντέλο των Συνόλων Κανόνων που αντιστοιχούν σε DTDs που περιγράφουν τη Δομή Παραδοσιακών Δεδομένων	211

Πίνακας Διαγραμμάτων

Διάγραμμα 3-1: Απλοποιημένο Διάγραμμα Παράταξης για το Σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση».....	28
Διάγραμμα 4-1: UML Διάγραμμα Παράταξης του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».....	100
Διάγραμμα 4-2: Διάγραμμα Περιπτώσεων Χρήσης του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».....	103
Διάγραμμα 5-1: Διάγραμμα Παράταξης για την Αρχιτεκτονική του Διαμεσολαβητή.....	109
Διάγραμμα 5-2: Σειριακό Διάγραμμα για την Απάντηση Ερώτησης.....	114
Διάγραμμα 5-3: Σειριακό Διάγραμμα για την Ανάκτηση Αντικειμένου.....	116
Διάγραμμα 5-4: Σειριακό Διάγραμμα για την Ειδοποίηση Χρηστών.....	117
Διάγραμμα 5-5: E-R Διάγραμμα που περιγράφει τη Βάση Δεδομένων του Διαμεσολαβητή.....	129
Διάγραμμα 5-6: Διάγραμμα Κλάσεων που περιγράφει τη Βάση Δεδομένων του Διαμεσολαβητή, με έμφαση στην περιγραφή της Μεταπληροφορίας που σχετίζεται με τις Πηγές Πληροφορίας.....	133
Διάγραμμα 5-7: Διάγραμμα Περιπτώσεων Χρήσης για τη Λειτουργικότητα της Βάσης Δεδομένων του Διαμεσολαβητή κατά την Επεξεργασία Ερωτήσεων.....	137
Διάγραμμα 5-8: Διάγραμμα Περιπτώσεων Χρήσης για τη Λειτουργικότητα της Βάσης Δεδομένων του Διαμεσολαβητή κατά την Εξυπηρέτηση Αιτήσεων.....	139
Διάγραμμα 5-9: Διάγραμμα Περιπτώσεων Χρήσης για τη Λειτουργικότητα της Βάσης Δεδομένων του Διαμεσολαβητή κατά την Εισαγωγή Πληροφορίας στις Πηγές Πληροφορίας.....	140
Διάγραμμα 6-1: Διάγραμμα Παράταξης για τη Γενικευμένη Αρχιτεκτονική των Πηγών Πληροφορίας.....	144
Διάγραμμα 6-2: Γενικευμένο Διάγραμμα Περιπτώσεων Χρήσης για τις Πηγές Πληροφορίας.....	146
Διάγραμμα 6-3: Διάγραμμα Παράταξης για την Αρχιτεκτονική των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα.....	149
Διάγραμμα 6-4: Αφαιρετικό E-R Διάγραμμα για το Σχήμα της Βάσης Δεδομένων των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα.....	160
Διάγραμμα 6-5: Αφαιρετικό Μοντέλο Αναπαράστασης Εγγράφων στις Βάσεις Δεδομένων των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα.....	165
Διάγραμμα 6-6: E-R Διάγραμμα της Βάσης Δεδομένων πριν υποστηριχθεί κάποιος DTD ή κάποιος XML Σχήμα.....	177
Διάγραμμα 6-7: E-R Διάγραμμα της Βάσης Δεδομένων, μετά την υποστήριξη του DTD Άρθρο.dtd.....	179
Διάγραμμα 6-8: E-R διάγραμμα της Βάσης Δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αμέσως μετά την αρχικοποίησή τους.....	186
Διάγραμμα 6-9: Διάγραμμα Κλάσεων της Βάσης Δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αμέσως μετά την αρχικοποίησή τους, με έμφαση στην αναπαράσταση των αντικειμένων.....	188
Διάγραμμα 6-10: E-R διάγραμμα της Βάσης Δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αμέσως μετά την υποστήριξη του DTD Πρόσωπο.dtd.....	191
Διάγραμμα 7-1: Διάγραμμα Δραστηριοτήτων για το Πρόγραμμα Εγκατάστασης Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα.....	196
Διάγραμμα 7-2: Διάγραμμα Δραστηριοτήτων για το Πρόγραμμα Εγκατάστασης Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα.....	198
Διάγραμμα 7-3: Διάγραμμα Δραστηριοτήτων για το εργαλείο SupportDTD.....	201
Διάγραμμα 7-4: Διάγραμμα Δραστηριοτήτων για τον Ορισμό Κανόνων Αναπαράστασης, Διαχείρισης & Δεικτοδότησης των Στιγμιότυπων των Κλάσεων Εγγράφων.....	206
Διάγραμμα 7-5: Διάγραμμα Δραστηριοτήτων για τον Ορισμό Κανόνων Αναπαράστασης, Διαχείρισης & Δεικτοδότησης των Στιγμιότυπων των Κλάσεων Παραδοσιακών Δεδομένων.....	207
Διάγραμμα 7-6: Διάγραμμα Δραστηριοτήτων του Μηχανισμού που υποστηρίζει την Εισαγωγή Δομημένων Εγγράφων στις Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα.....	229
Διάγραμμα 7-7: Διάγραμμα Δραστηριοτήτων του Μηχανισμού που υποστηρίζει την Εισαγωγή Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα.....	231

Διάγραμμα 7-8: UML Διάγραμμα Δραστηριοτήτων που περιγράφει τη λειτουργία του Μηχανισμού Δεικτοδότησης Απομακρυσμένων Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα	238
Διάγραμμα 7-9: UML Διάγραμμα Δραστηριοτήτων που περιγράφει τη λειτουργία του Μηχανισμού Δεικτοδότησης Απομακρυσμένων Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα	239
Διάγραμμα 7-10: UML Διάγραμμα Δραστηριοτήτων που περιγράφει τη λειτουργία του Εργαλείου ISAgent.....	245
Διάγραμμα 8-1: UML Διάγραμμα Παράταξης για την Εφαρμογή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»	249
Διάγραμμα 8-2: Αφαιρετικό Διάγραμμα Κλάσεων που απεικονίζει το Μοντέλο Πληροφορίας που εξάγεται στους Χρήστες μέσω του Διαμεσολαβητή	272

Πίνακας Πινάκων

Πίνακας 3-1: Τα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Παράταξης.....	27
Πίνακας 3-2 Τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Περιπτώσεων Χρήσης.....	29
Πίνακας 3-3: Τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Κλάσεων	30
Πίνακας 3-4: Τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Σειριακά Διαγράμματα.....	31
Πίνακας 3-5: Τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Δραστηριοτήτων	31
Πίνακας 3-6: Ορισμοί XML Εγγράφων.....	49
Πίνακας 3-7: Συγκριτική παρουσίαση SQL/XQL	93
Πίνακας 3-8: Σύνταξη για τους Όρους και τα σχετικά με αυτούς Κριτήρια	94
Πίνακας 3-9: Σύνταξη για τις Namespaces και τα σχετικά με αυτές Κριτήρια	95
Πίνακας 3-10: Τελεστές για την υποστήριξη Συγκρίσεων και Περιορισμών από την XQL	95
Πίνακας 3-11: Οι Ιεραρχικές Σχέσεις που υποστηρίζονται από την XQL	96
Πίνακας 3-12: Οι Λογικοί Τελεστές και οι Τελεστές Συνόλων που υποστηρίζονται από την XQL	96
Πίνακας 3-13: Οι Τελεστές Διάταξης που υποστηρίζονται από την XQL.....	97
Πίνακας 5-1: Η επιπλέον της SQL Λειτουργικότητα που παρέχεται από τη Γλώσσα του Διαμεσολαβητή	121
Πίνακας 6-1: Οι συναρτήσεις που παρέχουν την επιπλέον της XQL λειτουργικότητα για την υποστήριξη ερωτήσεων και αιτήσεων που παρέχεται από τη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα	152
Πίνακας 6-2: Οι συναρτήσεις που υλοποιούν τις λειτουργίες υποστήριξης των εργαλείων διαχείρισης των πηγών πληροφορίας για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και ενσωματώνονται στη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.....	153
Πίνακας 6-3: Κανόνες Αντιστοίχισης Ερωτήσεων και Αιτήσεων μεταξύ της Γλώσσας του Διαμεσολαβητή και της Γλώσσας των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα	157
Πίνακας 6-4: Περιγραφή Στοιχείων Εγγράφου ή Ανεξάρτητων Στοιχείων Εγγράφου στον Περιγραφέα του Εγγράφου στο οποίο περιέχονται	168
Πίνακας 6-5: Η Έξοδος του Επιπέδου Διεπαφής με XML Έγγραφα	181
Πίνακας 6-6: Η επιπλέον της SQL Λειτουργικότητα που παρέχεται από τη Γλώσσα των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα	183
Πίνακας 6-7: Οι λειτουργίες υποστήριξης των εργαλείων διαχείρισης των πηγών πληροφορίας που παρέχονται από τη γλώσσα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα	184
Πίνακας 7-1: Οι υποστηριζόμενες από την ΚΑΔΔΑ Μέθοδοι	209
Πίνακας 7-2: Τα Στοιχεία που απαρτίζουν ένα ΚΑΔΔΑ Σύνολο Κανόνων Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Δομημένων Εγγράφων	213
Πίνακας 7-3: Τα Στοιχεία που απαρτίζουν ένα ΚΑΔΔΑ Σύνολο Κανόνων Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Παραδοσιακών Δεδομένων	218

Πίνακας RDF Σχημάτων

RDF Σχήμα 6-1: RDF Σχήμα που περιγράφει τη Δομή της Βάσης Δεδομένων μιας Πηγής Πληροφορίας που περιέχει Δομημένα Έγγραφα πριν την υποστήριξη κάποιας Κλάσης Εγγράφων	174
RDF Σχήμα 6-2: RDF Σχήμα που περιγράφει τη Δομή της Βάσης Δεδομένων μιας Πηγής Πληροφορίας που περιέχει Παραδοσιακά Δεδομένα πριν την υποστήριξη κάποιας κλάσης.....	190
RDF Σχήμα 7-1: Η Σύνταξη της Γλώσσας ΚΑΔΔΑ	224
RDF Σχήμα 7-2: Το RDF Σχήμα που αντιστοιχεί στο DTD 6-2	226
RDF Σχήμα 7-3: RDF σχήμα που περιγράφει τη Δομή της Κλάσης Πρόσωπο.....	227
RDF Σχήμα 8-1: Το RDF Σχήμα που αντιστοιχεί στο DTD 8-1	255
RDF Σχήμα 8-2: Το RDF Σχήμα που αντιστοιχεί στο DTD 8-2	263
RDF Σχήμα 8-3: RDF σχήμα που περιγράφει τη Δομή της Κλάσης Περιγραφή_Βιβλίου	266

Πίνακας DTDs

DTD 5-1: Ο DTD Result.dtd	126
DTD 6-1: Ο DTD SourceResult.dtd	154
DTD 6-2: Ο DTD Άρθρο.dtd που περιγράφει τη δομή και τη σύνταξη Άρθρων	176
DTD 6-3: Ο DTD Πρόσωπο.dtd	190
DTD 8-1: Ο DTD Αναφορά.dtd που περιγράφει τη δομή και τη σύνταξη Αναφορών	251
DTD 8-2: Ο DTD Διατριβή.dtd που περιγράφει τη δομή και τη σύνταξη Διατριβών	258
DTD 8-3: Ο DTD ΠεριγραφήΒιβλίου.dtd που περιγράφει τη δομή και τη σύνταξη Περιγραφών Βιβλίων	263

Πίνακας Παραδειγμάτων

Παράδειγμα 3-1: Ένας DTD για την αναπαράσταση σημειωμάτων	41
Παράδειγμα 3-2: Ένα έγγραφο σύμφωνο με το DTD για την αναπαράσταση σημειωμάτων	41
Παράδειγμα 3-3: DCD περιγραφή των χαρακτηριστικών του στοιχείου DL της HTML	46
Παράδειγμα 3-4: Σχήμα για την αναπαράσταση Σημειωμάτων	51
Παράδειγμα 3-5: Ορισμός και χρήση XML namespace.....	54
Παράδειγμα 3-6: Γενικευμένο παράδειγμα χρήσης της RDF	79
Παράδειγμα 3-7: RDF Σχήμα που περιγράφει κάποιο Πρόσωπο	80
Παράδειγμα 3-8: XML-QL Ερώτηση για την επιλογή των σημειωμάτων που στον τίτλο τους περιέχεται η συμβολοσειρά WWW	92
Παράδειγμα 3-9: Παράδειγμα Ομαδοποίησης Αποτελεσμάτων Ερώτησης.....	96
Παράδειγμα 3-10: Παράδειγμα Κλήσης Συνάρτησης.....	97
Παράδειγμα 3-11: XQL Ερώτηση για την επιλογή των σημειωμάτων που στον τίτλο τους περιέχεται η συμβολοσειρά WWW	98
Παράδειγμα 5-1: Ερώτηση όπου καθορίζεται η Πηγή Πληροφορίας στα περιεχόμενα της οποίας θα πραγματοποιηθεί η αναζήτηση.....	121
Παράδειγμα 5-2: Ερώτηση που πραγματοποιεί αναζήτηση μεταξύ των στιγμιοτύπων της κλάσης «Άρθρο» και των υποκλάσεων της	122
Παράδειγμα 5-3: Ερώτηση Ομοιότητας για το περιεχόμενο Αντικειμένων	122
Παράδειγμα 5-4: Ερώτηση με βάση Ασαφή Χαρακτηριστικά.....	123
Παράδειγμα 5-5: Ερώτηση με βάση τη Δομή των Αντικειμένων, για «Παιδιά»Αντικειμένων.....	123
Παράδειγμα 5-6: Ερώτηση με βάση τη Δομή των Αντικειμένων, για «Απογόνους» Αντικειμένων ...	124
Παράδειγμα 5-7: Ερώτηση με βάση τη Δομή των Αντικειμένων, για συγκεκριμένους «Απογόνους» Αντικειμένων	124
Παράδειγμα 5-8: Ερώτηση με βάση τη Δομή των Αντικειμένων, για συγκεκριμένους «Απογόνους» Αντικειμένων	124
Παράδειγμα 5-9: Ερώτηση με κριτήρια που περιέχουν Σχέσεις Διάταξης.....	124
Παράδειγμα 5-10: Αίτηση Ανάκτησης Περιεχομένου Αντικειμένου	125
Παράδειγμα 5-11: Αίτηση Μερικής Ανάκτησης Περιεχομένου Αντικειμένου.....	125
Παράδειγμα 5-12: Αίτηση Ανάκτησης του Περιεχομένου Δομικού Στοιχείου Αντικειμένου.....	125
Παράδειγμα 5-13: Αίτηση Ανάκτησης Περιεχομένου και Χαρακτηριστικών Αντικειμένου.....	126
Παράδειγμα 5-14: Παράδειγμα Πληροφορίας που επιστρέφεται ως Αποτέλεσμα Ερώτησης	128
Παράδειγμα 5-15: Παράδειγμα Πληροφορίας που επιστρέφεται ως Αποτέλεσμα Αίτησης	128
Παράδειγμα 6-1: Η Ερώτηση που φαίνεται στο Παράδειγμα 5-1 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα	158
Παράδειγμα 6-2: Η Ερώτηση που φαίνεται στο Παράδειγμα 5-3 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα	158
Παράδειγμα 6-3: Η Ερώτηση που φαίνεται στο Παράδειγμα 5-4 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα	158
Παράδειγμα 6-4: Η Ερώτηση που φαίνεται στο Παράδειγμα 5-11 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα	158
Παράδειγμα 6-5 Η Ερώτηση που φαίνεται στο Παράδειγμα 5-12 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα	159
Παράδειγμα 6-6: Αίτηση επεξεργασμένη από το Επίπεδο Διεπαφής με XML Έγγραφα	181
Παράδειγμα 7-1: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 6-2.....	216
Παράδειγμα 7-2: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 6-3.....	220
Παράδειγμα 8-1: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 8-1.....	254
Παράδειγμα 8-2: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 6-2.....	258
Παράδειγμα 8-3: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 8-2.....	261
Παράδειγμα 8-4: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 8-3.....	265
Παράδειγμα 8-5: XML Έγγραφο που δομείται με βάση το DTD 6-2.....	267
Παράδειγμα 8-6: Αντικείμενο που δομείται με βάση το DTD 8-3.....	270
Παράδειγμα 8-7: Αντικείμενο που δομείται με βάση το DTD 8-3.....	270

Παράδειγμα 8-8: Ερώτηση για την ανάκτηση των Τίτλων και των Ταυτοτήτων Εγγράφων που το περιεχόμενό τους σχετίζεται με Επικοινωνία Ανθρώπων-Υπολογιστών και βρίσκονται στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης	272
Παράδειγμα 8-9: Η Ερώτηση που φαίνεται στο Παράδειγμα 8-8 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα	273
Παράδειγμα 8-10: Ερώτηση για την ανάκτηση των Τίτλων και των Ταυτοτήτων Εγγράφων που το περιεχόμενό τους σχετίζεται με Επικοινωνία Ανθρώπων-Υπολογιστών και βρίσκονται στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης ή την Πηγή Πληροφορίας του Πανεπιστημίου Κρήτης	273
Παράδειγμα 8-11: Ερώτηση για την ανάκτηση των Τίτλων και των Ταυτοτήτων Εγγράφων που το περιεχόμενό τους σχετίζεται με Επικοινωνία Ανθρώπων-Υπολογιστών και βρίσκονται σε οποιαδήποτε Πηγή Πληροφορίας γίνεται διαθέσιμη μέσω του Διαμεσολαβητή	273
Παράδειγμα 8-12: Η Πληροφορία που επιστρέφεται από την Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης ως Αποτέλεσμα της Ερώτησης που φαίνεται στο Παράδειγμα 8-11	274
Παράδειγμα 8-13: Η Πληροφορία που επιστρέφεται από την Πηγή Πληροφορίας του Πανεπιστημίου Κρήτης ως Αποτέλεσμα της Ερώτησης που φαίνεται στο Παράδειγμα 8-11	275
Παράδειγμα 8-14: Η Πληροφορία που επιστρέφεται από την Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης ως Αποτέλεσμα της Ερώτησης που φαίνεται στο Παράδειγμα 8-11	275
Παράδειγμα 8-15: Η Πληροφορία που επιστρέφεται από τη Γλώσσα του Διαμεσολαβητή ως Αποτέλεσμα της Ερώτησης που φαίνεται στο Παράδειγμα 8-11	276
Παράδειγμα 8-16: Αίτηση Ανάκτησης του Περιεχομένου του Άρθρου που βρίσκεται στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης και έχει ταυτότητα 125	276
Παράδειγμα 8-17: Η Πληροφορία που επιστρέφεται ως Αποτέλεσμα της Αίτησης που φαίνεται στο Παράδειγμα 8-16	277

1^ο Κεφάλαιο

Εισαγωγή

Είναι κοινά αποδεκτό το γεγονός ότι τα τελευταία χρόνια η χρήση του *Διαδικτύου (Internet)* έχει εξαπλωθεί σε πολλούς τομείς της ανθρώπινης ζωής. Επιπλέον, επιστημονικές κοινότητες που ήταν μέχρι πρόσφατα ξένες στους υπολογιστές έχουν έρθει κοντά στους υπολογιστές και το Διαδίκτυο: Στην πραγματικότητα, η χρήση των υπολογιστών και του Διαδικτύου έχει γίνει απαραίτητη σε όλους τους επιστήμονες. Σημαντικό ρόλο στο γεγονός αυτό έπαιξε η εξάπλωση του *Παγκοσμίου Ιστού (World-Wide Web – WWW)* [WWW92]. Ο Παγκόσμιος Ιστός έδωσε τις εξής δυνατότητες σε όλες τις ομάδες χρηστών (επιστημόνων και απλών χρηστών):

- Να βλέπουν πληροφορία σχετική με τα ενδιαφέροντά τους σε μορφή *Υπερκειμένου (Hypertext)* και να πλοηγούνται μέσα σ' αυτή την πληροφορία.
 - Να αναζητούν πληροφορία που τους ενδιαφέρει μέσω μηχανών αναζήτησης.
 - Να προσπελαίνουν την πληροφορία που διαθέτουν μέσω του Παγκοσμίου Ιστού κάποιες παραδοσιακές βιβλιοθήκες. Η πληροφορία αυτή μπορεί να είναι τόσο περιγραφές των περιεχομένων των βιβλιοθηκών αυτών όσο και μέρος των περιεχομένων αυτών.
 - Να εκτελούν τις παραπάνω λειτουργίες μέσω ενός ομοιόμορφου συστήματος διεπαφής (interface) που παρέχεται από τα *Προγράμματα Πλοήγησης του Παγκοσμίου Ιστού (Web Browsers)*.
-

Με τον τρόπο αυτό δημιουργήθηκε μια τεράστια παγκόσμια βιβλιοθήκη σε μορφή υπερκειμένου, που επιτρέπει σε επιστήμονες και μη να ανακαλύπτουν πληροφορία σχετική με τα ενδιαφέροντά τους.

Παράλληλα, η δυνατότητα έκδοσης και διάδοσης της πληροφορίας έγινε πολύ εύκολη, καθώς:

- Μέσω του Παγκοσμίου Ιστού η πληροφορία και οι ιδέες διαδίδονται άμεσα σε ένα ευρύ κοινό
- Δεν είναι απαραίτητη η κάλυψη ενός εκδοτικού οίκου για να δημιουργηθεί μια «έκδοση» στον Παγκόσμιο Ιστό. Έτσι, ο αριθμός των *Ιστοσελίδων (Web Pages)* υπολογίζεται σήμερα γύρω στις 300.000.000.

Παρ' όλα αυτά, η παγκόσμια βιβλιοθήκη του Παγκοσμίου Ιστού δε μπορεί να αξιοποιηθεί εύκολα, καθώς ο Παγκόσμιος Ιστός δεν έχει αρκετά από τα βασικά χαρακτηριστικά που είναι συνδεδεμένα με την έννοια της βιβλιοθήκης. Η διαχειρισσιμότητα της κατάστασης στον Παγκόσμιο Ιστό μειώνεται καθημερινά καθώς η κατάσταση σε ότι αφορά την ανεύρεση πληροφορίας είναι χαοτική, η οργάνωση μηδενική και η έννοια της συντήρησης ανύπαρκτη. Τα κυριότερα προβλήματα που αντιμετωπίζει σήμερα ο Παγκόσμιος Ιστός είναι τα εξής:

- Οι απαιτήσεις των χρηστών από τις παρεχόμενες υπηρεσίες είναι υψηλές, παρ' όλο που τα εργαλεία και οι μηχανισμοί που χρησιμοποιούνται για την παροχή τους είναι σχεδόν πρωτόγονα.
- Η ποιότητα των παρεχόμενων υπηρεσιών, η εμπιστοσύνη στην παρεχόμενη πληροφορία και η ακεραιότητα υπηρεσιών και πληροφοριών είναι αβέβαιες.
- Δεν αντιμετωπίζεται το πρόβλημα της ποιότητας των παρεχόμενων υπηρεσιών (quality of service) παρά μόνο περιστασιακά.
- Παρ' όλο που η διάδοση δυναμικής πληροφορίας ή πληροφορίας με μικρή διάρκεια ζωής είναι εύκολη στον Παγκόσμιο Ιστό, η παροχή πρόσβασης σε πόρους (resources) είναι προβληματική.
- Δεν αντιμετωπίζονται τα ζητήματα των εκδόσεων (versions) και των προηγούμενων «τευχών» (back issues).
- Οι πόροι και τα ονόματά τους αλλάζουν πολύ συχνά. Αν κάποιος πόρος αλλάξει τοποθεσία ή όνομα, πρέπει να ενημερωθούν όλες οι ιστοσελίδες που αναφέρονται σ' αυτόν. Αν αυτό δε γίνει, αντιμετωπίζονται τα γνωστά προβλήματα των «σπασμένων συνδέσμων» (broken links), που δεν οδηγούν πουθενά.
- Το πρόβλημα της αρχειοθέτησης δεν έχει λυθεί ακόμα.
- Δεν αντιμετωπίζεται το πρόβλημα των πνευματικών δικαιωμάτων.
- Η υποστήριξη που παρέχεται για *πολύγλωσσο (multilingual) περιεχόμενο* (ειδικά όταν ένας πόρος είναι σε γλώσσα που δε χρησιμοποιεί το λατινικό αλφάβητο) γίνεται με μη-τυποποιημένο (non-standard) τρόπο.

Επιπλέον, παρέχονται περιορισμένες δυνατότητες *διαγλωσσικής ανάκτησης (cross-language retrieval)*, που θα επέτρεπε οι τιμές των όρων που θα περιέχονται στις ερωτήσεις να μπορούν να εκφραστούν σε οποιαδήποτε γλώσσα, ενώ θα επιστρέφονται

τα αντικείμενα που περιέχουν τους όρους αυτούς σε όλες τις γλώσσες που ενδιαφέρουν το χρήστη.

- Η ανάκτηση (recall) για κάθε ερώτηση είναι πολύ μεγάλη ενώ η ακρίβεια (precision) περιορισμένη.
- Οι ερωτήσεις που υποστηρίζονται ψάχνουν σε ολόκληρα τα έγγραφα (full text search) και θεωρούν ισοδύναμες τις εμφανίσεις όρων σε οποιοδήποτε σημείο του εγγράφου (έτσι, π.χ., μια λέξη που εμφανίζεται στον τίτλο ενός εγγράφου θεωρείται ισοδύναμη με αυτές που εμφανίζονται σε μια παράγραφο του).
- Συχνά οι απαντήσεις που επιστρέφονται από τις μηχανές αναζήτησης είναι υπερβολικά πολλές και μη ικανοποιητικές τόσο λόγω των περιορισμών που θέτει η HTML (που δεν υποστηρίζει δομημένα έγγραφα) όσο και λόγω του ότι υποστηρίζονται ερωτήσεις με βάση μόνο το περιεχόμενο των εγγράφων και όχι με βάση τη δομή ή τα χαρακτηριστικά τους.
- Οι μηχανές αναζήτησης βασίζονται στη διαφήμιση, με αποτέλεσμα να επιστρέφουν, μαζί με τα αντικείμενα που πληρούν τα κριτήρια που έθεσε ο χρήστης, και αρκετά άσχετα ή ελαχίστως σχετικά αντικείμενα, που αναφέρονται σε πελάτες του οργανισμού που παρέχει τη μηχανή αναζήτησης.

Επιπλέον, οι σελίδες των πελατών του οργανισμού, όταν σχετίζονται, έστω και λίγο, με την ερώτηση του χρήστη, επιστρέφονται ως αντικείμενα που ικανοποιούν σε μεγάλο βαθμό τα κριτήρια της ερώτησης.

- Η κάλυψη σε ότι αφορά τη δεικτοδότηση (indexing) των υπαρχόντων πόρων είναι περιορισμένη και μη προβλέψιμη (περίπου το 1/3 των υπαρχόντων πόρων είναι δεικτοδοτημένο).
- Από τις μηχανές αναζήτησης δεικτοδοτούνται μόνο οι στατικές σελίδες σε μορφή HTML και απλού κειμένου. Έτσι, δε δεικτοδοτούνται τα περιεχόμενα των πηγών δυναμικής πληροφορίας ούτε των συστημάτων που βασίζονται σε βάσεις δεδομένων, όπως είναι τα συστήματα των βιβλιοθηκών. Αυτό σημαίνει ότι πρέπει να γίνουν ξεχωριστές ερωτήσεις στα συστήματα αυτά για να εντοπιστούν αντικείμενα που ίσως ενδιαφέρουν τους χρήστες.
- Δεν παρέχονται μηχανισμοί σχολιασμού (annotators), οι οποίοι θα πρότειναν στους χρήστες αντικείμενα με περιεχόμενο σχετικό με το περιεχόμενο του αντικειμένου που τους παρουσιάζεται κάθε φορά και θα ενεργοποιούνταν και θα απενεργοποιούνταν κάθε φορά από τον ίδιο το χρήστη.
- Δεν υπάρχει κάποιος αυτόματος τρόπος ειδοποίησης των χρηστών για ιστοσελίδες που περιέχουν πληροφορία σχετική με τα ενδιαφέροντά τους.

Για την αντιμετώπιση κάποιων από τα προβλήματα αυτά έχουν αναπτυχθεί τρεις κατηγορίες συστημάτων:

- Οι *Ψηφιακές Βιβλιοθήκες (Digital Libraries)*, που είναι οργανωμένες συλλογές υλικού, το οποίο μπορεί να βρίσκεται σε μορφή πολλών και διαφορετικών μέσων. Οι συλλογές αυτές έχουν σχεδιαστεί έτσι ώστε να καλύπτουν τις ανάγκες μιας συγκεκριμένης

κοινότητας χρηστών. Η δόμησή τους εγγυάται την πρόσβαση στα περιεχόμενά τους και είναι εξοπλισμένες με μέσα πλοήγησης προς άλλες πηγές πληροφορίας.

- Τα *Συστήματα Νέων κατ' Απαίτηση (News on Demand Systems)*, που έχουν σχεδιαστεί για να παρέχουν στους χρήστες τους άμεση και όσο είναι δυνατόν πληρέστερη ενημέρωση. Οι χρήστες των συστημάτων αυτών έχουν τις εξής δυνατότητες:
 - Να επιλέξουν, ανάλογα με τα ενδιαφέροντά τους, από ποιες από τις διαθέσιμες κατηγορίες νέων θα τους παρουσιάζονται νέα.
 - Να καθορίσουν την περίοδο ανανέωσής των νέων.
 - Να καθορίσουν το χρόνο και τον τρόπο παραλαβής των νέων.
 - Να οργανώσουν την παρουσίαση των νέων σύμφωνα με τις προτιμήσεις τους.
 - Να χρησιμοποιήσουν ένα μηχανισμό ειδοποίησης για να ενημερώνονται για το ότι κάποια νέα γίνονται διαθέσιμα.
 - Να αναζητήσουν στις ιστορικές βάσεις δεδομένων των παροχών ενημέρωσης παλαιότερες ειδήσεις που τους ενδιαφέρουν.
- Οι *Διαμεσολαβητές (Mediators)*, που είναι συστήματα που επιτρέπουν στους χρήστες τους ομοιόμορφη πρόσβαση σε ετερογενείς πηγές πληροφορίας. Αυτό σημαίνει ότι οι χρήστες αυτών των συστημάτων θα μπορούν να ψάξουν ταυτόχρονα σε πολλές πηγές πληροφορίας (ακόμα κι αν η δομή και η οργάνωσή τους διαφέρει από πηγή σε πηγή) και να δουν με ομοιόμορφο τρόπο τα αποτελέσματα των ερωτήσεών τους.

Αυτό που πρέπει πάντως να σημειωθεί είναι ότι και οι τρεις κατηγορίες συστημάτων που αναφέρθηκαν παραπάνω αποτελούν επιμέρους λύσεις για την κάλυψη συγκεκριμένων αναγκών των χρηστών και σε καμία περίπτωση δε λύνουν τα προβλήματα που αναφέρθηκαν.

Προς την κατεύθυνση της λύσης των προβλημάτων αυτών, τα τελευταία χρόνια ορίζεται μια «νέα γενιά» προτύπων για τον Παγκόσμιο Ιστό, τα οποία έχουν ως στόχο την επίλυση των προβλημάτων που καθιστούν την κατάσταση στον Παγκόσμιο Ιστό προβληματική και θα επιτρέψουν την παροχή υψηλού επιπέδου υπηρεσιών μέσω αυτού. Τα πρότυπα αυτά όμως δεν έχουν δεν έχουν ενσωματωθεί ακόμα πλήρως στον Παγκόσμιο Ιστό ούτε υποστηρίζονται πλήρως από τα υπάρχοντα συστήματα Ψηφιακών Βιβλιοθηκών, Συστημάτων Νέων κατ' Απαίτηση και Διαμεσολαβητών.

1.1. Αναγκαιότητα – Συνεισφορά

Όπως έχει ήδη φανεί, παρ' όλο που ο Παγκόσμιος Ιστός έχει κάνει ευκολότερη την πρόσβαση των χρηστών, ειδικών και μη, σε πληροφορία που τους ενδιαφέρει, υποφέρει από ένα σύνολο προβλημάτων και περιορισμών που δεν επιτρέπουν τον εύκολο και γρήγορο εντοπισμό της πληροφορίας αυτής. Αυτό σημαίνει ότι το χαρακτηριστικό που λείπει από τον Παγκόσμιο Ιστό είναι η παροχή λειτουργικότητας αντίστοιχης με αυτή που παρέχουν οι βάσεις δεδομένων, που διαθέτουν μηχανισμούς που εγγυώνται τη γρήγορη και εύκολη πρόσβαση στην πληροφορία που αποθηκεύεται σ' αυτές. Η λειτουργικότητα αυτή μπορεί να δοθεί από ένα σύστημα που θα έχει τα εξής χαρακτηριστικά:

- *Σταθερότητα*: Οι χρήστες του συστήματος δεν πρέπει να έρχονται αντιμέτωποι με εμπόδια, όπως είναι οι σπασμένοι σύνδεσμοι, που δεν τους επιτρέπουν να έχουν πρόσβαση στην πληροφορία που τους ενδιαφέρει.
- *Παροχή Ομοιόμορφης Πρόσβασης σε Ετερογενείς Πηγές Πληροφορίας*: Οι χρήστες του συστήματος πρέπει να μπορούν να προσπελάσουν ταυτόχρονα και με ομοιόμορφο τρόπο πληροφορία αποθηκευμένη σε (τοπικές ή απομακρυσμένες ως προς το σύστημα) ετερογενείς πηγές πληροφορίας (π.χ. παραδοσιακές βάσεις δεδομένων, πληροφοριακά συστήματα πολυμέσων, πηγές ημι-δομημένης ή αδόμητης πληροφορίας κ.α.). Μ' αυτό τον τρόπο θα δοθεί ταυτόχρονη πρόσβαση τόσο σε στατικά έγγραφα όσο και σε πηγές πληροφορίας που βασίζονται σε βάσεις δεδομένων, όπως είναι οι Ψηφιακές Βιβλιοθήκες.
- *Υποστήριξη Ανάκτησης Πληροφορίας (Information Retrieval) μέσω της υποβολής ερωτήσεων*: Οι χρήστες του συστήματος πρέπει να μπορούν να ανακτήσουν την πληροφορία που τους ενδιαφέρει μέσω ερωτήσεων που θα περιέχουν τόσο *όρους δίτιμης λογικής (boolean terms)* όσο και *όρους ομοιότητας (similarity terms)*. Οι ερωτήσεις αυτές μπορεί να υποβάλλονται σε περισσότερες από μια πηγές πληροφορίας κάθε φορά
- *Υποστήριξη Ανάκτησης Αντικειμένων με βάση το Περιεχόμενο ή τη Δομή τους*: Για τα δομημένα αντικείμενα (όπως είναι τα XML έγγραφα), πρέπει να παρέχεται στο χρήστη η δυνατότητα να ανακτήσει αυτά που τον ενδιαφέρουν με βάση τόσο το περιεχόμενο όσο και τη δομή τους.

Με την ιδιαίτερη μεταχείριση των δομημένων αντικειμένων θα αυξηθεί η ακρίβεια σε ότι αφορά τα αποτελέσματα των ερωτήσεων, καθώς ο τόπος εμφάνισης ενός όρου μέσα σ' ένα αντικείμενο θα καθορίζει τη βαρύτητά του για το αντικείμενο αυτό (π.χ. μια λέξη που θα εμφανίζεται στον τίτλο ενός εγγράφου θα θεωρείται μεγαλύτερης βαρύτητας από αυτές που εμφανίζονται σε μια παράγραφο του). Επιπλέον, οι χρήστες θα μπορούν να εκφράσουν με μεγαλύτερη ακρίβεια τα κριτήρια που πρέπει να πληρούν τα αντικείμενα που τους ενδιαφέρουν.

- *Κατηγοριοποίηση και Περιγραφή των Αντικειμένων*: Με την κατάταξη των αντικειμένων σε κατηγορίες βάσει του περιεχομένου τους είναι ευκολότερο για τους χρήστες να εντοπίσουν τα αντικείμενα που τους ενδιαφέρουν. Η περιγραφή των αντικειμένων θα τα εμπλουτίσει με ένα σύνολο χαρακτηριστικών που θα βοηθήσουν τους χρήστες να εκφράσουν με μεγαλύτερη ακρίβεια τα κριτήρια αναζήτησης για κάθε ερώτηση.
- *Υποστήριξη Παροχής Πληροφοριών κατ' Απαίτηση (Information on Demand)*: Στους χρήστες πρέπει να παρέχεται η δυνατότητα να ανακτούν την πληροφορία που τους ενδιαφέρει με βάση τα ενδιαφέροντα και τις προτιμήσεις τους, όπως αυτά καταγράφονται στο διάγραμμα (profile) τους.
- *Δυνατότητες Εναλλακτικής Παρουσίασης Αντικειμένων*: Ο χρήστης του συστήματος πρέπει να έχει τη δυνατότητα να προσπελάσει κάθε αντικείμενο με εναλλακτικούς τρόπους: ολόκληρο το αντικείμενο, σε μια συμπυκνόμενη μορφή (που θα περιέχει συνδέσμους προς την υπόλοιπη πληροφορία του αντικειμένου) ή μόνο κάποια επιλεγμένα δομικά στοιχεία του αντικειμένου.

- *Υποστήριξη Αυτόματης Ειδοποίησης*: Το σύστημα πρέπει να διαθέτει μηχανισμούς ειδοποίησης των χρηστών, οι οποίοι θα τους γνωστοποιούν ότι κάποιοι πόροι που σχετίζονται με τα ενδιαφέροντά τους έγιναν διαθέσιμοι.

Από τα παραπάνω φαίνεται ότι ένα πραγματικά χρήσιμο σύστημα πρέπει να συνδυάζει χαρακτηριστικά από τις ψηφιακές βιβλιοθήκες, τους διαμεσολαβητές και τα συστήματα νέων κατ' απαίτηση και να έχει ενσωματωθεί σ' αυτό η αιχμή της τεχνολογίας των προτύπων για τον Παγκόσμιο Ιστό. Ένα τέτοιο σύστημα είναι το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» (*«Information Provision on Demand»*) που περιγράφεται στην παρούσα διατριβή. Μια πρώτη περιγραφή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» γίνεται στο τμήμα 1.2.

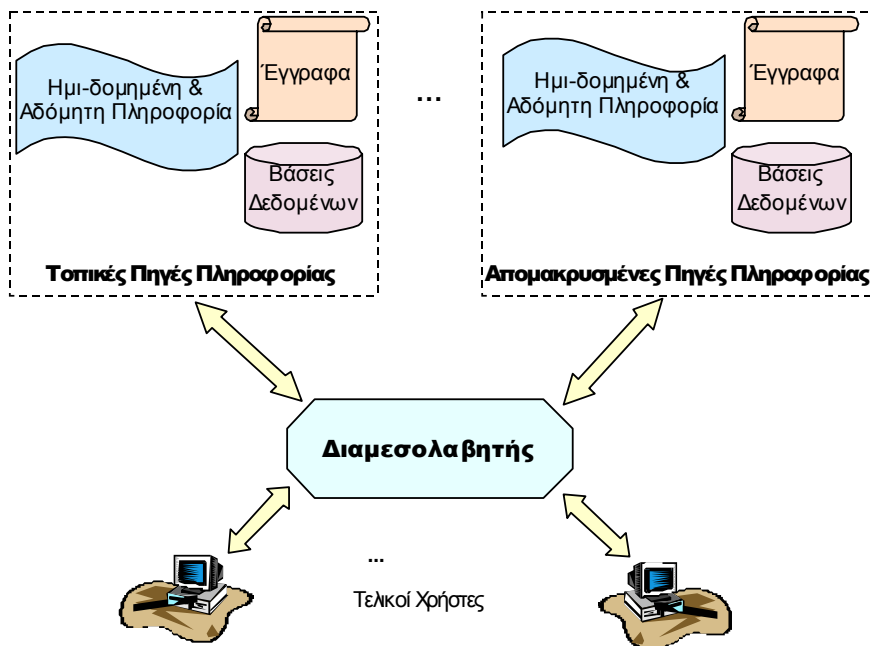
1.2. Περιγραφή του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»

Στο τμήμα αυτό θα γίνει μια πρώτη περιγραφή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Τα στιγμιότυπα του συστήματος αυτού φιλοδοξούν να παίξουν το ρόλο του παροχέα θεματικής πληροφορίας στον Παγκόσμιο Ιστό. Για την κάλυψη των αναγκών αυτών, το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» διαθέτει χαρακτηριστικά τόσο των ψηφιακών βιβλιοθηκών, όσο και των διαμεσολαβητών και των συστημάτων νέων κατ' απαίτηση. Επιπλέον, στο σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» έχει ενσωματωθεί η αιχμή της τεχνολογίας των προτύπων για τον Παγκόσμιο Ιστό. Οι υπηρεσίες που παρέχονται στους χρήστες του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» είναι οι εξής:

- Παροχή ομοιόμορφης πρόσβασης σε ετερογενείς πηγές πληροφορίας, με ιδιαίτερη έμφαση στην υποστήριξη πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
- Υποστήριξη ισχυρών ερωτήσεων στα περιεχόμενα των πηγών πληροφορίας. Πιο συγκεκριμένα, δίνονται στους χρήστες οι εξής δυνατότητες:
 - Δυνατότητες υποβολής ερωτήσεων που θα περιέχουν τόσο όρους δίτιμης λογικής όσο και όρους ομοιότητας.
 - Δυνατότητες ανάκτησης αντικειμένων με βάση το περιεχόμενο και/ή τη δομή τους.
 - Δυνατότητες ανάκτησης αντικειμένων με βάση την κατηγορία στην οποία έχει ενταχθεί το περιεχόμενό τους.
- Υποστήριξη ανάκτησης πληροφορίας κατ' απαίτηση, με βάση τις προτιμήσεις των χρηστών όπως αυτές καταγράφονται στα διαγράμματά τους.
- Υποστήριξη μηχανισμών σχολιασμού, που θα προτείνουν στους χρήστες αντικείμενα σχετικά με τα ενδιαφέροντά τους.
- Υποστήριξη πολύγλωσσου περιεχομένου, με ιδιαίτερη έμφαση στην υποστήριξη της ελληνικής γλώσσας.
- Δυνατότητες εναλλακτικής παρουσίασης αντικειμένων, ανάλογα με τις προτιμήσεις των χρηστών.

- Υποστήριξη αυτόματης ειδοποίησης των χρηστών για το ότι γίνονται διαθέσιμα αντικείμενα που τους ενδιαφέρουν.

Στο σημείο αυτό πρέπει να τονιστεί ότι το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» είναι σταθερό, με την έννοια ότι έχει ληφθεί ιδιαίτερη μέριμνα ώστε να εξαλειφθεί (ή έστω να ελαχιστοποιηθεί) το φαινόμενο των «σπασμένων συνδέσμων». Μια απλουστευμένη άποψη της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» δίνεται στο *Σχήμα 1-1*.



Σχήμα 1-1: Αρχιτεκτονική του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»

Όπως φαίνεται και στο *Σχήμα 1-1*, η αρχιτεκτονική του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» απαρτίζεται από ένα *Διαμεσολαβητή (Mediator)* και ένα σύνολο *Πηγών Πληροφορίας (Information Sources)*. Ο χρήστης προσπελαίνει την πληροφορία που τον ενδιαφέρει μέσω του Διαμεσολαβητή, ο οποίος παρέχει ομοιόμορφη προσπέλαση σε ετερογενείς πηγές πληροφορίας που είναι είτε τοπικές είτε απομακρυσμένες σε σχέση με το διαμεσολαβητή. Οι πηγές πληροφορίας είναι τα πληροφοριακά συστήματα που περιέχουν την πληροφορία που γίνεται διαθέσιμη μέσω του Διαμεσολαβητή και μπορεί να είναι παραδοσιακές βάσεις δεδομένων, πληροφοριακά συστήματα πολυμέσων, πηγές ημι-δομημένης ή αδόμητης πληροφορίας κ.α. Ο χρήστης μπορεί να βρει εύκολα και γρήγορα την πληροφορία που τον ενδιαφέρει, καθώς βλέπει ένα ενιαίο μοντέλο πληροφορίας που προέρχεται από τη σύνθεση των μοντέλων πληροφορίας των πηγών πληροφορίας που το περιεχόμενό τους ανταποκρίνεται στα ενδιαφέροντά του. Ιδιαίτερη έμφαση στα πλαίσια του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» δίνεται στην υποστήριξη πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Όπως φαίνεται από τα παραπάνω, η «καρδιά» του συστήματος είναι ο διαμεσολαβητής, ο οποίος αποτελεί το σύστημα διεπαφής χρήστη-συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»: Ο χρήστης εκφράζει τις ερωτήσεις και τις αιτήσεις ανάκτησης πληροφορίας με

γραφικό τρόπο, και αυτές μετατρέπονται σε αιτήσεις εκφρασμένες στη γλώσσα του διαμεσολαβητή. Στη συνέχεια, ο διαμεσολαβητής στέλνει στις (τοπικές ή απομακρυσμένες σε σχέση με αυτόν) πηγές πληροφορίας τις αιτήσεις και, όταν λάβει τις απαντήσεις, τις παρουσιάζει στο χρήστη μορφοποιημένες ανάλογα με τις προτιμήσεις του τελευταίου.

Στο σημείο αυτό πρέπει να αναφερθεί ότι σημαντικό ρόλο στο σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» έπαιξε η εμπειρία σε Ψηφιακές Βιβλιοθήκες και Διαμεσολαβητές που αποκτήθηκε στα πλαίσια του προγράμματος VENIVA (VENetlan Virtual Archive), στο οποίο συμμετείχε το Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων και Εφαρμογών του Πολυτεχνείου Κρήτης.

1.3. Στόχοι της Διατριβής

Στα πλαίσια της παρούσας διατριβής και της διατριβής [Πουτ2000] έλαβε χώρα ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».

Ο σχεδιασμός τόσο της συνολικής αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» όσο και των τμημάτων από τα οποία αυτή απαρτίζεται, βασίστηκε στο συνδυασμό των ώριμων τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό (XML, RDF, XQL κλπ.). Επιπλέον, ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» βασίστηκε στη χρήση της UML για την περιγραφή της αρχιτεκτονικής και της συμπεριφοράς του συστήματος.

Η παρούσα διατριβή ασχολήθηκε με τη συνολική αρχιτεκτονική του συστήματος, τους διαμεσολαβητές, τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, ενώ η διατριβή [Πουτ2000] επικεντρώνεται στην υποστήριξη τεχνικών ανάκτησης πληροφορίας και την υποστήριξη πολύγλωσσου περιεχομένου και διαγλωσσικής ανάκτησης. Έτσι, οι στόχοι της παρούσας διατριβής ήταν οι εξής:

- Ο σχεδιασμός, με βάση το συνδυασμό των τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό, της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Η αρχιτεκτονική του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» απαρτίζεται από ένα *Διαμεσολαβητή (Mediator)* και ένα σύνολο *Πηγών Πληροφορίας (Information Sources)*, τα περιεχόμενα των οποίων γίνονται διαθέσιμα μέσω του Διαμεσολαβητή.
- Ο σχεδιασμός του Διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Ο στόχος αυτός περιλαμβάνει:
 - Την περιγραφή του διαμεσολαβητή.
 - Τον καθορισμό της λειτουργικότητας του διαμεσολαβητή.
 - Τον καθορισμό της αρχιτεκτονικής του διαμεσολαβητή.
 - Τον καθορισμό των συστημάτων διεπαφής του διαμεσολαβητή με τις πηγές πληροφορίας και με τους χρήστες.

- Τον καθορισμό, την περιγραφή και το σχεδιασμό των δομικών στοιχείων του διαμεσολαβητή. Ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας του διαμεσολαβητή και τη διαχείριση των δομημένων μεταδεδομένων που χρησιμοποιούνται στο διαμεσολαβητή για την περιγραφή της δομής και του περιεχομένου των πηγών πληροφορίας που γίνονται διαθέσιμες μέσω του διαμεσολαβητή αυτού.
- Ο σχεδιασμός των Πηγών Πληροφορίας που περιέχουν δομημένα έγγραφα. Ο στόχος αυτός περιλαμβάνει:
 - Την περιγραφή των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Τον καθορισμό της λειτουργικότητας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Τον καθορισμό της αρχιτεκτονικής των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Τον καθορισμό, την περιγραφή και το σχεδιασμό των δομικών στοιχείων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας και της βάσης δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Τον καθορισμό της αναπαράστασης δομημένων εγγράφων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, με τρόπο που να επιτρέπει την αποδοτική αναπαράσταση XML εγγράφων.
 - Τον καθορισμό του τρόπου διαχείρισης των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Στον καθορισμό του τρόπου διαχείρισης των πηγών πληροφορίας εμπεριέχεται και ο σχεδιασμός των εργαλείων και των μηχανισμών που υποστηρίζουν τη διαχείριση των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Στον όρο «διαχείριση των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα» περιλαμβάνονται οι εξής λειτουργίες:
 - Η δημιουργία πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Η υποστήριξη κλάσεων δομημένων εγγράφων από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και η δυνατότητα εξέλιξης του σχήματος της βάσης δεδομένων τους ώστε να μπορούν να υποστηριχθούν νέες κλάσεις δομημένων εγγράφων.
 - Η εισαγωγή δομημένων εγγράφων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Η αναζήτηση δομημένων εγγράφων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Η ανάκτηση δομημένων εγγράφων από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Η ενημέρωση των πηγών πληροφορίας για έγγραφα που γίνονται διαθέσιμα και το περιεχόμενό τους σχετίζεται με το περιεχόμενο των πηγών

πληροφορίας. Η ενημέρωση αυτή έχει ως συνέπεια είτε την εισαγωγή είτε τη δεικτοδότηση, μέσα στις πηγές πληροφορίας, των εγγράφων αυτών.

- Ο σχεδιασμός των Πηγών Πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Ο στόχος αυτός περιλαμβάνει:
 - Την περιγραφή των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
 - Τον καθορισμό της λειτουργικότητας των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
 - Τον καθορισμό της αρχιτεκτονικής των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
 - Τον καθορισμό, την περιγραφή και το σχεδιασμό των δομικών στοιχείων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας και της βάσης δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
 - Τον καθορισμό του τρόπου διαχείρισης των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Στον καθορισμό του τρόπου διαχείρισης των πηγών πληροφορίας εμπεριέχεται και ο σχεδιασμός των εργαλείων και των μηχανισμών που υποστηρίζουν τη διαχείριση των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Στην έννοια «διαχείριση των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα» περιλαμβάνονται, όπως και για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, οι λειτουργίες της δημιουργίας των πηγών πληροφορίας, της εισαγωγής πληροφορίας σ' αυτές, της αναζήτησης πληροφορίας, της ανάκτησης πληροφορίας και της ενημέρωσης για αντικείμενα που γίνονται διαθέσιμα και σχετίζονται με το περιεχόμενο των συγκεκριμένων πηγών πληροφορίας.

1.4. Δομή της Διατριβής

Η παρούσα διατριβή συνεχίζεται στο 2^ο κεφάλαιο, όπου παρουσιάζεται η σχετική με την παρούσα διατριβή έρευνα. Πιο συγκεκριμένα, γίνεται αναφορά σε συστήματα που έρχονται να καλύψουν τις ίδιες ή παρόμοιες ανάγκες (ή μέρος των αναγκών) που καλύπτει το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» που αναπτύχθηκε στα πλαίσια της παρούσας διατριβής και της διατριβής [Πουτ2000] και στην έρευνα που γίνεται σε θέματα που πραγματεύεται η παρούσα διατριβή (Ψηφιακές Βιβλιοθήκες, Διαμεσολαβητές, Συστήματα Νέων κατ' Απαίτηση, Μηχανές Αναζήτησης και Πράκτορες). Επιπλέον, δίνεται μια σύντομη περιγραφή του συστήματος που αναπτύχθηκε στα πλαίσια του ευρωπαϊκού ερευνητικού προγράμματος VENIVA, καθώς ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» βασίστηκε στην εμπειρία που αποκτήθηκε από το πρόγραμμα αυτό.

Στη συνέχεια, στο 3^ο κεφάλαιο γίνεται αναφορά στη συσχετιζόμενη με την παρούσα διατριβή εργασία, στην οποία συμπεριλαμβάνονται τα πρότυπα που καθοδήγησαν το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Έτσι, γίνεται αναφορά:

- Στη UML (Unified Modeling Language), που χρησιμοποιήθηκε για το σχεδιασμό και την περιγραφή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».
- Σε πρότυπα αναπαράστασης εγγράφων, που επηρέασαν το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» και ενσωματώθηκαν σ' αυτόν.
- Σε πρότυπα αναπαράστασης μεταδεδομένων, που χρησιμοποιήθηκαν στο σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση».
- Στο DOM (Document Object Model) που επηρέασε το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».
- Σε γλώσσες αναζήτησης και διαχείρισης εγγράφων, που επηρέασαν το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» και ενσωματώθηκαν, ολικά ή μερικά, σ' αυτόν.

Στο 4^ο κεφάλαιο γίνεται μια παρουσίαση της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Πιο συγκεκριμένα, δίνεται μια γενική περιγραφή της αρχιτεκτονικής του συστήματος και γίνονται σύντομες αναφορές στα τμήματα που την απαρτίζουν. Ιδιαίτερη έμφαση δίνεται στα τμήματα το σχεδιασμό των οποίων πραγματεύεται η παρούσα διατριβή.

Στο 5^ο κεφάλαιο περιγράφεται ο Διαμεσολαβητής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Πιο συγκεκριμένα, δίνεται μια γενική περιγραφή του διαμεσολαβητή και καθορίζονται η λειτουργικότητά του, η αρχιτεκτονική του και τα συστήματα διεπαφής του διαμεσολαβητή με τις πηγές πληροφορίας και με τους χρήστες. Ιδιαίτερη έμφαση δίνεται στην περιγραφή της γλώσσας και της βάσης δεδομένων του διαμεσολαβητή, που αποτελούν συνεισφορά της παρούσας διατριβής.

Στο 6^ο κεφάλαιο περιγράφονται οι πηγές πληροφορίας. Πιο συγκεκριμένα, περιγράφονται η λειτουργικότητά και η αρχιτεκτονική των πηγών πληροφορίας, ενώ ιδιαίτερη έμφαση δίνεται στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Τόσο στην περιγραφή των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα όσο και στην περιγραφή των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα, ιδιαίτερη έμφαση δίνεται στην περιγραφή της γλώσσας και της βάσης δεδομένων τους. Ειδικά για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, έμφαση δίνεται στην αναπαράσταση των εγγράφων στις σχεσιακές βάσεις δεδομένων όπου αποθηκεύονται τα περιεχόμενα των πηγών πληροφορίας αυτής της κλάσης πηγών πληροφορίας.

Στο 7^ο κεφάλαιο περιγράφεται η διαχείριση των πηγών πληροφορίας, με ιδιαίτερη έμφαση στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Πιο συγκεκριμένα, περιγράφονται τα εργαλεία και οι μηχανισμοί που υποστηρίζουν τις λειτουργίες:

- Της δημιουργίας πηγών πληροφορίας.
- Της υποστήριξης κλάσεων και της δυνατότητα εξέλιξης του σχήματος και του μοντέλου πληροφορίας της βάσης δεδομένων των πηγών πληροφορίας ώστε να μπορούν να υποστηριχθούν από αυτές νέες κλάσεις.
- Της εισαγωγής αντικειμένων στις πηγές πληροφορίας.

- Της αναζήτησης αντικειμένων και τμημάτων αντικειμένων στις πηγές πληροφορίας.
- Της ανάκτησης αντικειμένων και τμημάτων αντικειμένων από τις πηγές πληροφορίας.
- Της δεικτοδότησης, μέσα στις πηγές πληροφορίας, απομακρυσμένης ως προς αυτές πληροφορίας.
- Της ενημέρωσης των πηγών πληροφορίας για πληροφορία που γίνεται διαθέσιμη και σχετίζεται με το περιεχόμενο των συγκεκριμένων πηγών πληροφορίας.

Στο 8^ο κεφάλαιο δίνεται ένα παράδειγμα εφαρμογής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» σε ακαδημαϊκό περιβάλλον, με στόχο να γίνουν εμφανείς η χρησιμότητα και η εφαρμοσιμότητα του συστήματος.

Στο 9^ο κεφάλαιο γίνεται μια ανακεφαλαίωση της παρούσας διατριβής, τονίζεται η συνεισφορά της και η διατριβή κλείνει με μια προσπάθεια καταγραφής των μελλοντικών επεκτάσεών της.

2^ο Κεφάλαιο

Σχετική Έρευνα

Στο κεφάλαιο αυτό γίνεται μια αναφορά στη σχετική με την παρούσα διατριβή έρευνα. Πιο συγκεκριμένα, περιγράφονται τα συστήματα που έρχονται να καλύψουν τις ίδιες ή παρόμοιες ανάγκες (ή μέρος των αναγκών) που καλύπτει το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση». Τα συστήματα αυτά κατατάσσονται στις εξής κατηγορίες:

- *Ψηφιακές Βιβλιοθήκες (Digital Libraries)*. Οι ψηφιακές βιβλιοθήκες είναι οργανωμένες συλλογές υλικού των οποίων η δόμηση εγγυάται την πρόσβαση στα περιεχόμενά τους και την πλήρη κάλυψη των αναγκών μιας τουλάχιστον κοινότητας χρηστών. Οι ψηφιακές βιβλιοθήκες περιγράφονται εκτενέστερα στο τμήμα 2.1.
 - *Συστήματα Νέων κατ' Απαίτηση (News on Demand Systems)*. Τα συστήματα Νέων κατ' Απαίτηση είναι συστήματα που παρέχουν στους χρήστες τους υπηρεσίες άμεσης ενημέρωσης βάσει των ενδιαφερόντων τους. Τα συστήματα Νέων κατ' Απαίτηση περιγράφονται εκτενέστερα στο τμήμα 2.3.
 - *Διαμεσολαβητές (Mediators)*. Οι διαμεσολαβητές είναι συστήματα που παρέχουν στους χρήστες τους ομοιόμορφη πρόσβαση στα περιεχόμενα ετερογενών πηγών πληροφορίας. Εκτενέστερη αναφορά στους διαμεσολαβητές γίνεται στο τμήμα 2.2.
 - *Μηχανές Αναζήτησης (Search Engines)*. Οι μηχανές αναζήτησης είναι συστήματα που γίνονται διαθέσιμα μέσω του Παγκοσμίου Ιστού και επιτρέπουν τον εντοπισμό αντικειμένων που το περιεχόμενό τους πληρεί συγκεκριμένα κριτήρια και γίνονται
-

διαθέσιμα μέσω του Παγκοσμίου Ιστού. Οι μηχανές αναζήτησης εξετάζονται στο τμήμα 2.4.

- *Πράκτορες (Agents)*. Οι πράκτορες είναι προγράμματα που ενεργούν για λογαριασμό των χρηστών τους, με σκοπό κυρίως τον εντοπισμό πληροφορίας που ενδιαφέρει τους τελευταίους. Εκτενέστερη αναφορά στους πράκτορες γίνεται στο τμήμα 2.5.

Επιπλέον, αναφορά γίνεται στην έρευνα σε θέματα που πραγματεύεται η παρούσα διατριβή όπως είναι η αναπαράσταση δομημένων εγγράφων σε βάσεις δεδομένων στο τμήμα 2.6.

Αναφορά γίνεται και στο πρόγραμμα VENIVA (τμήμα 2.7), καθώς ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» βασίστηκε στην εμπειρία που αποκτήθηκε από το πρόγραμμα αυτό.

Τέλος, το κεφάλαιο κλείνει με μια σύντομη *Ανακεφαλαίωση* στο τμήμα 2.8.

2.1. Ψηφιακές Βιβλιοθήκες

Σ' αυτό το τμήμα γίνεται αναφορά στις *Ψηφιακές Βιβλιοθήκες (Digital Libraries)* και στην έρευνα που διεξάγεται στα σχετικά με αυτές θέματα. Οι *Ψηφιακές Βιβλιοθήκες* είναι οργανωμένες συλλογές υλικού που μπορεί να βρίσκεται σε μορφή πολλών και διαφορετικών μέσων. Οι ψηφιακές βιβλιοθήκες έχουν σχεδιαστεί έτσι ώστε να καλύπτουν πλήρως τις ανάγκες μιας τουλάχιστον κοινότητας χρηστών. Η δόμησή τους εγγυάται την πρόσβαση στα περιεχόμενά τους τόσο μέσω μηχανισμών πλοήγησης όσο και μέσω μηχανισμών αναζήτησης πληροφορίας. Επιπλέον, οι ψηφιακές βιβλιοθήκες είναι εξοπλισμένες με μέσα πλοήγησης προς άλλες πηγές πληροφορίας.

Όπως φαίνεται από τον παραπάνω ορισμό, οι ψηφιακές βιβλιοθήκες είναι συστήματα που για την ικανοποιητική κάλυψη των απαιτήσεων των χρηστών τους θέτουν την ανάγκη εξέλιξης όπου αυτό είναι απαραίτητο και ενσωμάτωσης σε αυτές της αιχμής της τεχνολογίας (state of the art) σε αρκετά ερευνητικά πεδία: Ο μεγάλος όγκος δεδομένων που αυτές περιέχουν κάνει απαραίτητη την υποστήριξή τους από πολύ μεγάλες βάσεις δεδομένων, το γεγονός ότι το περιεχόμενο των ψηφιακών βιβλιοθηκών μπορεί να απαρτίζεται από αντικείμενα πολυμέσων (και μάλιστα από *αντικείμενα ευαίσθητα σε καθυστερήσεις – delay sensitive objects*) θέτει απαιτήσεις σε ότι αφορά την αποθήκευση και τη διαχείριση των αντικειμένων αυτών, ενώ η ανάγκη για ευκολότερη και αποτελεσματικότερη επικοινωνία με τους χρήστες απαιτεί την παροχή εξελιγμένων *γραφικών επιπέδων διεπαφής (graphical user interfaces)*. Επιπλέον, η απαίτηση παροχής στους χρήστες δυνατοτήτων αναζήτησης των αντικειμένων που τους ενδιαφέρουν στις ψηφιακές βιβλιοθήκες και πλοήγησης μεταξύ των περιεχομένων τους θέτει την ανάγκη υποστήριξης τόσο δομών υπερμέσων όσο και εξελιγμένων δυνατοτήτων αναζήτησης αντικειμένων βάσει του περιεχομένου και/ή της δομής τους μέσω ερωτήσεων που περιέχουν όρους δίτιμης λογικής και/ή όρους ομοιότητας. Αυτό μπορεί να επιτευχθεί με την παροχή δυνατοτήτων *ανάκτησης πληροφορίας (information retrieval)* και περιγραφής των αντικειμένων με τη χρήση *μεταδεδομένων (metadata)*. Τέλος, καθώς με την εξάπλωση του Διαδικτύου έχει γίνει πραγματικότητα η παγκοσμιοποίηση της γνώσης, απαραίτητη είναι η παροχή *πολύγλωσσης υποστήριξης (multilanguage support)* από τις ψηφιακές βιβλιοθήκες: Η παροχή δυνατοτήτων *διαγλωσσικής ανάκτησης (cross-language retrieval)*, η προσπάθεια

ανάπτυξης μηχανισμών αυτόματης μετάφρασης και η ανάπτυξη και η παροχή πρόσβασης σε διαγλωσσικούς θησαυρούς είναι οι συχνότερα χρησιμοποιούμενοι τρόποι για την επίτευξη του στόχου αυτού.

Στη συνέχεια αυτού του τμήματος έμφαση θα δοθεί σε θέματα που αφορούν το σχεδιασμό και την αρχιτεκτονική των ψηφιακών βιβλιοθηκών, καθώς τα υπόλοιπα θέματα που σχετίζονται με τις ψηφιακές βιβλιοθήκες είτε δε σχετίζονται άμεσα με την παρούσα διατριβή είτε εξετάζονται σε άλλα σημεία της διατριβής.

Η ανάγκη ενσωμάτωσης στις ψηφιακές βιβλιοθήκες της αιχμής της τεχνολογίας σε αρκετά ερευνητικά πεδία θέτει τεράστιες απαιτήσεις σε ότι αφορά τη *διαλειτουργικότητα (interoperability)* σε θέματα που σχετίζονται με τις ψηφιακές βιβλιοθήκες [Pae98]. Στην κατεύθυνση αυτή, ο σχεδιασμός των ψηφιακών βιβλιοθηκών τα τελευταία χρόνια στηρίζεται στη χρήση κοινών πρακτικών και *προτύπων (standards)*.

Μια από τις βασικότερες τάσεις που υπάρχουν σε ότι αφορά την αρχιτεκτονική των ψηφιακών βιβλιοθηκών είναι η παροχή πρόσβασης στα περιεχόμενά τους μέσω του Παγκοσμίου Ιστού [AMM97] [FLM98] [NCSTRL]. Ο κυριότερος λόγος γι αυτό είναι το ότι η προσέγγιση αυτή παρέχει τη δυνατότητα πρόσβασης στα περιεχόμενα των βιβλιοθηκών σε απομακρυσμένους χρήστες, οι οποίοι επιπλέον θα χρησιμοποιούν το οικείο σ' αυτούς επίπεδο διεπαφής του προγράμματος πλοήγησης του Παγκοσμίου Ιστού που προτιμούν.

Καθώς στην παρούσα φάση υπάρχει σε εξέλιξη η διαδικασία αναδιοργάνωσης και επανασχεδιασμού του Παγκοσμίου Ιστού μέσω του ορισμού νέων προτύπων που θα επιτρέψουν την αποδοτικότερη λειτουργία του και τη δυνατότητα παροχής επιπλέον υπηρεσιών μέσω αυτού, τα τελευταία συστήματα ψηφιακών βιβλιοθηκών που αναπτύσσονται ενσωματώνονται σιγά-σιγά τα πρότυπα αυτά. Παραδείγματα τέτοιων συστημάτων είναι τα συστήματα FEDORA [PL98], Aurora [MPP98] και Alexandria [FFFH+98].

Στην κατεύθυνση της υποστήριξης της διαλειτουργικότητας σε περιβάλλοντα ψηφιακών βιβλιοθηκών γίνεται μια προσπάθεια ορισμού γλωσσών που καθορίζουν τους κανόνες αναπαράστασης, διαχείρισης και αποθήκευσης αντικειμένων στις βάσεις δεδομένων που υποστηρίζουν τις ψηφιακές βιβλιοθήκες [AMM97] [Chan98] [Bour99] [ST98].

Τέλος, ιδιαίτερο ενδιαφέρον παρουσιάζουν οι προσπάθειες χρήσης της τεχνολογίας των πρακτόρων στις ψηφιακές βιβλιοθήκες τόσο για την εξασφάλιση της δεικτοδότησης πληροφορίας [WN99] [HNBG+98] που αποθηκεύεται σε άλλες πηγές πληροφορίας όσο και για την ενημέρωση δομών που υποστηρίζουν την ανάκτηση πληροφορίας στις ψηφιακές βιβλιοθήκες, όπως είναι τα μεταδεδομένα [DLP98] [Norg98] [Wein98] και τα λεξικά καταλογογράφησης [Norg98].

2.2. Διαμεσολαβητές

Το θέμα αυτού του τμήματος είναι οι *Διαμεσολαβητές (Mediators)*. Οι *Διαμεσολαβητές* είναι συστήματα που παρέχουν στους χρήστες τους ομοιόμορφη πρόσβαση στα περιεχόμενα ετερογενών πηγών πληροφορίας. Αυτό σημαίνει ότι οι χρήστες αυτών των συστημάτων θα μπορούν να ψάξουν ταυτόχρονα σε πολλές πηγές πληροφορίας (ακόμα κι αν η δομή και η

οργάνωσή τους διαφέρει από πηγή σε πηγή) και να δουν με ομοιόμορφο τρόπο τα αποτελέσματα των ερωτήσεών τους.

Η ανάπτυξη των διαμεσολαβητών έγινε αναγκαία μετά την εκρηκτική εξάπλωση του Διαδικτύου, η οποία είχε ως συνέπεια τη δημιουργία ενός μεγάλου αριθμού πηγών πληροφορίας, καθεμιά από τις οποίες παρείχε διαφορετικό *επίπεδο διεπαφής (interface)* στους χρήστες, τόσο στο επίπεδο του *γραφικού επιπέδου διεπαφής (graphical user interface)* όσο και στο επίπεδο της λειτουργικότητας που παρέχει και της γλώσσας μέσω της οποίας παρέχει τη λειτουργικότητα αυτή. Το γεγονός αυτό δημιούργησε την ανάγκη παροχής δυνατοτήτων ενοποίησης της πρόσβασης σε διαφορετικές πηγές πληροφορίας, την οποία κάλυψαν οι διαμεσολαβητές [BJN97] [AMM98.2].

Η λειτουργία των διαμεσολαβητών βασίζεται στην παρουσίαση στους χρήστες ενός ιδεατού μοντέλου πληροφορίας, το οποίο προέρχεται από το συνδυασμό των μοντέλων πληροφορίας των πηγών πληροφορίας των οποίων τα περιεχόμενα παρέχεται πρόσβαση μέσω του διαμεσολαβητή¹. Οι ερωτήσεις και οι αιτήσεις των χρηστών ικανοποιούνται από τις πηγές πληροφορίας: Ο διαμεσολαβητής τις παραπέμπει σε αυτές μέσω των *περιτυλιγμάτων (wrappers)* τους [AK97] [RS97] [HGN+97] [Sah98] [KWD97] [CG98], τα οποία και είναι υπεύθυνα για τη μετάφραση των ερωτήσεων και των αιτήσεων από μορφή σύμφωνη με τη γλώσσα και το μοντέλο πληροφορίας που εξάγεται από το διαμεσολαβητή σε μορφή σύμφωνη με τις γλώσσες και τα μοντέλα πληροφορίας των πηγών πληροφορίας. Στη συνέχεια, αφού οι ερωτήσεις και οι αιτήσεις εκτελεστούν στις πηγές πληροφορίας, τα αποτελέσματα που επιστρέφουν οι πηγές πληροφορίας μετατρέπονται σε μορφή σύμφωνη με τη γλώσσα και το μοντέλο πληροφορίας του διαμεσολαβητή και κανονικοποιούνται τα βάρη των επιστρεφόμενων αντικειμένων. Τα περιτυλίγματα των πηγών πληροφορίας μπορεί να δημιουργούνται χειρωνακτικά, αυτόματα ή ημιαυτόματα.

Ανοικτά ερευνητικά θέματα που σχετίζονται με τους διαμεσολαβητές [AFLM99] είναι οι τεχνικές μετάφρασης που χρησιμοποιούνται από τα περιτυλίγματα [FKL97] [MBP98] και θέματα που σχετίζονται με τις δυσκολίες της εκτέλεσης ερωτήσεων σε διαφορετικές πηγές πληροφορίας, όπως είναι οι ερωτήσεις σε συχνά ημιδομημένες πηγές πληροφορίας [GMHW99] και η κανονικοποίηση των βαρών των αντικειμένων που επιστρέφονται ως αποτελέσματα των ερωτήσεων που τίθενται σε διαφορετικές πηγές πληροφορίας [GGM97] [VP97] [HKWY97].

Διαμεσολαβητές αναπτύχθηκαν στα πλαίσια των γνωστών συστημάτων TSIMMIS [CGMH+94] [PGMU96] [GM95], Informia [Bar98], Garlic [RS97], Disco [FRV96] [TRV97], Hermes [ACPS96], SIMS [AKS96], Internet SoftBot [EW94] [KW96], Aquarelle [Vel99] [Mic98], VENIVA [Tsin98] και Information Manifold [LRO96].

2.3. Συστήματα Νέων κατ' Απαίτηση

Σ' αυτό το τμήμα περιγράφονται τα *Συστήματα Νέων κατ' Απαίτηση (News on Demand Systems)*. Τα *Συστήματα Νέων κατ' Απαίτηση* έχουν σχεδιαστεί για να παρέχουν στους

¹ Το μοντέλο πληροφορίας που παρουσιάζεται σε κάποιο χρήστη μπορεί να προέρχεται από το συνδυασμό μόνο των μοντέλων πληροφορίας των πηγών πληροφορίας των οποίων τα περιεχόμενα ενδιαφέρουν το συγκεκριμένο χρήστη.

χρήστες τους άμεσα και όσο είναι δυνατόν πληρέστερη ενημέρωση, βάσει του περιεχομένου που γίνεται διαθέσιμο από κάποιο παροχέα ενημέρωσης (π.χ. κάποιο τηλεοπτικό ή ραδιοφωνικό σταθμό, κάποια εφημερίδα κλπ.).

Τα σύγχρονα συστήματα Νέων κατ' Απαίτηση χρησιμοποιούν βάσεις δεδομένων για την αποθήκευση και διαχείριση των ειδήσεων. Επιπλέον των βάσεων δεδομένων, χρησιμοποιούν μεθόδους *φιλτραρίσματος (filtering)* της πληροφορίας, οι οποίες επιτρέπουν την αυτόματη προώθηση στους χρήστες μόνο των ειδήσεων που ενδιαφέρουν τους τελευταίους. Η δυνατότητα αυτή, μαζί με τη δυνατότητα καθορισμού του τρόπου παρουσίασης της πληροφορίας στους χρήστες βάσει των προτιμήσεών τους και της υπολογιστικής πλατφόρμας² που χρησιμοποιούν, αποτελούν την υπηρεσία της *εξατομίκευσης (personalization)*.

Με βάση τα παραπάνω, τα συστήματα Νέων κατ' Απαίτηση βασίζονται σε δύο κατηγορίες μηχανισμών:

- Τους *Μηχανισμούς Προώθησης (Push Mechanisms)* [Mins99], που είναι υπεύθυνοι για το φιλτράρισμα των ειδήσεων βάσει των προτιμήσεων των χρηστών. Οι μηχανισμοί αυτοί απαρτίζονται από *φίλτρα (filters)* και *ιεραρχίες κατηγοριών (categorization hierarchies)*.
- Τους *Μηχανισμούς Ανέλκυσης (Pull Mechanisms)*, οι οποίοι παρέχουν στους χρήστες τη δυνατότητα ανάκτησης αρχειοθετημένων ειδήσεων. Οι μηχανισμοί ανέλκυσης είναι, στην πραγματικότητα, οι ιστορικές βάσεις δεδομένων των παροχέων πληροφορίας.

Έτσι, οι χρήστες των συστημάτων Νέων κατ' Απαίτηση έχουν τις εξής δυνατότητες:

- Να επιλέξουν, ανάλογα με τα ενδιαφέροντά τους, από ποιες από τις διαθέσιμες κατηγορίες νέων θα τους παρουσιάζονται νέα.
- Να καθορίσουν την περίοδο ανανέωσής των νέων.
- Να καθορίσουν το χρόνο και τον τρόπο παραλαβής των νέων.
- Να οργανώσουν την παρουσίαση των νέων σύμφωνα με τις προτιμήσεις τους.
- Να χρησιμοποιήσουν ένα μηχανισμό ειδοποίησης για να ενημερώνονται για το ότι κάποια νέα γίνονται διαθέσιμα.
- Να αναζητήσουν στις ιστορικές βάσεις δεδομένων των παροχέων ενημέρωσης παλαιότερες ειδήσεις που τους ενδιαφέρουν.

Παραδείγματα συστημάτων Νέων κατ' Απαίτηση στην Αμερική είναι το CNN (www.cnn.com), το MSNBC (www.msnbc.com), και στην Ελλάδα τα συστήματα αθλητικών ειδήσεων SportLine (www.sport.gr) και SportNews (www.sportnews.gr). Ένα κοινό χαρακτηριστικό των συστημάτων αυτών είναι ότι υποστηρίζουν μία ή, το πολύ, δύο γλώσσες. Το θέμα της υποστήριξης πολλαπλών γλωσσών σε συστήματα Νέων κατ' Απαίτηση αντιμετωπίζεται τόσο στο σύστημα που αναπτύχθηκε στα πλαίσια του ερευνητικού προγράμματος Hynode [Hynode] [Σκον98], όπου υποστηρίζονται τρεις

² Η υπολογιστική πλατφόρμα κάποιου χρήστη απαρτίζεται από το υλικό (hardware), το λειτουργικό σύστημα και τις εφαρμογές που αυτός χρησιμοποιεί.

γλώσσες, όσο και στο σύστημα “Hypermedia Custom News System” [Καρασ98] [Πετρ98], που υποστηρίζει οποιονδήποτε αριθμό γλωσσών.

Ιδιαίτερο ενδιαφέρον παρουσιάζει το σύστημα Νέων κατ’ Απαίτηση “Krakatoa Cronicle” [Bha98], που υποστηρίζει την εξισορρόπηση των ατομικών ενδιαφερόντων με αυτά των κοινοτήτων χρηστών, επιτρέποντας στους χρήστες να πλοηγηθούν σε ένα αριθμό εφημερίδων χρησιμοποιώντας διαφορετικές σκοπίες (viewpoints). Επιπλέον υποστηρίζει τη δυναμική ενημέρωση των διαγραμμάτων προτίμησης των χρηστών με τη χρήση τεχνικών *ανάδρασης συσχέτισης (relevance feedback)* και *εξαγωγής συμπερασμάτων (inference)*.

Μια διαφορετική προσέγγιση συστήματος Νέων κατ’ Απαίτηση, που επιτρέπει τη δημιουργία μιας «δια-εφημεριδιακής» ηλεκτρονικής εφημερίδας με το συνδυασμό πληροφορίας που περιέχεται σε διαφορετικές εφημερίδες μέσω της χρήσης πρακτόρων (agents) προτείνεται από το σύστημα Hypernews [Mor96]. Το σύστημα, για την αποφυγή προβλημάτων πνευματικών δικαιωμάτων, επιβάλλει στους χρήστες την πολιτική της «πληρωμής ανά χρήση» (“pay-per-use”).

Τέλος, θέματα που αποτελούν πεδίο έρευνας στην περιοχή των συστημάτων Νέων κατ’ Απαίτηση (ή σχετίζονται πολύ μ’ αυτά) είναι η αυτόματη κατηγοριοποίηση εγγράφων [Coh95] [Coh96] [Bour99], ώστε τα νέα που γίνονται διαθέσιμα να κατηγοριοποιούνται άμεσα βάσει του περιεχομένου τους και ο εντοπισμός της αλλαγής των ενδιαφερόντων των χρηστών, ώστε να τροποποιηθούν κατάλληλα τα διαγράμματα προτιμήσεών τους [Cra98].

2.4. Μηχανές Αναζήτησης

Το θέμα αυτού του τμήματος είναι οι *Μηχανές Αναζήτησης (Search Engines)*. Οι μηχανές αναζήτησης είναι συστήματα διαθέσιμα στον Παγκόσμιο Ιστό, που επιτρέπουν τον εντοπισμό αντικειμένων που γίνονται διαθέσιμα μέσω του Παγκοσμίου Ιστού και το περιεχόμενό τους πληρεί συγκεκριμένα κριτήρια.

Σχεδόν όλοι οι χρήστες του Παγκοσμίου Ιστού έχουν εμπειρία από τη χρήση μηχανών αναζήτησης. Ιδιαίτερα δημοφιλείς στο ευρύ κοινό μηχανές αναζήτησης είναι οι Altavista (www.altavista.com), Lycos (www.lycos.com), Excite (www.excite.com), Yahoo! (www.yahoo.com), InfoSeek (www.infoseek.com), HotBot (www.hotbot.com), ενώ ιδιαίτερα δημοφιλής στους ακαδημαϊκούς κύκλους είναι η μηχανή αναζήτησης Google (www.google.com) [Page98] [Cho98].

Ο σημαντικότερες διαφοροποιήσεις που συναντώνται μεταξύ των μηχανών αναζήτησης εμφανίζονται στα εξής χαρακτηριστικά τους:

- Την πολιτική δημιουργίας και ενημέρωσης των *δομών δεικτοδότησης (indexing structures)* που υποστηρίζουν τη μηχανή αναζήτησης. Οι πολιτικές που ακολουθούνται στον τομέα αυτό είναι δύο: Η πολιτική της *αυτόματης ενημέρωσης* των δομών δεικτοδότησης και η πολιτική της *χειρωνακτικής (manual) ενημέρωσης*.

Η πολιτική της αυτόματης ενημέρωσης περιλαμβάνει την ενημέρωση των δομών δεικτοδότησης από κάποιο *ρομπότ (robot)* [Kos95], το οποίο διασχίζει το Διαδίκτυο και δεικτοδοτεί, βάσει του περιεχομένου τους, τα αντικείμενα που συναντά.

Η πολιτική της χειρονακτικής δεικτοδότησης περιλαμβάνει την ενημέρωση των δομών δεικτοδότησης από άτομα εξειδικευμένα στο χαρακτηρισμό του περιεχομένου των αντικειμένων και την κατηγοριοποίηση των αντικειμένων βάσει του χαρακτηρισμού αυτού.

Το βασικό πλεονέκτημα της αυτόματης ενημέρωσης είναι η πλήρης (ή έστω σχεδόν πλήρης) κάλυψη του Διαδικτύου, η οποία είναι πολύ δύσκολο να επιτευχθεί με τη χειρονακτική δεικτοδότηση. Στον αντίποδα, η χειρονακτική δεικτοδότηση παρέχει ακριβέστερη δεικτοδότηση, καθώς οι ειδικοί στο χαρακτηρισμό του περιεχομένου των αντικειμένων είναι πολύ πιο δύσκολο να εξαπατηθούν απ' ό,τι τα ρομπότ.

- Τον αριθμό των αντικειμένων που έχουν δεικτοδοτηθεί από τη μηχανή αναζήτησης.
- Την *κάλυψη (coverage)*, θεματική ή γεωγραφική, που παρέχει η μηχανή αναζήτησης.
- Τη συχνότητα ενημέρωσης και αναδιοργάνωσης των δομών δεικτοδότησης που υποστηρίζουν τη μηχανή αναζήτησης.
- Τη μέθοδο διάσχισης του Διαδικτύου, η οποία και καθορίζει τη σειρά επίσκεψης των διευθύνσεων στο Διαδίκτυο. Οι μέθοδοι που ακολουθούνται είναι δύο: Η μέθοδος *διάσχισης-πρώτα-σε-βάθος (depth-first)* και η μέθοδος *διάσχισης-πρώτα-σε-μήκος (breadth-first)*.

Η μέθοδος διάσχισης-πρώτα-σε-βάθος επισκέπτεται την πρώτη διεύθυνση της αρχικής λίστας διευθύνσεων που διαθέτει και συνεχίζει επαναληπτικά με τις διευθύνσεις προς τις οποίες αυτή έχει συνδέσμους και τις διευθύνσεις προς τις οποίες έχουν συνδέσμους οι διευθύνσεις με τις οποίες συνδέεται η αρχική διεύθυνση πριν επισκεφθεί την επόμενη διεύθυνση της αρχικής λίστας διευθύνσεων.

Αντίθετα, η μέθοδος διάσχισης-πρώτα-σε-μήκος επισκέπτεται πρώτα όλες τις διευθύνσεις της αρχικής λίστας, στη συνέχεια τις διευθύνσεις προς τις οποίες αυτές έχουν συνδέσμους κ.ο.κ.

- Τη μέθοδο δεικτοδότησης, που καθορίζει αν θα δεικτοδοτούνται τα έγγραφα βάσει ολόκληρου του περιεχομένου τους ή βάσει λέξεων-κλειδιών. Καθώς όμως η τελευταία προσέγγιση είχε ως συνέπεια την ευκολότερη εξαπάτηση των ρομπότ, οι περισσότερες μηχανές αναζήτησης χρησιμοποιούν σήμερα δεικτοδότηση βάσει ολόκληρου του περιεχομένου των αντικειμένων.
- Τη μέθοδο βαθμολόγησης των αντικειμένων που επιστρέφονται ως αποτελέσματα των ερωτήσεων που θέτουν οι χρήστες.
- Το επίπεδο αλληλεπίδρασης με τους χρήστες, που συνίσταται συνήθως από HTML φόρμες και δυναμικά παραγόμενες HTML σελίδες.
- Τον τρόπο παρουσίασης των αποτελεσμάτων, που καθορίζει τόσο τον αριθμό των επιστρεφόμενων αντικειμένων όσο και την πληροφορία που παρουσιάζεται στα αποτελέσματα σχετικά με τα αντικείμενα αυτά (π.χ. τίτλος μόνο, τίτλος και περιγραφή κλπ.).

Τέλος, πρέπει να σημειωθεί ότι το σημαντικότερο πρόβλημα που παρουσιάζουν οι μηχανές αναζήτησης είναι ο τεράστιος αριθμός αποτελεσμάτων που επιστρέφουν, τα περισσότερα

από τα οποία σχετίζονται ελάχιστα με τα κριτήρια που έθεσε ο χρήστης. Το γεγονός αυτό οφείλεται τόσο στην εξαπάτηση των ρομπότ που χρησιμοποιούνται για την ενημέρωση των δομών δεικτοδότησης που υποστηρίζουν τις μηχανές αναζήτησης όσο και στο ότι συχνά προωθούνται σε υψηλή θέση στα αποτελέσματα των ερωτήσεων (ακόμα και των ελάχιστα σχετικών) σελίδες με διαφημίσεις πελατών του οργανισμού που παρέχει τη μηχανή αναζήτησης.

2.5. Πράκτορες

Σ' αυτό το τμήμα γίνεται αναφορά στους *Πράκτορες (Agents)*. Ο όρος πράκτορας χρησιμοποιείται στο χώρο της πληροφορικής τόσο με τεχνική όσο και με μεταφορική έννοια και αναφέρεται στο λογισμικό που ενεργεί αντίστοιχα με τους παραδοσιακούς πράκτορες για λογαριασμό του χρήστη του [Gil97].

Οι παραδοσιακοί πράκτορες είναι αξιόπιστα άτομα ή οργανισμοί στα οποία έχουν μεταβιβαστεί η ευθύνη και η αρμοδιότητα για την πραγματοποίηση συγκεκριμένων ενεργειών, ως αντιπροσώπων των πελατών τους και για λογαριασμό των τελευταίων.

Οι πράκτορες που χρησιμοποιούνται σε υπολογιστικά συστήματα ονομάζονται *έξυπνοι πράκτορες πληροφόρησης (intelligent information agents)*. Οι έξυπνοι πράκτορες πληροφόρησης είναι υπολογιστικά συστήματα που έχουν πρόσβαση σε πολλαπλές ετερογενείς και γεωγραφικά κατανεμημένες πηγές πληροφορίας.

Με βάση τα παραπάνω και θέτοντας ως πεδίο δράσης των πρακτόρων το Διαδίκτυο, οι υπηρεσίες που αυτοί μπορούν να προσφέρουν είναι οι εξής [Klu98]:

- Αναζήτηση, για λογαριασμό των χρηστών τους, πληροφορίας σε πολλές και ετερογενείς πηγές πληροφορίας και στη συνέχεια ανάκτηση, ανάλυση, ολοκλήρωση και αποθήκευση της πληροφορίας αυτής [AKH96] [Chu96] [DP96] [DS83] [GM95] [KH96].
- Πληροφόρηση των χρηστών τους για νέα αντικείμενα που γίνονται διαθέσιμα [SK97].
- Διαπραγμάτευση, πληρωμή και παραλαβή, για λογαριασμό των χρηστών τους, πληροφορίας, νέων, αγαθών και υπηρεσιών [Ker97.1] [Ker97.2].
- Βαθμολόγηση της συσχέτισης με τα ενδιαφέροντα των χρηστών τους, της ποιότητας και της αξιοπιστίας της πληροφορίας που παρέχεται από κάποια πηγή πληροφορίας [Mot95] [MS96].
- Προσαρμογή σε συνθήκες που αλλάζουν και εξέλιξη βάσει των αλλαγών αυτών [DSW97] [Gai97] [IK97] [KLMW94] [LPH97] [RGK98].

Έτσι, για παράδειγμα, ένας «τουριστικός πράκτορας» ψάχνει να βρει πακέτα διακοπών που καλύπτουν τα ενδιαφέροντα και ικανοποιούν τις απαιτήσεις του χρήστη του. Στη συνέχεια, παρουσιάζει βαθμολογημένα τα αποτελέσματα της έρευνάς του στο χρήστη, και εφ' όσον ο τελευταίος επιλέξει κάποιο πακέτο, διαπραγματεύεται και πραγματοποιεί την αγορά του.

Ανάλογα με τις προδιαγραφές, το ρόλο, το πεδίο εφαρμογών και τον τρόπο λειτουργίας τους οι έξυπνοι πράκτορες πληροφόρησης χωρίζονται στις εξής κατηγορίες:

- *Συνεργάσιμοι (Cooperative)* και *Μη Συνεργάσιμοι (Non-Cooperative)* πράκτορες πληροφόρησης. Οι πράκτορες αυτής της κατηγορίας χαρακτηρίζονται ανάλογα με το αν

έχουν ή όχι τη δυνατότητα να συνεργαστούν με άλλους πράκτορες κατά τη διάρκεια της εκτέλεσης των καθηκόντων τους.

Οι συνεργάσιμοι πράκτορες είτε μοιράζονται με τους πράκτορες με τους οποίους μπορούν να συνεργαστούν κοινές *οντολογίες (ontologies)*³ είτε «καταλαβαίνουν» την ίδια γλώσσα. Οι δημοφιλέστερες γλώσσες επικοινωνίας μεταξύ πρακτόρων (*agent communication languages – ACLs*) είναι η KQML [KQML1] [KQML2] και η FIPA [FIPA1] [FIPA2].

- *Ορθολογιστικοί (Rational)* πράκτορες πληροφόρησης. Οι πράκτορες αυτής της κατηγορίας χρησιμοποιούνται σε οικονομικό επίπεδο και βρίσκουν εφαρμογή σε περιοχές όπως το *ηλεκτρονικό εμπόριο (electronic commerce ή e-commerce)*, οι *αυτόματες εμπορικές συναλλαγές (automated trading)* κλπ.
- *Προσαρμόσιμοι (Adaptive)* πράκτορες πληροφόρησης. Οι πράκτορες αυτής της κατηγορίας έχουν τη δυνατότητα προσαρμογής στις αλλαγές που λαμβάνουν χώρα στα δίκτυα και τα πληροφοριακά περιβάλλοντα στα οποία δραστηριοποιούνται. Παράδειγμα πρακτόρων αυτής της κατηγορίας αποτελούν οι *προσωπικοί βοηθοί εκμάθησης στον Παγκόσμιο Ιστό (personal learning assistants on the Web)*.
- *Κινητοί (Mobile)* πράκτορες πληροφόρησης. Οι πράκτορες αυτής της κατηγορίας έχουν τη δυνατότητα να ταξιδεύουν αυτόνομα στα όρια ενός τοπικού δικτύου ή ενός δικτύου ευρείας περιοχής. Παράδειγμα πρακτόρων αυτής της κατηγορίας αποτελούν οι πράκτορες που εξασφαλίζουν την εξισορρόπηση του φόρτου των κόμβων ενός δικτύου μεγάλης κλίμακας.

2.6. Αναπαράσταση Δομημένων Εγγράφων σε Βάσεις Δεδομένων

Στο τμήμα αυτό γίνεται αναφορά στην αιχμή της τεχνολογίας σχετικά με την αναπαράσταση δομημένων εγγράφων σε βάσεις δεδομένων. Για να υπάρξουν αποδοτικοί μηχανισμοί διαχείρισης εγγράφων, πρέπει η αναπαράσταση των εγγράφων στις βάσεις δεδομένων των πηγών πληροφορίας όπου αποθηκεύονται τα δομημένα έγγραφα να είναι ευέλικτη και να εκμεταλλεύεται όσο είναι δυνατόν καλύτερα τις δυνατότητες του ΣΔΒΔ (Συστήματος Διαχείρισης Βάσεων Δεδομένων) χωρίς να υποφέρει από πλευράς απόδοσης. Επιπλέον, η αναπαράσταση των δομημένων εγγράφων πρέπει να παρέχει δυνατότητες αναφοράς και ανάκτησης των επιμέρους δομικών στοιχείων που απαρτίζουν τα έγγραφα. Καθώς η παρούσα διατριβή αντιμετωπίζει το πρόβλημα της αποθήκευσης δομημένων εγγράφων σε σχεσιακές βάσεις δεδομένων, το τμήμα αυτό επικεντρώνει στην παρουσίαση προσεγγίσεων αναπαράστασης δομημένων εγγράφων που θα μπορούσαν να εφαρμοστούν σε σχεσιακές βάσεις δεδομένων.

Μια πρώτη προσέγγιση στο θέμα αυτό θα ήταν η αναπαράσταση των διαφορετικών κλάσεων δομημένων εγγράφων από οντότητες (*entities*) όπου, ανεξάρτητα από την κλάση στην οποία ανήκει κάθε έγγραφο, το περιεχόμενο κάθε εγγράφου αποθηκεύεται σε ένα

³ Οντολογία (*Ontology*) είναι ένα σύνολο λεξικογραφικών ορισμών που εκφράζουν την άποψη και τη γνώση μιας κοινότητας σχετικά με κάποιο θέμα ή κάποιο πεδίο γνώσης ή έρευνας [WB98].

πεδίο κειμένου (content) και η λογική δομή του εγγράφου αποθηκεύεται σε ένα άλλο πεδίο κειμένου (descriptor). Μ' αυτή την προσέγγιση, η λογική δομή των εγγράφων είναι αυτή που επιτρέπει την προσπέλαση των επιμέρους δομικών στοιχείων που συνθέτουν κάθε έγγραφο [Κοτταν96].

Η παραπάνω προσέγγιση είναι αποδοτική για έγγραφα μικρού και μεσαίου μεγέθους, δεν είναι όμως αρκετά ευέλικτη καθώς προϋποθέτει ότι όλο το περιεχόμενο ενός εγγράφου αποθηκεύεται μαζεμένο. Επίσης, για πολύ μεγάλα έγγραφα μια τέτοια προσέγγιση δεν είναι ιδιαίτερα αποδοτική καθώς η λογική δομή ενός μεγάλου εγγράφου είναι μεγάλη σε μέγεθος και η ανάκτηση και η διαχείρισή της είναι χρονοβόρες διαδικασίες.

Η αντιδιαμετρική αυτής της προσέγγισης προτείνει την εξής αναπαράσταση:

- Κάθε κλάση δομικών στοιχείων που απαρτίζουν δομημένα έγγραφα αναπαρίσταται από μια οντότητα.
- Κάθε κλάση δομημένων εγγράφων αναπαρίσταται από μια οντότητα.
- Όλα τα έγγραφα, ανεξάρτητα από την κλάση στην οποία ανήκουν, διαθέτουν ένα πεδίο που αναπαριστά τη λογική δομή τους και περιέχει τις ταυτότητες των δομικών στοιχείων που τα απαρτίζουν.
- Όλα τα δομικά στοιχεία εγγράφων διαθέτουν ένα πεδίο, όπου αποθηκεύεται η ταυτότητα του εγγράφου το οποίο περιέχει το τρέχον δομικό στοιχείο.
- Όλα τα δομικά στοιχεία εγγράφων που δεν αποτελούν φύλλα της λογικής δομής, ανεξάρτητα από την κλάση στην οποία ανήκουν, διαθέτουν ένα πεδίο που περιέχει τις ταυτότητες των υποστοιχείων τους.
- Όλα τα δομικά στοιχεία εγγράφων που αποτελούν φύλλα της λογικής δομής, ανεξάρτητα από την κλάση στην οποία ανήκουν, διαθέτουν ένα πεδίο όπου αποθηκεύεται το περιεχόμενό τους.

Παρ' όλο που η προσέγγιση αυτή δεν παρουσιάζει τα προβλήματα της πρώτης προσέγγισης, υλοποιείται αποτελεσματικά μόνο σε αντικειμενοστραφή περιβάλλοντα [Τσιν95] [Chris97]. Αντίθετα, σε σχεσιακά περιβάλλοντα η προσέγγιση αυτή υποφέρει από πλευράς απόδοσης.

Μια τρίτη προσέγγιση προτείνει την αποθήκευση των εγγράφων σε απλά αρχεία που διαχειρίζεται το λειτουργικό σύστημα, ενώ στη βάση δεδομένων αποθηκεύεται μόνο μια σχέση-περιγραφέας (descriptor) για κάθε έγγραφο. Ο περιγραφέας αυτός περιέχει τη λογική δομή των εγγράφων (με άμεσες αναφορές στο αρχείο) και τα χαρακτηριστικά τους.

Η προσέγγιση αυτή δεν προσφέρει τίποτα περισσότερο από πλευράς λειτουργικότητας απ' ό,τι η πρώτη προσέγγιση, ενώ αφαιρεί τη δυνατότητα αξιοποίησης εξελιγμένων χαρακτηριστικών των βάσεων δεδομένων (όπως π.χ. η χρήση ειδικών τελεστών για δεδομένα κειμένου).

Τέλος, μια παραλλαγή της 2^{ης} προσέγγισης προτείνει την εξής αναπαράσταση:

- Όλα τα έγγραφα αναπαρίστανται ως στιγμιότυπα της οντότητας «Έγγραφο».
- Όλα τα δομικά στοιχεία εγγράφων αναπαρίστανται ως στιγμιότυπα της οντότητας «Τμήμα Εγγράφου».

- Όλα τα χαρακτηριστικά των εγγράφων και των δομικών στοιχείων από τα οποία απαρτίζονται τα έγγραφα αναπαρίστανται ως στιγμιότυπα της οντότητας «Χαρακτηριστικό».

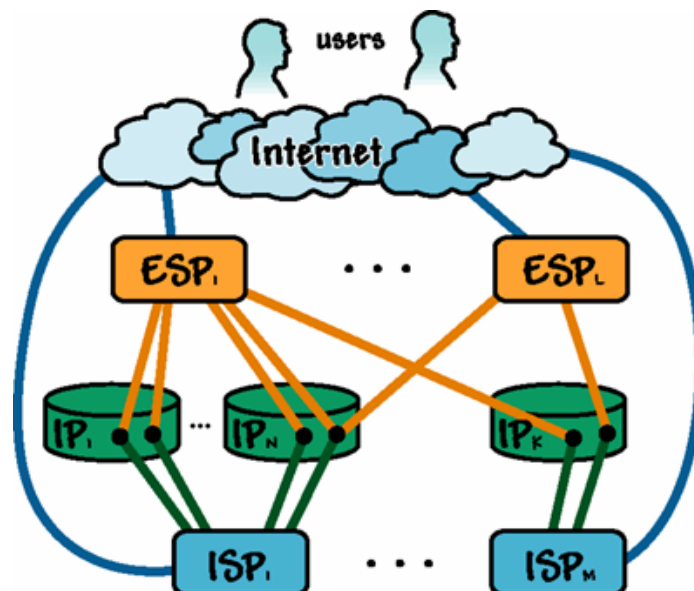
Η προσέγγιση αυτή παρουσιάζει τα ίδια προβλήματα με τη 2^η προσέγγιση, ενώ παρουσιάζει και επιπλέον προβλήματα απόδοσης όσο μεγαλώνει το μέγεθος των πινάκων όπου αποθηκεύονται τα στιγμιότυπα των παραπάνω οντοτήτων.

2.7. Το πρόγραμμα VENIVA

Σ' αυτό το τμήμα δίνεται μια σύντομη περιγραφή του συστήματος που αναπτύχθηκε στα πλαίσια του ευρωπαϊκού ερευνητικού προγράμματος VENIVA (VENetlan Virtual Archive), στο οποίο συμμετείχε το Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων και Εφαρμογών του Πολυτεχνείου Κρήτης. Ο κυριότερος λόγος της αναφοράς στο πρόγραμμα VENIVA είναι το ότι ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» βασίστηκε στην εμπειρία που αποκτήθηκε από το πρόγραμμα αυτό [Tsin98].

Στα πλαίσια του προγράμματος VENIVA αναπτύχθηκε ένα κατακεντρωμένο ιδεατό ιστορικό αρχείο, το οποίο επιτρέπει την ομοιόμορφη πρόσβαση στα περιεχόμενα των βάσεων δεδομένων που διαθέτουν ιδρύματα (και πιο συγκεκριμένα βιβλιοθήκες και ιστορικά αρχεία) που κατέχουν ιστορικό υλικό που σχετίζεται με τη Δημοκρατία της Βενετίας. Επιπλέον, πάνω στο υλικό που διατίθεται από τα ιδρύματα είναι δυνατό να παρασχεθούν από εκδοτικούς οίκους *υπηρεσίες προστιθέμενης αξίας (added value services)*, όπως είναι οι καθοδηγούμενες περιηγήσεις, οι μελέτες, οι εκδόσεις πολυμέσων κλπ.

Η αρχιτεκτονική του συστήματος που αναπτύχθηκε στα πλαίσια του προγράμματος VENIVA απεικονίζεται στο Σχήμα 2-1.



Σχήμα 2-1: Η Αρχιτεκτονική του Συστήματος που αναπτύχθηκε στα πλαίσια του προγράμματος VENIVA

Όπως φαίνεται στο Σχήμα 2-1, η αρχιτεκτονική του συστήματος που αναπτύχθηκε στα πλαίσια του προγράμματος VENIVA απαρτίζεται από:

- *Παροχείς Πληροφορίας (Information Providers - IPs)*. Οι παροχείς πληροφορίας είναι τα ιδρύματα που κατέχουν το ιστορικό υλικό και έχουν την ευθύνη της ψηφιοποίησης του υλικού τους και την εισαγωγή του στις βάσεις δεδομένων τους. Το σχήμα της βάσης δεδομένων όλων των βιβλιοθηκών είναι το ίδιο και βασίζεται στο πρότυπο ISBD [ISBD], ενώ και το σχήμα της βάσης δεδομένων όλων των αρχείων είναι κοινό και βασίζεται στο πρότυπο ISAD [ISAD].

- *Εσωτερικούς Παροχείς Υπηρεσιών (Internal Service Providers - ISPs)*. Οι εσωτερικοί παροχείς υπηρεσιών είναι υπεύθυνοι για την υποστήριξη των παροχών πληροφορίας, μέσω της παροχής εργαλείων και της μεταφοράς τεχνογνωσίας, σε ότι αφορά τη διαχείριση των βάσεων δεδομένων τους και την παροχή πρόσβασης στα περιεχόμενά τους μέσω του Παγκοσμίου Ιστού. Ένας εσωτερικός παροχέας υπηρεσιών μπορεί να υποστηρίξει ένα ή περισσότερους *ομογενείς (homogeneous)*⁴ παροχείς πληροφορίας.

Μέσω ενός εσωτερικού παροχέα υπηρεσιών οι χρήστες του συστήματος μπορούν να πλοηγηθούν στην πληροφορία που γίνεται διαθέσιμη από τους παροχείς πληροφορίας που συνεργάζονται με τον τρέχοντα παροχέα υπηρεσιών και να αναζητήσουν αντικείμενα που τους ενδιαφέρουν μέσω απλών ερωτήσεων.

- *Εξωτερικούς Παροχείς Υπηρεσιών (External Service Providers - ESPs)*. Οι εξωτερικοί παροχείς υπηρεσιών είναι εταιρείες που ενεργούν ως εκδότες και έχουν συμφωνίες με συγκεκριμένους παροχείς πληροφορίας, οι οποίες τους επιτρέπουν να παρέχουν πρόσβαση μέσω του Παγκοσμίου Ιστού στο υλικό που κατέχουν οι παροχείς πληροφορίας με τους οποίους έχουν κάποια συμφωνία. Επιπλέον, οι εξωτερικοί παροχείς υπηρεσιών μπορεί να παρέχουν υπηρεσίες προστιθέμενης αξίας για το υλικό των παροχών πληροφορίας με τους οποίους συνεργάζονται.

Οι εξωτερικοί παροχείς υπηρεσιών, στα πλαίσια του VENIVA, παίζουν το ρόλο του διαμεσολαβητή, καθώς επιτρέπουν στους χρήστες τους να προσπελάσουν με ομοιόμορφο τρόπο τα περιεχόμενα αρχείων και βιβλιοθηκών και να θέσουν ερωτήσεις που περιέχουν *όρους δίτημης λογικής (boolean terms)* και/ή *όρους ομοιότητας (similarity terms)* ταυτόχρονα σε όλες τις βάσεις δεδομένων των παροχών πληροφορίας με τους οποίους συνεργάζονται, ανεξάρτητα από τον τύπο του ιδρύματος του κάθε παροχέα πληροφορίας.

Το Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων και Εφαρμογών του Πολυτεχνείου Κρήτης ήταν υπεύθυνο, στα πλαίσια του προγράμματος VENIVA, για το σχεδιασμό και την υλοποίηση των εξωτερικών παροχών υπηρεσιών. Στο σημείο αυτό πρέπει να σημειωθεί ότι τόσο οι περιορισμοί που εντοπίστηκαν στις προδιαγραφές του συστήματος που αναπτύχθηκε στα πλαίσια του προγράμματος VENIVA⁵ όσο και τα προβλήματα που αντιμετωπίστηκαν λόγω της ανωριμότητας της τεχνολογίας του Παγκοσμίου Ιστού ξεπεράστηκαν στο σχεδιασμό του συστήματος «Παροχής Πληροφοριών

⁴ Ο όρος ομογενείς σημαίνει ότι όλα τα ιδρύματα είναι του ίδιου τύπου (είτε όλα αρχεία είτε όλα βιβλιοθήκες) και οι βάσεις δεδομένων τους έχουν το ίδιο σχήμα.

⁵ Τέτοιοι περιορισμοί είναι η υποστήριξη μόνο δυο προκαθορισμένων τύπων ιδρυμάτων, η υποστήριξη προκαθορισμένων σχημάτων για τις βάσεις δεδομένων των ιδρυμάτων αυτών, η ελλιπής μοντελοποίηση και υποστήριξη δεδομένων κειμένου κλπ.

κατ' Απαίτηση», ο οποίος χρησιμοποιεί πιο ώριμες τεχνολογίες στον Παγκόσμιο Ιστό και φιλοδοξεί να είναι ανοιχτός και χωρίς περιορισμούς.

2.8. Ανακεφαλαίωση

Στο κεφάλαιο αυτό έγινε μια αναφορά στη σχετική με το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» έρευνα. Πιο συγκεκριμένα, έγινε μια σύντομη ανασκόπηση των συστημάτων που έρχονται να καλύψουν τις ίδιες ή παρόμοιες ανάγκες (ή μέρος των αναγκών) που καλύπτει το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» και της έρευνας σε θέματα που πραγματεύεται η παρούσα διατριβή. Τέλος, δόθηκε μια σύντομη περιγραφή του συστήματος που αναπτύχθηκε στα πλαίσια του ευρωπαϊκού ερευνητικού προγράμματος VENIVA, καθώς ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» βασίστηκε στην εμπειρία που αποκτήθηκε από το πρόγραμμα αυτό.

3^ο Κεφάλαιο

Συσχετιζόμενη Εργασία

Στο κεφάλαιο αυτό γίνεται μια αναφορά στη συσχετιζόμενη με την παρούσα διατριβή εργασία. Πιο συγκεκριμένα, αναφορά γίνεται στα πρότυπα που καθοδήγησαν το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» και ενσωματώθηκαν σε αυτόν.

Έτσι, στο τμήμα 3.1 περιγράφεται η *UML (Unified Modeling Language)*, στο τμήμα 3.2 γίνεται αναφορά σε *Πρότυπα Αναπαράστασης Εγγράφων*, στο τμήμα 3.3 περιγράφονται τα σημαντικότερα *Πρότυπα Αναπαράστασης Μεταδεδομένων*, στο τμήμα 3.4 γίνεται αναφορά στο *Document Object Model* και στην ενότητα 3.5 γίνεται αναφορά σε *Γλώσσες Διαχείρισης & Αναζήτησης Εγγράφων*. Το κεφάλαιο κλείνει με μια σύντομη *Ανακεφαλαίωση* στο τμήμα 3.6.

3.1. UML (Unified Modeling Language)

Στο τμήμα αυτό γίνεται μια σύντομη αναφορά στη *UML (Unified Modeling Language)* [UML] [BRJ98] [RJB98] [Fow97], η οποία χρησιμοποιήθηκε για την περιγραφή τόσο του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» όσο και των τμημάτων από τα οποία αυτό απαρτίζεται. Η UML είναι μια γλώσσα μοντελοποίησης που χρησιμοποιείται από τους μηχανικούς λογισμικού και παρέχει ένα σύνολο τύπων διαγραμμάτων για την περιγραφή συστημάτων λογισμικού. Τα διαγράμματα αυτά, που χρησιμοποιούνται κατά το σχεδιασμό των συστημάτων λογισμικού, επιτρέπουν την περιγραφή της αρχιτεκτονικής, της δομής και της συμπεριφοράς των συστημάτων και των συστατικών από τα οποία απαρτίζονται.

Καθένας από τους τύπους διαγραμμάτων που παρέχονται από τη UML χρησιμοποιείται σε διαφορετική φάση του σχεδιασμού κάποιου συστήματος και για την κάλυψη διαφορετικών αναγκών. Έτσι, στη συνέχεια θα γίνει αναφορά στα χαρακτηριστικά εκείνα της UML που χρησιμοποιήθηκαν στα πλαίσια της παρούσας διατριβής: τα *Διαγράμματα Παράταξης (Deployment Diagrams)* (βλέπε ενότητα 3.1.1), τα *Διαγράμματα Περιπτώσεων Χρήσης (Use Case Diagrams)* (βλέπε ενότητα 3.1.2), τα *Διαγράμματα Κλάσεων (Class Diagrams)* (βλέπε ενότητα 3.1.3), τα *Διαγράμματα Αλληλεπίδρασης (Interaction Diagrams)* (βλέπε ενότητα 3.1.4) και τα *Διαγράμματα Δραστηριοτήτων (Activity Diagrams)* (βλέπε ενότητα 3.1.5).

3.1.1. Διαγράμματα Παράταξης (Deployment Diagrams)

Σ' αυτή την ενότητα γίνεται μια σύντομη αναφορά στα *Διαγράμματα Παράταξης (Deployment Diagrams)*. Τα διαγράμματα παράταξης είναι ένα εργαλείο που παρέχεται από τη UML για την απεικόνιση της «χωρικής» διάταξης του συστήματος που περιγράφεται: Τα συστατικά λογισμικού απεικονίζονται πάνω στους κόμβους υλικού στους οποίους βρίσκονται. Με τον τρόπο αυτό απεικονίζονται οι φυσικές σχέσεις μεταξύ συστατικών λογισμικού και υλικού. Ο Πίνακας 3-1 περιέχει τα στοιχεία από τα οποία απαρτίζονται τα διαγράμματα παράταξης.

Όνομα Στοιχείου	Σύμβολο	Περιγραφή
Κόμβος (Node)		Υπολογιστής ή τμήμα υπολογιστή (συνήθως υλικό (hardware))
Συστατικό (Component)		Συστατικό λογισμικού που αποτελεί τμήμα προγράμματος ή εφαρμογής.
Επίπεδο Διεπαφής (Interface)		Επίπεδο διεπαφής μεταξύ συστατικών λογισμικού
Πολλαπλότητα	*	Πολλαπλές εμφανίσεις κόμβου ή συστατικού λογισμικού
Εξάρτηση (Dependency)		Εξάρτηση και/ή αλληλεπίδραση μεταξύ συστατικών λογισμικού
Σύνδεση (Connection)		Μονοπάτι σύνδεσης που επιτρέπει την αλληλεπίδραση

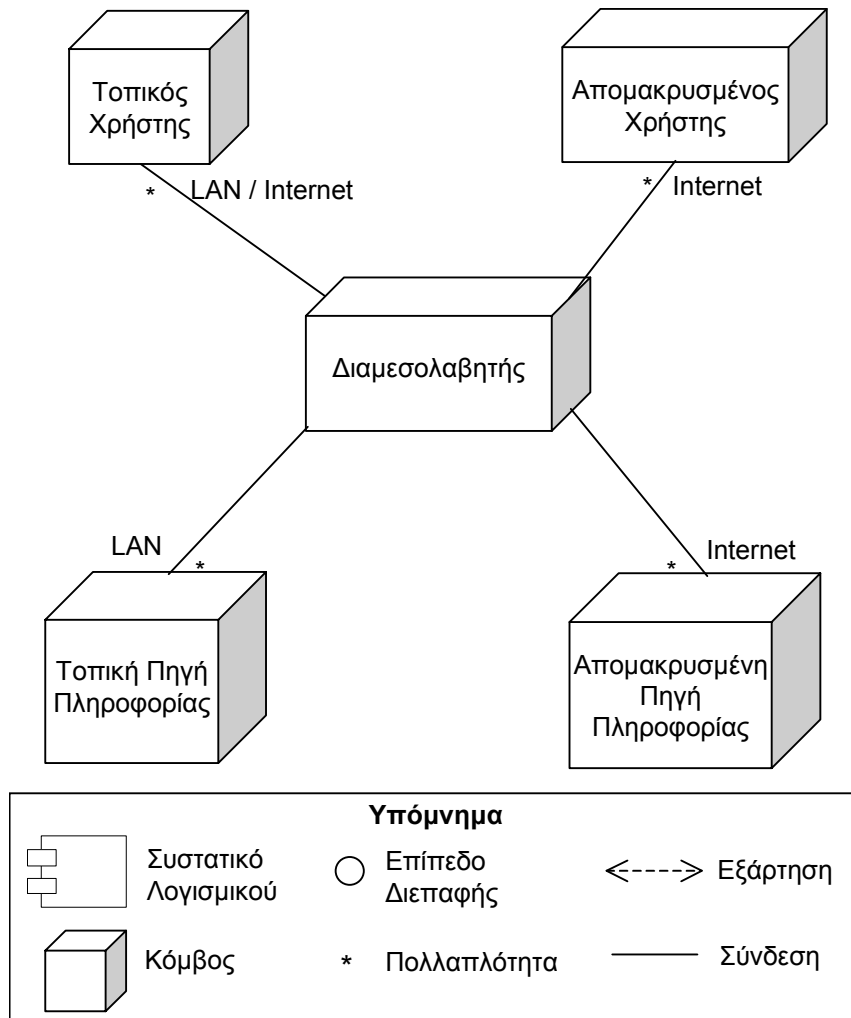
Πίνακας 3-1: Τα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Παράταξης

Όπως φαίνεται στον παραπάνω πίνακα, τα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Παράταξης είναι τα εξής:

- Οι *Κόμβοι (Nodes)*, που αντιπροσωπεύουν τους υπολογιστές ή τα τμήματα υλικού υπολογιστών (hardware) που παίρνουν μέρος στο διάγραμμα.
- Τα *Συστατικά (Components)*, που αντιπροσωπεύουν τα συστατικά λογισμικού που αποτελούν τμήματα των προγραμμάτων ή των εφαρμογών που εμφανίζονται στο διάγραμμα.
- Τα *Επίπεδα Διεπαφής (Interfaces)*, που αναπαριστούν τα επίπεδα διεπαφής μεταξύ συστατικών λογισμικού που εμφανίζονται στο διάγραμμα.
- Τους τελεστές *Πολλαπλότητας*, που αναπαριστούν τις πολλαπλές εμφανίσεις κόμβων ή συστατικών λογισμικού που εμφανίζονται στο διάγραμμα.

- Τις *Εξαρτήσεις (Dependencies)*, που αναπαριστούν τις εξαρτήσεις και/ή την αλληλεπίδραση μεταξύ των συστατικών λογισμικού που εμφανίζονται στο διάγραμμα.
- Τις *Συνδέσεις (Connections)*, που αναπαριστούν τα μονοπάτια σύνδεσης μέσα στο διάγραμμα, τα οποία επιτρέπουν την αλληλεπίδραση.

Ένα απλοποιημένο διάγραμμα παράταξης, αντίστοιχο με το *Σχήμα 1-1*, που περιγράφει το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση», φαίνεται στο *Διάγραμμα 3-1*.



Διάγραμμα 3-1: Απλοποιημένο Διάγραμμα Παράταξης για το Σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση»

3.1.2. Διαγράμματα Περιπτώσεων Χρήσης (Use Case Diagrams)

Το θέμα αυτής της ενότητας είναι η περιγραφή των *Διαγραμμάτων Περιπτώσεων Χρήσης (Use Case Diagrams)*. Τα διαγράμματα περιπτώσεων χρήσης περιγράφουν τη λειτουργικότητα που παρέχεται στους χρήστες και βασίζονται στις απαιτήσεις που αυτοί εκφράζουν. Στα διαγράμματα περιπτώσεων χρήσης φαίνεται επίσης η ροή της πληροφορίας τόσο μεταξύ των χρηστών και των συστατικών του συστήματος όσο και μεταξύ των συστατικών του συστήματος. Ο *Πίνακας 3-2* περιέχει τα στοιχεία από τα οποία απαρτίζονται τα διαγράμματα περιπτώσεων χρήσης.

Όνομα Στοιχείου	Σύμβολο	Περιγραφή
Περίπτωση Χρήσης (Use Case)		Αναπαράσταση της απαιτούμενης λειτουργικότητας
Δρών (Actor)		Ρόλος, στα πλαίσια του συστήματος, κάποιου χρήστη, μιας ομάδας χρηστών ή κάποιων συστατικών λογισμικού
Μονόδρομη Σχέση (Unidirectional Relationship)		Σχέση αλληλεπίδρασης μεταξύ δύο δρώντων με μονόδρομη φορά
Αμφίδρομη Σχέση (Bidirectional Relationship)		Αμφίδρομη σχέση αλληλεπίδρασης μεταξύ δύο δρώντων

Πίνακας 3-2 Τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Περιπτώσεων Χρήσης

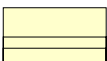
Όπως φαίνεται στον παραπάνω πίνακα, τα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Περιπτώσεων Χρήσης είναι τα εξής:

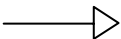
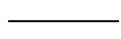
- Οι *Περιπτώσεις Χρήσης (Use Cases)*, που αναπαριστούν στοιχειώδη τμήματα λειτουργικότητας.
- Οι *Δρώντες (Actors)*, που αναπαριστούν τους ρόλους, στα πλαίσια του περιγραφόμενου από το διάγραμμα συστήματος, κάποιου χρήστη, μιας ομάδας χρηστών ή κάποιων συστατικών λογισμικού.
- Οι *Μονόδρομες Σχέσεις (Unidirectional Relationships)*, που αναπαριστούν τις σχέσεις αλληλεπίδρασης με μονόδρομη φορά μεταξύ δύο δρώντων.
- Οι *Αμφίδρομες Σχέσεις (Bidirectional Relationships)*, που αναπαριστούν τις αμφίδρομες σχέσεις αλληλεπίδρασης μεταξύ δύο δρώντων.

3.1.3. Διαγράμματα Κλάσεων (Class Diagrams)

Τα *Διαγράμματα Κλάσεων (Class Diagrams)* είναι το θέμα αυτής της ενότητας. Τα διαγράμματα κλάσεων περιγράφουν τους τύπους αντικειμένων (objects) που υπάρχουν στο περιγραφόμενο σύστημα και τις στατικές σχέσεις που υπάρχουν μεταξύ τους. Έτσι, τα διαγράμματα κλάσεων αναπαριστούν τη δομή εννοιών (concepts), τύπων (types) και κλάσεων (classes). Οι έννοιες αναπαριστούν την άποψη των χρηστών για το σύστημα. Οι τύποι αναπαριστούν το επίπεδο διεπαφής μεταξύ των συστατικών λογισμικού που απαρτίζουν το σύστημα. Τέλος, οι κλάσεις αναπαριστούν την υλοποίηση των συστατικών λογισμικού.

Επιπλέον, τα διαγράμματα κλάσεων περιέχουν τα χαρακτηριστικά (attributes) και τους περιορισμούς που ισχύουν για τον τρόπο διασύνδεσης των αντικειμένων. Ο Πίνακας 3-3 περιέχει τα σημαντικότερα από τα στοιχεία που απαρτίζουν τα διαγράμματα περιπτώσεων χρήσης.

Όνομα Στοιχείου	Σύμβολο	Περιγραφή
Κλάση (Class)		Κλάση Αντικειμένων

Είναι (IsA)		Σχέση «Είναι»
Σχετίζεται (Relates)		Σχέση συσχέτισης
Ομάδα (Aggregation)		Σχέση «Έχει Λίστα Από»

Πίνακας 3-3: Τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Κλάσεων

Όπως φαίνεται στον παραπάνω πίνακα, τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Κλάσεων είναι οι κλάσεις και οι σχέσεις συσχέτισης, «Είναι» και «Έχει Λίστα Από».

3.1.4. Διαγράμματα Αλληλεπίδρασης (Interaction Diagrams)

Σ' αυτή την ενότητα θα περιγραφούν τα Διαγράμματα Αλληλεπίδρασης (*Interaction Diagrams*). Τα διαγράμματα αλληλεπίδρασης χρησιμοποιούνται για την απεικόνιση της αλληλεπίδρασης μεταξύ των αντικειμένων σε κάθε περίπτωση χρήσης. Στις υποενότητες που ακολουθούν περιγράφονται οι δυο τύποι διαγραμμάτων αλληλεπίδρασης που ορίζονται από τη UML: Τα Σειριακά Διαγράμματα (*Sequence Diagrams*) στην υποενότητα 3.1.4.1 και τα Διαγράμματα Συνεργασίας (*Collaboration Diagrams*) στην υποενότητα 3.1.4.2.

3.1.4.1. Σειριακά Διαγράμματα (Sequence Diagrams)

Τα Σειριακά Διαγράμματα (*Sequence Diagrams*) περιγράφονται σ' αυτή την υποενότητα. Τα σειριακά διαγράμματα απεικονίζουν τις αλληλεπιδράσεις που λαμβάνουν χώρα σε κάποια περίπτωση χρήσης με τη μορφή δισδιάστατου γραφήματος. Η κάθετη διάσταση αναπαριστά το χρόνο, ενώ η οριζόντια περιέχει τους δρώντες και τα αντικείμενα που παίρνουν μέρος στην περίπτωση χρήσης που περιγράφεται από το διάγραμμα. Καθένας από τους συμμετέχοντες (αντικείμενο ή δρων) διαθέτει μια διπλή κάθετη γραμμή, που αναπαριστά τις χρονικές στιγμές κατά τις οποίες είναι ενεργός και ονομάζεται *γραμμή της ζωής (lifeline)*. Οι γραμμές της ζωής των συμμετεχόντων συνδέονται μεταξύ τους με μηνύματα, που αναπαρίστανται ως κατευθυνόμενα βέλη από τον αποστολέα προς τον παραλήπτη. Ο Πίνακας 3-4 περιέχει τα σημαντικότερα από τα στοιχεία των σειριακών διαγραμμάτων: Τους Δρώντες, τα Αντικείμενα, τις Γραμμές της Ζωής, τα Μηνύματα και τους τελεστές Πολλαπλότητας.

Όνομα Στοιχείου	Σύμβολο	Περιγραφή
Δρων (Actor)		Δρων που παίρνει μέρος στην περίπτωση χρήσης που περιγράφεται από το διάγραμμα
Αντικείμενο (Object)		Αντικείμενο που παίρνει μέρος στην περίπτωση χρήσης που περιγράφεται από το διάγραμμα
Γραμμή Ζωής (Lifeline)		Γραμμή της ζωής ενός από τους συμμετέχοντες στην περίπτωση χρήσης που περιγράφεται από το διάγραμμα
Μήνυμα (Message)		Μήνυμα που αποστέλλεται από ένα από τους συμμετέχοντες στην περίπτωση χρήσης σε κάποιο άλλο από τους συμμετέχοντες

Πίνακας 3-4: Τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Σειριακά Διαγράμματα

3.1.4.2. Διαγράμματα Συνεργασίας (Collaboration Diagrams)

Το θέμα αυτής της υποενότητας είναι τα *Διαγράμματα Συνεργασίας (Collaboration Diagrams)*. Τα Διαγράμματα Συνεργασίας είναι, στην πραγματικότητα, διαγράμματα κλάσεων που περιέχουν, εκτός από κλάσεις και σχέσεις μεταξύ κλάσεων, ρόλους κλάσεων και σχέσεων. Τα διαγράμματα συνεργασίας μοιάζουν πολύ με τα σειριακά διαγράμματα, με τη διαφορά ότι τα αντικείμενα και οι δρώντες δε διαθέτουν γραμμή ζωής και η χρονική σειρά των αλληλεπιδράσεων καθορίζεται από την αρίθμηση των μηνυμάτων που ανταλλάσσονται. Καθώς για την περιγραφή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» προτιμήθηκε η χρήση σειριακών διαγραμμάτων, δε θα γίνει παραπέρα συζήτηση για τα διαγράμματα συνεργασίας.

3.1.5. Διαγράμματα Δραστηριοτήτων (Activity Diagrams)

Σ' αυτή την υποενότητα δίνεται μια σύντομη περιγραφή των *Διαγραμμάτων Δραστηριοτήτων (Activity Diagrams)*. Τα Διαγράμματα Δραστηριοτήτων απεικονίζουν το συνδυασμό συμπεριφοράς (behavior) και δομής ελέγχου (control structure). Ενθαρρύνουν την παράλληλη συμπεριφορά των αντικειμένων και έχουν τη δυνατότητα απεικόνισης τόσο της συμπεριφορά αντικειμένων σε διαφορετικές περιπτώσεις χρήσης όσο και την υλοποίηση μιας μεθόδου. Ο Πίνακας 3-5 περιέχει τα σημαντικότερα από τα στοιχεία των διαγραμμάτων δραστηριοτήτων.

Όνομα Στοιχείου	Σύμβολο	Περιγραφή
Αρχή (Start)	●	Αρχή της διαδικασίας που περιγράφεται από το διάγραμμα
Τέλος (End)	⦿	Τέλος της διαδικασίας που περιγράφεται από το διάγραμμα
Δραστηριότητα (Activity)	□	Κατάσταση στην οποία είτε γίνεται κάποια διεργασία στον πραγματικό κόσμο είτε εκτελείται κάποια διεργασία λογισμικού
Διαίρεση (Fork)	⌋	Διαίρεση του διαγράμματος σε λειτουργίες που μπορούν να γίνουν παράλληλα
Σύμπτυξη (Join)	⌋	Σύμπτυξη του διαγράμματος μετά από λειτουργίες που μπορούν να γίνουν παράλληλα
Διακλάδωση (Branch) – Συνένωση (Join)	◊	Διακλάδωση ή μετέπειτα συνένωση του διαγράμματος μετά τον έλεγχο κάποιας συνθήκης
Έλεγχος (Guard)	[...]	Έλεγχος για κάποια συνθήκη
Πολλαπλότητα	*	Πολλαπλές εκτελέσεις δραστηριότητας

Πίνακας 3-5: Τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Δραστηριοτήτων

Όπως φαίνεται στον παραπάνω πίνακα, τα σημαντικότερα στοιχεία από τα οποία απαρτίζονται τα Διαγράμματα Δραστηριοτήτων είναι τα εξής:

- Το σύμβολο της *Αρχής* του διαγράμματος, που δηλώνει την αρχή της διαδικασίας που περιγράφεται από το διάγραμμα.
- Το σύμβολο του *Τέλους* του διαγράμματος, που δηλώνει το τέλος της διαδικασίας που περιγράφεται από το διάγραμμα.
- Οι *Δραστηριότητες* (*Activities*), που αναπαριστούν τις καταστάσεις στις οποίες είτε γίνεται κάποια διεργασία στον πραγματικό κόσμο είτε εκτελείται κάποια διεργασία λογισμικού.
- Οι *Διαιρέσεις* (*Forks*), που αναπαριστούν τις διαιρέσεις του διαγράμματος σε λειτουργίες που μπορούν να γίνουν παράλληλα.
- Οι *Συμπήξεις* (*Joins*), που αναπαριστούν τις συμπήξεις του διαγράμματος μετά από λειτουργίες που μπορούν να γίνουν παράλληλα
- Το σύμβολο της *Διακλάδωσης* (*Branch*) και της *Συνένωσης* (*Join*), που αναπαριστά τις διακλαδώσεις και τις μετέπειτα συνενώσεις του διαγράμματος μετά τον έλεγχο κάποιας συνθήκης.
- Οι *Έλεγχοι* (*Guards*), που αναπαριστούν τους ελέγχους που γίνονται σε διάφορα σημεία του διαγράμματος.
- Οι τελεστές *Πολλαπλότητας*, που αναπαριστούν τις πολλαπλές εκτελέσεις κάποιας δραστηριότητας.

Πρέπει να σημειωθεί ότι, για να είναι πιο ξεκάθαρος ο ρόλος κάθε αντικειμένου σε ένα διάγραμμα δραστηριοτήτων, χρησιμοποιούνται οι swimlanes, που χωρίζουν τα διαγράμματα δραστηριοτήτων σε ζώνες. Για τις δραστηριότητες που λαμβάνουν χώρα σε καθεμιά από τις ζώνες αυτές υπεύθυνα είναι τα στιγμιότυπα μιας συγκεκριμένης κλάσης ή κάποιοι συγκεκριμένοι δρώντες.

3.2. Πρότυπα Αναπαράστασης Εγγράφων

Σ' αυτό το τμήμα θα γίνει μια αναφορά στα σημαντικότερα πρότυπα αναπαράστασης εγγράφων πολυμέσων. Ιδιαίτερη έμφαση θα δοθεί στην περιγραφή της XML, καθώς αυτή αποτελεί την επιλογή που καθοδήγησε το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» και ενσωματώθηκε σ' αυτόν.

Η σημασία των *Εγγράφων Πολυμέσων* (*Multimedia Documents*)⁶ για μια σειρά εφαρμογών, όπως είναι τα *Συστήματα Αυτοματισμού Γραφείου* (*Office Automation Systems*), τα *Συστήματα Υπερκειμένου και Υπερμέσων* (*Hypertext and Hypermedia Systems*), τα *Συστήματα Νέων κατ' Απαίτηση* (*News on Demand Systems*), ο Παγκόσμιος Ιστός, οι *Ψηφιακές Βιβλιοθήκες* (*Digital Libraries*) κ.α. οδήγησαν στον καθορισμό προτύπων αναπαράστασης εγγράφων πολυμέσων. Αυτό συνέβη γιατί η χρήση προτύπων για την αναπαράσταση εγγράφων πολυμέσων παρέχει τα εξής πλεονεκτήματα:

⁶ Στο σημείο αυτό πρέπει να σημειωθεί ότι, παρ' όλο που γίνεται αναφορά σε έγγραφα πολυμέσων, αυτά που παίζουν κεντρικό ρόλο είναι τα τμήματά τους που περιέχουν κείμενο, καθώς αυτά καθορίζουν και τη δομή των εγγράφων.

- Τα πρότυπα υποστηρίζουν *δομημένα έγγραφα (structured documents)*. Με τον όρο δομημένα έγγραφα περιγράφονται τα έγγραφα που διαθέτουν τόσο περιεχόμενο (με τη μορφή κειμένου, εικόνων κλπ.) όσο και ενδείξεις για το ρόλο που έχει για το έγγραφο το περιεχόμενο αυτό⁷. Έτσι τυποποιούνται οι κανόνες στους οποίους βασίζεται η δόμηση των εγγράφων.
- Παρέχονται τυποποιημένοι τρόποι αναπαράστασης εγγράφων.
- Υποστηρίζεται η διαλειτουργικότητα (interoperability) περιεχομένου και παρουσίασης μεταξύ εφαρμογών και πλατφορμών [Bos98].
- Οι παραγωγοί περιεχομένου απελευθερώνονται από τον έλεγχο των προμηθευτών λογισμικού (software vendors).
- Οι χρήστες μπορούν να επιλέξουν τον τρόπο με τον οποίο θα τους παρουσιαστεί το περιεχόμενο.
- Είναι εύκολη η δημιουργία ισχυρών εργαλείων για την ευρείας κλίμακας διαχείριση περιεχομένου.
- Το έδαφος γίνεται πρόσφορο για ανεξάρτητους παραγωγούς λογισμικού.
- Γίνεται δυνατή η πραγματικά παγκόσμια έκδοση σε όλα τα μέσα ενημέρωσης.
- Καθορίζονται η λογική δομή και η δομή μορφοποίησης των εγγράφων, καθώς και ο συγχρονισμός μεταξύ των αντικειμένων που τα απαρτίζουν.
- Καθορίζονται στοιχειώδεις αρχές (primitives), όπως είναι οι αρχές αλληλεπίδρασης (interaction primitives), με έγγραφα πολυμέσων.
- Τα πρότυπα εκδίδονται από αναγνωρισμένους οργανισμούς, είναι καταληπτά απ' όλους και είναι κατάλληλα τόσο για ανοικτά περιβάλλοντα όσο και για εκδόσεις σε CD-ROM και στο Διαδίκτυο.
- Τα πρότυπα μπορούν να χρησιμοποιηθούν:
 - Από επεξεργαστές κειμένου, έτσι ώστε να επιταχυνθεί η διαδικασία δημιουργίας δομημένων εγγράφων.
 - Από μηχανές διαφόρων κατασκευαστών που αναλύουν και επεξεργάζονται τα έγγραφα που υπακούουν σε πρότυπα.
 - Από προγράμματα πλοήγησης για παρουσίαση.
 - Από χρήστες που εξάγουν τμήματα εγγράφων και τα ενσωματώνουν σε δικά τους έγγραφα.
 - Από μηχανές αναζήτησης που χρησιμοποιούν τη δομή των εγγράφων για να παρέχουν πιο εξελιγμένες υπηρεσίες στους χρήστες κ.α.

Τα σημαντικότερα και ευρύτερα χρησιμοποιούμενα πρότυπα αναπαράστασης εγγράφων πολυμέσων είναι τα εξής:

⁷ Π.χ., το περιεχόμενο του τίτλου ενός εγγράφου έχει διαφορετικό ρόλο από το περιεχόμενο μιας παραγράφου ή μιας υποσημείωσης.

- *ODA (Office ή Open Document Architecture)* [ODA89], [Ros91]: Πρότυπο που σα στόχο είχε την πρότυπη (standard) αναπαράσταση της λογικής και της φυσικής δομής των εγγράφων πολυμέσων. Το γεγονός αυτό επιτρέπει τη χρήση της ίδιας αναπαράστασης των εγγράφων πολυμέσων σε διαφορετικά ανοικτά περιβάλλοντα, όπου θα επιτρέπεται η ανταλλαγή εγγράφων.

Το μοντέλο της ODA είναι ιεραρχικό και αντικειμενοστραφές. Ένα ODA έγγραφο περιγράφεται από τη λογική του δομή και από τη δομή μορφοποίησής του.

Η *λογική δομή (logical structure)* ενός εγγράφου είναι ένα δέντρο, καθέναν από τους κόμβους του οποίου αναπαριστά μια *οντότητα (entity)* που έχει νόημα για το συγγραφέα ή τον αναγνώστη. Οι οντότητες αυτές μπορεί να είναι *γενικές (generic)* (π.χ. κεφάλαια, τμήματα, παράγραφοι κ.α.) ή *ειδικές (specific)* (π.χ. τιμές, σύμβολα κ.α.).

Η *δομή μορφοποίησης (layout structure)* ενός εγγράφου έχει να κάνει με την παρουσίασή του. Πρόκειται για μια δενδρική δομή, οι κόμβοι της οποίας αναπαριστούν σύνολα σελίδων, σελίδες και ορθογώνιες περιοχές μέσα στις σελίδες. Οι ορθογώνιες περιοχές που μπορούν να περιέχουν άλλες ορθογώνιες περιοχές ονομάζονται *πλαίσια (frames)*, ενώ οι στοιχειώδεις ορθογώνιες περιοχές ονομάζονται *περιοχές (blocks)*. Τόσο τα πλαίσια όσο και οι στοιχειώδεις περιοχές τοποθετούνται στις σελίδες με τρόπο τέτοιο, ώστε οι ακμές τους να είναι παράλληλες με τις ακμές της σελίδας στην οποία περιέχονται.

- *SGML (Standard Generalized Markup Language)* [SGML90], [SGML86]: Μια ISO πρότυπη μεταγλώσσα για τον ορισμό γλωσσών που περιγράφουν κλάσεις δομημένης πληροφορίας. Οι γλώσσες που ορίζονται με βάση την SGML για την περιγραφή κλάσεων δομημένης πληροφορίας ονομάζονται *DTDs (Document Type Definitions)*. Οι DTDs ορίζουν τους κανόνες δόμησης της πληροφορίας παρέχοντας έτσι πρόσβαση στη λογική δομή των SGML εγγράφων, αλλά δεν επηρεάζουν την παραπέρα επεξεργασία των εγγράφων.

Ως *Έγγραφο SGML (SGML Document)* ορίζεται ένα έγγραφο που απαρτίζεται από μια σειρά χαρακτήρων που οργανώνονται φυσικά σε μια *δομή οντοτήτων (entity structure)* και λογικά σε μια *δομή στοιχείων (element structure)*.

Στοιχείο (Element) είναι ένα τμήμα της ιεραρχικής δομής που περιγράφεται από ένα DTD. Οι ιδιότητες των στοιχείων (element properties) προσδιορίζουν όχι μόνο το περιεχόμενο των στοιχείων, αλλά και τη στοιχειοθεσία τους (markup).

Οντότητα (Entity) είναι μια συλλογή χαρακτήρων που περιέχονται σε ένα SGML έγγραφο. Ένα SGML έγγραφο απαρτίζεται από χαρακτήρες περιεχομένου, οι οποίοι αποτελούν το περιεχόμενό του, και από χαρακτήρες στοιχειοθεσίας, οι οποίοι αναπαριστούν τόσο τη δομή του περιεχομένου του εγγράφου όσο και την πληροφορία που σχετίζεται με τη στοιχειοθεσία του και μπορεί να χρησιμοποιηθεί κατά την επεξεργασία του εγγράφου.

Για τον καθορισμό της μορφοποίησης και της παρουσίασης των SGML εγγράφων, ακόμα και σε διαφορετικά μέσα, όπως π.χ. σε παραδοσιακές εκδόσεις και σε εκδόσεις στο Διαδίκτυο, καθώς και για άλλους μετασχηματισμούς που μπορεί να εφαρμοστούν

σε ένα SGML έγγραφο, χρησιμοποιείται η *DSSSL (Document Style Semantics and Specifications Language)* [DSSSL96].

- *HTML (Hypertext Markup Language)* [HTML]: Ένα de facto πρότυπο για έγγραφα υπερκειμένου ευρείας περιοχής που χρησιμοποιείται στον Παγκόσμιο Ιστό από το 1990. Η HTML είναι εφαρμογή της SGML: Κάθε έκδοση της HTML απαρτίζεται από ένα ή περισσότερους SGML DTDs.

Η HTML είναι μια απλή γλώσσα στοιχειοθεσίας, η οποία υποστηρίζει έγγραφα υπερκειμένου που διαθέτουν τίτλο, 6 επίπεδα επικεφαλίδων, λίστες (αριθμημένες, κουκκίδων (bulleted) και φωλιασμένες (nested)), προ-μορφοποιημένο κείμενο, συνδέσμους υπερκειμένου (hyperlinks), ενσωματωμένες εικόνες, προσδιορισμούς στυλ (style specifications), πίνακες, μαθηματικά σύμβολα, αντικείμενα πολυμέσων και διεθνοποίηση (internationalization) [YNA] μέσα από ένα προκαθορισμένο σύνολο εντολών (tags). Οι τελευταίες εκδόσεις της HTML [HTML4] έχουν εμπλουτιστεί και υποστηρίζουν χαρακτηριστικά (scripting, style sheets, Java applets, ActiveX controls κ.α.) που δίνουν δυνατότητες αλληλεπίδρασης με τις ιστοσελίδες.

Ένα από τα κυριότερα προβλήματα της HTML είναι ότι είναι κατάλληλη μόνο για την παρουσίαση της πληροφορίας, καθώς τα στοιχεία που παρέχει δεν αρκούν για την ικανοποιητική περιγραφή της λογικής δομής των εγγράφων.

Για τον καθορισμό της παρουσίασης των HTML εγγράφων χρησιμοποιείται η *CSSL (Cascading Style Sheet Language)* [CSS1] [CSS2]. Η CSSL είναι ένας μηχανισμός καθορισμού στυλ που έχει σχεδιαστεί ειδικά για να καλύπτει τις ανάγκες τόσο των χρηστών του Παγκόσμιου Ιστού όσο και των δημιουργών ιστοσελίδων.

Πιο συγκεκριμένα, η CSSL είναι μια απλή δηλωτική γλώσσα που επιτρέπει τόσο στους συγγραφείς όσο και στους αναγνώστες να καθορίσουν την παρουσίαση δομημένων εγγράφων που είναι γραμμένα σε HTML ή XML. Η CSSL επιτρέπει τον καθορισμό του τρόπου παρουσίασης των στοιχείων που απαρτίζουν τα έγγραφα μέσω της διασύνδεσής τους με ιδιότητες και τιμές.

- *XML (eXtensible Markup Language)* [XML]: Μια απλή και ευέλικτη μεταγλώσσα που βασίζεται στην SGML. Στην πραγματικότητα η XML έχει σχεδιαστεί για να γίνει δυνατή η χρήση της SGML στον Παγκόσμιο Ιστό. Όπως και η SGML, η XML δεν είναι μια απλή γλώσσα στοιχειοθεσίας, αλλά μια μεταγλώσσα που επιτρέπει την περιγραφή γλωσσών στοιχειοθεσίας (markup languages).

Μια πιο αναλυτική αναφορά στην XML γίνεται στην ενότητα 3.2.1, καθώς η XML καθόρισε σε σημαντικό βαθμό το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» και ενσωματώθηκε σ' αυτόν.

- *SPDL (Standard Page Description Language)* [SPDL]: Ένα ανεπίσημο πρότυπο για την αντιστοίχιση εγγράφων που συντάσσονται με βάση κάποιο πρότυπο σε μορφή σύμφωνη με μια γλώσσα περιγραφής σελίδων που θα τυπωθούν σε συσκευές εξόδου. Έτσι, ένα SGML έγγραφο θα περάσει από διεργασίες που καθορίζονται από τη DSSSL και την SPDL πριν τυπωθεί.

- *HyTime (Hypermedia – Time based Structuring Language)* [HyTime92], [Buf94], [New91]: Μια ISO πρότυπη μεταγλώσσα για την αναπαράσταση εγγράφων υπερκειμένου και πολυμέσων, τα οποία μπορεί να περιέχουν *χωρο-χρονικά αντικείμενα (spatio-temporal objects)* με βάση τη δομή τους. Η HyTime αποτελεί επέκταση της SGML και παρέχει ένα υπερσύνολο της λειτουργικότητας που παρέχει η HTML.

Για να είναι δυνατός ο ορισμός DTDs που υποστηρίζουν τη HyTime παρέχονται οι *αρχιτεκτονικές φόρμες (architectural forms)*. Κάθε αρχιτεκτονική φόρμα καθορίζει πως μια κλάση SGML στοιχείων μπορεί να οριστεί σε ένα DTD, τα έγγραφα-στιγμιότυπα του οποίου περιέχουν πληροφορία που αφορά συγκεκριμένες HyTime έννοιες ή σύνολα HyTime εννοιών. Έτσι, οι αρχιτεκτονικές φόρμες απαρτίζουν ένα μετα-DTD που καθορίζει πως θα δημιουργηθεί ένας συμβατός με τη HyTime DTD.

- *MHEG (Multimedia & Hypermedia Expert Group)* [MHEG], [Ste95], [Pri93]: Ένα ISO πρότυπο για την ανταλλαγή μεταφέρσιμης (portable) πληροφορίας μεταξύ ανεξάρτητων περιβαλλόντων παραγωγής και εκτέλεσης. Η MHEG παρέχει λειτουργικότητα συγκρίσιμη με αυτή που παρέχεται από τη HyTime, με την επιπρόσθετη υποστήριξη γλωσσών scripting. Ο ορισμός της MHEG απαρτίζεται από τα εξής 6 τμήματα:

MHEG-1: Το τμήμα αυτό ασχολείται με την *Αναπαράσταση Αντικειμένων χρησιμοποιώντας ASN.1 Σύνταξη (Object Representation – Base Notation ASN.1)*.

MHEG-2: Το τμήμα αυτό ασχολείται με την *Αναπαράσταση Αντικειμένων χρησιμοποιώντας SGML Σύνταξη (Object Representation – Base Notation SGML)*.

MHEG-3: Το τμήμα αυτό ασχολείται με την *Αναπαράσταση της Ανταλλαγής MHEG Scripts (MHEG Script Interchange Representation)*.

MHEG-4: Το τμήμα αυτό ασχολείται με τη *Διαδικασία MHEG Πιστοποίησης (MHEG Registration Procedure)*.

MHEG-5: Το τμήμα αυτό ασχολείται με την *Υποστήριξη για Αλληλεπιδραστικές Εφαρμογές Βασικού Επιπέδου (Support for Base – Level Interactive Applications)*.

MHEG-6: Το τμήμα αυτό ασχολείται με την *Υποστήριξη Ενισχυμένων Αλληλεπιδραστικών Εφαρμογών (Support for Enhanced Interactive Applications)*.

Τα MHEG έγγραφα απαρτίζονται από MHEG αντικείμενα. Ένα *MHEG αντικείμενο (MHEG object)* είναι στιγμιότυπο μιας MHEG κλάσης. Τα MHEG αντικείμενα αναπαρίστανται με τη χρήση ASN.1 ή SGML σύνταξης και ανταλλάσσονται αφού υποστούν κωδικοποίηση *BER (Binary Encoding Rules)*.

Μια *MHEG κλάση (MHEG class)* είναι ένας αφαιρετικός (abstract) ορισμός της δομής ανταλλάξιμων αντικειμένων πολυμέσων και υπερμέσων.

Η *MHEG μηχανή (MHEG engine)* παίζει τον κύριο ρόλο σε ένα MHEG σύστημα: Είναι μια διεργασία ή ένα σύνολο διεργασιών που κωδικοποιεί, αποκωδικοποιεί, αναλύει συντακτικά, μεταφράζει, και υλοποιεί την αναπαράσταση των MHEG αντικειμένων.

3.2.1. eXtensible Markup Language (XML)

Σ' αυτή την ενότητα θα γίνει μια αναλυτικότερη συζήτηση για την XML που, όπως αναφέρθηκε και παραπάνω, έπαιξε καθοριστικό ρόλο στο σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», στον οποίο και ενσωματώθηκε.

Όπως έχει ήδη αναφερθεί, η XML είναι μια εξαιρετικά απλή κι ευέλικτη «διάλεκτος» της SGML, που σχεδιάστηκε έτσι ώστε να εγγυάται την ευκολία στην υλοποίηση συστημάτων που την υποστηρίζουν και τη διαλειτουργικότητα τόσο με την SGML όσο και με την HTML [XML] [MFDG98] [Gott98] [Wal98]. Η XML δημιουργήθηκε για να γίνει δυνατή η χρήση δομημένων εγγράφων στον Παγκόσμιο Ιστό. Πιο συγκεκριμένα, οι ανάγκες που κάλυψε η XML συνοψίζονται στις εξής:

- *Επεκτασιμότητα (extensibility)*, ώστε να δίνεται η δυνατότητα ορισμού νέων εντολών όταν αυτό είναι απαραίτητο.
- *Δομή (structure)*, ώστε να είναι δυνατή η αναπαράσταση δεδομένων με δομή που παρουσιάζει οποιοδήποτε επίπεδο πολυπλοκότητας.
- Δυνατότητα *επαλήθευσης (validation)* της σωστής δόμησης των εγγράφων.
- *Ανεξαρτησία από μέσα (media independence)*, ώστε να είναι δυνατή η έκδοση περιεχομένου σε πολλαπλές μορφές.
- *Ανεξαρτησία από πλατφόρμες και προμηθευτές*, ώστε να είναι δυνατή η επεξεργασία οποιουδήποτε συμβατού εγγράφου με τη χρήση τόσο εμπορικού λογισμικού όσο και απλών εργαλείων επεξεργασίας κειμένου.

Η XML είναι, όπως και η SGML, μεταγλώσσα και δημιουργήθηκε καθώς υπήρχε η ανάγκη για μια γλώσσα απλούστερη από την SGML και πιο επεκτάσιμη από την HTML για την περιγραφή εγγράφων στον Παγκόσμιο Ιστό. Σύμφωνα με το W3C (WWW Consortium), που είναι υπεύθυνο για την XML, η XML θα παρέχει «περισσότερο από το 80% της λειτουργικότητας της SGML για λιγότερο από το 20% της πολυπλοκότητάς της». Καθώς η XML είναι γνήσιο υποσύνολο της SGML, οποιοδήποτε σύστημα υποστηρίζει SGML έγγραφα μπορεί να υποστηρίξει και XML έγγραφα.

Αυτό που κάνει την XML τόσο ισχυρή είναι το γεγονός ότι οποιοσδήποτε τύπος δεδομένων – ακόμα και κάποια αφηρημένη έννοια – μπορεί να αποκτήσει μορφή και δομή με τη χρήση της XML.

Καθώς η XML χρησιμοποιείται για τον ορισμό της δομής των εγγράφων κι όχι για το πώς τα έγγραφα θα παρουσιαστούν από ένα πρόγραμμα πλοήγησης, θα δώσει τόσο στους δημιουργούς ιστοσελίδων όσο και στους χρήστες πολύ μεγαλύτερη ευελιξία από την HTML. Παρ' όλο που η XML και η HTML παρουσιάζουν μια σειρά από ομοιότητες, έχουν και μια σειρά διαφορών:

- Η XML είναι μεταγλώσσα ενώ η HTML γλώσσα. Έτσι, μέσω της XML μπορούν να οριστούν πολλές γλώσσες, με συνέπεια να μην έχουν όλα τα XML έγγραφα την ίδια σύνταξη. Μέσω της XML ορίζονται τόσο η σύνταξη των γλωσσών περιγραφής εγγράφων, όσο και η δόμηση των στοιχείων που απαρτίζουν τα έγγραφα που υπακούουν σ' αυτές. Από την άλλη, η HTML διαθέτει ένα συγκεκριμένο σύνολο

εντολών, με συνέπεια όλα τα HTML έγγραφα να υπακούουν στους ίδιους συντακτικούς κανόνες.

- Η XML καθορίζει τη δομή των εγγράφων κι όχι το πώς θα έπρεπε αυτά να παρουσιαστούν από ένα πρόγραμμα πλοήγησης, σε αντίθεση με την HTML που αναμιγνύει τη δομή των εγγράφων με τους κανόνες παρουσίασής τους. Το γεγονός αυτό δίνει στους δημιουργούς ιστοσελίδων μεγαλύτερη ευελιξία.
- Καθώς η XML παρέχει πρόσβαση στη δομή των εγγράφων, οι μηχανές αναζήτησης θα γίνουν πιο αποτελεσματικές καθώς θα μπορούν να υποστηρίξουν ερωτήσεις που θα επιτρέπουν στους χρήστες να εκφράσουν με σαφέστερο τρόπο ποια έγγραφα τους ενδιαφέρουν (π.χ. δώσε μου τα έγγραφα που η λέξη «υπολογιστής» αναφέρεται σε μια τουλάχιστον παράγραφο τους).
- Η XML εισάγει έννοιες που θα κάνουν πιο σταθερές τις εφαρμογές στον Παγκόσμιο Ιστό. Τέτοιες έννοιες είναι οι *αμφίδρομοι σύνδεσμοι υπερκειμένου (bi-directional links)* και οι *έμμεσοι σύνδεσμοι (indirect links)* που αποθηκεύονται ανεξάρτητα από τα έγγραφα.

Στο σημείο αυτό είναι σημαντικό να τονιστεί ότι η HTML και η XML δεν είναι αντίζηλοι, αλλά η μια είναι συμπλήρωμα της άλλης. Τα προγράμματα πλοήγησης υποστηρίζουν και τις δυο και οι μελλοντικές εκδόσεις της HTML θα επιτρέπουν την ανάμιξη XML με HTML. Στο σημείο αυτό πρέπει να αναφερθεί ότι τα HTML έγγραφα, όπως και κάθε κατηγορία SGML εγγράφων, δεν είναι καταληπτά από ένα σύστημα που υποστηρίζει XML έγγραφα. Για το λόγο αυτό έχει οριστεί η μεθοδολογία μετατροπής HTML εγγράφων σε XML από το W3C [XHTML]. Η μεθοδολογία αυτή είναι γνωστή με το όνομα XHTML και περιγράφεται παρακάτω, μαζί με τις υπόλοιπες *Εφαρμογές της XML*.

Στη συνέχεια θα συζητηθούν τα σημαντικότερα από τα χαρακτηριστικά της XML: *Η Δομή και το Περιεχόμενο των XML Εγγράφων*, *η Πληροφορία που περιέχεται σε XML Έγγραφα*, *η παρεχόμενη από την XML Υποστήριξη Συνδέσμων*, *η Σύνδεση Στυλ με XML Έγγραφα*, *η Ανταλλαγή Τμημάτων XML Εγγράφων*, *η Κωδικοποίηση XML Εγγράφων* καθώς και κάποιες από τις υπάρχουσες *Εφαρμογές της XML*.

3.2.1.1. Η Δομή και το Περιεχόμενο των XML Εγγράφων

Από τη στιγμή που η XML δε βασίζεται σε προκαθορισμένες εντολές και χαρακτηριστικά, ο συγγραφέας έχει πλήρη έλεγχο στη δομή των δεδομένων που απαρτίζουν ένα XML έγγραφο: Μπορεί να ορίσει τόσο τις εντολές σύνταξής και τα χαρακτηριστικά για τα στοιχεία που απαρτίζουν το έγγραφο όσο και τους κανόνες δόμησης των στοιχείων αυτών.

Τα XML έγγραφα αποτελούνται από *περιεχόμενο (content)* και από *στοιχεία τυποποίησης (markup)*, τα οποία περιγράφουν το ρόλο που παίζει το περιεχόμενο, ανάλογα με την κατηγορία στην οποία αυτό εντάσσεται. Τα στοιχεία τυποποίησης διακρίνονται σε *στοιχεία (elements)*, *οντότητες (entities)*, *σημειογραφίες (notations)*, *σχόλια (comments)*, *εντολές επεξεργασίας (processing instructions)*, *επιλεγμένες περιοχές (marked sections)* και *Δηλώσεις Τύπων Εγγράφων (Document Type Declarations)*.

Όπως και τα SGML έγγραφα, τα XML έγγραφα οργανώνονται λογικά σε μια δομή στοιχείων και φυσικά σε μια δομή οντοτήτων. Τα στοιχεία είναι οι κόμβοι της λογικής δομής των

εγγράφων, ενώ οι οντότητες, οι οποίες παράλληλα συνιστούν τη φυσική δομή των XML εγγράφων, είναι οι ομάδες χαρακτήρων από τους οποίους απαρτίζονται τα έγγραφα. Καθώς τα στοιχεία από τα οποία απαρτίζονται τα XML έγγραφα οργανώνονται σε ιεραρχίες, η λογική δομή ενός XML εγγράφου είναι δένδροειδής.

Τα στοιχεία από τα οποία απαρτίζονται τα XML έγγραφα μπορούν να διαθέτουν χαρακτηριστικά (attributes) τα οποία εμφανίζονται με τη μορφή ζευγών τιμής-ονόματος (name-value pairs). Τα χαρακτηριστικά των στοιχείων από τα οποία απαρτίζονται τα XML έγγραφα όμως, παρέχουν συνήθως επιπλέον πληροφορία για τη δομή των εγγράφων και όχι για τη μορφοποίησή τους όπως συμβαίνει με τα στοιχεία από τα οποία απαρτίζονται τα HTML έγγραφα.

Οι σημειογραφίες καθορίζουν τις εφαρμογές που θα χρησιμοποιούνται για την παρουσίαση του περιεχομένου των XML εγγράφων που δεν είναι σε μορφή κειμένου (π.χ. εικόνες, ήχους κλπ.).

Τα σχόλια μπορούν να υπάρχουν διάσπαρτα μέσα στα έγγραφα, χωρίς να αποτελούν τμήμα του περιεχομένου τους. Έτσι, οι επεξεργαστές XML εγγράφων δεν απαιτείται να τα περνούν στις εφαρμογές.

Οι εντολές επεξεργασίας είναι ένα ποσό πληροφορίας που δεν αποτελεί τμήμα του περιεχομένου των XML εγγράφων, αλλά σε αντίθεση με τα σχόλια, οι επεξεργαστές XML εγγράφων τα περνούν χωρίς να τα επεξεργαστούν στις εφαρμογές, οι οποίες και είναι υπεύθυνες για την εκτέλεσή τους.

Οι επιλεγμένες περιοχές είναι τμήματα κειμένου μέσα στα οποία οι συντακτικοί αναλυτές XML εγγράφων αγνοούν τους ειδικούς χαρακτήρες.

Οι Δηλώσεις Τύπων Εγγράφων είτε περιέχουν τους ορισμούς που καθορίζουν τους κανόνες δόμησης μιας κλάσης XML εγγράφων είτε αναφέρονται σ' αυτούς. Οι ορισμοί αυτοί συνιστούν τη γραμματική βάσει της οποίας συντάσσονται τα στιγμιότυπα μιας κλάσης εγγράφων. Η γραμματική αυτή ονομάζεται *Ορισμός Τύπου Εγγράφων (Document Type Definition - DTD)*.

Τα αντικείμενα που περιγράφονται από την XML είναι τα XML έγγραφα. Ένα αντικείμενο είναι XML έγγραφο όταν η σύνταξή του είναι συμβατή με την XML. Ένα XML έγγραφο χαρακτηρίζεται *Σωστό (Valid)* όταν τα στοιχεία του είναι σωστά ορισμένα και τοποθετημένα σε σημείο που επιτρέπεται μέσα στις ιεραρχίες που αυτά σχηματίζουν. Ένα XML έγγραφο χαρακτηρίζεται *Καλά Ορισμένο (Well-Formed)* αν είναι μεν δομημένο σύμφωνα με τους κανόνες της XML, δεν έχει γίνει όμως έλεγχος για το αν τα στοιχεία που περιέχει ορίζονται στο DTD (ή στους DTDs) που αυτό υπακούει.

Για την υποστήριξη εφαρμογών και συσκευών που θα υποστηρίζουν ένα υποσύνολο της XML προτείνεται η χρήση των *διαγραμμάτων XML (XML profiles)*. Τα διαγράμματα XML είναι υποσύνολα των χαρακτηριστικών της XML, τα οποία θα υποστηρίζονται από συγκεκριμένες συσκευές και εφαρμογές.

Επιπλέον, για να γίνει δυνατή η χρήση *ψηφιακών υπογραφών (digital signatures)* για XML έγγραφα, είναι σε εξέλιξη η διαδικασία καθορισμού των κανόνων *κανονικοποίησης (canonicalizing)* XML εγγράφων. Για το λόγο αυτό έχει σχεδιαστεί από το W3C η *Canonical*

XML [CanXML], μια γλώσσα που χρησιμοποιείται για την κωδικοποίηση της λογικής δομής των XML εγγράφων.

Από τη διαδικασία της κανονικοποίησης για κάθε έγγραφο παράγεται το *κανονικό (canonical)* του μέσω της ελαχιστοποίησης των κενών, της χρήσης ομοιόμορφων εισαγωγικών κλπ. Δυο έγγραφα των οποίων οι κανονικές μορφές είναι ίδιες θεωρούνται, από αρκετές εφαρμογές, ισοδύναμα.

Για την περιγραφή της δομής των XML εγγράφων, αρχικά προτάθηκαν οι DTDs. Καθώς όμως στη συνέχεια φάνηκε ότι οι DTDs δεν καλύπτουν όλες τις ανάγκες περιγραφής της δομής των XML εγγράφων, έγιναν αρκετές προσπάθειες για τον ορισμό μιας γλώσσας XML σχημάτων. Η γλώσσα αυτή θα είχε ως στόχο την κάλυψη των αναγκών που υπάρχουν για την περιγραφή της δομής των XML εγγράφων. Αρχικά, στα πλαίσια αυτά κινήθηκαν οι προτάσεις XML-Data, DDML, SOX και DCD και στη συνέχεια το W3C συγκρότησε την ομάδα εργασίας XML Schema που εξετάζει το σκοπό, τα σενάρια χρήσης, τις αρχές σχεδιασμού και τις βασικές απαιτήσεις για μια γλώσσα περιγραφής της δομής των XML εγγράφων. Στη συνέχεια γίνεται αναφορά τόσο στους DTDs όσο και στις γλώσσες περιγραφής XML σχημάτων.

Document Type Declaration (DTD)

Οι *DTDs (Document Type Definitions)* είναι ένα χαρακτηριστικό αντίστοιχο με τους SGML DTDs. Όταν χρησιμοποιούνται DTDs για την περιγραφή της δομής των XML εγγράφων, το σύνολο των κανόνων που ισχύουν για ένα έγγραφο ορίζεται σε ένα DTD: Όπως και στην περίπτωση της SGML, ο DTD είναι μια γλώσσα στοιχειοθεσίας που έχει οριστεί με βάση την XML. Έτσι, όπως και οι SGML DTDs, οι XML DTDs χρησιμοποιούνται για την παροχή μεταπληροφορίας σχετικής με τη δομή των εγγράφων που δομούνται με βάση αυτούς. Σε ένα DTD καθορίζονται:

- Οι οντότητες, τα στοιχεία και οι σημειογραφίες που ορίζονται για μια κλάση XML εγγράφων.
- Η σύνταξη, το περιεχόμενο και τα χαρακτηριστικά των στοιχείων που ορίζονται για μια κλάση XML εγγράφων. Το περιεχόμενο ενός στοιχείου χαρακτηρίζεται ως:
 - Empty, όταν το στοιχείο είναι κενό (δεν επιτρέπεται να έχει περιεχόμενο).
 - PCDATA, όταν το στοιχείο επιτρέπεται να έχει μόνο χαρακτήρες ως περιεχόμενο.
 - Children, όταν το στοιχείο επιτρέπεται να έχει μόνο υποστοιχεία ως περιεχόμενο.
 - Mixed, όταν το στοιχείο επιτρέπεται να έχει τόσο χαρακτήρες όσο και υποστοιχεία ως περιεχόμενο.
 - Any, όταν το στοιχείο επιτρέπεται να έχει οποιοδήποτε περιεχόμενο.
- Οι σχέσεις μεταξύ των στοιχείων και των χαρακτηριστικών που ορίζονται για μια κλάση XML εγγράφων.
- Η λογική δομή των εγγράφων που υπακούουν στο DTD και πιο συγκεκριμένα ποιοι τύποι κόμβων επιτρέπεται να είναι παιδιά ποιων κόμβων μέσα στην ιεραρχία που σχηματίζουν οι κόμβοι που αναπαριστούν τα στοιχεία από τα οποία απαρτίζονται τα

έγγραφα. Κατά συνέπεια, σε κάθε DTD ορίζεται η επιτρεπτή μορφή της λογικής δομής των εγγράφων που υπακούουν σ' αυτό το DTD.

Στην ιδανική περίπτωση, οι ορισμοί σε ένα DTD πρέπει να είναι προσανατολισμένοι στην περιγραφή της δομής των εγγράφων και όχι στο πως αυτά θα παρουσιάζονται.

Στο *Παράδειγμα 3-1* παρουσιάζεται ένας DTD για την αναπαράσταση απλών σημειωμάτων.

```
<!ELEMENT Memo (Title, Body)>
<!ELEMENT Title (#PCDATA)>
<!ELEMENT Body ANY>
<!ENTITY Greater ">">
```

Παράδειγμα 3-1: Ένας DTD για την αναπαράσταση σημειωμάτων

Στο DTD του παραδείγματος ορίζονται τα στοιχεία Memo, Title και Body, που αναπαριστούν το έγγραφο, τον τίτλο και το περιεχόμενό του αντίστοιχα, και η οντότητα Greater, που αναπαριστά το χαρακτήρα «>». Στο *Παράδειγμα 3-2* φαίνεται ένα έγγραφο σύμφωνο με αυτό το DTD.

```
<memo>
  <title>Μάθημα</title>
  <body>
    Αύριο στις 10:00 π.μ. έχω μάθημα αγγλικών.
  </body>
</memo>
```

Παράδειγμα 3-2: Ένα έγγραφο σύμφωνο με το DTD για την αναπαράσταση σημειωμάτων

Ένας DTD μπορεί να αποθηκεύεται σε ένα ή περισσότερα αρχεία. Οι DTDs δεν είναι απαραίτητο να ορίζονται από τους συγγραφείς των εγγράφων: Οι τελευταίοι μπορούν να χρησιμοποιούν ήδη υπάρχοντες DTDs ή και να μη χρησιμοποιήσουν καθόλου DTD για κάποιο έγγραφο.

Γλώσσες Περιγραφής XML Σχημάτων

Για να είναι αξιόπιστη η επεξεργασία εγγράφων, είναι απαραίτητες η περιγραφή κλάσεων εγγράφων και η επαλήθευση της δυνατότητας συμμετοχής των εγγράφων ως στιγμιότυπων σ' αυτές. Πρέπει δηλαδή να οριστούν κλάσεις εγγράφων μέσω της έκφρασης περιορισμών για το περιεχόμενο και τη δομή των εγγράφων που θα ανήκουν σε αυτές. Για το σκοπό αυτό παρέχεται από τους DTDs περιορισμένη υποστήριξη, καθώς οι τελευταίοι διαθέτουν περιορισμένη εκφραστικότητα και είναι απαραίτητος ο επιπλέον πειραματισμός στο σχεδιασμό σχημάτων με νέες ιδέες.

Επιπρόσθετα, καθώς η XML, μετά τον ορισμό της άρχισε να χρησιμοποιείται για την περιγραφή όχι μόνο δομημένων εγγράφων, αλλά και κάθε είδους δομημένης πληροφορίας, έγινε φανερό ότι οι DTDs δεν επαρκούσαν για την περιγραφή XML πληροφορίας. Τα

βασικότερα χαρακτηριστικά που λείπουν από τους DTDs έτσι ώστε αυτοί να επαρκούν για την περιγραφή XML πληροφορίας είναι τα εξής:

- Η παροχή της δυνατότητας διαχείρισης δομημένης πληροφορίας.
- Η παροχή τύπων δεδομένων (datatypes), καθώς η XML δεν υποστηρίζει τύπους δεδομένων (όλα τα δεδομένα αντιμετωπίζονται από την XML ως συμβολοσειρές (strings)).
- Η παροχή ενσωματωμένης τεκμηρίωσης (documentation).
- Η υποστήριξη κληρονομικότητας (inheritance).
- Η υποστήριξη καθολικών χαρακτηριστικών (global attributes) στα πλαίσια ενός σχήματος.
- Ο διαχωρισμός της φυσικής από τη λογική δομή της πληροφορίας.

Για την κάλυψη των αναγκών αυτών προτάθηκε η δημιουργία μιας γλώσσας περιγραφής XML σχημάτων. Έτσι, σε πρώτη φάση προτάθηκαν οι εξής γλώσσες περιγραφής σχημάτων:

DDML: Η *DDML (Document Definition Markup Language)* [DDML] είναι μια πειραματική γλώσσα που σχεδιάστηκε για την παροχή ενός σημείου εκκίνησης για πειραματισμούς σε ότι αφορά τη χρήση μιας σύνταξης που μοιάζει πολύ με το περιεχόμενο των XML εγγράφων, συγκεκριμένους τρόπους επέκτασης και αυστηρότερο διαχωρισμό συντακτικής ανάλυσης και επαλήθευσης.. Πρόκειται για μια γλώσσα που ασχολείται με τον ορισμό σχημάτων περιγραφής της δομής των στοιχείων και των χαρακτηριστικών των XML εγγράφων. Στον καθορισμό της DDML περιγράφεται η μετατροπή των σχημάτων που τη χρησιμοποιούν σε DTDs, ώστε να εξασφαλιστεί η χρησιμοποίησιμότητά της.

XML-Data: Η *XML-Data* είναι μια γλώσσα περιγραφής της δομής των δεδομένων που απαρτίζουν τα XML έγγραφα, που στόχο έχει να φέρει την XML πίσω στις ρίζες της ως γλώσσας που περιγράφει δομές. Η XML-Data είναι ένα XML λεξικό *σχημάτων (schemas)*, ασχολείται δηλαδή με τον ορισμό και την περιγραφή κλάσεων αντικειμένων.

Τα σχήματα ορίζουν τα χαρακτηριστικά των κλάσεων αντικειμένων και μπορούν να χρησιμοποιηθούν τόσο για τον ορισμό αυστηρά συντακτικών κλάσεων (όπως π.χ. οι κλάσεις XML εγγράφων) όσο και για τον ορισμό κλάσεων που υποδηλώνουν έννοιες και σχέσεις μεταξύ εννοιών (όπως π.χ. στις σχεσιακές βάσεις δεδομένων). Τα σχήματα της πρώτης κατηγορίας ονομάζονται *Συντακτικά Σχήματα (Syntactic Schemas)*, ενώ τα σχήματα της δεύτερης κατηγορίας *Εννοιολογικά Σχήματα (Conceptual Schemas)*.

Το κενό της XML σε ότι αφορά την υποστήριξη τύπων δεδομένων καλύπτεται από την XML-Data, καθώς αυτή παρέχει το μηχανισμό τυποποίησης που λείπει από την XML.

Μέσω της XML-Data είναι δυνατή η περιγραφή των χαρακτηριστικών διαφόρων δομών: στα XML-Data σχήματα ορίζονται οι τρόποι χρήσης και ομαδοποίησης

των δομών καθώς και ο τρόπος συσχέτισης των διαφορετικών δομών μεταξύ τους. Τα σχήματα απαρτίζονται από δηλώσεις:

- *Εννοιών (concepts).*
- *Κλάσεων αντικειμένων* (που περιλαμβάνουν ιεραρχίες κλάσεων, ιδιότητες (properties), περιορισμούς (constraints), σχέσεις (relationships) κ.α.).
- *Συμβατότητας* με γραμματικές που ορίζονται από DTDs. Πιο συγκεκριμένα, οι δηλώσεις αυτές περιλαμβάνουν:
 - *Γραμματικούς κανόνες* που καθορίζουν τη σωστή δόμηση στοιχείων και χαρακτηριστικών.
 - *Χαρακτηριστικά* στοιχείων.
 - *Εσωτερικά (στα έγγραφα) και εξωτερικά ορισμένες οντότητες.*
 - *Σημειογραφίες.*
- *Τύπων δεδομένων*, που καθορίζουν τη μορφή της υλοποίησης και τη συντακτική ανάλυση.
- *Κανόνων αντιστοίχισης* που επιτρέπουν την αντιστοίχιση γραμματικών σε μοντέλα δεδομένων.

Έτσι, μέσω των XML-Data σχημάτων μπορούν να περιγραφούν οι τύποι XML εγγράφων χρησιμοποιώντας την ίδια την XML αντί για DTDs.

Από τα παραπάνω φαίνεται ότι η XML-Data παρέχει λειτουργικότητα παρόμοια με τους DTDs. Καθώς όμως τα σχήματα που ορίζονται μέσω της XML-Data γράφονται σε XML, τα σχήματα αυτά είναι εύχρηστα και πλήρως επεκτάσιμα, γεγονός που τα κάνει πιο ισχυρά από τους DTDs. Επιπλέον, τα XML-Data σχήματα παρέχουν ένα κοινό λεξικό για ιδέες που επικαλύπτονται τόσο μεταξύ συντακτικών και εννοιολογικών σχημάτων όσο και μεταξύ σχημάτων βάσεων δεδομένων.

SOX: Η SOX (*Schema for Object-oriented XML*) [SOX] αποτελεί μια πρόταση για τον ορισμό σχημάτων (schemas), τα οποία θα καθορίζουν τη δομή, το περιεχόμενο και τη σημασιολογία των συμβατών με αυτά XML εγγράφων. Έτσι, γίνεται δυνατός ο αυτόματος έλεγχος του περιεχομένου των XML εγγράφων σε υψηλότερο επίπεδο.

Η SOX προτείνεται όχι μόνο ως υποκατάστατο των DTDs, αλλά ως μια γλώσσα μοντελοποίησης της πληροφορίας (information modeling language). Έτσι, η SOX παρέχει επιπλέον αυτών που παρέχουν οι DTDs:

- *Εγγενείς τύπους δεδομένων (intrinsic datatypes).*
- *Ένα εκτεταμένο (και επεκτάσιμο) μηχανισμό υποστήριξης τύπων δεδομένων.*
- *Κληρονομικότητα (inheritance)* σε ότι αφορά το μοντέλο δεδομένων και τα χαρακτηριστικά των στοιχείων.
- *Ένα ισχυρό μηχανισμό υποστήριξης namespaces.*

- Υποστήριξη *πολύμορφου περιεχομένου (polymorphic content)*.
- Ενσωματωμένη *τεκμηρίωση (documentation)* σε μορφή HTML.
- Χαρακτηριστικά που επιτρέπουν τη σωστή διαχείριση κατανεμημένων σχημάτων.

Τα χαρακτηριστικά αυτά υποστηρίζονται από ένα αυστηρό έλεγχο σύνταξης και τύπων δεδομένων.

Ένα SOX έγγραφο (ή SOX σχήμα) είναι ένα συντακτικά σωστό XML έγγραφο που υπακούει στο SOX DTD, ο οποίος αναπαριστά πλήρως τη δομή των XML εγγράφων που υποστηρίζουν οι XML DTDs. Η επεξεργασία των SOX εγγράφων γίνεται από SOX επεξεργαστές (SOX processors), οι οποίοι μπορούν να παράγουν ως έξοδο μια πληθώρα μορφών (XML DTDs και έγγραφα, κλάσεις που αναπαριστούν τα έγγραφα και τα στοιχεία τους σε αντικειμενοστραφή περιβάλλοντα κλπ.).

Η SOX προτάθηκε για να υποστηριχθεί η ανάπτυξη ευρείας κλίμακας κατανεμημένων εφαρμογών ηλεκτρονικού εμπορίου, αλλά μπορεί να χρησιμοποιηθεί και για οποιαδήποτε εφαρμογή στοιχειοθεσίας.

Σε σχέση με τους DTDs, η SOX απλοποιεί την υποστήριξη διαλειτουργικότητας μεταξύ ετερογενών εφαρμογών μέσω:

- Της υποστήριξης της αντιστοίχισης, σε επίπεδο λογισμικού, XML δομών δεδομένων.
- Της έκφρασης αφαιρέσεων και κοινών σχέσεων άμεσα σε κάθε περιοχή.
- Τη δυνατότητα επαναχρησιμοποίησης στα επίπεδα του σχεδιασμού και του προγραμματισμού εφαρμογών.
- Την υποστήριξη της δημιουργίας κοινών συστατικών εφαρμογών.

DCD: Η *DCD (Document Content Description)* [DCD], είναι ένα λεξικό μεταδεδομένων που ασχολείται με τη δομική περιγραφή κλάσεων XML εγγράφων μέσω του καθορισμού των κανόνων που ισχύουν για τη δομή και το περιεχόμενό τους. Η DCD περιλαμβάνει ένα τμήμα της XML-Data, το οποίο εκφράζει με τρόπο συμβατό με την RDF (βλέπε υποενότητα 3.3.1.3): Στην πραγματικότητα η DCD είναι ένα RDF λεξικό.

Στόχος της DCD είναι ο ορισμός των περιορισμών που ισχύουν για XML έγγραφα με τη χρήση XML σύνταξης. Οι περιορισμοί που καθορίζονται με τη χρήση της DCD μπορούν να χρησιμοποιηθούν όπως οι DTDs, αλλά η DCD παρέχει επιπλέον λειτουργικότητα από τους DTDs (π.χ. υποστήριξη βασικών τύπων δεδομένων).

Με τον όρο DCD αναφέρονται, εκτός από το πρότυπο, και οι DCD περιγραφές που αναφέρονται σε κλάσεις XML εγγράφων. Μ' αυτή την έννοια, DCD είναι ένα σύνολο ιδιοτήτων που χρησιμοποιούνται για τη δήλωση περιορισμών στο περιεχόμενο των στοιχείων και στα ονόματα των χαρακτηριστικών που μπορεί να

εμφανιστούν σε μια κλάση XML εγγράφων και στα στοιχεία που απαρτίζουν τα έγγραφα αυτής της κλάσης.

Οι ορισμοί των DCD ιδιοτήτων και των DCD πόρων δίνονται από την XML namespace DCD που χαρακτηρίζεται από το URI <http://www.w3.org/Schemas/DCD>. Οι βασικές έννοιες που ορίζονται στη DCD είναι οι εξής:

- DCD, που περιγράφει έναν DCD πόρο. DCD πόρος είναι μια δομή εγγράφου που θέτει περιορισμούς στη δομή και το περιεχόμενο μιας κλάσης XML εγγράφων. Ένα XML έγγραφο μπορεί να υπακούει σε περισσότερες από μια DCDs. Το γεγονός αυτό παρέχει τα εξής πλεονεκτήματα:
 - Μια DCD μπορεί να εφαρμόζεται από ένα μεγάλο αριθμό XML εγγράφων.
 - Η DCD παρέχει το κατάλληλο επίπεδο διακριτότητας για την επιβολή namespace μηχανισμών.
- ElementDef, που περιγράφει τα στοιχεία των XML εγγράφων που είναι σύμφωνα με την τρέχουσα DCD. Το όνομα του στοιχείου που περιγράφει ένας ElementDef δίνεται από το Type χαρακτηριστικό του, ενώ το περιεχόμενό του περιγράφεται από το χαρακτηριστικό Model. Οι τιμές που μπορεί να πάρει το χαρακτηριστικό Model είναι οι εξής:
 - Empty, όταν το στοιχείο είναι κενό (δεν επιτρέπεται να έχει περιεχόμενο).
 - Data, όταν το στοιχείο επιτρέπεται να έχει μόνο χαρακτήρες ως περιεχόμενο.
 - Children, όταν το στοιχείο επιτρέπεται να έχει μόνο υποστοιχεία ως περιεχόμενο.
 - Mixed, όταν το στοιχείο επιτρέπεται να έχει τόσο χαρακτήρες όσο και υποστοιχεία ως περιεχόμενο.
 - Any, όταν το στοιχείο επιτρέπεται να έχει οποιοδήποτε περιεχόμενο.

Τα ElementDefs μπορούν να περιέχουν επεξηγηματικές περιγραφές που αναπαρίστανται από την έννοια Description. Στιγμιότυπα στοιχείων που έχουν ήδη οριστεί ή ορίζονται παρακάτω σε μια DCD αναπαρίστανται από την έννοια Element.

- Group, που περιγράφει τα δοχεία στοιχείων, τα οποία περιέχουν τα υποστοιχεία των στοιχείων που έχουν σύνθετη δομή. Τα Groups διαθέτουν τα εξής χαρακτηριστικά:
 - Occurs, που εκφράζει τον αριθμό των εμφανίσεων του Group. Το χαρακτηριστικό Occurs μπορεί να πάρει τις εξής τιμές:
 - Required, αν το Group πρέπει να εμφανιστεί οπωσδήποτε.
 - Optional, αν δεν είναι απαραίτητο να εμφανιστεί το Group.
 - OneOrMore, αν το Group πρέπει να εμφανιστεί μία ή περισσότερες φορές.

- ZeroOrMore, αν το Group πρέπει να εμφανιστεί μηδέν ή περισσότερες φορές.
- Order, που εκφράζει τη σειρά με την οποία μπορούν να εμφανιστούν τα περιεχόμενα του Group. Το χαρακτηριστικό Order μπορεί να πάρει τις εξής τιμές:
 - Seq, αν τα μέλη του Group θα εμφανιστούν με την ίδια σειρά με την οποία εμφανίζονται μέσα στο Group.
 - Alt, αν τα μέλη του Group είναι εναλλακτικά το ένα του άλλου.
- AttributeDef, που περιγράφει τα χαρακτηριστικά των στοιχείων των XML εγγράφων που είναι σύμφωνα με την τρέχουσα DCD. Όπως και τα ElementDefs, τα AttributeDefs μπορούν να περιέχουν επεξηγηματικές περιγραφές. Στιγμιότυπα χαρακτηριστικών που έχουν ήδη οριστεί ή ορίζονται παρακάτω σε μια DCD αναπαρίστανται από την έννοια Attribute.
- InternalEntityDef, που περιγράφει τις οντότητες που ορίζονται στην τρέχουσα DCD.
- ExternalEntityDef, που περιγράφει τις οντότητες που ορίζονται έξω από την τρέχουσα DCD.

Ένα παράδειγμα χρήσης DCD φαίνεται στο *Παράδειγμα 3-3*, όπου έχουμε τη DCD που περιγράφει τα χαρακτηριστικά του στοιχείου DL της HTML.

```
<DCD>
  <ElementDef Type="DL" Model="Elements" Content="Closed">
    <Description> A simple 'definition list' construct, which
      contains paired 'DT' (DL Term) and 'DD' (DL
      Definition) Elements
    </Description>
    <Group Occurs="OneOrMore" RDF:Order="Seq">
      <Element>DT</Element>
      <Group Occurs="Optional">
        <Element>DD</Element>
      </Group>
    </Group>
  </ElementDef>
  <ElementDef Type="DT" Model="Data" Content="Closed">
    <Description> The term being defined in a DL list item
    </Description>
  </ElementDef>
  <ElementDef Type="DD" Model="Mixed" Content="Open">
    <Description> A term's definition in a DL list item
    </Description>
    <!--Open because lots of markup can be in a DL-->
  </ElementDef>
</DCD>
```

Παράδειγμα 3-3: DCD περιγραφή των χαρακτηριστικών του στοιχείου DL της HTML

Όπως φαίνεται στο *Παράδειγμα 3-3*, αρχικά ορίζεται το στοιχείο DL, το οποίο περιέχει DTs και DDs. Κάθε DT αντιστοιχεί σε οποιοδήποτε αριθμό DDs, ενώ κάθε DD αντιστοιχεί σε ένα ακριβώς DT.

XML Schema Language

Το γεγονός ότι, εκτός από τους XML DTDs, έχουν γίνει αρκετές ακόμα προτάσεις για γλώσσες περιγραφής της δομής των XML εγγράφων, οι περισσότερες από τις οποίες παρέχουν πλουσιότερη λειτουργικότητα και διαθέτουν ευρύτερη σημασιολογία από τους DTDs, δείχνει ότι υπάρχει η ανάγκη για μια γλώσσα περιγραφής σχημάτων XML εγγράφων. Αυτό οδήγησε το W3C στη συγκρότηση της ομάδας εργασίας XML Schema [XMLSche]. Τα αποτελέσματα της ομάδας αυτής στοχεύουν στην υποστήριξη εφαρμογών που χρειάζονται διαφορετική⁸ λειτουργικότητα και σημασιολογία σε ότι αφορά τη δομή των XML εγγράφων.

Η ομάδα εργασίας XML Schema ασχολείται με τα εξής θέματα:

- *Δομικά Σχήματα (Structural Schemas)*: Μηχανισμός αντίστοιχος με τους DTDs, που χρησιμοποιείται για τον καθορισμό της δομής (σειρά εμφάνισης στοιχείων, επιτρεπτός αριθμός εμφανίσεων κάθε στοιχείου, χαρακτηριστικά) και του περιεχομένου (τύποι δεδομένων, οντότητες, σημειογραφίες) των XML εγγράφων. Επιπλέον της λειτουργικότητας των DTDs, ο μηχανισμός αυτός παρέχει τα εξής:
 - Συνδυασμό XML εγγράφων με namespaces (για τις namespaces βλέπε την υποενότητα 3.2.1.3).
 - Συνδυασμό δομικών σχημάτων με πρωτογενείς τύπους δεδομένων (primitive data types).
 - Κληρονομικότητα σε ότι αφορά ορισμούς στοιχείων, χαρακτηριστικών και τύπων δεδομένων.
 - Τμηματικό ορισμό των περιορισμών για το περιεχόμενο συγκεκριμένων τύπων στοιχείων. Τα σχήματα επιτρέπουν τον έλεγχο περιορισμών που καλύπτουν τις ανάγκες μιας ευρείας γκάμας XML εφαρμογών: τα σχήματα όχι μόνο αναδομούν τους περιορισμούς που παρέχουν οι XML DTDs με τη χρήση XML σύνταξης, αλλά επιτρέπουν τον ορισμό νέων τύπων περιορισμών.
 - Την παροχή «υπονοούμενης» (implicit) πληροφορίας, όπως είναι οι προκαθορισμένες τιμές των χαρακτηριστικών των στοιχείων.
 - Μηχανισμούς για την υποστήριξη ενσωματωμένης τεκμηρίωσης: Τα ίδια τα σχήματα τεκμηριώνουν τη σημασία, τη χρήση και τη λειτουργία τους μέσα από ένα κοινό λεξικό τεκμηρίωσης.
 - Μηχανισμούς για την υποστήριξη της εξέλιξης των σχημάτων (schema evolution).
 - Μηχανισμούς για την υποστήριξη περιορισμών και περιγραφών «δεμένων» με συγκεκριμένες εφαρμογές.
- *Τύποι Δεδομένων (Data Types)*: Μηχανισμός υποστήριξης τύπων δεδομένων που καλύπτει τις ανάγκες των συγγραφέων XML εγγράφων και εφαρμογών που βασίζονται

⁸ Ο όρος διαφορετική που μπορεί να σημαίνει περισσότερο ή λιγότερο πλούσια

στην XML για ένα σταθερό και επεκτάσιμο σύστημα υποστήριξης τύπων δεδομένων για την XML, που θα μπορούσε να ενσωματωθεί σε XML επεξεργαστές. Ο μηχανισμός αυτός παρέχει την εξής λειτουργικότητα:

- Υποστήριξη πρωτογενών τύπων δεδομένων (που περιλαμβάνουν bytes, ημερομηνίες, ακέραιους, ακολουθίες και πρωτογενείς τύπους δεδομένων που υποστηρίζονται από την SQL και τη Java).
 - Υποστήριξη τύπων που ορίζονται από τους χρήστες με το συνδυασμό και τη θέση περιορισμών (που περιλαμβάνουν περιοχή τιμών, ακρίβεια, μήκος κλπ.) στα χαρακτηριστικά των πρωτογενών τύπων.
 - Τον ορισμό ενός συστήματος τύπων που υποστηρίζει την ανταλλαγή δεδομένων με ΣΔΒΔ (Συστήματα Διαχείρισης Βάσεων Δεδομένων).
 - Το διαχωρισμό των απαιτήσεων που σχετίζονται με την αναπαράσταση λεξιλογιακών δεδομένων από τις απαιτήσεις που σχετίζονται με την αναπαράσταση απλών συνόλων δεδομένων.
- **Συμβατότητα (Conformance):** Καθορισμός των προϋποθέσεων που πρέπει να τηρούνται έτσι ώστε:
- Να είναι συντακτικά σωστό ένα σχήμα.
 - Να υπακούει ένα XML έγγραφο σε ένα σχήμα.

Η περιγραφή της XML Schema Language αποτελείται από δυο τμήματα:

1. Στο πρώτο τμήμα (XML Schema Part 1: Structures [XML-Sche1]) περιέχονται οι δυνατότητες περιγραφής της δομής των XML εγγράφων και η δυνατότητα να τεθούν περιορισμοί στο περιεχόμενό τους.
2. Στο δεύτερο τμήμα (XML Schema Part 2: Structures [XML-Sche2]) περιγράφεται μια γλώσσα που επιτρέπει τον ορισμό τύπων δεδομένων που μπορούν να χρησιμοποιηθούν σε XML σχήματα.

Τα σχήματα που ορίζονται με βάση την XML Schema Language είναι XML έγγραφα που υπακούουν στο DTD που έχει οριστεί για τον ορισμό σχημάτων. Ο σκοπός των σχημάτων είναι ο ορισμός και η τεκμηρίωση (documentation) κλάσεων XML εγγράφων με τη χρήση δομών που περιορίζουν και περιγράφουν τη σημασία, τη χρήση και τις σχέσεις των τμημάτων από τα οποία τα έγγραφα απαρτίζονται (στοιχείων, χαρακτηριστικών, οντοτήτων, σημειογραφιών και τύπων δεδομένων). Επιπλέον, στόχος είναι η παροχή ενός καταλόγου δομών XML στοιχειοθεσίας που θα επιτρέπουν τον ορισμό XML σχημάτων.

Καθώς τα ίδια τα σχήματα είναι XML έγγραφα, στη συνέχεια (Πίνακας 3-6) παρατίθενται ορισμοί οι οποίοι ξεκαθαρίζουν τη σχέση μεταξύ κάποιων συγκεκριμένων τύπων XML εγγράφων.

Όνομα	Ορισμός
Στιγμιότυπο	Ένα XML έγγραφο που η δομή του υπακούει σε κάποιο σχήμα
Σχήμα	Ένα σύνολο κανόνων, που καθορίζουν τους περιορισμούς που τίθενται για τη δομή XML εγγράφων που ανήκουν σε μια συγκεκριμένη κλάση και

	σκιαγραφούν το σύνολο της πληροφορίας που μπορεί να περιέχουν τα έγγραφα αυτής της κλάσης
Ορισμός Σχήματος	Ένα XML έγγραφο που περιέχει τον ορισμό ενός σχήματος. Αυτό σημαίνει ότι κάθε ορισμός σχήματος είναι ταυτόχρονα και στιγμιότυπο

Πίνακας 3-6: Ορισμοί XML Εγγράφων

Ένας ορισμός σχήματος απαρτίζεται από τα εξής συστατικά στοιχεία:

Σχήμα (Schema):	Ένα περιτύλιγμα που περιέχει όλους τους ορισμούς και τις δηλώσεις που απαρτίζουν τον ορισμό ενός σχήματος.
Ορισμός Τύπου Δεδομένων (Datatype Definition):	Ένας τύπος ⁹ (π.χ. ακέραιος), που εφαρμόζεται σε συμβολοσειρές που περιέχονται σε κάποιο στιγμιότυπο (είτε οι συμβολοσειρές αυτές αποτελούν τιμές χαρακτηριστικών είτε αποτελούν περιεχόμενο).
Αρχέτυπος Ορισμός (Archetype Definition):	Ένα πλήρες σύνολο περιορισμών που εφαρμόζονται τόσο στα χαρακτηριστικά όσο και στο περιεχόμενο των στοιχείων που περιέχονται στα στιγμιότυπα.
Δήλωση Τύπου Στοιχείων (Element Type Declaration):	Συσχέτιση μεταξύ ενός ονόματος στοιχείου και ενός αρχετύπου, αντίστοιχη με τον ορισμό τύπων στοιχείων στους DTDs.
Δήλωση Χαρακτηριστικού (Attribute Declaration):	Συσχέτιση μεταξύ ενός ονόματος χαρακτηριστικού και ενός αρχετύπου, αντίστοιχη με τον ορισμό χαρακτηριστικών στους DTDs.
Τύπος Περιεχομένου (Content Type):	Τύπος δεδομένων ή μοντέλο δεδομένων, που παρέχει μια ενσπονημένη αφαίρεση για τους περιορισμούς που τίθενται στο περιεχόμενο των στοιχείων. Ένας τύπος περιεχομένου εφαρμόζεται στο περιεχόμενο των στοιχείων ενός στιγμιότυπου, αλλά όχι και στις τιμές των χαρακτηριστικών του.
Μοντέλο Περιεχομένου Στοιχείων (Element Content Model):	Ένας τύπος ⁹ που εφαρμόζεται στο περιεχόμενο των στοιχείων ενός στιγμιότυπου. Τα μοντέλα περιεχομένου στοιχείων δεν περιλαμβάνουν πληροφορία για τους ορισμούς των χαρακτηριστικών των στοιχείων.
Περιεχόμενο Στοιχείων (Element-Only Content):	Συστατικά για τη δημιουργία μοντέλων περιεχομένου, που περιλαμβάνουν την υποστήριξη ομαδοποίησης και ακολουθιών, καθώς και δηλώσεις τύπων στοιχείων και αναφορές στους τύπους αυτούς.
Ορισμός Ομάδων Χαρακτηριστικών (Attribute Group Definition):	Συσχέτιση μεταξύ ενός ονόματος και μιας επαναχρησιμοποιήσιμης συλλογής δηλώσεων χαρακτηριστικών.

⁹ Που αποτελεί περιορισμό περιεχομένου

<i>Επώνυμο Μοντέλο Ομάδας</i> (<i>Named Model Group</i>):	Συσχέτιση μεταξύ ενός ονόματος και ενός μοντέλου ομάδας ¹⁰ , που γίνεται έτσι επαναχρησιμοποιήσιμο.
<i>Τροποποίηση Αρχετύπου</i> (<i>Archetype Refinement</i>):	Προσδιορισμός ενός αρχετύπου με βάση ένα ή περισσότερα άλλα αρχέτυπα, από τα οποία παίρνει ορισμούς τόσο για το περιεχόμενο, όσο και για τα χαρακτηριστικά των στοιχείων που περιέχονται στα στιγμιότυπα.
<i>Εισαγωγή Σχήματος</i> (<i>Schema Import</i>):	Επέκταση του τρέχοντος σχήματος με ορισμούς και/ή δηλώσεις από κάποιο εξωτερικό σχήμα, διατηρώντας τη συσχέτιση με το αρχικό εξωτερικό σχήμα.
<i>Ενσωμάτωση Σχήματος</i> (<i>Schema Inclusion</i>):	Ενσωμάτωση ορισμών και/ή δηλώσεων στο υπό ορισμό σχήμα από ένα εξωτερικό σχήμα, σαν να ήταν τοπικά ορισμένοι/ες.

Το πλαίσιο υποστήριξης τύπων δεδομένων από την XML Schema Language επηρεάστηκε τόσο από τους τύπους δεδομένων της SQL και των γλωσσών προγραμματισμού όσο και από το πρότυπο ISO 11404 [ISO11404] για τύπους δεδομένων ανεξάρτητους από γλώσσες.

Στα πλαίσια της XML Schema Language, ένας τύπος δεδομένων διαθέτει ένα σύνολο διακριτών τιμών, που αποτελούν το πεδίο τιμών του, και χαρακτηρίζεται από απόψεις (facets)¹¹ και/ή ιδιότητες αυτών των τιμών, καθώς και από λειτουργίες πάνω σε αυτές τις τιμές ή που είναι αποτέλεσμά τους. Επιπλέον, κάθε τύπος δεδομένων χαρακτηρίζεται από ένα πεδίο που αποτελείται από αποδεκτές λεξικογραφικές αναπαραστάσεις κάθε τιμής που ανήκει σε αυτό το πεδίο τιμών.

Ένα παράδειγμα σχήματος, που χρησιμοποιείται για την αναπαράσταση σημειωμάτων, αντίστοιχου με το DTD που περιέχεται στο Παράδειγμα 3-1, φαίνεται στο Παράδειγμα 3-4.

```
<!DOCTYPE schema PUBLIC "-//W3C//DTD XML Schema Version 1.0//EN"
SYSTEM 'http://www.w3.org/1999/05/06-xmlschema-1/structures.dtd'>
<schema name='file://usr/schemas/xmlschema/memo.xsd'
version='M.n'
xmlns='http://www.w3.org/1999/05/06-xmlschema-1/structures.xsd'>
<elementType name='memo'>
<sequence>
<elementTypeRef name='title'>
<elementTypeRef name='body'>
</sequence>
</elementType>
<elementType name='body'>
<any/>
</elementType>
<elementType name='title'>
<empty/>
```

¹⁰ Τα μοντέλα ομάδας αποτελούν συστατικά των μοντέλων περιεχομένου.

¹¹ Μια άποψη (facet) είναι μια μόνο συνιστώσα του ορισμού ενός αντικειμένου. Εν γένει, μια άποψη χαρακτηρίζει το αντικείμενο σε ανεξάρτητες διαστάσεις και χώρους.

```
</elementType>
<textEntity name='greater'>
  ">"
</textEntity>
</schema>
```

Παράδειγμα 3-4: Σχήμα για την αναπαράσταση Σημειωμάτων

3.2.1.2. Πληροφορία που περιέχεται σε XML Έγγραφα

Σύμφωνα με τον ορισμό της, η XML περιγράφει τη φυσική αναπαράσταση των XML εγγράφων. Για να γίνει δυνατή η χρήση προτύπων βασισμένων στην XML που θα επιτρέπουν την υψηλότερου επιπέδου περιγραφή της λογικής αναπαράστασης των εγγράφων, συγκροτήθηκε από το W3C η ομάδα εργασίας XML Information Set [XMLISRQ] [XMLInfoSet]. Βασική απαίτηση ήταν το αποτέλεσμα των εργασιών της ομάδας αυτής να εγγυάται τη διαλειτουργικότητα μεταξύ των εφαρμογών και των εργαλείων που βασίζονται στην XML. Η ομάδα εργασίας XML Information Set όρισε το *Σύνολο XML Πληροφορίας (XML Information Set)*, το οποίο είναι ένα αφηρημένο σύνολο δεδομένων για την περιγραφή του της πληροφορίας που περιέχουν τα καλά ορισμένα XML έγγραφα.

Το σύνολο πληροφορίας ενός XML εγγράφου απαρτίζεται από δύο τουλάχιστον *αντικείμενα πληροφορίας (information items)*. Ένα αντικείμενο πληροφορίας είναι μια αφηρημένη αναπαράσταση ενός συστατικού (component) που ανήκει σε ένα XML έγγραφο. Κάθε αντικείμενο πληροφορίας διαθέτει ένα σύνολο ιδιοτήτων, κάποιες από τις οποίες είναι απαραίτητο να είναι διαθέσιμες μέσω του συνόλου πληροφορίας ενώ κάποιες άλλες είναι προαιρετικές. Το σύνολο πληροφορίας ενός XML εγγράφου μπορεί να περιέχει 11 διαφορετικούς τύπους αντικειμένων πληροφορίας. Οι τύποι αυτοί είναι οι εξής:

1. Το αντικείμενο πληροφορίας του εγγράφου, που είναι απαραίτητο για κάθε σύνολο πληροφορίας XML εγγράφου, καθώς όλα τα υπόλοιπα αντικείμενα πληροφορίας που πιθανώς περιέχονται στο σύνολο πληροφορίας συνδέονται άμεσα ή έμμεσα με το αντικείμενο πληροφορίας του εγγράφου.
2. Τα αντικείμενα πληροφορίας που αντιστοιχούν στα στοιχεία που απαρτίζουν το έγγραφο. Για κάθε στοιχείο του εγγράφου είναι απαραίτητο να υπάρχει στο σύνολο πληροφορίας του εγγράφου ένα αντικείμενο πληροφορίας.
3. Τα αντικείμενα πληροφορίας που αντιστοιχούν στα χαρακτηριστικά των στοιχείων που απαρτίζουν το έγγραφο. Για κάθε χαρακτηριστικό στοιχείου του εγγράφου είναι απαραίτητο να υπάρχει στο σύνολο πληροφορίας του εγγράφου ένα αντικείμενο πληροφορίας.
4. Τα αντικείμενα πληροφορίας που αντιστοιχούν στις εντολές επεξεργασίας που περιέχονται στο έγγραφο. Για κάθε εντολή επεξεργασίας του εγγράφου είναι απαραίτητο να υπάρχει στο σύνολο πληροφορίας του εγγράφου ένα αντικείμενο πληροφορίας.

5. Τα αντικείμενα πληροφορίας που αντιστοιχούν σε αναφορές σε άγνωστες οντότητες¹² που περιέχονται στο έγγραφο. Για κάθε άγνωστη οντότητα του εγγράφου είναι απαραίτητο να υπάρχει στο σύνολο πληροφορίας του εγγράφου ένα αντικείμενο πληροφορίας.
6. Τα αντικείμενα πληροφορίας που αντιστοιχούν στους χαρακτήρες πληροφορίας που περιέχονται στο έγγραφο. Για κάθε χαρακτήρα πληροφορίας του εγγράφου είναι απαραίτητο να υπάρχει στο σύνολο πληροφορίας του εγγράφου ένα αντικείμενο πληροφορίας.
7. Τα αντικείμενα πληροφορίας που αντιστοιχούν στα σχόλια που περιέχονται στο έγγραφο. Δεν είναι απαραίτητο για κάθε σχόλιο που υπάρχει στο έγγραφο να υπάρχει στο σύνολο πληροφορίας του εγγράφου ένα αντικείμενο πληροφορίας.
8. Τα αντικείμενα πληροφορίας που αντιστοιχούν στις οντότητες που περιέχονται στο έγγραφο. Για καθεμιά από τις οντότητες που τόσο αυτές όσο και η περιγραφή τους περιέχονται στο έγγραφο δεν είναι απαραίτητο να υπάρχει αντικείμενο πληροφορίας στο σύνολο πληροφορίας του εγγράφου. Αντίθετα, είναι απαραίτητο να υπάρχει στο σύνολο πληροφορίας του εγγράφου ένα αντικείμενο πληροφορίας για καθεμιά από τις οντότητες που περιέχονται στο έγγραφο αλλά που η περιγραφή τους δεν περιέχεται σ' αυτό.
9. Το αντικείμενο πληροφορίας που αντιστοιχεί στο DTD του εγγράφου, αν το έγγραφο διαθέτει DTD. Αν το έγγραφο διαθέτει DTD, η ύπαρξη του αντίστοιχου αντικειμένου πληροφορίας στο σύνολο πληροφορίας του εγγράφου είναι προαιρετική.
10. Τα αντικείμενα πληροφορίας που αντιστοιχούν στις σημειογραφίες που περιέχονται στο έγγραφο. Για κάθε σημειογραφία του εγγράφου είναι απαραίτητο να υπάρχει στο σύνολο πληροφορίας του εγγράφου ένα αντικείμενο πληροφορίας.
11. Τα αντικείμενα πληροφορίας που αντιστοιχούν στις λίστες χαρακτηριστικών των στοιχείων που μπορεί να περιέχονται στο έγγραφο. Για κάθε λίστα χαρακτηριστικών η ύπαρξη του αντίστοιχου αντικειμένου πληροφορίας στο σύνολο πληροφορίας του εγγράφου είναι προαιρετική.

Όπως αναφέρθηκε παραπάνω, το σύνολο πληροφορίας κάθε καλά ορισμένου εγγράφου διαθέτει τουλάχιστον δυο αντικείμενα: το αντικείμενο πληροφορίας του εγγράφου και το αντικείμενο πληροφορίας που αντιστοιχεί στο στοιχείο που αναπαριστά ολόκληρο το έγγραφο.

Το σύνολο πληροφορίας ενός XML εγγράφου μπορεί να υλοποιηθεί με οποιοδήποτε επίπεδο διεπαφής (interface). Τέτοια επίπεδα διεπαφής είναι το δένδρικό (tree-based), το βασισμένο σε γεγονότα (event-based), το βασισμένο σε ερωτήσεις (query-based) κλπ.

Στο σημείο αυτό πρέπει να σημειωθεί ότι για την υποστήριξη της δυνατότητας να εσωκλείεται σε ένα XML έγγραφο κάποιο άλλο XML έγγραφο και παράλληλα να εσωκλείεται το Σύνολο XML Πληροφορίας του δεύτερου εγγράφου στο Σύνολο XML

¹² Οι οντότητες αυτές είναι οντότητες που δεν αναγνωρίζονται από τους συντακτικούς αναλυτές που ελέγχουν μόνο αν τα έγγραφα που αναλύουν είναι καλά ορισμένα (και όχι αν είναι σωστά).

Πληροφορίας του πρώτου εγγράφου, έχει προταθεί από το W3C η σύνταξη XInclude [XIncl].

3.2.1.3. XML Namespaces

Όπως αναφέρθηκε παραπάνω, οι DTDs και οι άλλες προτάσεις για τον ορισμό της δομής των XML εγγράφων δίνουν τη δυνατότητα περιγραφής κλάσεων δομημένων εγγράφων και δομικών στοιχείων εγγράφων. Αυτό σημαίνει ότι, κατά την κατασκευή ενός εγγράφου, ο συγγραφέας μπορεί να ορίσει την ύπαρξη σε αυτό δομικών στοιχείων που ορίζονται σε περισσότερους από ένα DTDs ή σχήματα που χρησιμοποιούνται από το έγγραφο. Το πρόβλημα που υπάρχει σ' αυτή την περίπτωση είναι η «σύγκρουση» (*conflict*) των ονομάτων που ορίζονται στους DTDs ή τα σχήματα: αν ο συγγραφέας ενός εγγράφου χρησιμοποιήσει το στοιχείο `body` που ορίζεται σε περισσότερους από ένα από τους DTDs που χρησιμοποιεί, σε ποιο από τους δύο ορισμούς πρέπει να υπακούει το στοιχείο αυτό;

Τα προβλήματα αυτά λύνονται από μια τεχνολογία συμπληρωματική στην XML τις *XML Namespaces* [XML-Name]. Ο στόχος των XML Namespaces είναι η παροχή ενός ακριβέστερου μηχανισμού αναγνώρισης των στοιχείων που απαρτίζουν τα έγγραφα που υπακούουν σε κάποιο DTD ή σχήμα. Έτσι θα αποφεύγονται πιθανές «συγκρούσεις» ονομάτων στοιχείων και θα μπορούν να χρησιμοποιούνται από τα έγγραφα πολλαπλά DTDs και σχήματα, καθώς κάθε στοιχείο θα έχει ένα μοναδικό όνομα.

Οι XML namespaces παρέχουν μια απλή μέθοδο πιστοποίησης των ονομάτων που χρησιμοποιούνται από τα στοιχεία και τα χαρακτηριστικά των XML εγγράφων. Αυτό επιτυγχάνεται με τη σύνδεση των ονομάτων με namespaces που χαρακτηρίζουν συγκεκριμένους URIs (*Uniform Resource Identifiers*), οι οποίοι περιγράφουν τα αντικείμενα (DTDs, σχήματα ή έγγραφα) στα οποία ορίζονται τα ονόματα αυτά.

Έτσι, οι XML namespaces ορίζονται ως συλλογές ονομάτων, τα οποία χρησιμοποιούνται ως τύποι στοιχείων και ονόματα χαρακτηριστικών στα XML έγγραφα. Τα ονόματα αυτά χαρακτηρίζονται από τους URIs με τους οποίους συνδέονται κατά τον ορισμό των namespaces.

Οι namespaces έχουν ισχύ στα όρια των στοιχείων στα οποία ορίζονται, καθώς και στα όρια των υποστοιχείων τους.

Παράδειγμα ορισμού namespace έχουμε στο *Παράδειγμα 3-5*, όπου για ένα έγγραφο που υπακούει στο DTD που ορίστηκε στο *Παράδειγμα 3-1* ορίζεται η `htm` namespace. Η namespace αυτή χαρακτηρίζεται από το URI που περιγράφει τον ορισμό της HTML. Στη συνέχεια, μέσα στο στοιχείο `body` του εγγράφου εμφανίζεται μια `html` επικεφαλίδα επιπέδου 1 (`h1`).

```
<memo xmlns:htm='http://www.w3.org/TR/REC-html40'>
  <title>Μάθημα</title>
  <body>
    <htm:h1> Αυριανό Μάθημα </htm:h1>
    Αύριο στις 10:00 π.μ. έχω μάθημα αγγλικών.
  </body>
</memo>
```

3.2.1.4. Υποστήριξη Συνδέσμων

Είναι κοινά αποδεκτό ότι ένα από τα βασικότερα προβλήματα που συναντώνται στα HTML έγγραφα είναι οι «σπασμένοι» σύνδεσμοι υπερκειμένου, οι οποίοι δεν οδηγούν πουθενά. Το πρόβλημα αυτό γίνεται ιδιαίτερα σημαντικό, καθώς πολλές φορές οι σύνδεσμοι είναι σημαντικότεροι από την πληροφορία που συνδέουν: Ένας σύνδεσμος δεν είναι απλά μια εντολή που επιτρέπει τη μεταπήδηση από έγγραφο σε έγγραφο (ή σε κάποιο άλλο σημείο μέσα στο ίδιο έγγραφο), αλλά περιγράφει τη σαφή σχέση που υπάρχει μεταξύ δύο ή περισσότερων αντικειμένων [DuCha98]¹³. Το γεγονός αυτό, εκτός του ότι πιστοποιεί την αξία των συνδέσμων, σημαίνει ότι:

- Οι σύνδεσμοι μπορεί να παρέχουν πληροφορία επιπλέον αυτής που παρέχουν τα συνδεόμενα αντικείμενα (π.χ. ο τύπος της σχέσης που περιγράφεται από κάποιο σύνδεσμο).
- Είναι χρήσιμη η δυνατότητα δημιουργίας συνδέσμων χωρίς να απαιτούνται αλλαγές στα συνδεόμενα αντικείμενα. Για παράδειγμα, είναι χρήσιμο να μπορεί να δημιουργηθεί ένας σύνδεσμος που συνδέει ένα βιβλίο με μια κριτική γι αυτό χωρίς να γίνει οποιαδήποτε αλλαγή στην ηλεκτρονική έκδοση του βιβλίου ή της κριτικής.

Εκτός αυτού, η HTML υποστηρίζει, μέσω των URLs (Uniform Resource Locators) [URL94] και των URIs [URI97], μόνο την απλούστερη μορφή συνδέσμων. Οι σύνδεσμοι αυτοί:

- Είναι μιας κατεύθυνσης.
- Μπορούν να συνδέουν ακριβώς δύο αντικείμενα (την πηγή και το στόχο του κάθε συνδέσμου).
- Ορίζονται σε ένα τουλάχιστον από τα συνδεόμενα αντικείμενα.
- Οι δυνατότητες των χρηστών να γυρίσουν σε σελίδες που επισκέφθηκαν πριν, να ανοίξουν νέα παράθυρα στα προγράμματα πλοήγησης και να διατρέχουν συνδέσμους που αλλάζουν το περιεχόμενο μερικών μόνο πλαισίων (frames) καθορίζονται από τα προγράμματα πλοήγησης και όχι από τους συνδέσμους.

Καθώς είναι ξεκάθαρο ότι τα παραπάνω περιορίζουν τη λειτουργικότητα υπερκειμένου που παρέχεται από τα HTML έγγραφα, η XML οραματίζεται ένα σύστημα υπερκειμένου που δε θα υποφέρει από τους περιορισμούς αυτούς. Για την επίλυση αυτών των προβλημάτων, το W3C προτείνει την *XLL* ή *XLink* (*eXtensible Link Language*) [XLink], που φιλοδοξεί να αποτελέσει ένα πρότυπο μοντέλο συνδέσμων για XML έγγραφα. Μια λεπτομερέστερη αναφορά τόσο στην XLink όσο και στην XPointer και την XPath, από τις οποίες η XLink υποστηρίζεται, γίνεται στη συνέχεια.

XLL ή XLink (eXtensible Link Language)

Η *XLL* ή *XLink* (*eXtensible Link Language*) είναι μια γλώσσα περιγραφής σταθερών συνδέσμων υπερκειμένου. Μέσω της XLink ορίζονται δομές που μπορούν να εισαχθούν σε

¹³ Μια τέτοια κατηγορία σχέσεων είναι οι σχέσεις δεδομένων και εμφανίζονται όταν η XML χρησιμοποιείται για τη μεταφορά δεδομένων και μεταδεδομένων που περιέχουν συνδέσμους.

XML DTDs, ώστε να μπορούν να περιγραφούν σύνδεσμοι υπερκειμένου μέσα στα στιγμιότυπα των εγγράφων που η δομή τους είναι σύμφωνη με αυτούς τους DTDs. Η σύνταξη της XLink βασίζεται στην XML.

Η XLink υποστηρίζει τόσο τους απλούς συνδέσμους που υπάρχουν σήμερα στο Web όσο και πιο εξελιγμένους συνδέσμους. Για την XLink *σύνδεσμος (link)* είναι μια σαφής σχέση μεταξύ δύο ή περισσότερων αντικειμένων ή τμημάτων αντικειμένων. Έτσι, οι συλλογές XLink συνδέσμων μπορούν να αντιμετωπιστούν σαν ανεξάρτητες βάσεις δεδομένων.

Στην XLink ορίζονται δυο τύποι συνδέσμων:

- Οι *απλοί σύνδεσμοι*, οι οποίοι είναι μιας κατεύθυνσης και ορίζονται συνήθως μέσα σε κάποιο από τα συνδεόμενα αντικείμενα.

Οι απλοί σύνδεσμοι παρέχουν αντίστοιχη λειτουργικότητα με τους συνδέσμους που συναντώνται στα HTML έγγραφα και, επιπλέον, μπορούν να περιέχουν σημασιολογική πληροφορία σχετική με τους συνδέσμους και τα συνδεόμενα απ' αυτούς αντικείμενα.

- Οι πιο γενικευμένοι *επεκταμένοι σύνδεσμοι (extended links)*, οι οποίοι μπορεί να ορίζονται εντός ή εκτός των συνδεομένων αντικειμένων και έχουν πιο σύνθετη δομή.

Οι επεκταμένοι σύνδεσμοι διαφέρουν από τους απλούς στο ότι μπορούν να συνδέουν οποιοδήποτε αριθμό πόρων και στο ότι ορίζονται συχνότερα εκτός των συνδεομένων αντικειμένων. Οι επιπλέον δυνατότητες που παρέχουν οι επεκταμένοι σύνδεσμοι είναι απαραίτητες για:

- Την υποστήριξη συνδέσμων από και προς έγγραφα που δε μπορούν να μετατραπούν έτσι ώστε να οριστεί μέσα σ' αυτά κάποιος σύνδεσμος.
- Τη δημιουργία συνδέσμων από και προς πόρους που η μορφή στην οποία αποθηκεύονται δεν παρέχει υποστήριξη για την ενσωμάτωση συνδέσμων (όπως π.χ. τα αντικείμενα πολυμέσων).
- Την επιλογή και την ενεργοποίηση συνόλων συσχετιζόμενων συνδέσμων κατ' απαίτηση.
- Την ενεργοποίηση εξελιγμένων δυνατοτήτων των υπερμέσων.

Τόσο οι απλοί όσο και οι επεκταμένοι σύνδεσμοι μπορούν να ενσωματώσουν τους εξής σχετικούς με αυτούς τύπους πληροφορίας:

- Πληροφορία που προσδιορίζει την *ταυτότητα* καθενός από τους απομακρυσμένους πόρους που συμμετέχουν στον κάθε σύνδεσμο.
- Πληροφορία *σχετική με το σύνδεσμο*, που καθορίζει το ρόλο του συνδέσμου και το αν ο σύνδεσμος είναι ορισμένος εντός ή εκτός των συνδεόμενων αντικειμένων.
- Πληροφορία *σχετική με τους απομακρυσμένους πόρους που συμμετέχουν στο σύνδεσμο*, που καθορίζει το ρόλο του πόρου, τον τίτλο του και τη συμπεριφορά του όταν διατρέχεται ο σύνδεσμος προς τον πόρο αυτό.
- Πληροφορία *σχετική με τον τοπικό πόρο*, αν ο σύνδεσμος ορίζεται εντός κάποιου από τα συνδεόμενα αντικείμενα, που καθορίζει το ρόλο και τον τίτλο του πόρου.

Τα επιπλέον χαρακτηριστικά που παρέχει η XLink υποστηρίζονται από την *XPointer (XML Pointer Language)* [XPointer], η οποία παρέχει ένα σύνθετο και πλήρη μηχανισμό διευθυνσιοδότησης (addressing). Μια εκτενέστερη αναφορά στην XPointer γίνεται στη συνέχεια.

Η XLink δεν ασχολείται με τον έλεγχο της μορφοποίησης των συνδέσμων, καθώς γι αυτό υπεύθυνα είναι τα στυλ (βλέπε υποενότητα 3.2.1.5). Σε ότι αφορά τη συμπεριφορά των συνδέσμων, η XLink παρέχει μερικούς γενικούς μηχανισμούς συμπεριφοράς, που επιτρέπουν τον καθορισμό της συμπεριφοράς συνδέσμων και πόρων σε ότι αφορά το συγχρονισμό και τα εφέ κατά τη διάσχιση των συνδέσμων.

XPointer (XML Pointer Language)

Η XPointer είναι μια γλώσσα που υποστηρίζει τη διευθυνσιοδότηση βάσει της εσωτερικής δομής των XML εγγράφων. Πιο συγκεκριμένα, η XPointer εκτός από τη διευθυνσιοδότηση αυτόνομων αντικειμένων (π.χ. έγγραφα, εικόνες κλπ.), δίνει τη δυνατότητα καθορισμού ενός σημείου μέσα σε ένα έγγραφο, το οποίο αποτελεί την πηγή ή το στόχο ενός ή περισσότερων συνδέσμων χωρίς να απαιτείται οποιαδήποτε αλλαγή στο ίδιο το έγγραφο. Επιπλέον, δίνει τη δυνατότητα αναφοράς σε στοιχεία, συμβολοσειρές χαρακτήρων και σε οποιοδήποτε άλλο τμήμα XML εγγράφου, ανεξάρτητα από το αν το τμήμα αυτό διαθέτει ή όχι αναγνωριστικό.

Η XPointer βασίζεται στη γλώσσα περιγραφής εκφράσεων XPath [XPath], η οποία περιγράφεται στη συνέχεια. Οι XPointer εκφράσεις εντοπίζουν την πληροφορία μέσω της διάσχισης της λογικής δομής των εγγράφων, έτσι ώστε να είναι δυνατή η επιλογή των τμημάτων από τα οποία αυτή απαρτίζεται με βάση ιδιότητες όπως είναι οι τύποι των στοιχείων, οι τιμές των χαρακτηριστικών τους, το περιεχόμενό τους και η σχετική θέση και η σειρά τους μέσα στη λογική δομή της πληροφορίας.

Η XPointer παρέχει τα μέσα για την αναγνώριση τοποθεσιών σε XML έγγραφα, ανεξάρτητα από το αν αυτές πρόκειται να χρησιμοποιηθούν ως στόχοι συνδέσμων ή για άλλους σκοπούς, που διαφέρουν από εφαρμογή σε εφαρμογή.

Καθώς οι πόροι συνήθως προσδιορίζονται με βάση ένα URI, οι υποπόροι τους προσδιορίζονται με βάση την XPointer. Έτσι, το αναγνωριστικό κάθε τμήματος XML εγγράφου είναι ένας XPointer. Δεν υπάρχει όμως περιορισμός για τη σύνταξη των αναγνωριστικών των τμημάτων της μη XML πληροφορίας.

Η XPointer ενεργεί πάνω στο σύνολο πληροφορίας των XML εγγράφων με βάση ένα δενδρικό επίπεδο διεπαφής, επιλέγοντας τμήματα του δέντρου που αναπαριστά τη λογική δομή κάθε εγγράφου και των υποδέντρων του μέσω της δομικής σχέσης τους με ήδη αναγνωρισμένους κόμβους. Θεωρητικά, οι XPointers είναι επαναληπτικοί, καθώς μπορούν να εκφράσουν πολλαπλές τέτοιες επιλογές, καθεμιά από τις οποίες ενεργεί στο αποτέλεσμα της προηγούμενης.

Η επιλογή των τμημάτων των δέντρων γίνεται με τη χρήση αξόνων και κατηγορημάτων (predicates). Ένας άξονας ορίζει μια σειρά κόμβων ή τμημάτων που δεν περιέχουν πληροφορία ως υποψήφια για επιλογή. Τα κατηγορήματα εξετάζουν την πληροφορία του συνόλου πληροφορίας με βάση τους άξονες.

Από το συνδυασμό XLink και XPointer υποστηρίζονται:

- *Έμμεσοι Σύνδεσμοι (Indirect Links)*, που θα βάλουν τέλος στους «σπασμένους» συνδέσμους. Οι σύνδεσμοι αυτοί καθορίζονται και η διαχείρισή τους γίνεται εκτός των συνδεόμενων αντικειμένων.
- Ο *τελεστής ']*, που θα έχει ως αποτέλεσμα να ανακτάται από τον εξυπηρετητή μόνο το σχετικό με το τρέχον έγγραφο τμήμα των αντικειμένων προς ανάκτηση.
- *Ονομασίες ανεξάρτητες της θέσης των περιγραφόμενων αντικειμένων στο Διαδίκτυο*.
- *Αμφίδρομοι Σύνδεσμοι (Bi-directional Links)*, που θα επιτρέπουν στους χρήστες να επιστρέψουν στο αντικείμενο-πηγή του συνδέσμου χωρίς να επαφίνται γι αυτό στις δυνατότητες του προγράμματος πλοήγησης που χρησιμοποιούν.
- *Πολλαπλοί σύνδεσμοι υπερκειμένου*, που θα επιτρέπουν ο στόχος να είναι περισσότερα του ενός αντικείμενα (που παρουσιάζονται, π.χ., σε πολλαπλά παράθυρα).
- *Συνοθυλεύματα Συνδέσμων (Aggregate Links)*, που θα συνδέουν περισσότερα από δύο αντικείμενα.
- *Κλειστότητα (Transclosure)*, που συμβαίνει όταν ο στόχος ενός συνδέσμου εμφανίζεται να είναι τμήμα του εγγράφου-πηγής του συνδέσμου.
- *Ομάδες Επεκταμένων Συνδέσμων (Extended Link Groups)*, που αποτελούν ένα ειδικό τύπο επεκταμένου συνδέσμου για την αναπαράσταση συνόλων συνδέσμων προς έγγραφα που αποτελούν μια διασυνδεδεμένη ομάδα.
- *Σύνδεσμοι με πολύπλοκη δομή*, που επιτρέπουν τον ορισμό περίπλοκων δομών συνδέσμων που υποστηρίζουν τον ορισμό σχέσεων μεταξύ των τμημάτων των αντικειμένων που παίρνουν μέρος στο σύνδεσμο.

Ο σχεδιασμός της XLink και της XPointer επηρεάστηκε κατά μείζονα λόγο από τα εξής πρότυπα:

HTML: Από την HTML η XLink επηρεάστηκε σε ότι αφορά τους ορισμούς αρκετών SGML στοιχείων για την αναπαράσταση συνδέσμων. Η XPointer χρησιμοποίησε κάποιες από τις σχετικές με τους URLs ιδέες.

HyTime: Από τη HyTime η XLink επηρεάστηκε σε ότι αφορά τις δομές για την αναπαράσταση συνδέσμων που ορίζονται τόσο εντός όσο και εκτός των συνδεόμενων αντικειμένων (inline και out-of-line links) και χαρακτηριστικά όπως ο έλεγχος της διάσχισης των συνδέσμων και η παρουσίαση των αντικειμένων. Η XPointer επηρεάστηκε από τη HyTime σε ότι αφορά τους *προσδιορισμούς θέσης (location specifiers)* για κάθε τύπο δεδομένων.

TEI: Από την TEI (Text Encoding Initiative) [TEI95] η XLink επηρεάστηκε σε ότι αφορά δομές για τη δημιουργία συνδέσμων, συνοθυλεωμάτων αντικειμένων και συλλογών συνδέσμων. Η XPointer επηρεάστηκε από την TEI σε ότι αφορά τη σύνταξη για τους προσδιορισμούς θέσης που αναφέρονται σε δομημένα δεδομένα, γραφικά και άλλους τύπους δεδομένων.

XPath (XML Path Language)

Η XPath (*XML Path Language*) είναι μια γλώσσα που επιτρέπει τη διευθυνσιοδότηση των τμημάτων που απαρτίζουν ένα XML έγγραφο. Η XPath παρέχει κοινή σύνταξη και σημασιολογία λειτουργικότητας τόσο για την XPointer όσο και για την XSLT (βλέπε υποενότητα 3.2.1.5).

Εκτός από την υποστήριξη διευθυνσιοδότησης, η XPath παρέχει και ένα σύνολο βασικών δυνατοτήτων διαχείρισης συμβολοσειρών, αριθμών και δίτιμων μεταβλητών (booleans).

Η σύνταξη της XPath είναι συμπαγής και δε βασίζεται στην XML, έτσι ώστε να είναι δυνατή η χρήση της σε URIs και σε τιμές XML χαρακτηριστικών.

Η XPath λειτουργεί περισσότερο πάνω στην αφηρημένη λογική δομή των XML εγγράφων παρά στη σύνταξή τους.

Η βασική συντακτική δομή της XPath είναι η *Έκφραση (Expression)*. Μια έκφραση υπολογίζεται και αποδίδει κάποιο αποτέλεσμα. Το αποτέλεσμα μιας έκφρασης είναι ένα αντικείμενο που μπορεί να ανήκει σε κάποιον από τους εξής τέσσερις βασικούς τύπους:

1. *Σύνολο κόμβων (node set)*, που είναι μια μη διατεταγμένη συλλογή κόμβων που αναπαριστούν στοιχεία XML εγγράφων, όπου δεν υπάρχουν διπλές εμφανίσεις κόμβων.
2. *Δίτιμη μεταβλητή (boolean)*, που μπορεί να πάρει μια από τις τιμές αληθής και ψευδής.
3. *Αριθμός*, που μπορεί να πάρει οποιαδήποτε τιμή αντιστοιχεί σε πραγματικό αριθμό.
4. *Συμβολοσειρά*, που είναι ένας συνδυασμός χαρακτήρων.

Ο υπολογισμός της τιμής μιας έκφρασης γίνεται με βάση τις παραμέτρους (context) της έκφρασης, οι οποίες καθορίζονται από την XPointer και την XSLT. Οι παράμετροι μιας έκφρασης είναι οι εξής:

- Ο κόμβος των παραμέτρων της έκφρασης (context node).
- Το σύνολο κόμβων των παραμέτρων της έκφρασης (context list node), μέλος του οποίου αποτελεί ο κόμβος των παραμέτρων της έκφρασης.
- Ένα σύνολο αντιστοιχήσεων μεταβλητών (variable bindings), το οποίο αποτελείται από τις αντιστοιχίσεις των ονομάτων των μεταβλητών στις τιμές των μεταβλητών αυτών. Καθεμία από τις μεταβλητές είναι ένα αντικείμενο είτε κάποιου από τους τέσσερις βασικούς τύπους είτε κάποιου άλλου τύπου.
- Μια βιβλιοθήκη λειτουργιών (function library), η οποία είναι ένα σύνολο επωνύμων συναρτήσεων. Κάθε συνάρτηση δέχεται 0 ή περισσότερα ορίσματα και επιστρέφει κάποιο αποτέλεσμα. Η XPath διαθέτει ένα πυρήνα βασικών συναρτήσεων, των οποίων τόσο τα ορίσματα όσο και το αποτέλεσμα ανήκουν στους βασικούς τύπους. Τόσο η XPointer όσο και η XSLT ορίζουν επιπρόσθετες συναρτήσεις, κάποιες από τις οποίες λειτουργούν με τους τέσσερις βασικούς τύπους ενώ κάποιες άλλες ενεργούν σε επιπρόσθετους τύπους που ορίζονται από την XPointer και την XSLT.
- Το σύνολο δηλώσεων των namespaces που εφαρμόζονται στην έκφραση.

3.2.1.5. Σύνδεση Στυλ με XML Έγγραφα

Παρ' όλο που η XML επιτρέπει τον καθορισμό της δομής των εγγράφων, δε μπορεί να χρησιμοποιηθεί για τον καθορισμό της παρουσιάσής τους. Σύμφωνα και με τη γενικότερη φιλοσοφία της XML, η δομή και το περιεχόμενο των εγγράφων καθορίζονται ανεξάρτητα από τον τρόπο με τον οποίο αυτά θα παρουσιαστούν στα μέσα παρουσίασης (χαρτί, οθόνη, ήχος κλπ.) [McGra98]. Έτσι, για την παρουσίαση των εγγράφων προτείνεται η *δημιουργία style sheets*, τα οποία θα καθορίζουν πως θα μορφοποιούνται τα έγγραφα και τα στοιχεία από τα οποία απαρτίζονται και θα παρέχουν πλήρη έλεγχο στην παρουσίασή τους. Η φιλοσοφία αυτή μπορεί να περιγραφεί σαν *καθυστερημένη σύνδεση της σημασιολογίας παρουσίασης (late binding of presentation semantics)*. Αυτό σημαίνει ότι η πληροφορία για το πώς θα φαίνεται ένα έγγραφο κατά την παρουσίασή του αποθηκεύεται χωριστά από το έγγραφο, σε ένα style sheet. Η διαδικασία μορφοποίησης του εγγράφου γίνεται αργότερα, όταν το έγγραφο παρουσιάζεται. Η προσέγγιση αυτή έχει τα εξής πλεονεκτήματα:

- Η παρουσίαση ενός εγγράφου (ή και περισσότερων, αν όλα είναι συνδεδεμένα με το ίδιο style sheet) μπορεί να αλλάξει αυτόματα με τη μετατροπή του style sheet με το οποίο αυτό συνδέεται.
- Πολλαπλές μορφές του ίδιου περιεχομένου μπορούν να παραχθούν με την εφαρμογή διαφορετικών style sheets στο περιεχόμενο αυτό.
- Το περιεχόμενο είναι επαναχρησιμοποιήσιμο και παρέχεται η δυνατότητα να παρουσιαστεί χρησιμοποιώντας πιο εξελιγμένους τρόπους όταν αυτοί γίνουν διαθέσιμοι.
- Η διατήρηση του περιεχομένου ελεύθερου από στοιχεία μορφοποίησης κάνει ευκολότερη την επεξεργασία του.

Για την παρουσίαση των εγγράφων θα μπορούσαν να χρησιμοποιηθούν CSSL style sheets, αλλά η προσέγγιση που προτιμάται είναι η χρήση της XSL (*eXtensible Style sheet Language*) [XSL]. Αυτό συμβαίνει γιατί η XSL παρέχει πολύ περισσότερες δυνατότητες από τη CSSL: Η CSSL υποστηρίζει έγγραφα με καθορισμένη δομή, ενώ η XSL επιτρέπει την αναδιοργάνωση της δομής των εγγράφων και την πολλαπλή επεξεργασία τους. Η XSL αναπτύσσεται από το W3C με στόχο να γίνει δυνατή η αντιστοίχιση στυλ παρουσίασης με τα στοιχεία που ορίζονται σε ένα XML DTD.

Η XSL είναι μια γλώσσα περιγραφής style sheets που έχει σχεδιαστεί έτσι ώστε να ικανοποιεί τις απαιτήσεις των δημιουργών XML ιστοσελίδων. Οι περιοχές εφαρμογών της XSL είναι δύο:

1. Η μετατροπή των XML εγγράφων σε HTML, έτσι ώστε να είναι δυνατή η παρουσίασή τους από προγράμματα πλοήγησης.
2. Ο καθορισμός των στυλ που εφαρμόζονται στα στοιχεία που απαρτίζουν τα XML έγγραφα με τη χρήση των αντικειμένων μορφοποίησης που περιγράφονται στο *Μοντέλο Μορφοποίησης του W3C (W3C Formatting Model)*.

Η επεξεργασία των XML εγγράφων προς παρουσίαση και των αντίστοιχων style sheets [XML-SS] γίνεται από τους XSL *Επεξεργαστές (XSL Processors)*. Η επεξεργασία αυτή

απαρτίζεται από δυο διεργασίες: το μετασχηματισμό δέντρων (*tree transformation*) και τη μορφοποίηση (*formatting*). Έτσι, ο ορισμός της XSL αποτελείται από δύο τμήματα:

- Από την XSLT (*XSL Transformations*) [XSLT], που είναι μια γλώσσα μετασχηματισμού των XML εγγράφων. Η γλώσσα αυτή χρησιμοποιεί την XPath και δίνει τη δυνατότητα έκφρασης κανόνων για τη μετατροπή ενός δέντρου (του δέντρου πηγής – *source tree*) σε ένα άλλο δέντρο (το δέντρο αποτελέσματος – *result tree*).
- Από ένα XML λεξικό για τον καθορισμό της σημασιολογίας της μορφοποίησης.

Ένα XSL style sheet καθορίζει, μέσω της XSLT¹⁴, την παρουσίαση μιας κλάσης XML εγγράφων περιγράφοντας πως ένα στιγμιότυπο της κλάσης μετατρέπεται σε XML έγγραφο που χρησιμοποιεί το λεξικό μορφοποίησης. Η XSL παρέχει τη δυνατότητα σύνδεσης της πληροφορίας μορφοποίησης με στοιχεία των εγγράφων πηγής, έτσι ώστε να παραχθεί μορφοποιημένη έξοδος.

Με τη χρήση της XSL καθορίζονται το πώς εξάγεται η πληροφορία από ένα XML έγγραφο και το πώς μορφοποιείται η πληροφορία αυτή κατά τη διάρκεια της παρουσίασης του εγγράφου. Συχνά, όπως συμβαίνει και με τον XSL επεξεργαστή της Microsoft (τον MSXSL), τα XML έγγραφα μετατρέπονται σε άλλη μορφή (π.χ. σε HTML).

Η XSL μπορεί, χάρη στην επεκτασιμότητά της, να διαχειριστεί απεριόριστο αριθμό στοιχείων. Παράλληλα μπορεί να διαχειριστεί το καθένα από τα στοιχεία αυτά με απεριόριστο αριθμό τρόπων. Η XSL εμπλουτίζει τις δυνατότητες μορφοποίησης των σελίδων στον Παγκόσμιο Ιστό με περιστρεμμένο κείμενο, πολλαπλές στήλες, ανεξάρτητες περιοχές και όλες τις μορφές οριζόντιας (αριστερή, δεξιά, κεντραρισμένη και πλήρης) και κάθετης στοίχισης (στην κορυφή, στη βάση, κεντραρισμένη και πλήρης).

Οι έννοιες που αποτελούν τη βάση της XSL είναι οι εξής:

- *Αντικείμενα Ροής (Flow Objects)*: Τα αντικείμενα (π.χ. παράγραφοι, πίνακες κ.α.) που απαρτίζουν τα έγγραφα της μορφής (π.χ. RTF, HTML κ.α.) στην οποία μετατρέπονται τα XML έγγραφα κατά την παρουσίασή τους.
- *Αντικείμενα Ροής Ανεξάρτητα από Πλατφόρμα (Platform-Independent Flow Objects)*: Ένα σύνολο πρότυπων αντικειμένων ροής (π.χ. παράγραφος, σελίδα, ακολουθία σελίδων κλπ.), που ορίζονται από την XSL. Η χρήση τους επιτρέπει τη δημιουργία πολλών μορφών εξόδου για κάθε έγγραφο με τη χρήση ενός και μόνο style sheet.
- *Αντικείμενα Ροής βασισμένα στην HTML (HTML-Specific Flow Objects)*: Ένα σύνολο αντικειμένων ροής που αναπαριστούν τα HTML στοιχεία. Με τη χρήση των στοιχείων αυτών γίνεται δυνατή η μετατροπή XML εγγράφων σε HTML με τη χρήση XSL style sheets.
- *Κανόνες Παραγωγής (Production Rules)*: Κανόνες που περιέχονται στα style sheets και ελέγχουν τη δημιουργία αντικειμένων ροής. Πιο συγκεκριμένα, οι κανόνες παραγωγής καθορίζουν ποια αντικείμενα ροής πρέπει να δημιουργηθούν σε κάθε περίπτωση και τι αυτά θα περιέχουν.

¹⁴ Η XSLT έχει σχεδιαστεί έτσι ώστε να μπορεί να χρησιμοποιηθεί και ανεξάρτητα από την XSL. Ο βασικός στόχος των σχεδιαστών της πάντως ήταν η χρήση της ως τμήματος της XSL.

- *Χαρακτηριστικά (Characteristics)*: Χαρακτηριστικά που συνδέονται με τα αντικείμενα ροής και καθορίζουν, ανάλογα με την κατηγορία στην οποία τα τελευταία ανήκουν, την παρουσίασή τους.
- *Scripting*: Η ECMAScript, που είναι ένα περιορισμένο υποσύνολο της Javascript, καθορίζεται ως η γλώσσα scripting για το περιβάλλον της XSL. Η χρήση μιας γλώσσας scripting στο περιβάλλον μιας γλώσσας περιγραφής style sheets είναι επιβεβλημένη, καθώς καμία γλώσσα περιγραφής style sheets που παρέχει ένα προκαθορισμένο σύνολο δυνατοτήτων δε μπορεί να εγγυηθεί την επεκτασιμότητα και τη μελλοντική επάρκειά της.

Τα XSL style sheets απαρτίζονται από κανόνες οι οποίοι αποτελούνται από δύο τμήματα: Το πρώτο τμήμα των XSL κανόνων, το *τμήμα υπόθεσης (condition part)*, καθορίζει το μοτίβο (pattern) που αντιπροσωπεύει ένα στοιχείο στο συντακτικό δέντρο των XML εγγράφων. Το δεύτερο τμήμα των κανόνων, το *τμήμα ενέργειας (action part)*, καθορίζει την αντιστοίχιση των στοιχείων με αντικείμενα παρουσίασης. Πιο συγκεκριμένα, το τμήμα ενέργειας παράγει νέα στοιχεία και χαρακτηριστικά τα οποία απαρτίζουν το *δέντρο αποτελέσματος (result tree)*. Τέλος, το λογισμικό που υποστηρίζει τα XSL style sheets μετατρέπει το δέντρο αποτελέσματος και το περιεχόμενό του σε μορφή κατάλληλη για παρουσίαση (π.χ. HTML, RTF κλπ.).

Όπως η XML βρίσκεται στο ενδιάμεσο μεταξύ SGML και HTML, η XSL βρίσκεται μεταξύ CSSL και DSSSL: Η XSL βασίζεται τόσο στη DSSSL όσο και στη CSSL και στοχεύει στην υποστήριξη εγγράφων με περίπλοκη δομή.

3.2.1.6. Κωδικοποίηση XML Εγγράφων

Για την αποτελεσματική χρήση XML εγγράφων σε «στενής περιοχής» (narrowband) κανάλια επικοινωνίας έχει προταθεί από το WAP (Wireless Application Protocol) Forum η *WBXML (WAP Binary XML)* [WBXML]. Η κωδικοποίηση αυτή σχεδιάστηκε έτσι ώστε να επιτρέπει την κωδικοποιημένη μεταφορά XML εγγράφων χωρίς να μεταβάλλονται η λειτουργικότητα και η σημασιολογία τους.

Η WBXML έχει σχεδιαστεί με τέτοιο τρόπο, ώστε από τη μια να διατηρείται η δομή των στοιχείων των XML εγγράφων και από την άλλη, όταν ένα πρόγραμμα πλοήγησης συναντά άγνωστα στοιχεία ή χαρακτηριστικά, να μπορεί να τα αγνοήσει. Κατά την κωδικοποίηση ενός XML εγγράφου σε WBXML κωδικοποιούνται η δομή και το περιεχόμενο του εγγράφου, ενώ η σχετική με αυτό μεταπληροφορία (π.χ. DTDs) παραλείπεται.

3.2.1.7. Ανταλλαγή Τμημάτων XML Εγγράφων

Όπως αναφέρθηκε πριν, η XML υποστηρίζει δομημένα έγγραφα που αποτελούνται από στοιχεία και οντότητες. Σε κάποιες περιπτώσεις, μπορεί οι χρήστες να ενδιαφέρονται να δουν ή να επεξεργαστούν ένα ή περισσότερα από τα στοιχεία που απαρτίζουν ένα έγγραφο κι όχι ολόκληρο το έγγραφο. Τέτοιες περιπτώσεις είναι οι εξής [XML-Frag]:

- Ο χρήστης επιλέγει ένα τμήμα εγγράφου, το οποίο θέλει να διαχειριστεί ως ανεξάρτητη οντότητα (π.χ. να το στείλει σε κάποιον άλλο χρήστη, να το ενσωματώσει σε άλλο έγγραφο ή να το αποθηκεύσει ξεχωριστά).

- Ο χρήστης θέλει να διαχειριστεί ξεχωριστά συγκεκριμένα σύνολα στοιχείων και οντοτήτων.
- Σε πολύ μεγάλα έγγραφα, ο χρήστης θέλει να διαχειριστεί και να δει ξεχωριστά συγκεκριμένα τμήματά τους και να δημιουργήσει δομές πλοήγησης μέσα σ' αυτά.

Με την ανταλλαγή και τη διαχείριση τμημάτων XML εγγράφων στις παραπάνω περιπτώσεις (αλλά και σε άλλες που μπορεί να εντοπισθούν στο μέλλον) ασχολείται η ομάδα εργασίας του W3C "XML Fragment WG". Πιο συγκεκριμένα, η ομάδα αυτή ασχολείται με τον ορισμό ενός τρόπου ανταλλαγής τμημάτων XML εγγράφων – ανεξάρτητα από το αν τα τμήματα αυτά είναι ή όχι προκαθορισμένες οντότητες – χωρίς να πρέπει να σταλεί όλο το τμήμα του εγγράφου που προηγείται των τμημάτων αυτών. Τα τμήματα που θα στέλνονται μ' αυτό τον τρόπο θα μπορούν να παρουσιαστούν άμεσα ή να αποθηκευθούν για να χρησιμοποιηθούν αργότερα.

Ιδιαίτερη προσοχή απαιτείται όταν ένα τμήμα XML εγγράφου δεν περιέχει αρκετή πληροφορία ώστε να γίνει σωστά η συντακτική ανάλυσή του. Στην περίπτωση αυτή, στόχος είναι να δίνεται η δυνατότητα στους αποστολείς να δίνουν την επιπλέον πληροφορία που απαιτείται για να είναι δυνατή η σωστή επεξεργασία των τμημάτων των XML εγγράφων από τον παραλήπτη.

Για την επίτευξη των στόχων που αναφέρθηκαν πριν, η ομάδα εργασίας του W3C θα ορίσει:

- Ακριβείς περιορισμούς για το ποια δομικά στοιχεία ενός XML εγγράφου μπορούν να αποτελέσουν τμήματα.
- Το σύνολο της πληροφορίας που απαιτείται για την επιτυχή συντακτική ανάλυση, την παρουσίαση και την επεξεργασία ενός XML εγγράφου.
- Τη σύνταξη αυτής της πληροφορίας.
- Μηχανισμούς για τη διασύνδεση αυτής της πληροφορίας με τα τμήματα.

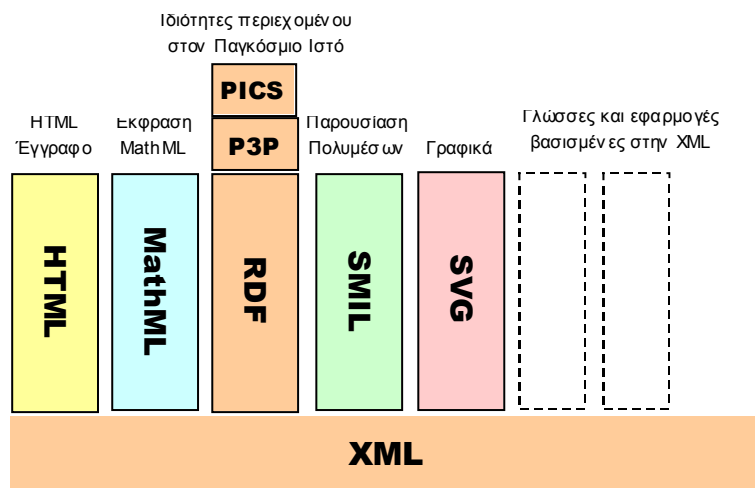
Η διαδικασία δημιουργίας του τμήματος έχει ως εξής: Ο κάτοχος του ολοκληρωμένου εγγράφου πηγής δημιουργεί τον *Καθορισμό «Συμφραζομένων» του Τμήματος (Fragment Context Specification)*. Το περιεχόμενο του τμήματος του εγγράφου αποτελεί το *Σώμα του Τμήματος (Fragment Body)*. Το αντικείμενο που περιέχει ολόκληρο το τμήμα του εγγράφου (τόσο το περιεχόμενο όσο και τα «συμφραζόμενα»), είναι ένα στιγμιότυπο της οντότητας που αναπαριστά το τμήμα (fragment entity).

Η σύνταξη που χρησιμοποιείται για την περιγραφή τμημάτων XML εγγράφων είναι XML: Στην πραγματικότητα, εκτός από το περιεχόμενο του τμήματος που είναι σε μορφή συμβατή με την XML, και η περιγραφή των «συμφραζομένων» του αναπαρίσταται ως ένα XML έγγραφο.

3.2.1.8. Εφαρμογές της XML

Όπως έχει αναφερθεί παραπάνω, η XML είναι μια μεταγλώσσα που επιτρέπει τον ορισμό γλωσσών για την περιγραφή των αντικειμένων που χρησιμοποιούνται σε συγκεκριμένες περιοχές εφαρμογών. Έτσι, οι εφαρμογές της XML είναι γλώσσες που η σύνταξή τους βασίζεται στην XML. Η σχέση της XML με τις XML εφαρμογές φαίνεται στο Σχήμα 3-1.

Στην υποενότητα αυτή θα γίνει μια προσπάθεια απαρίθμησης των σημαντικότερων εφαρμογών της XML. Η σημασία της ανάπτυξης εφαρμογών της XML είναι μεγάλη για την εξάπλωση της γλώσσας, καθώς η επικράτησή της σε πολλές περιοχές εφαρμογών θα προκαλέσει την ευρύτερη αποδοχή και επέκτασή της. Οι σημαντικότερες εφαρμογές της XML είναι οι εξής:



Σχήμα 3-1: XML και XML Εφαρμογές

- *CDF (Channel description Format)*: Μια XML εφαρμογή που προτάθηκε από τη Microsoft και επιτρέπει την έκδοση στον Παγκόσμιο Ιστό (Web publishing) με τη χρήση *Τεχνολογιών Προώθησης (Push Technologies)*.
- *CML (Chemical Markup Language)*: Μια εφαρμογή για την περιγραφή των ατομικών, μοριακών και κρυσταλλικών δομών, των φασμάτων ανάλυσης και άλλων αντικειμένων που χρησιμοποιούνται από τους χημικούς. Αρχικά η CML είχε χρησιμοποιήσει SGML σύνταξη, αλλά πρόσφατα προσάρμοσε τη σύνταξή της σε XML.
- *ICE (Information and Content Exchange)*: Πρωτόκολλο για το συντονισμό των δεδομένων που χρησιμοποιούνται στον Παγκόσμιο Ιστό.
- *MathML (Mathematical Markup Language)* [MathML]: Μια εφαρμογή της XML για την περιγραφή της δομής και του περιεχομένου μαθηματικών εκφράσεων. Μέσω της MathML γίνεται για πρώτη φορά δυνατή η σωστή παρουσίαση μαθηματικών εκφράσεων στον Παγκόσμιο Ιστό.
- *NVML (NaVigation Markup Language)* [NVML]: Εφαρμογή της XML για την περιγραφή πληροφορίας πλοήγησης (π.χ. τοποθεσίες, διαδρομές κλπ.), η οποία έχει σχεδιαστεί έτσι ώστε να καλύπτει τις ανάγκες πλοήγησης με τη χρήση διαφόρων κινητών συσκευών¹⁵ και με οποιοδήποτε μεταφορικό μέσο (με αυτοκίνητο, τρένο, λεωφορείο και με τα πόδια).

Η NVML παρέχει τη δυνατότητα περιγραφής διαφόρων τύπων πληροφορίας πλοήγησης, όπως:

- Διαδρομές από το τρέχον σημείο προς ένα σημείο προορισμού.

¹⁵ Κινητές συσκευές είναι τα έξυπνα τηλέφωνα (smart phones), οι PDAs (Personal Data Assistants) που διαθέτουν GPS (Global Positioning System), τα συστήματα πλοήγησης αυτοκινήτων κλπ.

- Διαδρομές για καταστήματα από τον κοντινότερο στο τρέχον σημείο σταθμό.
- Προτάσεις για τη μεταφορά από το τρέχον σημείο σε κάποιο σημείο προορισμού.
- Προτάσεις για αξιοθέατα.
- Προγράμματα εκδρομών.

Η NVML κάνει δυνατή τη χρήση υπηρεσιών πλοήγησης οπουδήποτε και οποτεδήποτε, και μπορεί να χρησιμοποιηθεί για την υποστήριξη υπηρεσιών όπως οι οδηγοί αξιοθεάτων, ο σχεδιασμός ταξιδιών και οι τουριστικές εκδόσεις.

- *OFX (Open Financial Exchange)*: Πλαίσιο ανταλλαγής οικονομικών δεδομένων και εντολών μεταξύ οικονομικών οργανισμών και των πελατών τους.
- *OSD (Open Software Description format)*: Μια XML εφαρμογή για την περιγραφή πακέτων λογισμικού (software packages), των εξαρτήσεων τους από άλλα πακέτα, από λειτουργικά συστήματα κ.α. και για την αντιμετώπιση πιθανών προβλημάτων που αντιμετωπίζουν πακέτα που χρησιμοποιούνται σε πολλές πλατφόρμες.
- *PGML (Precision Graphics Markup Language)*: Εφαρμογή της XML για την περιγραφή διανυσματικών γραφικών.
- *PIDL (Personalized Information Description Language) [PIDL]*: Εφαρμογή της XML που στοχεύει στην υποστήριξη της εξατομίκευσης της πληροφορίας μέσω της παροχής αυξημένης διαλειτουργικότητας μεταξύ εφαρμογών εξατομίκευσης. Η PIDL παρέχει ένα κοινό πλαίσιο για τη σταδιακή επεξεργασία της πληροφορίας από τις εφαρμογές εξατομίκευσης και την επικόλληση στο τέλος, συμπτυγμένων, των εξατομικευμένων μορφών της. Η PIDL υποστηρίζει την εξατομίκευση διαφορετικών μέσων (π.χ. δομημένου και μη κειμένου, γραφικών κλπ.) πολλαπλές μεθόδους εξατομίκευσης (π.χ. φιλτράρισμα, διάταξη, αντικατάσταση κλπ.) και πολλαπλές μεθόδους διανομής (π.χ. μέσω πρωτοκόλλου SMTP, HTTP κλπ.).
- *PML (Procedural Markup Language) [PML]*: Εφαρμογή της XML για τη δημιουργία παρουσιάσεων με τη χρήση πολυμέσων, η οποία επιτρέπει την αποσύνδεση περιεχομένου και παρουσίασης.
- *POIX (Point Of Interest eXchange) [POIX]*: Εφαρμογή της XML που επιτρέπει την περιγραφή πληροφορίας σχετικής με συγκεκριμένες τοποθεσίες (π.χ. πληροφορία που βρίσκεται σε κάποιο χάρτη. Η πληροφορία αυτή χρησιμοποιείται από διάφορους τύπους κινητών συσκευών.

Η POIX δεν παρέχει απλώς την αναπαράσταση των συντεταγμένων των τοποθεσιών, αλλά δίνει τη δυνατότητα αναπαράστασης διάφορων μορφών πληροφορίας που συσχετίζεται με την κάθε τοποθεσία.

Η POIX υποστηρίζει την ανταλλαγή πληροφορίας σχετικής με τοποθεσίες τόσο μέσω ηλεκτρονικού ταχυδρομείου (e-mail) όσο και μέσω της ενσωμάτωσης της πληροφορίας αυτής σε XML και HTML έγγραφα.

- *RDF (Resource Description Framework)*: Πλαίσιο υποστήριξης μεταδεδομένων που επιτρέπει την περιγραφή και τη βαθμολόγηση του περιεχομένου των αντικειμένων στο Διαδίκτυο. Η RDF εξετάζεται αναλυτικότερα στην υποενότητα 3.3.1.3.

- *SMIL (Synchronized Multimedia Integration Language)* [SMIL]: Εφαρμογή που επιτρέπει τη δημιουργία αλληλεπιδραστικών παρουσιάσεων πολυμέσων μέσα από το συνδυασμό, το συγχρονισμό και τη διασύνδεση αντικειμένων πολυμέσων (κειμένου, εικόνων, ήχου, video και animation). Επιπλέον, η SMIL παρέχει δυνατότητες συσχετισμού υπερσυνδέσμων με αντικείμενα πολυμέσων, περιγραφής της χρονικής συμπεριφοράς μιας παρουσίασης και περιγραφής της μορφοποίησης της παρουσίασης στην οθόνη.

Η SMIL ορίζεται ως ένα σύνολο τμημάτων (modules), στα οποία ορίζονται μια XML σύνταξη και η σημασιολογία για καθεμιά από τις συγκεκριμένες περιοχές λειτουργικότητας της SMIL. Καθένα από τα τμήματα της SMIL συνδέεται με ένα DOM (Document Object Model – βλέπε ενότητα 3.4). Η βασική ιδέα της SMIL είναι:

1. Η δημιουργία συστατικών των παρουσιάσεων πολυμέσων (σε μορφή κειμένου, εικόνων, video και ήχου) τα οποία διαθέτουν μοναδικά αναγνωριστικά που βασίζονται στη διεύθυνσή τους.
 2. Ο σχεδιασμός της παρουσίασης που περιέχει τα συστατικά, έτσι ώστε αυτά να παρουσιάζονται σειριακά, παράλληλα ή και συγχρονισμένα.
 3. Η δυνατότητα επαναχρησιμοποίησης της σύνταξης και της σημασιολογίας της SMIL από άλλες γλώσσες που βασίζονται στην XML, ιδιαίτερα από αυτές που χρειάζεται να αναπαραστήσουν χρονικές σχέσεις και συγχρονισμό. Έτσι, π.χ., τμήματα της SMIL θα μπορούσαν να ενσωματωθούν στην XHTML, ώστε να παρέχεται η δυνατότητα συγχρονισμού των αντικειμένων που απαρτίζουν τα XHTML έγγραφα.
- *SVG (Scalable Vector Graphics)*: Εφαρμογή της XML, που χρησιμοποιείται τόσο για την περιγραφή διδιάστατων διανυσματικών γραφικών όσο και για την περιγραφή ανάμικτων διανυσματικών και raster γραφικών [SVG99]. Η SVG μπορεί να χρησιμοποιηθεί με δυο τρόπους:
 1. Για τη δημιουργία ανεξάρτητων SVG αρχείων γραφικών.
 2. Για τη δημιουργία SVG τμημάτων (SVG fragments), τα οποία αποτελούν τμήματα (συνήθως ενσωματωμένα) XML εγγράφων.

Τόσο τα ανεξάρτητα SVG αρχεία γραφικών όσο και τα SVG τμήματα μπορεί να περιέχουν τρία είδη γραφικών αντικειμένων: διανυσματικά γραφικά, εικόνες και κείμενο. Τα γραφικά αντικείμενα που συνθέτουν τα SVG γραφικά έχουν τη δυνατότητα να ομαδοποιηθούν, να μετασχηματιστούν και να εφαρμοστούν πάνω τους στυλ. Επιπλέον, για τα SVG γραφικά αντικείμενα υποστηρίζεται ένα πλούσιο σύνολο γεγονότων (π.χ. onmouseover, onclick κλπ).

Τα SVG γραφικά έχουν τη δυνατότητα να είναι αλληλεπιδραστικά (interactive) και δυναμικά. Οι animations που περιέχονται στα SVG γραφικά είναι δυνατό να ορίζονται και να ενεργοποιούνται είτε δηλωτικά (π.χ. με την ενσωμάτωση στοιχείων SVG animation σ' ένα SVG έγγραφο) είτε με τη χρήση scripting.

Οι προδιαγραφές της SVG, εκτός από την πλήρη υποστήριξη διανυσματικών γραφικών, περιλαμβάνουν τόσο την υποστήριξη πλούσιας μορφοποίησης όσο και την υποστήριξη επιπέδων (layering) και μασκών (masking) για τη δημιουργία περίπλοκων

εφέ. Το σύνολο χαρακτηριστικών ενός γραφικού αντικειμένου περιλαμβάνει αλληπαλλήλους μετασχηματισμούς (nested transformations), μονοπάτια αποκοπής (clipping paths), μάσκες, φίλτρα και αντικείμενα προτύπων (template objects).

Οι πιο περίπλοκες εφαρμογές της SVG χρησιμοποιούν επιπρόσθετα γλώσσες scripting και αξιοποιούν την πλήρη πρόσβαση που παρέχει το DOM (Document Object Model – βλέπε ενότητα 3.4) σε όλα τα στοιχεία, τα χαρακτηριστικά και τις ιδιότητες των SVG εγγράφων.

- *VML (Vector Markup Language)*: Μια εφαρμογή της XML για την περιγραφή διανυσματικών γραφικών.
- *UCLP (Universal Commerce Language and Protocol)*: Ένα συμβατό με την XML σχήμα που περιέχει εντολές για την περιγραφή μεταδεδομένων. Η UCLP [UCLP] μπορεί να χρησιμοποιηθεί από εμπορικές εφαρμογές για την αναγνώριση και την ανάκτηση δεδομένων στο Διαδίκτυο. Η UCLP επίσης παρέχει ένα επίπεδο υποστήριξης τύπων δεδομένων (data typing), που επιτρέπει τον ορισμό ονομάτων ειδικά για εμπορικές εφαρμογές και υποστηρίζει με τον τρόπο αυτό τη δυνατότητα εξειδίκευσης των επιλογών των χρηστών.

Η UCLP φιλοδοξεί να εισάγει ένα νέο παράδειγμα, τη δυναμική δημιουργία εντολών για την περιγραφή δεδομένων (dynamic data tagging), για το οποίο η υποστήριξη τύπων δεδομένων είναι απαραίτητη.

- *WDDX (Web Distributed Data eXchange)*: Εφαρμογή που επιτρέπει την ανταλλαγή δομών δεδομένων μεταξύ εφαρμογών πάνω από το HTTP (HyperText Transfer Protocol) χωρίς να βασίζεται σε RPCs (Remote Procedure Calls) [WDDX]. Η WDDX χρησιμοποιεί XML σύνταξη για την περιγραφή κατανεμημένων αντικειμένων (distributed objects), και παρέχει μια μέθοδο δομικής περιγραφής των αντικειμένων ανεξάρτητα από γλώσσες προγραμματισμού.
- *WML (Wireless Markup Language)*: Γλώσσα [WAP98b] που βασίζεται στην XML και προορίζεται για τον καθορισμό του περιεχομένου και του επιπέδου διεπαφής των ασύρματων συσκευών με τους χρήστες. Η WML έχει σχεδιαστεί λαμβάνοντας υπ' όψιν τους περιορισμούς που θέτουν οι ασύρματες συσκευές (μικρές οθόνες, περιορισμένες δυνατότητες εισαγωγής δεδομένων από τους χρήστες, χαμηλής ταχύτητας σύνδεση με το δίκτυο, περιορισμένη μνήμη και μικρή υπολογιστική ισχύς).

Η WML επικεντρώνει σε τέσσερις περιοχές:

1. Στην αναπαράσταση και τη μορφοποίηση δεδομένων σε μορφή κειμένου.
2. Στην υποστήριξη της μεταφοράς της κάρτας (card) και της στοίβας (deck): Μια κάρτα αναπαριστά ένα στοιχείο αλληλεπίδρασης με το χρήστη, ενώ οι κάρτες ομαδοποιούνται σε στοίβες. Επιπλέον, η στοίβα αποτελεί τη μονάδα μεταφοράς περιεχομένου σε περιβάλλον φορητών συσκευών.
3. Στην πλοήγηση μεταξύ καρτών (inter-card navigation) και τη διασύνδεση καρτών (card linking).
4. Στην παραμετροποίηση συμβολοσειρών (string parameterization) και τη διαχείριση των καταστάσεων τους (state management).

- *XFA (XML Forms Architecture)*: Πλαίσιο για την υποστήριξη ηλεκτρονικών φορμών και των εφαρμογών που τις χρησιμοποιούν. Η XFA έχει εντοπίσει τις ανάγκες των οργανισμών για ασφαλή απόκτηση, παρουσίαση, μετακίνηση, έξοδο και εκτύπωση της πληροφορίας που συνδέεται με τις ηλεκτρονικές φόρμες και προσπαθεί να καλύψει τις ανάγκες αυτές με ένα σύνολο προσδιορισμών. Οι σημαντικότεροι από τους προσδιορισμούς αυτούς είναι οι εξής:

- Η *XFA-Template* [XFA-Templ], που είναι μια XML εφαρμογή για τη μοντελοποίηση προτύπων ηλεκτρονικών φορμών. Πιο συγκεκριμένα, η XFA-Template υποστηρίζει την ανοιχτή και επεκτάσιμη δημιουργία ασφαλών φορμών με υψηλής πιστότητας σύνθεση, αυτοματοποιημένους υπολογισμούς και ελέγχους και ευέλικτη διαχείριση δεδομένων. Επιπλέον, στις φόρμες αυτές είναι δυνατή η διασύνδεση στοιχείων διεπαφής (interface components) με τους χρήστες.
- Η *XFA-FormCalc* [XFA-FC], που είναι μια απλή γλώσσα scripting, βελτιστοποιημένη για τον ορισμό λογικής και υπολογισμών που χρησιμοποιούνται σε ηλεκτρονικές φόρμες.

Η XFA-FormCalc είναι γλώσσα που δεν υποστηρίζει τύπους, βασίζεται σε εκφράσεις και διαθέτει περιορισμένες δυνατότητες ελέγχου ροής (control flow). Διαθέτει ένα ευρύ σύνολο αριθμητικών, επιστημονικών και οικονομικών συναρτήσεων, ώστε να καλύπτει τις ανάγκες υποστήριξης ηλεκτρονικών φορμών για εφαρμογές που μπορεί να διαφέρουν τόσο ως προς το πεδίο όσο και ως προς την πολυπλοκότητα.

- *XHTML (eXtensible HyperText Markup Language)*: Μια εφαρμογή της XML, που παρέχει μια συμβατή με την XML σύνταξη της HTML 4.0.

Οι διαφορές που υπάρχουν μεταξύ ενός HTML και ενός XHTML εγγράφου με το ίδιο περιεχόμενο είναι οι εξής:

1. Τα XHTML έγγραφα πρέπει να είναι συντακτικά σωστά.
2. Τα ονόματα στοιχείων και χαρακτηριστικών πρέπει να είναι γραμμένα με πεζούς χαρακτήρες.
3. Όλα τα μη κενά στοιχεία διαθέτουν και εντολές τέλους. Έτσι, π.χ., ενώ στην HTML δεν είναι απαραίτητο να υπάρχει η εντολή τέλους `</p>` για τις παραγράφους, είναι απαραίτητη η ύπαρξή της στην XHTML.
4. Στην XHTML οι τιμές των χαρακτηριστικών των στοιχείων πρέπει πάντα να περικλείονται από αγγλικά εισαγωγικά (").
5. Δεν επιτρέπεται στην XHTML η «συμπυκνόμενη» σύνταξη για χαρακτηριστικά που μπορούν να έχουν μόνο μια τιμή.
6. Τα κενά στοιχεία πρέπει να τελειώνουν με `/>`. Έτσι, π.χ., το στοιχείο `<br>` της HTML συντάσσεται `<br/>` στην XHTML.

Η XHTML παρέχει τη βάση για μια οικογένεια DTDs που θα αποτελέσουν υπο- και υπερ- σύνολα της XHTML, μέσω του ορισμού τμημάτων και του καθορισμού ενός μηχανισμού για το συνδυασμό αυτών των τμημάτων. Πιο συγκεκριμένα, ορίζονται

σημασιολογικά τμήματα (*semantic modules*), τα οποία επιτρέπουν τόσο την επέκταση της XHTML όσο και τη δημιουργία υποσυνόλων της XHTML για την κάλυψη των αναγκών συγκεκριμένων περιοχών εφαρμογών. Έτσι, με την τμηματοποίηση της XHTML σε σημασιολογικά τμήματα και την υλοποίηση των τελευταίων από DTDs εξασφαλίζεται η επεκτασιμότητα της XHTML [ModXHTML].

Επιπλέον, μέσω της XHTML ορίζονται *προφίλ εγγράφων (document profiles)*, τα οποία είναι συλλογές από τις απαιτήσεις που υπάρχουν για την επεξεργασία και την παρουσίαση διαφορετικών οικογενειών εγγράφων.

- **XML/EDI:** Μια προσπάθεια για την παροχή μιας πρότυπης μορφής (format) για την περιγραφή εμπορικών δεδομένων, που θα επιτρέψει την αυτοματοποίηση των εμπορικών συναλλαγών μέσω της *EDI (Electronic Data Interchange)*.
- **XML-RPC [XML-RPC]:** Μια εφαρμογή της XML, που αποτελεί ένα πρωτόκολλο απομακρυσμένων κλήσεων συναρτήσεων (remote procedure calls – RPCs) που τρέχει πάνω από το πρωτόκολλο HTTP και χρησιμοποιεί XML σύνταξη τόσο για τα RPC μηνύματα – κλήσεις όσο και για τα RPC μηνύματα – αποτελέσματα.
- **XMTP (eXtensible Mail Transport Protocol):** Μια εφαρμογή της XML για την αναπαράσταση μηνυμάτων ηλεκτρονικού ταχυδρομείου, που επιτρέπει τη διαχείριση των μηνυμάτων από συστήματα που υποστηρίζουν την XML[XMTP98]. Η XMTP βασίζεται στην αντιστοίχιση της MIME/SMTP με την XML.

3.3. Πρότυπα Αναπαράστασης Μεταδεδομένων

Η χρήση των *Μεταδεδομένων (Metadata)* για τη δεικτοδότηση ψηφιοποιημένων και μη πόρων (resources), έτσι ώστε να υποστηριχθεί η αποτελεσματικότερη αναζήτηση πόρων σε περιβάλλοντα δικτύου, υλοποιείται σταδιακά σ' όλο τον κόσμο. Αυτό συμβαίνει γιατί η χρήση των μεταδεδομένων εγγυάται μεγαλύτερη ακρίβεια στην ανάκτηση της πληροφορίας που ενδιαφέρει το χρήστη απ' ότι η χρήση των παραδοσιακών *Μηχανών Αναζήτησης (Search Engines)*. Επιπλέον, τα μεταδεδομένα μπορούν να χρησιμοποιηθούν και για το «φιλτράρισμα» των αποτελεσμάτων που επιστρέφονται σε μια ερώτηση από μια μηχανή αναζήτησης. Τα μεταδεδομένα αποτελούν την ελπίδα επιβολής τάξης στο χάος του Παγκοσμίου Ιστού, καθώς η κατάσταση στον Παγκόσμιο Ιστό αρχίζει να παίρνει τις διαστάσεις προβλήματος.

Πριν προχωρήσουμε όμως στην καταγραφή των αναγκών που καλύπτουν (ή θα καλύψουν) τα μεταδεδομένα, των προβλημάτων που καλούνται να επιλύσουν και των υπαρχόντων και υπό εξέλιξη προτύπων για την αναπαράσταση και διαχείρισή τους, θα δώσουμε έναν ορισμό των μεταδεδομένων:

Τα *Μεταδεδομένα (Metadata)* είναι δομημένα δεδομένα που περιγράφουν άλλα δεδομένα και αντικείμενα.

Τα μεταδεδομένα που περιγράφουν πόρους μπορούν να καλύπτουν διάφορες πτυχές του κάθε πόρου και διάφορες οπτικές γωνίες. Έτσι τα μεταδεδομένα μπορούν να καταταγούν στις εξής 7 κατηγορίες:

1. Μεταδεδομένα που εξυπηρετούν την αναγνώριση και την ανακάλυψη πόρων (resource identification and discovery).
2. Μεταδεδομένα που περιγράφουν τις συνθήκες πρόσβασης (access conditions) και τις απαιτήσεις χρήσης (usage requirements).
3. Μεταδεδομένα που περιγράφουν τη δομή (structure) των πόρων.
4. Μεταδεδομένα που περιγράφουν τα «συμφραζόμενα» (context) των πόρων.
5. Μεταδεδομένα που περιγράφουν το περιεχόμενο (content) των πόρων.
6. Μεταδεδομένα που περιγράφουν την ως τώρα χρήση (use history) των πόρων.
7. Μεταδεδομένα που έχουν να κάνουν με τη διαχείριση των πόρων (administrative metadata).

Υπεύθυνες για την περιγραφή των μεταδεδομένων που καλύπτουν τις απαιτήσεις των διαφόρων κοινοτήτων χρηστών είναι οι *Κοινότητες Περιγραφής Πόρων (Resource Description Communities)*. Μια κοινότητα περιγραφής πόρων χαρακτηρίζεται από κοινή αντίληψη σε ότι αφορά τη σημασιολογία, τη δομή και τη σύνταξη για την ανταλλαγή πληροφορίας περιγραφής πόρων.

Όπως αναφέρθηκε και παραπάνω, τα μεταδεδομένα χρησιμοποιούνται για την περιγραφή ψηφιοποιημένων και μη πόρων που βρίσκονται στο κατανεμημένο περιβάλλον του Διαδικτύου. Για να είναι αποτελεσματική η χρήση των μεταδεδομένων είναι απαραίτητο η αναπαράσταση και η διαχείρισή τους να βασίζονται σε κοινά αποδεκτά πρότυπα. Για το λόγο αυτό έχουν συσταθεί σ' όλο τον κόσμο ομάδες εργασίας που ασχολούνται με την αναπαράσταση και τη διαχείριση μεταδεδομένων.

Στο τμήμα αυτό θα γίνει μια σύντομη παρουσίαση των αποτελεσμάτων που έχουν βγάλει αυτές οι ομάδες εργασίας, καθώς σ' αυτά έχει βασιστεί ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Πιο συγκεκριμένα, στη συνέχεια περιγράφονται τόσο το πλαίσιο Warwick, το οποίο έθεσε τις βάσεις για τον ορισμό και τη διαχείριση λεξικών μεταδεδομένων από ομάδες χρηστών όσο και τα σημαντικότερα από τα λεξικά μεταδεδομένων.

3.3.1. Το Πλαίσιο Warwick

Όπως αναφέρθηκε παραπάνω, η κατάσταση στον Παγκόσμιο Ιστό είναι εν πολλοίς προβληματική και τα μεταδεδομένα αποτελούν μια από τις βασικότερες ελπίδες για την επιβολή τάξης στο χάος που τείνει να επικρατήσει. Για να επιλυθούν όμως τα υπάρχοντα προβλήματα χωρίς να δημιουργηθούν νέα, πρέπει να αντιμετωπιστούν τα εξής ζητήματα:

- Η υποστήριξη πολλαπλών διαστάσεων και μορφών μεταδεδομένων.
- Η διαχείριση της εξέλιξης των μορφών μεταδεδομένων (δημιουργία νέων μορφών μεταδεδομένων και αλλαγών στις υπάρχουσες).
- Η ανάγκη συγκερασμού των προσεγγίσεων που βασίζονται στην υποστήριξη αυξημένης λειτουργικότητας με προσεγγίσεις που βασίζονται στην απλότητα.
- Η ανάγκη συγκερασμού της διαλειτουργικότητας με την επεκτασιμότητα.

- Το γεγονός ότι τα μεταδεδομένα θα χρησιμοποιούνται τόσο από ανθρώπους όσο και από υπολογιστές.
- Η αποφυγή της επ' άπειρον αναδρομής (μετα-...-μετα-...-μεταδεδομένα).
- Η ανάγκη υποστήριξης της πρόσβασης, δημιουργίας και διαχείρισης «λεξικών» μεταδεδομένων από συγκεκριμένες κοινότητες χρηστών, καθώς κανένα σύνολο στοιχείων μεταδεδομένων δεν επαρκεί για την ταυτόχρονη κάλυψη των απαιτήσεων όλων των κοινοτήτων χρηστών. Αυτό συμβαίνει γιατί οι διαφορετικές κοινότητες χρηστών έχουν διαφορετικές απαιτήσεις σε ότι αφορά τόσο τη σημασιολογία των μεταδεδομένων όσο και το επίπεδο πολυπλοκότητάς τους.

Για να λυθούν τα προβλήματα αυτά πρέπει να καλυφθούν οι απαιτήσεις που θέτουν τα μοντέλα μεταδεδομένων που προτείνουν οι διαφορετικές κοινότητες χρηστών. Για να γίνει αυτό δυνατό και να μπορούν να συνδυαστούν, να αλληλοσυμπληρωθούν και να ανταλλάσσονται τα μεταδεδομένα που υπακούουν στα μοντέλα αυτά, απαιτείται μια αρχιτεκτονική που θα μπορεί να υποστηρίξει μια μεγάλη ποικιλία μοντέλων μεταδεδομένων. Μια τέτοια αρχιτεκτονική προτείνεται από το *Πλαίσιο Warwick (Warwick Framework)* [Lag96] [Dem96], το οποίο περιγράφεται στη συνέχεια αυτής της ενότητας.

Το πλαίσιο Warwick είναι μια αρχιτεκτονική που επιτρέπει τη συγκέντρωση πολλαπλών συνόλων μεταδεδομένων και υποστηρίζει την ανταλλαγή μεταδεδομένων μεταξύ των συνόλων αυτών. Το πλαίσιο Warwick έχει δύο βασικά συστατικά:

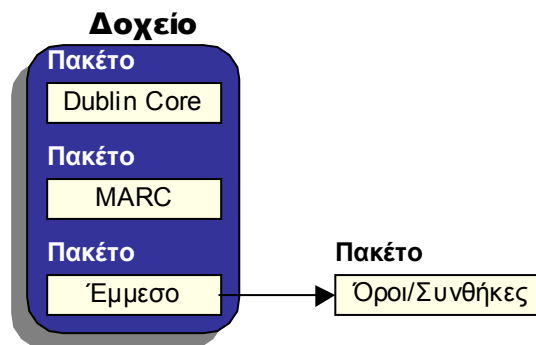
- Τα *Πακέτα (Packages)*, που αναπαριστούν τυποποιημένα (typed) σύνολα μεταδεδομένων, καθένα από τα οποία καλύπτει τις ανάγκες μιας ή περισσότερων συγκεκριμένων περιοχών.
- Τα *Δοχεία (Containers)*, που είναι οι μονάδες συγκέντρωσης πακέτων στο πλαίσιο Warwick. Ένα δοχείο μπορεί να είναι μόνιμο ή προσωρινό. Τα προσωρινά δοχεία είναι διαβιβαστικά αντικείμενα μεταξύ πηγών πληροφορίας, πελατών και διαμεσολαβητών. Τα μόνιμα δοχεία είναι πρώτης τάξης αντικείμενα στο δίκτυο πληροφορίας του οποίου κόμβοι είναι τα πακέτα μεταδεδομένων και τα περιγραφόμενα από τα μεταδεδομένα αντικείμενα. Αυτό σημαίνει ότι κάθε δοχείο αποθηκεύεται σε έναν ή περισσότερους εξυπηρετητές και μπορεί να προσπελαστεί απ' αυτούς με τη χρήση ενός αναγνωριστικού στο οποίο υπάρχει πρόσβαση από οποιοδήποτε σημείο. Η μόνη λειτουργία που ορίζεται για ένα δοχείο είναι η επιστροφή των πακέτων που αυτό περιέχει. Στο σημείο αυτό τονίζεται ότι ένα δοχείο μπορεί να αποτελεί τμήμα άλλου αντικειμένου (το οποίο, π.χ. περιέχει τόσο τα δεδομένα ενός αντικειμένου όσο και τα μεταδεδομένα που περιγράφουν το αντικείμενο αυτό).

Όπως αναφέρθηκε και παραπάνω, τα πακέτα είναι τυποποιημένα. Ένα πακέτο μπορεί να ανήκει σε κάποιον από τους εξής τύπους:

1. *Σύνολο Μεταδεδομένων (Metadata Set)*: Τα πρωτογενή πακέτα-σύνολα μεταδεδομένων, που περιέχουν πραγματικά μεταδεδομένα που ορίζονται σε κάποιο λεξικό (π.χ. Dublin Core, MARC κ.α.).

2. *Έμμεσο (Indirect)*: Τα έμμεσα πακέτα αποτελούν αναφορές σε άλλα αντικείμενα-κόμβους του δικτύου πληροφορίας τμήμα του οποίου αποτελούν τα περιγραφόμενα αντικείμενα και τα πακέτα που τα περιγράφουν.
3. *Δοχείο (Container)*: Τα πακέτα αυτά είναι τα ίδια και δοχεία. Δεν υπάρχει όριο στην αναδρομή αυτού του είδους (δεν ορίζεται δηλαδή μέχρι ποιο «βάθος» μπορεί να περιέχονται δοχεία μέσα σε άλλα δοχεία).

Στο Σχήμα 3-2 φαίνεται ένα απλό παράδειγμα δοχείου του πλαισίου Warwick. Το δοχείο του παραδείγματος περιέχει τρία πακέτα μεταδεδομένων. Τα δυο πρώτα, μια Dublin Core¹⁶ εγγραφή και μια MARC¹⁶ εγγραφή, περιέχονται μέσα στο δοχείο, ενώ το τρίτο, που καθορίζει τους όρους και τις συνθήκες πρόσβασης στο περιεχόμενο του περιγραφόμενου αντικειμένου, αναφέρεται από το δοχείο (είναι έμμεσο πακέτο).



Σχήμα 3-2: Δοχείο του πλαισίου Warwick

Μέσα από τα δοχεία και τα πακέτα, το πλαίσιο Warwick υποστηρίζει τον καθορισμό, τη συλλογή, την κωδικοποίηση και την ανταλλαγή διαφορετικών *Λεξικών Μεταδεδομένων (Metadata Vocabularies)*, στιγμιότυπα των οποίων αποτελούν τα πακέτα-σύνολα μεταδεδομένων. Η αρχιτεκτονική του πλαισίου Warwick έχει τα εξής χαρακτηριστικά:

- Συναρμολογισιμότητα, καθώς επιτρέπει τη χρήση διαφορετικών τύπων συνόλων μεταδεδομένων.
- Επεκτασιμότητα, καθώς επιτρέπει την υποστήριξη νέων τύπων συνόλων μεταδεδομένων.
- Αναδρομή, καθώς επιτρέπει την αντιμετώπιση των αντικειμένων μεταδεδομένων ως αντικειμένων πληροφορίας, τα οποία συνδέονται με αντικείμενα μεταδεδομένων που τα περιγράφουν.
- Είναι κατανεμημένη, καθώς επιτρέπει την αναφορά σε εξωτερικά αντικείμενα μεταδεδομένων.

Τα πλεονεκτήματα της προσέγγισης του πλαισίου Warwick είναι τα εξής:

- Παρέχεται ένα πλαίσιο που επιτρέπει τη συγκέντρωση και τη συνεπή ανταλλαγή αντικειμένων μεταδεδομένων.

¹⁶ Για τα Dublin Core και MARC βλέπε την ενότητα με τίτλο *Λεξικά Μεταδεδομένων*.

- Αποφεύγεται η «επανεφεύρεση του τροχού», καθώς τα υπάρχοντα πακέτα μεταδεδομένων μπορούν να εξειδικευτούν (με την προσθήκη νέων στοιχείων), έτσι ώστε να υποστηρίζουν νέες κοινότητες χρηστών.
- Η συγκέντρωση αντικειμένων μεταδεδομένων μπορεί να βελτιστοποιηθεί, ώστε να καλύπτει κάθε φορά τις ανάγκες μιας συγκεκριμένης κοινότητας χρηστών.
- Η αρχιτεκτονική είναι επεκτάσιμη και έχει τη δυνατότητα της κάλυψης απαιτήσεων που δεν ήταν γνωστές κατά το σχεδιασμό της.
- Οι σχεδιαστές των λεξικών μεταδεδομένων μπορούν να επικεντρώσουν στις συγκεκριμένες απαιτήσεις της κάθε κοινότητας χρηστών χωρίς να νοιάζονται για την κάλυψη γενικότερων αναγκών από το λεξικό που κάθε φορά σχεδιάζουν.
- Η σύνταξη των λεξικών μπορεί να διαφέρει ανάλογα με τις σημασιολογικές απαιτήσεις και τις απαιτήσεις λειτουργικότητας από το κάθε λεξικό καθώς και τις πρακτικές της κάθε κοινότητας.
- Η κάθε κοινότητα είναι υπεύθυνη για τη δημιουργία και τη διαχείριση καθενός από τα λεξικά.
- Η διαλειτουργικότητα και η επεκτασιμότητα ενισχύονται καθώς τα εργαλεία διαχείρισης μεταδεδομένων μπορούν να επεξεργάζονται συγκεκριμένα λεξικά και να αγνοούν άλλα.
- Διαφορετικά σύνολα μεταδεδομένων που περιγράφουν το ίδιο αντικείμενο μπορούν να ελέγχονται ξεχωριστά.
- Μελλοντικά λεξικά μεταδεδομένων μπορούν να υποστηριχθούν χωρίς να απαιτούνται αλλαγές στα υπάρχοντα λεξικά μεταδεδομένων και τα εργαλεία διαχείρισής τους.

Έτσι, το πλαίσιο Warwick μπορεί να οριστεί ως μια υψηλού επιπέδου αρχιτεκτονική που βασίζεται στην έννοια του δοχείου, δεν κάνει υποθέσεις για το περιεχόμενο των αντικειμένων μεταδεδομένων και δεν προϋποθέτει ότι οι πελάτες όλων των συστημάτων που υποστηρίζει θα μπορούν να κατανοήσουν το περιεχόμενο όλων των τύπων πακέτων μεταδεδομένων. Αυτό σημαίνει ότι πρέπει να υπάρχει αυστηρή τυποποίηση των πακέτων μεταδεδομένων και ότι απαιτείται ένας μηχανισμός πιστοποίησης των πακέτων μεταδεδομένων.

Στην παρούσα φάση αναπτύσσονται τρεις υλοποιήσεις του πλαισίου Warwick: η *DOA* (*Digital Object Architecture*), το *MOF/XMI* (*Meta-Object Facility/XML-based Metadata Interchange*) και η *RDF* (*Resource Description Framework*). Αναφορά στις υλοποιήσεις αυτές γίνεται στις υποενότητες που ακολουθούν.

3.3.1.1. Digital Object Architecture (DOA)

Το θέμα αυτής της υποενότητας είναι μια από τις προσπάθειες υλοποίησης του πλαισίου Warwick, η *DOA* (*Digital Object Architecture*) [DOA].

Ο στόχος του προγράμματος DOA, που είναι αποτέλεσμα της συνεργασίας του CNRI με το πανεπιστήμιο Cornell είναι η δημιουργία μιας αρχιτεκτονικής ψηφιακών αντικειμένων, η οποία να μπορεί να χρησιμοποιηθεί σε πληροφοριακά συστήματα στο Διαδίκτυο. Τα βασικά συστατικά της αρχιτεκτονικής αυτής είναι:

- Τα *Ψηφιακά Αντικείμενα (Digital Objects)* που αποτελούνται τόσο από δεδομένα όσο και από μεταδεδομένα.
- Τα *Συστήματα Αποθήκευσης (Repositories)* των ψηφιακών αντικειμένων, τα οποία παρέχουν αποθήκευση των ψηφιακών αντικειμένων και πρόσβαση στα ψηφιακά αντικείμενα μέσω του Διαδικτύου. Η πρόσβαση ελέγχεται από ένα απλό πρωτόκολλο και εξαρτάται από τους κανόνες που καθορίζει ο διαχειριστής του συστήματος
- Το *Σύστημα Εντοπισμού (Handle System)* των ψηφιακών αντικειμένων, το οποίο είναι υπεύθυνο για την παροχή μονίμων ονομάτων στους πόρους. Πρόκειται για ένα καταμετρημένο σύστημα υψηλής πιστότητας και απόδοσης.
- Ένα *Σύστημα Πιστοποίησης (Registry System)* για την επί μακρόν παραχώρηση ονόματος στα αντικείμενα. Το σύστημα αυτό είναι υπεύθυνο για την πιστοποίηση των αντικειμένων και την παροχή πρόσβασης σε αυτά μόνο στους χρήστες που έχουν αυτό το δικαίωμα.

3.3.1.2. Meta-Object Facility/XML-based Metadata Interchange (MOF/XMI)

Το *MOF (Meta-Object Facility)* [MOF] [Cra97] [Cra98], που μαζί με την *XMI (XML-based Metadata Interchange)* [XMI] θα συζητηθούν σ' αυτή την υποενότητα, είναι μια υλοποίηση του πλαισίου Warwick για τη διαχείριση και αναπαράσταση μεταδεδομένων σε ένα CORBA [CORBA] περιβάλλον. Πιο συγκεκριμένα, το MOF είναι ένα θεμελιώδες πλαίσιο για την περιγραφή μεταπληροφορίας και μετα-μεταπληροφορίας με βάση τα μετα-μετα-αντικείμενα. Οι βασικοί στόχοι του MOF είναι οι εξής:

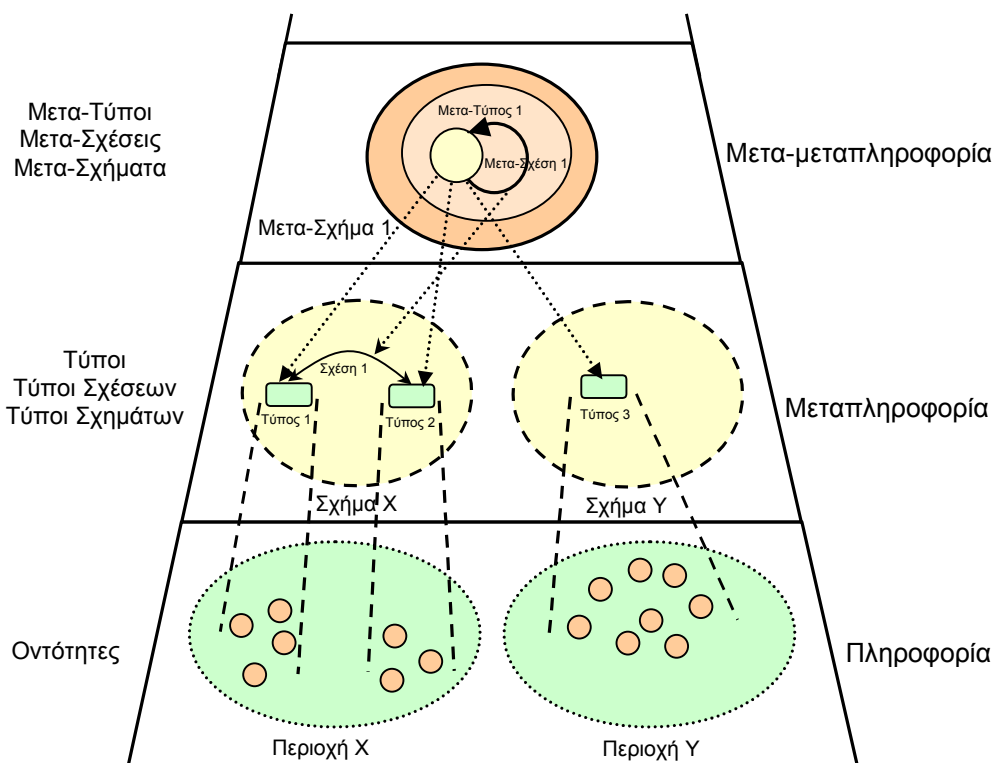
- Η αναπαράσταση σχημάτων, τύπων και σχέσεων μεταξύ τύπων ως μετα-αντικειμένων (meta-objects), σε μορφή εύκολα χρησιμοποιήσιμη από προγράμματα και εφαρμογές.
- Η αναπαράσταση MOF ορισμών συστημάτων τύπων (type system definitions) ως μετα-μετα-αντικειμένων (meta-meta-objects) και η υποστήριξη παραγωγής επιπέδων διεπαφής (interfaces) για τα μετα-αντικείμενα.
- Η παροχή «υπολογιζόμενων» (computed) τύπων σχέσεων, μεταξύ τύπων στο ίδιο ή σε διαφορετικά MOF συστήματα τύπων.

Οι τύποι αναπαρίστανται με δομημένο τρόπο στα πλαίσια του MOF, δίνοντας έτσι τη δυνατότητα να προστίθενται στο μοντέλο διαχείρισης τύπων καλά ορισμένα επίπεδα μετα-μεταπληροφορίας. Με τον τρόπο αυτό μπορούν να καλυφθούν οι πρόσθετες ανάγκες αναπαράστασης που έχουν τα διαφορετικά συστήματα.

Το MOF βασίζεται στο αφηρημένο μοντέλο αναπαράστασης της μεταπληροφορίας που απεικονίζεται στο Σχήμα 3-3. Όπως φαίνεται και στο σχήμα, το μοντέλο αυτό απαρτίζεται από 3 επίπεδα:

- Το *Επίπεδο Πληροφορίας (Information Level)*. Στο επίπεδο αυτό αποθηκεύεται η πληροφορία με τη μορφή οντοτήτων.
- Το *Επίπεδο Μεταπληροφορίας (Metainformation Level)*. Στο επίπεδο αυτό αποθηκεύεται μεταπληροφορία, πληροφορία δηλαδή σχετική με την πληροφορία που αποθηκεύεται στο πρώτο επίπεδο. Η μεταπληροφορία αυτή περιλαμβάνει:

- Ορισμούς των τύπων (type definitions) στους οποίους κατατάσσονται, ανάλογα με τα χαρακτηριστικά τους, οι οντότητες.
- Σχέσεις μεταξύ τύπων (type relations).
- Σχήματα τύπων (type schemas).
- Το *Επίπεδο Μετα-Μεταπληροφορίας (Meta-Metainformation Level)*. Στο επίπεδο αυτό αποθηκεύεται μετα-μεταπληροφορία, μεταπληροφορία δηλαδή σχετική με τη μεταπληροφορία που αποθηκεύεται στο δεύτερο επίπεδο. Η μεταπληροφορία αυτή περιλαμβάνει:
 - Μετα-Τύπους (Meta-Types).
 - Μετα-Σχέσεις (Meta-Relations).
 - Μετα-Σχήματα (Meta-Schemas).



Σχήμα 3-3: Πληροφορία, Μεταπληροφορία και Μετα-Μεταπληροφορία

Τα σημαντικότερα πλεονεκτήματα που προκύπτουν από το σχεδιασμό του MOF είναι τα εξής:

- Η υποστήριξη *ετερογένειας (heterogeneity)* σε ότι αφορά τα υποστηριζόμενα συστήματα τύπων.
- Η υποστήριξη *επεκτασιμότητας*, καθώς ο σχεδιασμός του MOF επιτρέπει την πρόσθεση νέων συστημάτων τύπων σε οποιαδήποτε χρονική στιγμή. Τα συστήματα που προστίθενται μπορεί να είναι είτε εντελώς καινούρια συστήματα είτε συστήματα που προέκυψαν από τη σύνθεση ή την επέκταση υπάρχοντων συστημάτων.

- Το σύστημα είναι ανοικτό, καθώς ο σχεδιασμός του MOF επιτρέπει τη διασύνδεση πολλαπλών MOF εξυπηρετητών, για τη δημιουργία ενός κατακευματισμένου συστήματος παροχής συστημάτων μεταπληροφωρίας.

Η XMI (*XML-based Metadata Interchange*) είναι μια μορφή κωδικοποίησης μεταδεδομένων που έχει οριστεί με βάση τις απαιτήσεις του MOF και τη σύνταξη της XML. Το γεγονός ότι η σύνταξη της XMI είναι σύμφωνη με την XML επιτρέπει τη χρήση της XMI για την ανταλλαγή μεταδεδομένων τόσο μεταξύ CORBA όσο και μεταξύ CORBA και μη-CORBA και μεταξύ μη-CORBA συστημάτων.

Η XMI απαρτίζεται από δυο τμήματα:

- Τους *Κανόνες Παραγωγής XML DTDs (XML DTD Production Rules)*, που επιτρέπουν την παραγωγή XML DTDs για την περιγραφή μεταδεδομένων κωδικοποιημένων σε XMI. Πιο συγκεκριμένα, οι κανόνες αυτοί επιτρέπουν την παραγωγή ενός DTD για οποιοδήποτε συντακτικά σωστό μοντέλο μεταδεδομένων που βασίζεται στο MOF. Έτσι, για κάθε πακέτο (με τη Warwick έννοια) μεταδεδομένων, μπορεί να παραχθεί ένας XML DTD.
- Τους *Κανόνες Παραγωγής XML Εγγράφων (XML Document Production Rules)*, που επιτρέπουν την κωδικοποίηση μεταδεδομένων σε μορφή συμβατή με την XML. Πιο συγκεκριμένα, οι κανόνες αυτοί περιγράφουν τον τρόπο παραγωγής XML εγγράφων που αναπαριστούν τα MOF μεταδεδομένα. Η αντίστροφη εφαρμογή των κανόνων αυτών επιτρέπει την επαναδημιουργία μεταδεδομένων από τα XML έγγραφα.

3.3.1.3. Resource Description Framework (RDF)

Σ' αυτή την υποενότητα θα συζητηθεί το πρότυπο *RDF (Resource Description Framework)*, το οποίο αποτελεί υλοποίηση του πλαισίου Warwick και προσπαθεί να υποστηρίξει τη διαλειτουργικότητα και την επεκτασιμότητα στα όρια του πλαισίου Warwick.

Για να υποστηριχθούν η διαλειτουργικότητα και η επεκτασιμότητα στα όρια του πλαισίου Warwick πρέπει:

- Να οριστούν λεξικά από τις κοινότητες περιγραφής πόρων, τα οποία θα περιγράφουν τη δομή (όπως είναι καταληπτή από τους ανθρώπους) και τη σημασιολογία των όρων περιγραφής πόρων.
- Να οριστούν μορφές (formats) που θα καλύπτουν τη σύνταξη και τη δόμηση (σε ότι αφορά την περιγραφή δομών με τρόπο καταληπτό από τους υπολογιστές) για την περιγραφή των μεταδεδομένων.
- Να καθοριστεί ο τρόπος ορισμού των σχημάτων μεταδεδομένων (metadata schemas).
- Να επιφορτιστούν κάποιοι οργανισμοί με το να πιστοποιούν τα λεξικά μεταδεδομένων και να παρέχουν σε όλα τα μέλη των κοινοτήτων χρηστών πρόσβαση στα εγκεκριμένα λεξικά μεταδεδομένων.
- Να καθοριστούν οι πολιτικές διαχείρισης των λεξικών μεταδεδομένων.
- Να υποστηριχθεί η βελτίωση των λεξικών μεταδεδομένων με την υποστήριξη υποδομών (sub-structures) μέσω της υποστήριξης υπο-στοιχείων (sub-elements) και σχημάτων.

- Να γίνεται προσεκτική επέκταση των λεξικών ώστε να διαφυλαχθεί η διαλειτουργικότητα.
- Να δοθεί η δυνατότητα να υποστηρίζονται συμπληρωματικά πακέτα μεταδεδομένων.
- Να δοθεί η δυνατότητα να ορίζονται επιπλέον στοιχεία για την υποστήριξη των ιδιαίτερων αναγκών των υπο-κοινοτήτων κάθε κοινότητας μέσω συμπληρωματικών πακέτων μεταδεδομένων.

Η RDF παρέχει το πλαίσιο υλοποίησης όσων προαναφέρθηκαν: Είναι μια υλοποίηση του πλαισίου Warwick για την επεξεργασία μεταδεδομένων, που υποστηρίζει τη διαλειτουργικότητα μεταξύ εφαρμογών που ανταλλάσσουν πληροφορία καταληπτή από υπολογιστές στον Παγκόσμιο Ιστό. Ιδιαίτερη έμφαση δίνεται στην παροχή της δυνατότητας αυτόματης επεξεργασίας των διαθέσιμων στον Παγκόσμιο Ιστό πόρων [RDF].

Στην πραγματικότητα, η RDF είναι μια γλώσσα για την κωδικοποίηση μεταδεδομένων με τρόπο προσαρμοσμένο στις ανάγκες του Παγκοσμίου Ιστού. Πιο συγκεκριμένα, η RDF είναι ένα είδος γλώσσας σχεσιακής βάσης δεδομένων που χρησιμοποιείται για την αναπαράσταση των σχέσεων που υπάρχουν μεταξύ αντικειμένων που είναι διαθέσιμα μέσω του Παγκοσμίου Ιστού.

Το πλαίσιο που παρέχει η RDF κάνει δυνατή την κωδικοποίηση, ανταλλαγή και επαναχρησιμοποίηση δομημένων μεταδεδομένων [Mil98]. Τα προαναφερθέντα υλοποιούνται μέσω του σχεδιασμού μηχανισμών που υποστηρίζουν μια κοινή αντίληψη σε ότι αφορά τη σημασιολογία, τη σύνταξη και τη δομή των μεταδεδομένων. Η RDF είναι μια εφαρμογή της XML που καθορίζει τους απαραίτητους δομικούς περιορισμούς για την παροχή σαφώς ορισμένων μεθόδων για την έκφραση της σημασιολογίας των μεταδεδομένων.

Ένας από τους βασικούς στόχους της RDF είναι ο ορισμός ενός μηχανισμού γενικού σκοπού για την περιγραφή πόρων χωρίς να γίνονται οποιεσδήποτε υποθέσεις που θα τον περιόριζαν σε μια συγκεκριμένη περιοχή εφαρμογών ούτε να ορίζεται εκ των προτέρων η σημασιολογία οποιασδήποτε περιοχής εφαρμογών. Μ' αυτό τον τρόπο, χωρίς ο ορισμός να είναι δεμένος με μια συγκεκριμένη περιοχή, ο μηχανισμός που προκύπτει είναι κατάλληλος για την υποστήριξη οποιασδήποτε περιοχής εφαρμογών.

Επιπλέον, η RDF δίνει τη δυνατότητα δημιουργίας λεξικών, που και οι άνθρωποι θα μπορούν να καταλάβουν και οι υπολογιστές θα μπορούν να επεξεργαστούν. Τα λεξικά είναι σύνολα ιδιοτήτων (properties) και στοιχείων μεταδεδομένων που ορίζονται από τις κοινότητες περιγραφής πόρων. Τα λεξικά θα προωθήσουν την επαναχρησιμοποίηση και την επέκταση της σημασιολογίας των μεταδεδομένων από διαφορετικές κοινότητες χρηστών. Η RDF έχει σχεδιαστεί έτσι ώστε να υποστηρίζει αυτό τον τύπο σημασιολογικής συναρμομοσιμότητας (modularity) με τη δημιουργία ενός πλαισίου που υποστηρίζει το συνδυασμό κατανεμημένων συστημάτων πιστοποίησης. Αυτό επιτρέπει στις κοινότητες περιγραφής πόρων να δηλώνουν λεξικά που μπορούν να ξαναχρησιμοποιηθούν, να επεκταθούν και να βελτιωθούν για να αντιμετωπίσουν τις απαιτήσεις θεματικών περιοχών και εφαρμογών.

Η RDF διαθέτει ένα αντικειμενοστραφές μοντέλο αναπαράστασης, όπου ορίζονται κλάσεις στοιχείων οι οποίες περιγράφουν συγκεκριμένα πεδία μεταδεδομένων. Οι συλλογές κλάσεων, οι οποίες βασίζονται στα προτεινόμενα από τις κοινότητες περιγραφής πόρων λεξικά, ονομάζονται *RDF σχήματα (RDF schemas)* [RDFS]. Οι κλάσεις μέσα σε ένα σχήμα οργανώνονται ιεραρχικά, ενώ δίνεται και η δυνατότητα εξειδίκευσης των κλάσεων (subclassing).

Οι δομικοί περιορισμοί που θέτει η RDF για την υποστήριξη συνεπούς κωδικοποίησης και ανταλλαγής μεταδεδομένων εγγυούνται τη διαλειτουργικότητα μεταξύ των πακέτων μεταδεδομένων που ορίζονται από διαφορετικές κοινότητες περιγραφής πόρων. Επιπλέον, η RDF υποστηρίζει τη χρήση συμβάσεων που κάνουν δυνατή τη διαλειτουργικότητα και την επαναχρησιμοποιησιμότητα (reusability) συνόλων μεταδεδομένων.

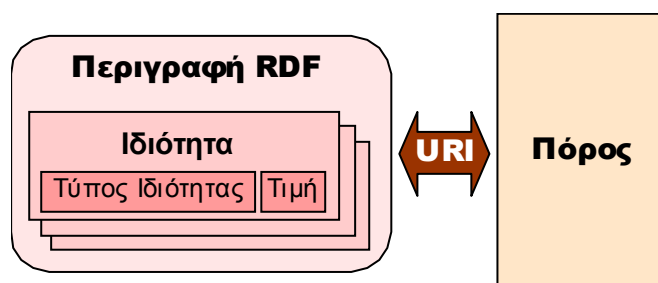
Η προτεινόμενη σύνταξη της RDF είναι σύμφωνη με την XML, καθώς ένας από τους στόχους της RDF είναι να παρέχει, με ένα πρότυπο και διαλειτουργικό τρόπο, τη δυνατότητα καθορισμού της σημασιολογίας των στοιχείων μεταδεδομένων με βάση την XML. Επιπλέον, η χρήση XML σύνταξης από την RDF επιτρέπει τη σαφή έκφραση της σημασιολογίας των μεταδεδομένων και, μέσω αυτής, τη συνεπή κωδικοποίηση, ανταλλαγή και επεξεργασία των μεταδεδομένων. Αυτό όμως δε σημαίνει ότι η XML σύνταξη είναι η μόνη που μπορεί να χρησιμοποιηθεί για την υποστήριξη της RDF: οποιαδήποτε σύνταξη ικανή να υποστηρίξει την αναπαράσταση του μοντέλου της RDF μπορεί να χρησιμοποιηθεί.

Στη συνέχεια θα επικεντρώσουμε στο μοντέλο της RDF, στην προτεινόμενη σύνταξή της και στα RDF σχήματα.

Το μοντέλο της RDF

Όπως αναφέρθηκε και παραπάνω, η RDF παρέχει ένα αντικειμενοστραφές μοντέλο για την περιγραφή πόρων: η καρδιά της RDF είναι ένα ανεξάρτητο από σύνταξη μοντέλο για την αναπαράσταση και την περιγραφή πόρων. Το μοντέλο αυτό θέτει δομικούς περιορισμούς στην έκφραση των μοντέλων δεδομένων των εφαρμογών, ώστε να είναι δυνατή η συνεπής κωδικοποίηση, ανταλλαγή και επεξεργασία των μεταδεδομένων. Παράλληλα, επιτρέπει στις κοινότητες περιγραφής πόρων να ορίσουν καθεμιά τη δική της σημασιολογία και θέτει τις βάσεις για δομική διαλειτουργικότητα μεταξύ διαφορετικών εφαρμογών.

Κεντρικό ρόλο στο μοντέλο της RDF παίζουν οι *Περιγραφές*. Μια γενικευμένη περιγραφή σύμφωνα με αυτό το μοντέλο έχει τη μορφή που φαίνεται στο Σχήμα 3-4.



Σχήμα 3-4: Γενικευμένη RDF περιγραφή πόρου:

Στα πλαίσια της RDF ως *πόρος (resource)* ορίζεται οποιοδήποτε αντικείμενο χαρακτηρίζεται μοναδικά από ένα URI [URI97]. Οι πόροι έχουν *ιδιότητες (properties)* οι

οποίες χαρακτηρίζονται από τον τύπο (*property type*) και την τιμή (*property value*) τους. Οι τύποι ιδιοτήτων εκφράζουν τις σχέσεις μεταξύ των τιμών και των πόρων με τους οποίους οι τελευταίες συνδέονται. Οι τιμές μπορεί να είναι ατομικές (συμβολοσειρές χαρακτήρων, αριθμοί κ.α.) ή να είναι πόροι, οι οποίοι με τη σειρά τους μπορούν να έχουν ιδιότητες. *Πρόταση (statement)* είναι ένας πόρος, μαζί με μια ιδιότητα και την τιμή της. Μια συλλογή προτάσεων ονομάζεται *περιγραφή (description)*.

Οι RDF ιδιότητες αναπαριστούν, εκτός από τα χαρακτηριστικά των πόρων, και τις σχέσεις μεταξύ πόρων. Έτσι, από μια αντικειμενοστραφή σκοπιά, οι πόροι αντιστοιχούν στα αντικείμενα και οι ιδιότητες στα χαρακτηριστικά των αντικειμένων.

Με τη χρήση του τριαδικού μοντέλου που απαρτίζεται από πόρους, τύπους ιδιοτήτων και τιμές, η RDF παρέχει μια σαφή μέθοδο για την έκφραση της σημασιολογίας των μεταδεδομένων με τρόπο καταληπτό από υπολογιστές. Οι σχέσεις μεταξύ πόρων, τύπων ιδιοτήτων και τιμών αναπαρίστανται από ένα κατευθυνόμενο γράφο: Οι πόροι αναπαρίστανται ως κόμβοι, οι τύποι ιδιοτήτων ως κατευθυνόμενα διανύσματα και οι συμβολοσειρές χαρακτήρων που αναπαριστούν τιμές περικλείονται από εισαγωγικά.

Οι ιδιότητες των πόρων συνδέονται σαφώς με τους αντίστοιχους πόρους με τη χρήση μοναδικών αναγνωριστικών ως ταυτοτήτων. Η σαφής ταυτότητα των πόρων επιτρέπει την επαναχρησιμοποίηση περιγραφικής πληροφορίας.

Το μοντέλο της RDF, το οποίο φαίνεται να μοιάζει με διάγραμμα οντοτήτων-σχέσεων (ER/Entity-Relationship diagrams), επιτρέπει τη δημιουργία πόρων σε πολλά επίπεδα.

Η σύνταξη της RDF

Για την υλοποίηση του μοντέλου της RDF είναι απαραίτητη μια σύνταξη που θα επιτρέπει την αποθήκευση των στιγμιotypών αυτού του μοντέλου σε μορφή καταληπτή από τους υπολογιστές και την ανταλλαγή τους μεταξύ εφαρμογών. Η σύνταξη που έχει επιλεγεί για το σκοπό αυτό είναι η XML, καθώς η RDF βασίζεται στην υποστηριζόμενη από την XML δομή για τη συνεπή αναπαράσταση της σημασιολογίας των λεξικών μεταδεδομένων και των μεταδεδομένων.

Η σύνταξη της RDF είναι αυστηρή και βασίζεται στη «σειριοποίηση» (serialization) των γράφων που αναπαριστούν το RDF μοντέλο.

Όπως έχει ήδη αναφερθεί, η RDF υποστηρίζει τη δημιουργία λεξικών από ανεξάρτητες κοινότητες χρηστών. Προβλήματα όμως μπορούν να υπάρξουν όταν κάποιες κοινότητες χρηστών χρησιμοποιούν τον ίδιο τύπο ιδιότητας για να περιγράψουν διαφορετικές έννοιες. Για να αντιμετωπιστεί το πρόβλημα αυτό, η RDF δίνει μοναδικά αναγνωριστικά στους τύπους ιδιοτήτων με τη χρήση του μηχανισμού των XML namespaces.

Η σύνταξη της RDF υποστηρίζει τη δημιουργία πόρων σε πολλαπλά επίπεδα καθώς περιλαμβάνει, εκτός από το σκελετό της RDF:

- *Πολλαπλά σχήματα.*
- *Δοχεία (Containers).* Τα δοχεία που υποστηρίζει η RDF είναι τα εξής:
 - *Bag*, που χρησιμοποιείται για την αναπαράσταση μη διατεταγμένων συνόλων.
 - *Sequence*, που χρησιμοποιείται για την αναπαράσταση διατεταγμένων συνόλων.

- *Alternatives*, που χρησιμοποιείται για την αναπαράσταση εναλλακτικών τιμών.
- *Περιγραφές περιγραφών*. Τόσο οι προτάσεις όσο και οι περιγραφές μπορούν να διαθέτουν τις δικές τους περιγραφές.
- *Σύνθετες Τιμές (Complex Values)*. Οι σύνθετες τιμές που υποστηρίζει η RDF είναι οι εξής:
 - *Συμβολοσειρές Χαρακτήρων (Strings)*.
 - *Σύνθετοι Τύποι Δεδομένων*.
 - *Ιδιότητες*.
 - *Πόροι*.
- *Πολύγλωσση Υποστήριξη (Multilingual Support)*. Η πολύγλωσση υποστήριξη παρέχεται στην RDF μέσω της XML σύμφωνα με το RFC1766 [RFC1766].

Για απλές εφαρμογές και για να είναι δυνατή μια σύνταξη απλούστερη της XML (π.χ. HTML) έχει οριστεί μια *συντημημένη σύνταξη (abbreviated syntax)* της RDF. Η σύνταξη αυτή όμως δεν επιτρέπει σύνθετες τιμές ούτε την επανάληψη ιδιοτήτων.

Ένα γενικευμένο παράδειγμα χρήσης της RDF φαίνεται στο *Παράδειγμα 3-5*. Στο παράδειγμα αυτό, εκτός από τις ιδιότητες που ορίζονται στον πυρήνα της RDF, χρησιμοποιούνται και οι ιδιότητες που περιγράφονται από τη namespace με αναγνωριστικό X που ορίζονται στο URI XURI. Στη συνέχεια δημιουργείται μια περιγραφή του XURI με τον ορισμό των ιδιοτήτων Property 1 και Property 2 που περιγράφονται στο XURI.

```
<? xml version = "1.0"?>
<RDF xmlns = http://www.w3.org/TR/WD-rdf-syntax#
  xmlns:X = "XURI">
<Description about = "XURI">
  <X:Property 1>
  <X:Property 2>
</Description>
</RDF>
```

Παράδειγμα 3-6: Γενικευμένο παράδειγμα χρήσης της RDF

RDF σχήματα

Τα *RDF σχήματα (RDF Schemas)* είναι μια γλώσσα καθορισμού σχημάτων μεταδεδομένων (*metadata schema specification language*) και χρησιμοποιούνται για τη δήλωση λεξικών και την αναπαράσταση της σημασιολογίας των τύπων ιδιοτήτων που ορίζονται από μια συγκεκριμένη κοινότητα περιγραφής πόρων. Στα RDF σχήματα ορίζονται οι επιτρεπτές ιδιότητες μιας RDF περιγραφής, καθώς και τα χαρακτηριστικά και οι περιορισμοί των τύπων ιδιοτήτων. Ο μηχανισμός των XML namespaces χρησιμοποιείται για το χαρακτηρισμό και την αναγνώριση των RDF σχημάτων.

Τα RDF σχήματα δομούνται με βάση το μοντέλο δεδομένων της RDF. Στα RDF σχήματα ορίζονται οι RDF κλάσεις (RDF classes) και οι RDF ιδιότητες (RDF properties), οι οποίες διαθέτουν ετικέτες (labels), σχόλια (comments) και περιορισμούς (constraints) με τη μορφή

περιοχών τιμών (ranges) και θεματικών περιοχών (domains). Τα RDF σχήματα είναι, στην πραγματικότητα, επεκταμένα με αντικειμενοστραφή χαρακτηριστικά διαγράμματα οντοτήτων-σχέσεων, τα οποία αποτελούν στιγμιότυπα RDF μοντέλων.

Αναμένεται ότι η δυνατότητα τυποποίησης λεξικών μεταδεδομένων καταληπτών τόσο από ανθρώπους όσο και από υπολογιστές που παρέχουν τα RDF σχήματα θα ενθαρρύνει τη χρήση, ανταλλαγή και επέκταση των λεξικών μεταδεδομένων από διαφορετικές κοινότητες περιγραφής πόρων.

Ένα παράδειγμα RDF σχήματος, όπου δίνεται η περιγραφή ενός προσώπου βάσει του ονόματος, του επωνύμου και της ηλικίας του φαίνεται στο Παράδειγμα 3-7.

```
<?xml version='1.0'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema"
  xmlns:Πρόσωπο="">
  <rdfs:Class ID="Πρόσωπο">
    <rdfs:label>Πρόσωπο</rdfs:label>
    <rdfs:SubclassOf
      rdf:resource=" http://www.w3.org/TR/PR-rdf-schema#Resource" />
  </rdfs:Class>
  <rdf:Property ID="ηλικία">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#Number" />
    <rdfs:domain rdf:resource="#Πρόσωπο" />
  </rdf:Property>
  <rdf:Property ID="όνομα">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#Πρόσωπο" />
  </rdf:Property>
  <rdf:Property ID="επώνυμο">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#Πρόσωπο" />
  </rdf:Property>
</rdf:RDF>
```

Παράδειγμα 3-7: RDF Σχήμα που περιγράφει κάποιο Πρόσωπο

3.3.2. Λεξικά Μεταδεδομένων

Στην ενότητα αυτή θα συζητηθούν τα πιο διαδεδομένα λεξικά μεταδεδομένων που έχουν προταθεί από διάφορες κοινότητες περιγραφής πόρων και χρησιμοποιούνται από κοινότητες χρηστών. Τα γνωστότερα λεξικά μεταδεδομένων είναι τα εξής [Rus98]:

- *Australian Government Metadata*: Λεξικό μεταδεδομένων που υποστηρίζει το σύστημα AGLS (Australian Government Locator Service) της κυβέρνησης της Αυστραλίας. Το λεξικό βασίζεται στη Dublin Core, ενώ χρησιμοποιεί 2 ακόμα στοιχεία.
- *CC/PP (Composite Capability/Preference Profiles)* [CC/PP]: RDF λεξικό μεταδεδομένων, που χρησιμοποιείται για την περιγραφή τόσο των ενδιαφερόντων και των προτιμήσεων των χρηστών όσο και των δυνατοτήτων της υπολογιστικής πλατφόρμας που χρησιμοποιεί κάθε χρήστης. Οι προτιμήσεις και τα ενδιαφέροντα των

χρηστών, μαζί με τα χαρακτηριστικά και τις δυνατότητες της υπολογιστικής πλατφόρμας που χρησιμοποιεί κάθε χρήστης, οργανώνονται σε διαγράμματα χρηστών (user profiles) γραμμένα σε σύνταξη συμβατή με τη CC/PP.

Η πληροφορία που περιέχεται στα διαγράμματα χρηστών μπορεί να χρησιμοποιηθεί τόσο από εξυπηρετητές (servers) όσο και από παροχείς υπηρεσιών (service providers), έτσι ώστε να διαμορφώνονται με βάση αυτήν τόσο οι υπηρεσίες όσο και το περιεχόμενο που παρέχονται σε κάθε χρήστη.

- *Dublin Core* [Lag96]: Το λεξικό Dublin Core είναι το αποτέλεσμα μιας προσπάθειας ορισμού ενός απλού αλλά παρ' όλα αυτά χρησιμοποιήσιμου συνόλου από 15 στοιχεία μεταδεδομένων που περιγράφουν τα βασικά χαρακτηριστικά των εγγράφων στο Διαδίκτυο. Η ολοένα αυξανόμενη χρήση του λεξικού Dublin Core καθώς και το ότι επηρέασε σε σημαντικό βαθμό το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» είναι οι βασικότεροι λόγοι για τους οποίους γίνεται παρακάτω μια πιο εκτεταμένη περιγραφή του.
- *DCD (Document Content Description)* [DCD]: Η DCD είναι ένα λεξικό που προτείνει το W3C για την περιγραφή μεταδεδομένων σχετικών με τη δομή (structural metadata) και το περιεχόμενο των XML εγγράφων. Η DCD είναι ένα RDF λεξικό. Αναλυτικότερη συζήτηση για τη DCD γίνεται στην υποενότητα 3.2.1.1.
- *DOI (Digital Object Identifier)*: Πρότυπο που αναπτύχθηκε από πέντε εκδότες και ενώσεις εκδοτών στην Αμερική και άρχισε να χρησιμοποιείται τον Οκτώβριο του 1997. Το DOI έχει υιοθετήσει τα στοιχεία του Dublin Core για τις βιβλιογραφικές και θεματικές περιγραφές, ενώ διαθέτει και μια ποικιλία επιπλέον στοιχείων χρήσιμων για τους εκδότες (π.χ. τιμές, συνδρομές κ.α.). Επιπλέον, διαθέτει μια σειρά στοιχείων που δηλώνουν τη θέση του περιγραφόμενου αντικειμένου στην αλυσίδα παραγωγής (π.χ. αν ένα βιβλίο είναι στη διαδικασία της έγκρισης, τότε κατατέθηκε από το συγγραφέα κ.α.).
- *EAD (Encoded Archival Description)*: Λεξικό που χρησιμοποιείται για την περιγραφή του υλικού των Αμερικάνικων ιστορικών αρχείων.
- *Πρότυπο FGDC για γεωγραφικά και χωρικά δεδομένα* [FGDC]: Το λεξικό αυτό καθορίζει το περιεχόμενο των μεταδεδομένων που περιγράφουν γεωγραφικά και χωρικά δεδομένα. Ο σκοπός του είναι η παροχή ενός κοινού συνόλου που αποτελείται από ορολογία και ορισμούς για θέματα που σχετίζονται με τα μεταδεδομένα. Στο λεξικό FGDC καθορίζονται τα ονόματα τόσο των απλών όσο και των σύνθετων στοιχείων που χρησιμοποιούνται για αυτό το σκοπό, οι ορισμοί των στοιχείων καθώς και πληροφορίες σχετικές με τις τιμές που μπορούν να πάρουν αυτά τα στοιχεία.
- *GEM (Gateway to Educational Materials)*: Λεξικό μεταδεδομένων που χρησιμοποιείται στο σύστημα που αναπτύσσεται από την κοινοπραξία GEM και στοχεύει στην παροχή στους εκπαιδευτές γρήγορης και εύκολης πρόσβασης σε εκπαιδευτικό υλικό που γίνεται διαθέσιμο από κρατικές υπηρεσίες, Πανεπιστήμια και κερδοσκοπικούς και μη οργανισμούς. Το λεξικό GEM περιέχει 23 στοιχεία, τα 15 από τα οποία είναι τα υποστηριζόμενα από τη Dublin Core στοιχεία, ενώ τα υπόλοιπα 8 καλύπτουν συγκεκριμένες ανάγκες της εκπαιδευτικής κοινότητας.

- *GILS (Government Information Locator Service)*: Ιδιαίτερα προσανατολισμένο προς τα δεδομένα κειμένου λεξικό που χρησιμοποιείται από τις κυβερνήσεις των ΗΠΑ και του Καναδά.
- *IMS [IMS]*: Το λεξικό IMS περιέχει τον ορισμό στοιχείων απαραίτητων για την υποστήριξη εκπαιδευτικών εφαρμογών στο Διαδίκτυο. Οι βασικότεροι στόχοι του IMS είναι οι εξής:
 - Η υποστήριξη αποτελεσματικής αναζήτησης υψηλής ποιότητας υλικού για συγκεκριμένους εκπαιδευτικούς σκοπούς στο Διαδίκτυο.
 - Η διαχείριση του εκπαιδευτικού υλικού, συμπεριλαμβανομένων των πνευματικών δικαιωμάτων, του εμπορίου και των εμπειριών μάθησης.
 - Η επαναχρησιμοποίηση του εκπαιδευτικού υλικού.

Το IMS αποτελείται από δύο τμήματα: από ένα λεξικό στοιχείων και τιμών και από την οργάνωση αυτών των στοιχείων σε ομάδες ή σύνολα που προσφέρονται για την περιγραφή διαφορετικών τύπων εκπαιδευτικών πόρων. Ο συνδυασμός αυτών των δυο τμημάτων αποτελεί το IMS σχήμα μεταδεδομένων.

- *INDECS*: Λεξικό μεταδεδομένων που βασίζεται στη Dublin Core και στοχεύει στην κάλυψη των αναγκών των εφαρμογών ηλεκτρονικού εμπορίου. Η σημασιολογία του καλύπτει τόσο τις ανάγκες του εμπορικού κόσμου όσο και του κόσμου της τέχνης και του πολιτισμού.
- *MARC (Machine Readable Catalogs) [Dels96]*: Οικογένεια λεξικών που χρησιμοποιείται σε πολλές χώρες του κόσμου για την περιγραφή του υλικού των βιβλιοθηκών. Υπάρχουν μικρές διαφοροποιήσεις στα λεξικά που χρησιμοποιούνται από χώρα σε χώρα (π.χ. USMARC, AUSMARC κ.α.), ενώ υπάρχει και ένα παγκόσμιο λεξικό, το UNIMARC. Όλες οι κατά τόπους διαφοροποιήσεις του MARC μπορούν να μετατραπούν σε UNIMARC με τη χρήση ενός συνόλου απλών κανόνων που καθορίζονται στη φάση του ορισμού κάθε επιμέρους λεξικού.

Τα MARC λεξικά ξεκίνησαν να χρησιμοποιούνται το 1966 από τη βιβλιοθήκη του Κογκρέσου, καθώς το MARC είναι αποτέλεσμα ενός προγράμματος της βιβλιοθήκης του Κογκρέσου. Από τότε τα MARC λεξικά εξελίσσονται διαρκώς.

- *P3P (Platform for Privacy Preferences Project)*: Πλατφόρμα που επιτρέπει στους χρήστες να ενημερώνονται για τις πρακτικές σε ότι αφορά την *εχεμύθεια (privacy)* των υπηρεσιών και των σελίδων που επισκέπτονται στον Παγκόσμιο Ιστό.

Ο καθορισμός της P3P περιλαμβάνει τη χρήση ενός ειδικού λεξικού που αποτελείται από λέξεις-κλειδιά και τιμές που συνδέονται με αυτές. Το λεξικό αυτό επιτρέπει την κωδικοποίηση πληροφορίας σχετικής με εχεμύθεια σε μορφή καταληπτή και επεξεργάσιμη από εξυπηρετητές και προγράμματα πλοήγησης του Παγκοσμίου Ιστού [P3P].

Η P3P είναι εφαρμογή της RDF και θα υποστηρίξει *ψηφιακές υπογραφές (digital signatures)* [DSig1] [DSig2] [DSig3] και *ψηφιακά πιστοποιητικά (digital certificates)* μόλις αυτά γίνουν διαθέσιμα.

Οι απαιτήσεις που καθοδήγησαν το σχεδιασμό της P3P ήταν οι εξής:

- Η δυνατότητα ενημέρωσης των χρηστών για την πολιτική εχεμύθειας που ακολουθείται για κάθε υπηρεσία ή ιστοσελίδα.
- Η δυνατότητα έκφρασης των προτιμήσεων των χρηστών σε σχέση με τις πολιτικές εχεμύθειας, μέσα από τον καθορισμό της σχέσης τους με τις υπηρεσίες και τις ιστοσελίδες που ακολουθούν κάποια πολιτική εχεμύθειας.
- Η επιβεβαίωση ότι όλες οι ανταλλαγές δεδομένων με την υπηρεσία ή την ιστοσελίδα γίνονται με βάση την προσυμφωνημένη με το χρήστη πολιτική.
- Ο καθορισμός της γραμματικής και των λεξικών για την ενημέρωση των χρηστών σχετικά με τις πολιτικές των υπηρεσιών και των ιστοσελίδων και την αποδοχή των πολιτικών αυτών από τους χρήστες.
- Η ανάπτυξη πρωτοκόλλων για την ανταλλαγή δεδομένων και τη σύναψη συμφωνιών με τους χρήστες.

Έτσι, οι υπηρεσίες και οι ιστοσελίδες που ακολουθούν κάποια πολιτική εχεμύθειας που συμφωνεί με τις προτιμήσεις κάποιου χρήστη θα μπορούν να προσπελαστούν απ' αυτόν χωρίς περαιτέρω διευκρινήσεις. Επιπρόσθετα, η σχέση του χρήστη με ιστοσελίδες και υπηρεσίες που ακολουθούν πολιτική εχεμύθειας διαφορετική από τις προτιμώμενες από αυτόν θα μπορούν να προσπελαστούν μόνο μετά από ενημέρωσή του σχετικά με την ακολουθούμενη πολιτική και αν αυτός τελικά την αποδεχθεί.

Έτσι, οι μεν χρήστες θα μπορούν να αποφασίζουν για το αν θα προσπελάσουν οι υπηρεσίες και οι ιστοσελίδες που ακολουθούν κάποια πολιτική εχεμύθειας με την οποία δε συμφωνούν (μερικά ή συνολικά), ενώ οι παροχείς υπηρεσιών και περιεχομένου θα μπορέσουν να αναβαθμίσουν την ποιότητα των προϊόντων που προσφέρουν.

- *PICS (Platform for Internet Content Selection)* [PICS1] [PICS2] [PICS3] [PICS4] [DSig]: Το λεξικό PICS αναπτύσσεται από την ομάδα DSig του W3C για τη δημιουργία ψηφιακά υπογεγραμμένων (digitally signed) βαθμολογήσεων και περιγραφών πόρων. Στην πραγματικότητα το λεξικό PICS έδωσε τη δυνατότητα ορισμού μεταδεδομένων σχετικών με το περιεχόμενο των πόρων στο Διαδίκτυο: Το λεξικό PICS ασχολήθηκε με τη βαθμολόγηση περιεχομένου πόρων (resource content rating) ώστε να είναι εύκολο να χτιστούν μηχανισμοί που θα επέτρεπαν στους γονείς να αποκλείουν από τα παιδιά τους την πρόσβαση σε βλαβερό για τη διαπαιδαγώγησή τους υλικό.
- *SPF (Site Preview Format)*: Λεξικό που χρησιμοποιείται από το Netscape Netcenter για την περιγραφή πόρων στο Διαδίκτυο. Η σύνταξη του λεξικού βασίζεται στην XML, η δομή του ορίζεται με βάση το μοντέλο δεδομένων της RDF και η σημασιολογία του βασίζεται στη Dublin Core.

Πιο συγκεκριμένα, η βάση του λεξικού είναι ένα υποσύνολο των στοιχείων της Dublin Core, τα οποία αντιστοιχίζονται σε «ετικέτες», ενώ έχουν προστεθεί μερικά επιπλέον στοιχεία που καλύπτουν συγκεκριμένες απαιτήσεις της Netscape. Η υλοποίηση του λεξικού γίνεται με τη χρήση RDF σχημάτων.

- *TEI (Text Encoding Initiative)*: Λεξικό που αναπτύχθηκε το 1988 από το Virginia Text Center. Το Λεξικό αυτό είναι προσανατολισμένο στην περιγραφή δεδομένων κειμένου.
- *WHOIS++ Templates*: Ένα από τα πρώτα λεξικά μεταδεδομένων με σαφή προσανατολισμό προς τα πρότυπα (templates) για την περιγραφή αντικειμένων στο περιβάλλον του Διαδικτύου. Χρησιμοποιήθηκε αρχικά για λίστες ηλεκτρονικού ταχυδρομείου (mailing lists) αλλά σύντομα επεκτάθηκε σε αρκετούς τύπους πόρων.

3.3.2.1. Dublin Core

Η πρωτοβουλία Dublin Core [DC97] [Bea99] είναι μια παγκόσμια προσπάθεια, επηρεασμένη από πολλές και διαφορετικές απόψεις, για τον ορισμό ενός πυρήνα στοιχείων που θα αποτελέσουν ένα λεξικό μεταδεδομένων για την υποστήριξη της αναζήτησης πόρων στο Διαδίκτυο [WeiHa98]. Το λεξικό αυτό, που αποτελεί το θέμα της τρέχουσας υποενότητας, είναι ένα από τα πιο διαδεδομένα λεξικά μεταδεδομένων.

Η αποτελεσματική ανταλλαγή πληροφορίας για την ανακάλυψη πόρων απαιτεί την ύπαρξη μιας αρχιτεκτονικής που υποστηρίζει συμβάσεις για τη σημασιολογία, τη δομή και τη σύνταξη των μεταδεδομένων. Για το λόγο αυτό η Dublin Core κοινότητα δουλεύει κοντά στην RDF κοινότητα, έτσι ώστε να αναπτυχθεί μια κοινή αρχιτεκτονική για την υποστήριξη γενικευμένων μεταδεδομένων.

Ο πυρήνας της Dublin Core είναι ένα σύνολο 15 στοιχείων, οι επιτρεπτές τιμές των οποίων είναι συμβολοσειρές χαρακτήρων. Το σύνολο αυτό είναι απλό, περιγραφικό, διαισθητικό, ευέλικτο και ταιριάζει σε πολλές και διαφορετικές ανάγκες. Τα 15 στοιχεία που αποτελούν τον πυρήνα της Dublin Core έχουν ήδη υιοθετηθεί από αρκετές κοινότητες χρηστών και υποστηρίζονται από πιλοτικά συστήματα. Τα στοιχεία αυτά χωρίζονται σε τρεις ομάδες, οι οποίες δηλώνουν την κλάση και την εμβέλεια της πληροφορίας που αποθηκεύεται σε κάθε στοιχείο. Οι ομάδες αυτές περιλαμβάνουν:

- Στοιχεία που σχετίζονται κυρίως με το *περιεχόμενο* του περιγραφόμενου πόρου. Στην ομάδα αυτή ανήκουν τα εξής στοιχεία:

Τίτλος (Title): Ο τίτλος που δίνεται στον πόρο από το Δημιουργό ή τον Εκδότη.

Θέμα (Subject): Το θέμα του πόρου. Συνήθως το θέμα είναι εκφρασμένο σαν ένα σύνολο από λέξεις-κλειδιά ή φράσεις που περιγράφουν το θέμα ή το περιεχόμενο του πόρου. Συνίσταται η χρήση καθορισμένων λεξικών και επίσημων σχημάτων ταξινόμησης για τον ορισμό της τιμής του θέματος του πόρου.

Περιγραφή (Description): Μια περιγραφή του περιεχομένου του πόρου, που περιέχει περιλήψεις όταν το περιεχόμενο του πόρου είναι σε μορφή κειμένου, την περιγραφή του περιεχομένου της εικόνας αν πρόκειται για εικόνες κλπ.

Κάλυψη (Coverage): Η χωρική ή χρονική περιοχή που καλύπτεται από το περιεχόμενο του τρέχοντος πόρου. Η χωρική κάλυψη χρησιμοποιεί συντεταγμένες ή ονόματα περιοχών. Η χρονική

κάλυψη περιγράφει τη χρονική περίοδο που καλύπτεται από τον πόρο σε μορφή συμβατή με το [Date].

Σχέση (Relation):

Ο κωδικός ενός άλλου πόρου και η σχέση του με τον τρέχοντα πόρο. Το στοιχείο αυτό επιτρέπει τη δήλωση σχέσεων και συνδέσμων μεταξύ συσχετιζόμενων πόρων και περιγραφών πόρων.

Το στοιχείο Σχέση διαθέτει δυο υποστοιχεία: το υποστοιχείο Σχέση.Ταυτότητα (Relation.Identifier) και το υποστοιχείο Σχέση.Τύπος (Relation.Type).

Το υποστοιχείο Σχέση.Ταυτότητα περιγράφει τα μεταδεδομένα του συσχετιζόμενου με τον περιγραφόμενο πόρο ή τον ίδιο το συσχετιζόμενο πόρο.

Το υποστοιχείο Σχέση.Τύπος περιγράφει τη φύση της σχέσης του περιγραφόμενου πόρου με το συσχετιζόμενο πόρο. Για να διασφαλιστεί η διαλειτουργικότητα, η τιμή αυτού του στοιχείου επιλέγεται από ένα σύνολο σχέσεων που υποστηρίζεται από τη Dublin Core. Το σύνολο αυτό περιλαμβάνει:

- Σχέσεις τμήματος/ολότητας, στις οποίες καθορίζεται αν ένας πόρος αποτελεί φυσικό ή λογικό τμήμα κάποιου άλλου πόρου. Οι αντίστοιχες σχέσεις που υποστηρίζονται από τη Dublin Core είναι οι σχέσεις *ΕίναιΤμήμαΤου* (*IsPartOf*) και *ΔιαθέτειΤμήμα* (*HasPart*).
- Σχέσεις που περιγράφουν την παραγωγή ενός πόρου από κάποιον άλλο μέσω της μετατροπής της μορφής στην οποία αποθηκεύεται ο πόρος. Οι αντίστοιχες σχέσεις που υποστηρίζονται από τη Dublin Core είναι οι σχέσεις *ΕίναιΤηςΜορφής* (*IsFormatOf*) και *ΔιαθέτειΜορφή* (*HasFormat*).
- Σχέσεις αναφοράς, που αναπαριστούν το γεγονός ότι ένας πόρος αναφέρεται από κάποιον άλλο πόρο. Οι σχέσεις αναφοράς που υποστηρίζονται από τη Dublin Core είναι οι σχέσεις *ΑναφέρεταιΑπό* (*IsReferencedBy*) και *ΑναφέρεταιΣε* (*References*).
- Σχέσεις δημιουργικότητας, που υπάρχουν όταν ένας πόρος αποτελεί τη βάση για κάποιον άλλο (που είναι π.χ. μελέτη πάνω σε αυτόν, μετάφραση του κ.α.). Οι αντίστοιχες σχέσεις που υποστηρίζονται από τη Dublin Core είναι οι σχέσεις *ΒασίζεταιΣε* (*IsBasedOn*) και *ΑποτελείΒάσηΓια* (*IsBasisFor*).

- Σχέσεις έκδοσης, που υπάρχουν όταν ένας πόρος αποτελεί νεότερη έκδοση κάποιου άλλου. Οι σχέσεις έκδοσης που υποστηρίζονται από τη Dublin Core είναι οι σχέσεις *ΑποτελείΈκδοσηΤου* (*IsVersionOf*) και *ΔιαθέτειΈκδοση* (*HasVersion*).

Πηγή (Source): Πληροφορία για κάποιον πόρο από τον οποίο ο τρέχων πόρος απορρέει. Συνιστάται στην πράξη να προτιμάται η χρήση του στοιχείου *Σχέση* αντί του στοιχείου *Πηγή*.

Γλώσσα (Language): Η γλώσσα στην οποία είναι γραμμένο το περιεχόμενο του τρέχοντος πόρου, εκφρασμένη σε μορφή συμβατή με το RFC 1766 [RFC1766].

- Στοιχεία που σχετίζονται κυρίως με θέματα *πνευματικής ιδιοκτησίας* του περιγραφόμενου πόρου. Στην ομάδα αυτή ανήκουν τα εξής στοιχεία:

Δημιουργός (Creator): Το άτομο ή ο οργανισμός που είναι κύρια υπεύθυνος για τη δημιουργία του πνευματικού περιεχομένου του πόρου (π.χ. οι συγγραφείς στην περίπτωση κειμένων, φωτογράφοι στην περίπτωση φωτογραφιών, εικονογράφοι στην περίπτωση σκίτσων κ.α.).

Εκδότης (Publisher): Ο οργανισμός που έκανε διαθέσιμο τον πόρο στην παρούσα μορφή του και είναι υπεύθυνος για τη διαχείρισή του. Ο οργανισμός αυτός μπορεί να είναι ένας εκδοτικός οίκος, ένα τμήμα πανεπιστημίου ή μια συνεταιρική οντότητα.

Συνεργάτης (Contributor): Πρόσωπο ή οργανισμός που δεν προσδιορίζεται στο στοιχείο *Δημιουργός* και έχει συμβάλει σημαντικά στη δημιουργία του πόρου, αλλά η συνεισφορά του είναι δευτερεύουσα σε σχέση με τη συνεισφορά οποιουδήποτε προσώπου ή οργανισμού που προσδιορίζεται στο στοιχείο *Δημιουργός*.

Δικαιώματα (Rights): Μια δήλωση δικαιωμάτων χρήσης, ένας κωδικός που οδηγεί σε μια δήλωση δικαιωμάτων χρήσης ή ένας κωδικός που οδηγεί σε μια υπηρεσία παροχής πληροφοριών σχετικά με τα δικαιώματα χρήσης του πόρου.

- Στοιχεία που σχετίζονται κυρίως με τη *δημιουργία* του περιγραφόμενου πόρου. Στην ομάδα αυτή ανήκουν τα εξής στοιχεία:

Κωδικός (Identifier): Μια συμβολοσειρά (π.χ. URL, URI) ή ένας αριθμός (π.χ. ISBN, ISSN) που χρησιμοποιείται για να προσδιορίσει μοναδικά τον πόρο.

Ημερομηνία (Date): Μια ημερομηνία σχετική με τη δημιουργία και τη διαθεσιμότητα του πόρου σε μορφή συμβατή με το [Date].

Μορφή (Format): Η μορφή στην οποία είναι αποθηκευμένα τα δεδομένα που απαρτίζουν το περιεχόμενο του πόρου. Χρησιμοποιείται για να πιστοποιήσει το λογισμικό (software) και ίσως και το υλικό

(hardware) που μπορεί να χρησιμοποιηθεί για την παρουσίαση ή τη λειτουργία του πόρου. Για να διασφαλιστεί η διαλειτουργικότητα, η τιμή της Μορφής πρέπει να επιλέγεται από ένα σύνολο υποστηριζόμενων μορφών, το οποίο βρίσκεται υπό κατάρτιση

Τύπος (Type): Ο τύπος του πόρου (π.χ. αναφορά εργασίας, μυθιστόρημα, λεξικό κ.α.). Για να διασφαλιστεί η διαλειτουργικότητα, η τιμή του Τύπου πρέπει να επιλέγεται από ένα σύνολο υποστηριζόμενων τύπων, το οποίο βρίσκεται υπό κατάρτιση.

Όλα τα στοιχεία που απαρτίζουν τον πυρήνα της Dublin Core είναι προαιρετικά και μπορούν να χρησιμοποιηθούν οσεςδήποτε φορές για κάποιο πόρο. Οποιοιδήποτε περιορισμοί (π.χ. αν ο τίτλος για κάποια κατηγορία εγγράφων πρέπει να είναι απαραίτητος ή/και μοναδικός) θα καθορίζονται μέσα από τις εφαρμογές.

Το σύνολο στοιχείων που παρέχει η Dublin Core είναι επεκτάσιμο, καθώς οι επιμέρους κοινότητες χρηστών, για να καλύψουν τις ιδιαίτερες ανάγκες τους, μπορούν να μετατρέψουν τη σημασιολογία των στοιχείων του πυρήνα της Dublin Core. Η σημασιολογία των στοιχείων της Dublin Core μπορεί να μετατραπεί με δυο τρόπους:

1. Με το δανεισμό *πιστοποιητών (qualifiers)* από υπάρχοντα σχήματα μεταδεδομένων, γεγονός που θα επιτρέψει την εισαγωγή λεπτομερέστερης και σαφέστερης σημασιολογίας από προϋπάρχοντα συστήματα στη Dublin Core. Τρεις πιστοποιητές, οι οποίοι επιτρέπουν στη Dublin Core να λειτουργεί σε παγκόσμιο επίπεδο και να καλύπτει υψηλότερου επιπέδου απαιτήσεις για την ανακάλυψη πόρων, έχουν ήδη γίνει αποδεκτοί. Οι πιστοποιητές αυτοί είναι οι εξής:

Lang: Πιστοποιητής που δηλώνει τη γλώσσα στην οποία είναι εκφρασμένα τα μεταδεδομένα. Μπορεί να χρησιμοποιηθεί τόσο κατά τη διαδικασία της ανακάλυψης πόρων όσο και κατά το φιλτράρισμα των αποτελεσμάτων.

Sheme: Πιστοποιητής που δηλώνει το σύνολο των κανονισμών, των προτύπων και των συμβάσεων απ' όπου προέρχεται ένας όρος της μεταπληροφορίας.

Sub-Element: Εξειδίκευση των στοιχείων της Dublin Core, έτσι ώστε αυτά να γίνουν ακριβέστερα και σαφέστερα.

Προτάσεις για τους πιστοποιητές scheme και subelement των στοιχείων του πυρήνα της Dublin Core γίνονται στο [Gue97]. Το γεγονός πάντως ότι σχεδόν όλες οι εφαρμογές που χρησιμοποιούν τον πυρήνα της Dublin Core χρησιμοποιούν κάποιους πιστοποιητές δείχνει την αναγκαιότητα των πιστοποιητών για την περιγραφή πόρων.

2. Με την ανάπτυξη επεκτάσεων και εξειδικεύσεων με συγκεκριμένο σκοπό (ad hoc), έτσι ώστε να αποσαφηνιστεί η σημασία των στοιχείων που αποτελούν τον πυρήνα της Dublin Core.

Στο σημείο αυτό πρέπει να αναφερθεί ότι στην 6^η συνεδρίαση της πρωτοβουλίας Dublin Core (DC6 - Νοέμβριος 1998) έγιναν οι εξής προτάσεις σε ότι αφορά το σύνολο των στοιχείων του πυρήνα της Dublin Core:

- Να αντικατασταθούν τα στοιχεία Δημιουργός, Συνεργάτης και Εκδότης από το στοιχείο *Διαχειριστής (Agent)*, το οποίο θα διαθέτει το υποστοιχείο *Διαχειριστής.Ρόλος (Agent.Role)*, το οποίο θα περιγράφει το ρόλο κάθε διαχειριστή.
- Να παραλειφθεί το στοιχείο *Πηγή* και να αντικατασταθεί από το στοιχείο *Σχέση*, αφού υποστηριχθεί ο επιπλέον τύπος σχέσης *Πηγή*.

Πρέπει επίσης να αναφερθεί ότι η κοινότητα Dublin Core προσανατολίζεται στη χρήση της RDF ως σύνταξης. Έτσι, μέσω της ισχυρής σε εκφραστικότητα σύνταξης της RDF θα υποστηριχθούν αποτελεσματικά τόσο το μοντέλο δεδομένων της Dublin Core (που περιλαμβάνει σύνθετες δομές όπως είναι οι πιστοποιητές) όσο και υπηρεσίες (όπως η πολύγλωσση υποστήριξη) στη σύνταξη της Dublin Core.

3.4. Document Object Model (DOM)

Σ' αυτό το τμήμα θα περιγραφεί το *DOM (Document Object Model)* [DOM1] [DOM2], το οποίο είναι ένα επίπεδο διεπαφής, ανεξάρτητο από πλατφόρμα και γλώσσα προγραμματισμού, που επιτρέπει σε προγράμματα και scripts να προσπελαίνουν και να ενημερώνουν δυναμικά τη δομή, το περιεχόμενο και τη μορφοποίηση των εγγράφων. Ο καθορισμός του DOM γίνεται σε δύο επίπεδα:

- Στο πρώτο επίπεδο του DOM (DOM Level 1 specification) ορίζεται ένα API (Application Programming Interface) για XML και HTML έγγραφα, που περιγράφει τη λογική δομή των εγγράφων και τον τρόπο προσπέλασης και διαχείρισής τους.
- Στο δεύτερο επίπεδο του DOM (DOM Level 2 specification) επεκτείνεται το API του πρώτου επιπέδου και ορίζονται επίπεδα διεπαφής για την υποστήριξη XML namespaces, για τη σύνδεση style sheets με έγγραφα, για το μοντέλο αντικειμένων των CSS (Cascading Style Sheets), για ένα μοντέλο γεγονότων και για ένα μοντέλο ερωτήσεων.

Ο στόχος του πρώτου επιπέδου του DOM είναι ο ορισμός προγραμματιστικών επιπέδων διεπαφής για την XML και την HTML. Το πρώτο επίπεδο του DOM παρέχει στους προγραμματιστές:

- Ένα σύνολο αντικειμένων για την αναπαράσταση HTML και XML εγγράφων.
- Ένα μοντέλο για το πώς τα αντικείμενα αυτά μπορούν να συνδυαστούν.
- Ένα επίπεδο διεπαφής για τη διαχείριση των αντικειμένων και την πρόσβαση σε αυτά.

Το πρώτο επίπεδο του DOM περιλαμβάνει 2 τμήματα: τον πυρήνα (core) και το HTML τμήμα. Ο πυρήνας παρέχει ένα χαμηλού επιπέδου σύνολο, που αποτελείται από βασικά επίπεδα διεπαφής, κατάλληλα για την αναπαράσταση οποιουδήποτε δομημένου εγγράφου και για τον ορισμό εκτεταμένων επιπέδων διεπαφής με XML έγγραφα. Επιπλέον, το HTML τμήμα του πρώτου επιπέδου του DOM παρέχει υψηλού επιπέδου επίπεδα διεπαφής που παρέχουν μια πιο εύχρηστη άποψη των HTML εγγράφων.

Οι στόχοι του δεύτερου επιπέδου του DOM, το οποίο βασίζεται στο πρώτο επίπεδο, είναι οι εξής:

- Η παροχή πρόσβασης με προγραμματιστικό τρόπο στις δομές των CSS. Έτσι, παρέχεται πρόσβαση στην πλούσια δομή των CSS και δίνεται η δυνατότητα μετατροπής της σε ότι αφορά τα στυλ και την παρουσίαση των εγγράφων. Με τον τρόπο αυτό ενισχύονται τα CSS με ένα μηχανισμό που ελέγχει δυναμικά την εξαίρεση και την ενσωμάτωση style sheets, καθώς και τη διαχείριση και την ενσωμάτωση κανόνων και ιδιοτήτων των CSS.
- Η παροχή, ανεξάρτητα από γλώσσες ερωτήσεων, επιπέδων διεπαφής για ερωτήσεις.
- Ο σχεδιασμός ενός γενικού συστήματος γεγονότων (event system) που θα επιτρέπει την πιστοποίηση των διαχειριστών γεγονότων (event handlers), θα περιγράφει τη ροή των γεγονότων μέσω μιας δενδρικής δομής και θα παρέχει βασικές πληροφορίες σχετικές με τα γεγονότα.
- Η παροχή βασικών συνόλων γεγονότων για τον έλεγχο των επιπέδων διεπαφής με τους χρήστες (user interfaces) και τη διαχείριση εγγράφων.
- Η παροχή ενός κοινού συνόλου γεγονότων, που θα υποστηρίζονται τόσο από τον MS-Internet Explorer όσο και από το Netscape Navigator.
- Η παροχή επιπέδων διεπαφής για τη διάσχιση δενδρικών δομών, λιστών και αποτελεσμάτων ερωτήσεων (με τη χρήση αντικειμένων όπως είναι τα αντικείμενα Filter και Iterator).
- Η αναγνώριση συνεχόμενων τμημάτων εγγράφων (με τη χρήση αντικειμένων όπως είναι το αντικείμενο Range).

Με τη χρήση του DOM δίνεται οι δυνατότητα να γίνουν προγραμματιστικά οι εξής λειτουργίες:

- Δημιουργία εγγράφων.
- Διάσχιση της λογικής δομής των εγγράφων.
- Πρόσθεση, διαγραφή και μετατροπή των στοιχείων που απαρτίζουν τη λογική δομή των εγγράφων.
- Δημιουργία style sheets.
- Πρόσβαση στα style sheets των εγγράφων.
- Πρόσθεση, διαγραφή και μετατροπή των στοιχείων που απαρτίζουν τα style sheets των εγγράφων.

Καθώς το DOM έχει αναπτυχθεί από το W3C, στόχος του είναι η παροχή ενός προγραμματιστικού επιπέδου διεπαφής που θα μπορεί να χρησιμοποιηθεί σε μια μεγάλη ποικιλία περιβαλλόντων και εφαρμογών. Για το λόγο αυτό το DOM έχει σχεδιαστεί έτσι ώστε να είναι δυνατή η χρησιμοποίησή του με οποιαδήποτε γλώσσα προγραμματισμού.

3.5. Γλώσσες Διαχείρισης & Αναζήτησης Εγγράφων

Το γεγονός ότι είναι διαθέσιμος ένας μεγάλος αριθμός XML εγγράφων θέτει ένα αριθμό τεχνικών ερωτήσεων που δεν απαντώνται από το XML πρότυπο. Οι σημαντικότερες από τις ερωτήσεις αυτές είναι οι εξής:

- Πως θα εξάγεται η πληροφορία από μεγάλα XML έγγραφα;
- Πως θα ανταλλάσσεται η XML πληροφορία;
- Πως θα ανταλλάσσονται δεδομένα μεταξύ κοινοτήτων χρηστών που χρησιμοποιούν XML έγγραφα με διαφορετική δομή (που υπακούουν δηλαδή σε διαφορετικά DTDs, XML-Data, XML Schema Language ή DDML σχήματα ή DCDs);
- Πως θα συνδυαστούν XML δεδομένα από πολλαπλές πηγές;

Σαν απάντηση στα τεχνικά αυτά ζητήματα έχουν προταθεί γλώσσες ερωτήσεων για XML έγγραφα. Αυτό συμβαίνει γιατί το μοντέλο δεδομένων της XML δεν περιγράφεται ικανοποιητικά ούτε από το σχεσιακό μοντέλο ούτε από το αντικειμενοστραφές μοντέλο, οπότε δε μπορούν να χρησιμοποιηθούν άμεσα ούτε οι γλώσσες ερωτήσεων για σχεσιακά συστήματα ούτε οι γλώσσες ερωτήσεων για αντικειμενοστραφή συστήματα.

Σ' αυτό το τμήμα θα γίνει μια ανασκόπηση των προτεινόμενων γλωσσών διαχείρισης και αναζήτησης XML εγγράφων. Πιο συγκεκριμένα, θα περιγραφούν η *XML-QL* στην υποενότητα 3.5.1 και η *XQL (XML Query Language)* στην υποενότητα 3.5.2.

3.5.1. XML-QL

Η *XML-QL*, που αποτελεί το θέμα αυτής της ενότητας, είναι μια γλώσσα ερωτήσεων που έρχεται να απαντήσει στα τεχνικά θέματα που τέθηκαν στην αρχή της ενότητας 3.5. Η *XML-QL* παρέχει δυνατότητες έκφρασης ερωτήσεων που εξάγουν δεδομένα από XML έγγραφα, μετατροπής XML δεδομένων και συνδυασμού πληροφορίας που προέρχεται από διαφορετικές πηγές.

Στο σημείο αυτό πρέπει να υπογραμμιστεί ότι η *XML-QL* διαφέρει από την *XSL*, που έχει ως κύριο στόχο τον καθορισμό των στυλ και της μορφοποίησης των XML εγγράφων και σ' αυτό το πλαίσιο ορίζει και χρησιμοποιεί κάποιες λειτουργίες γλώσσας ερωτήσεων. Παρ' όλο που η *XML-QL* έχει σε κάποια σημεία παρόμοια λειτουργικότητα με την *XSL*, η *XML-QL* υποστηρίζει περισσότερο λειτουργίες που είναι προσανατολισμένες προς τα δεδομένα (π.χ. joins) και παρέχει καλύτερη υποστήριξη για τη δημιουργία νέων XML εγγράφων μέσω των τροποποιήσεων (transformations).

Το προτεινόμενο από την *XML-QL* μοντέλο δεδομένων για την XML, που αναπαριστά τα XML έγγραφα ως *XML γράφους (XML graphs)*, είναι μια παραλλαγή του *Ημιδομημένου Μοντέλου Δεδομένων (Semistructured Data Model)* [Bun97]. Ένας XML γράφος απαρτίζεται από:

- Ένα γράφο *G*, κάθε κόμβος του οποίου αναπαρίσταται από μια μοναδική συμβολοσειρά που ονομάζεται *ταυτότητα αντικειμένου (object identifier – OID)*.
- Ο γράφος διαθέτει ένα κόμβο που χαρακτηρίζεται ως *ρίζα (root)* του γράφου.
- Οι ακμές του γράφου έχουν τα ονόματα των στοιχείων που απαρτίζουν το έγγραφο.
- Οι κόμβοι του γράφου έχουν τα ονόματα των συνόλων χαρακτηριστικών που συναντώνται στο έγγραφο.

- Τα φύλλα του γράφου έχουν τιμές με τη μορφή συμβολοσειρών χαρακτήρων. Μόνο τα φύλλα του γράφου μπορούν να έχουν τιμές, και κάθε φύλλο μπορεί να έχει μια μοναδική τιμή.

Σε ότι αφορά τη σειρά εμφάνισης των στοιχείων που απαρτίζουν τα XML έγγραφα, η XML-QL προτείνει δυο διαφορετικά μοντέλα δεδομένων:

- Το *μη διατεταγμένο μοντέλο δεδομένων (unordered data model)*, στο οποίο η σειρά με την οποία εμφανίζονται τα στοιχεία αγνοείται. Παρ' όλα αυτά, όταν ένα έγγραφο πρόκειται να υποστεί επεξεργασία ή να παρουσιαστεί, επιλέγεται μια διάταξη των στοιχείων του (που είτε είναι τυχαία είτε καθορίζεται από την εφαρμογή παρουσίασης ή επεξεργασίας των εγγράφων).
- Το *διατεταγμένο μοντέλο δεδομένων (ordered data model)*, που μοιάζει με το μη διατεταγμένο μοντέλο δεδομένων, αλλά μέσω των δομών που παρέχει η XML-QL μπορεί για κάθε κόμβο να δώσει μια λίστα των κόμβων που προηγούνται αυτού του κόμβου.

Στην παρούσα φάση η XML-QL παρέχει την εξής λειτουργικότητα:

- Υποστήριξη ερωτήσεων με τη χρήση των ονομάτων των στοιχείων που απαρτίζουν τα έγγραφα.
- Υποστήριξη *Εκφράσεων Μονοπατιού (Path Expressions)*, που επιτρέπουν τον προσδιορισμό φωλιασμένων και κυκλικών δομών (π.χ. γράφων, δέντρων κ.α.).
- Μετατροπή XML δεδομένων (π.χ. μετατροπή ενός εγγράφου που ανήκει στην κλάση A σε έγγραφο που ανήκει στην κλάση B).
- Συνδυασμό δεδομένων από διαφορετικές πηγές. Με τη χρήση της XML-QL μπορούν να τεθούν ταυτόχρονα ερωτήσεις σε πολλαπλές πηγές XML δεδομένων και τα αποτελέσματα να παραχθούν σε κοινή μορφή.
- Ενσωμάτωση ερωτήσεων στα δεδομένα.
- Ορισμό συναρτήσεων (functions).

Τα σχετικά με την XML-QL θέματα που παραμένουν ανοικτά είναι τα εξής:

- Αναγνώριση οντοτήτων (entities) σε XML-QL ερωτήσεις.
- Υποστήριξη κατηγορημάτων που θα ορίζονται από τους χρήστες.
- Υποστήριξη κανονικών εκφράσεων (regular expressions).
- Υποστήριξη συνοθυλευματών (aggregates).
- Υποστήριξη semijoins.
- Υποστήριξη namespaces.
- Επέκταση σε άλλα πρότυπα που σχετίζονται με την XML (XLink, XPointer, RDF κ.α.).
- Κωδικοποίηση των XML-QL ερωτήσεων ως XML έγγραφα ή τμήματα εγγράφων.

Η σύνταξη της XML-QL μοιάζει με SQL (είναι SQL-like). Έτσι, για την επιλογή των σημειωμάτων που στον τίτλο τους περιέχεται η συμβολοσειρά WWW εκφράζεται σε XML-QL η ερώτηση που φαίνεται στο Παράδειγμα 3-8:

```
WHERE <memo> $m
      <title>*WWW*</>
    </>
CONSTRUCT <result>$m</>
```

Παράδειγμα 3-8: XML-QL Ερώτηση για την επιλογή των σημειωμάτων που στον τίτλο τους περιέχεται η συμβολοσειρά WWW

3.5.2. XQL (XML Query Language)

Η XQL (*XML Query Language*) [Rob99] [Rob98] [XQL], που περιγράφεται σ' αυτή την ενότητα, είναι μια γενικού σκοπού γλώσσα ερωτήσεων για XML έγγραφα. Η XQL είναι απλή, ισχυρή και κομψή γλώσσα. Το μοντέλο δεδομένων της είναι αυτό του *διατεταγμένου δέντρου (ordered tree)*. Κατά τη διάρκεια του σχεδιασμού της, 4 περιοχές επιλέχθηκαν για την ανάλυση των απαιτήσεων που τέθηκαν για την XQL:

- Η υποστήριξη ερωτήσεων μέσα σε ένα έγγραφο. Η λειτουργία αυτή είναι χρήσιμη σε εργαλεία όπως είναι τα προγράμματα πλοήγησης και οι επεξεργαστές κειμένου, όπου ο χρήστης πρέπει να μπορεί να βρει το τμήμα του εγγράφου που τον ενδιαφέρει χωρίς να χρειαστεί να διασχίσει ολόκληρο το έγγραφο.
- Η υποστήριξη ερωτήσεων σε συλλογές εγγράφων. Η λειτουργία αυτή είναι χρήσιμη για την υποστήριξη ερωτήσεων βάσει του περιεχομένου και της δομής των εγγράφων σε βάσεις εγγράφων.
- Η διευθυνσιοδότηση εντός ενός εγγράφου ή μεταξύ διαφορετικών εγγράφων. Η λειτουργία αυτή είναι χρήσιμη για την αναφορά σε γνωστές τοποθεσίες μέσα στα έγγραφα με τη χρήση συνδέσμων υπερκειμένου.
- Τα XSL πρότυπα (XSL patterns). Η λειτουργία αυτή είναι χρήσιμη στους μετασχηματισμούς από-δέντρο-σε-δέντρο που λαμβάνουν χώρα κατά τη διάρκεια εφαρμογής style sheets σε έγγραφα.

Οι περιοχές αυτές είναι αρκετά διαφορετικές μεταξύ τους, καθώς διαφέρουν σε ότι αφορά την είσοδο, την έξοδο το μοντέλο επεξεργασίας και τη μορφή στην οποία θα επιστραφούν τα αποτελέσματα. Επιπλέον, σε κάποιες από αυτές οι ερωτήσεις δεν έχουν την παραδοσιακή μορφή τους. Από την άλλη, υπάρχει μια αρκετά ευρεία κοινή περιοχή, ειδικά σε ότι αφορά τον προσδιορισμό των XML δομών. Εκτός αυτού, και στις 4 περιοχές υπάρχει η ανάγκη να μπορεί να επιλεγεί ένα σύνολο κόμβων (οι οποίοι αναπαριστούν έγγραφα ή στοιχεία εγγράφων) με βάση κριτήρια σχετικά με το όνομα, το περιεχόμενο, τις τιμές των χαρακτηριστικών και τις σχέσεις μεταξύ συνόλων κόμβων.

Οι προσεγγίσεις που θα μπορούσαν να ακολουθηθούν είναι δύο:

1. Να σχεδιαστεί διαφορετική γλώσσα ερωτήσεων για κάθε περιοχή. Η προσέγγιση αυτή έχει το πλεονέκτημα ότι μπορεί να δώσει ιδιαίτερη σημασία στα σημεία που έχουν ιδιαίτερη βαρύτητα για κάθε περιοχή. Παράλληλα όμως υπάρχει το προηγούμενο ότι οι ειδικού σκοπού γλώσσες έχουν συνήθως μικρή διάρκεια ζωής.

2. Να σχεδιαστεί μια γενικού σκοπού γλώσσα για όλες αυτές τις περιοχές. Η προσέγγιση αυτή έχει το πλεονέκτημα ότι μια αρκετά γενική γλώσσα ερωτήσεων είναι χρήσιμη για την επίλυση διαφορετικών προβλημάτων, ιδιαίτερα αν αυτά στην πορεία ξεπεράσουν τα προκαθορισμένα όριά τους. Το πρόβλημα που προκύπτει από το ότι οι διαφορετικές περιοχές έχουν διαφορετικές απαιτήσεις μπορεί να λυθεί μέσω της χρήσης διαφορετικών υποσυνόλων της γλώσσας από κάθε περιοχή.

Η XQL αποτελεί υλοποίηση της δεύτερης προσέγγισης και έχει το επιπλέον πλεονέκτημα ότι είναι πλήρως ανεξάρτητη από τους παράγοντες που διαχωρίζουν τις περιοχές: Η XQL δε λέει τίποτα για τον τρόπο με τον οποίο θα δίνεται η είσοδος, για τη μορφή της εξόδου ή για το μοντέλο επεξεργασίας που χρησιμοποιείται για να ελεγχθεί αν κάποια στοιχεία πληρούν τα κριτήρια που έχουν εκφραστεί.

Για να γίνει περισσότερο κατανοητή η XQL, ακολουθεί μια συγκριτική με την SQL παρουσίασή της στον πίνακα που ακολουθεί:

SQL	XQL
Η Βάση Δεδομένων είναι ένα σύνολο πινάκων	Η Βάση Δεδομένων είναι ένα σύνολο XML εγγράφων
Οι ερωτήσεις εκφράζονται στην SQL, μια γλώσσα ερωτήσεων που το βασικό μοντέλο της χρησιμοποιεί πίνακες	Οι ερωτήσεις εκφράζονται στην XQL, μια γλώσσα ερωτήσεων που το βασικό μοντέλο της χρησιμοποιεί τη δομή της XML
Οι πίνακες που παίρνουν μέρος στην ερώτηση καθορίζονται στη FROM πρόταση της ερώτησης	Καθορίζεται ένα σύνολο κόμβων από ένα ή περισσότερα έγγραφα (search context), οι οποίοι, όπως και οι κόμβοι-απόγονοί τους παίρνουν μέρος στην ερώτηση
Το αποτέλεσμα της ερώτησης είναι ένας πίνακας οι γραμμές του οποίου αναπαριστούν τα επιστρεφόμενα αντικείμενα	Το αποτέλεσμα (result) της ερώτησης είναι ένα σύνολο XML κόμβων (result set), οι οποίοι ομαδοποιούνται από ένα κόμβο-ρίζα για να δημιουργηθεί έτσι ένα «καλά ορισμένο» XML έγγραφο ως απάντηση.

Πίνακας 3-7: Συγκριτική παρουσίαση SQL/XQL

Η XQL υποστηρίζει ερωτήσεις με βάση:

- Το περιεχόμενο των στοιχείων που απαρτίζουν τα XML έγγραφα.
- Τις τιμές των χαρακτηριστικών των στοιχείων που απαρτίζουν τα XML έγγραφα.
- Τον τύπο των στοιχείων που απαρτίζουν τα XML έγγραφα.
- Τις namespaces που χρησιμοποιούνται στα XML έγγραφα.
- Τις σχέσεις που υπάρχουν μεταξύ στοιχείων, οι οποίες και καθορίζουν τη λογική δομή των εγγράφων. Οι σχέσεις αυτές χωρίζονται σε 3 κατηγορίες:
 - Ιεραρχικές σχέσεις, που περιλαμβάνουν τις σχέσεις:
 - Πατέρα-παιδιού (Parent-Child)
 - Προγόνου-Απογόνου (Ancestor-Descendant)

- Σχέσεις διάταξης (μέσα σε ένα έγγραφο ή μεταξύ στοιχείων του ίδιου επιπέδου στη δένδροειδή λογική δομή), που περιλαμβάνουν τις σχέσεις:
 - Προηγείται (Precedes)
 - Προηγείται άμεσα(Immediately precedes)
- Σχέσεις θέσης (μέσα σε ένα έγγραφο ή μεταξύ στοιχείων του ίδιου επιπέδου στη δένδροειδή λογική δομή), που περιλαμβάνουν τις σχέσεις:
 - Απόλυτη (Absolute) Θέση.
 - Σχετική (Relevant) Θέση.
 - Περιοχή (Range) Θέση.

Οι συνθήκες που είναι σχετικές με στοιχεία και με τις σχέσεις μεταξύ στοιχείων συνδυάζονται σε ερωτήσεις μονοπατιού.

Οι XQL εκφράσεις απαρτίζονται από:

- Όρους (Terms) (όπως είναι τα ονόματα στοιχείων, τα ονόματα χαρακτηριστικών κλπ.) που παίρνουν μέρος στην ερώτηση. Ο Πίνακας 3-8 περιέχει τη σύνταξη που υποστηρίζει την αναφορά σε όρους και την έκφραση κριτηρίων σχετικά με αυτούς.

Όνομα	Σύνταξη	Περιγραφή
Όνομα Στοιχείου	<Όνομα Στοιχείου>	Τα στοιχεία με όνομα «Όνομα Στοιχείου»
Οποιοδήποτε Όνομα Στοιχείου	*	Τα στοιχεία με οποιοδήποτε όνομα
Όνομα Χαρακτηριστικού	<Όνομα Στοιχείου>@ <Όνομα Χαρακτηριστικού>	Τα χαρακτηριστικά των στοιχείων με όνομα «Όνομα Στοιχείου» που έχουν όνομα «Όνομα Χαρακτηριστικού»
Οποιοδήποτε Όνομα Χαρακτηριστικού	<Όνομα Στοιχείου>@*	Τα χαρακτηριστικά των στοιχείων με όνομα «Όνομα Στοιχείου» που έχουν οποιοδήποτε όνομα
Κόμβος Κειμένου	text()	Όλοι οι κόμβοι που περιέχουν μόνο κείμενο
Σχόλιο	comment()	Όλοι οι κόμβοι που περιέχουν μόνο σχόλια
Εντολή Επεξεργασίας	pi()	Όλοι οι κόμβοι που περιέχουν μόνο εντολές επεξεργασίας
Εντολή Επεξεργασίας με συγκεκριμένο στόχο	pi («Στόχος»)	Όλοι οι κόμβοι που περιέχουν μόνο εντολές επεξεργασίας με στόχο τα στοιχεία με όνομα «Στόχος»
Κόμβος – πατέρας των αποτελεσμάτων της έκφρασης	.	Ο (πραγματικός ή ιδεατός) κόμβος – πατέρας των αποτελεσμάτων της έκφρασης

Πίνακας 3-8: Σύνταξη για τους Όρους και τα σχετικά με αυτούς Κριτήρια

- Περιορισμούς σχετικά με τις namespaces, οι οποίοι καθορίζουν πότε οι namespaces πρέπει να εξετάζονται κατά την επεξεργασία των ερωτήσεων.

Σύνταξη	Περιγραφή
<Όνομα Στοιχείου>	Τα στοιχεία με όνομα «Όνομα Στοιχείου», ανεξάρτητα από τη namespace στην οποία το «Όνομα Στοιχείου» ορίζεται
<Όνομα Namespace>: <Όνομα Στοιχείου>	Τα στοιχεία με όνομα «Όνομα Στοιχείου», όπου το «Όνομα Στοιχείου» ορίζεται στη namespace με όνομα «Όνομα Namespace»
*	Τα στοιχεία με οποιοδήποτε όνομα, ανεξάρτητα από namespaces
: <Όνομα Στοιχείου>	Τα στοιχεία με όνομα «Όνομα Στοιχείου», όπου το «Όνομα Στοιχείου» δεν ορίζεται σε κάποια namespace
*: <Όνομα Στοιχείου>	Τα στοιχεία με όνομα «Όνομα Στοιχείου», όπου το «Όνομα Στοιχείου» ορίζεται σε κάποια namespace
<Όνομα Namespace>: *	Τα στοιχεία με οποιοδήποτε όνομα που ορίζεται στη namespace με όνομα «Όνομα Namespace»
: *	Τα στοιχεία με οποιοδήποτε όνομα που δεν ορίζεται σε κάποια namespace
*: *	Τα στοιχεία με οποιοδήποτε όνομα που ορίζεται σε κάποια namespace

Πίνακας 3-9: Σύνταξη για τις Namespaces και τα σχετικά με αυτές Κριτήρια

- Συγκρίσεις και περιορισμούς (constraints) που χρησιμοποιούνται για την έκφραση κριτηρίων που σχετίζονται με το περιεχόμενο των κόμβων που παίρνουν μέρος στην ερώτηση. Ο Πίνακας 3-10 περιέχει τους τελεστές που υποστηρίζονται για το σκοπό αυτό από την XQL.

Όνομα	Σύνταξη
Ισότητα	Τελεστής «=» Τελεστής «eq»
Σύγκριση (ανεξάρτητη κεφαλαίων – πεζών)	Τελεστής «ieq»
Ανισότητα	Τελεστής «!=» Τελεστής «ne»
Ύπαρξη προτύπου κειμένου	Τελεστής «contains»
Ύπαρξη προτύπου κειμένου (ανεξάρτητη κεφαλαίων – πεζών)	Τελεστής «icontains»
Μικρότερο	Τελεστής «<» Τελεστής «lt»
Μικρότερο ή ίσο	Τελεστής «<=» Τελεστής «lte»
Μεγαλύτερο	Τελεστής «>» Τελεστής «gt»
Μεγαλύτερο ή ίσο	Τελεστής «>=» Τελεστής «gte»

Πίνακας 3-10: Τελεστές για την υποστήριξη Συγκρίσεων και Περιορισμών από την XQL

- Ιεραρχικές σχέσεις, που επιτρέπουν την έκφραση κριτηρίων που σχετίζονται με τη δομή των κόμβων που παίρνουν μέρος στην ερώτηση. Ο Πίνακας 3-11 περιέχει τις ιεραρχικές σχέσεις που υποστηρίζονται από την XQL.

Περιγραφή	Σύνταξη
Σχέση πατέρα – παιδιού	<Όνομα Στοιχείου - Πατέρα> /<Όνομα Στοιχείου - Παιδιού>
Σχέση προγόνου – απογόνου	<Όνομα Στοιχείου - Προγόνου> //<Όνομα Στοιχείου - Απογόνου>
N-ιοστός απόγονος (ή N-ιοστό παιδί) του στοιχείου με όνομα «Όνομα Στοιχείου»	<Όνομα Στοιχείου>[N]

Πίνακας 3-11: Οι Ιεραρχικές Σχέσεις που υποστηρίζονται από την XQL

- Φίλτρα, που επιτρέπουν τον περιορισμό του συνόλου κόμβων που παίρνουν μέρος στην ερώτηση με τη χρήση τελεστών (operators), συγκρίσεων και περιορισμών. Τα φίλτρα περιέχουν συνθήκες, που περικλείονται μέσα σε «[]».
- Δυναμικούς Τελεστές (Binary Operators) και Τελεστές Συνόλων (Set Operators), που επιτρέπουν το συνδυασμό όρων και άλλων XQL εκφράσεων. Οι κανόνες προτεραιότητας για τους τελεστές ορίζονται από την XQL, ενώ τα αποτελέσματα των XQL εκφράσεων υπολογίζονται από αριστερά προς δεξιά. Ο Πίνακας 3-12 περιέχει τους τελεστές που υποστηρίζονται από την XQL.

Όνομα	Σύνταξη
Άρνηση (Negation)	not («Εκφραση»)
Ένωση (Union)	«Εκφραση 1» union «Εκφραση 2» «Εκφραση 1» «Εκφραση 2»
Τομή (Intersection)	«Εκφραση 1» intersect «Εκφραση 2»
Και τα δύο (Both)	«Εκφραση 1» ~ «Εκφραση 2»
Ή (Or)	«Εκφραση 1» or «Εκφραση 2»
Και (And)	«Εκφραση 1» and «Εκφραση 2»

Πίνακας 3-12: Οι Λογικοί Τελεστές και οι Τελεστές Συνόλων που υποστηρίζονται από την XQL

- Τελεστές Ομαδοποίησης (Grouping Operators), που επιτρέπουν την ομαδοποίηση των αποτελεσμάτων των ερωτήσεων. Ο τελεστής ομαδοποίησης που υποστηρίζεται από την XQL είναι ο τελεστής «{}». Για την ομαδοποίηση των αποτελεσμάτων της ερώτησης «Ερώτηση» βάσει της τιμής του όρου «Όρος» ακολουθείται η σύνταξη που φαίνεται στο Παράδειγμα 3-9.

«Όρος» {«Ερώτηση»}

Παράδειγμα 3-9: Παράδειγμα Ομαδοποίησης Αποτελεσμάτων Ερώτησης

- Τελεστές Διάταξης (Sequence Operators), που επιτρέπουν την έκφραση κριτηρίων που σχετίζονται με τη διάταξη των κόμβων που παίρνουν μέρος στην ερώτηση. Ο Πίνακας 3-13 περιέχει τους τελεστές διάταξης που υποστηρίζονται από την XQL.

Όνομα	Σύνταξη	Περιγραφή
Πριν (Before)	«Έκφραση 1» before «Έκφραση 2»	Όλα τα στιγμιότυπα της έκφρασης «Έκφραση 1» που ακολουθούνται από κάποιο στιγμιότυπο της έκφρασης «Έκφραση 2»
Μετά (After)	«Έκφραση 1» after «Έκφραση 2»	Όλα τα στιγμιότυπα της έκφρασης «Έκφραση 1» που ακολουθούν κάποιο στιγμιότυπο της έκφρασης «Έκφραση 2»
Συνένωση Λιστών (List Concatenation)	«Έκφραση 1», «Έκφραση 2»	Όλα τα στιγμιότυπα της έκφρασης «Έκφραση 1» ακολουθούμενα από όλα τα στιγμιότυπα της έκφρασης «Έκφραση 2»

Πίνακας 3-13: Οι Τελεστές Διάταξης που υποστηρίζονται από την XQL

- Συναρτήσεις (Functions), που ενισχύουν τη λειτουργικότητα της XQL. Με την υποστήριξη μάλιστα συναρτήσεων οριζόμενων από τους χρήστες (user defined functions) η XQL αποκτά το χαρακτηριστικό της επεκτασιμότητας. Η συνάρτηση με όνομα «Συνάρτηση», που δέχεται το όρισμα «Όρισμα 1» καλείται όπως φαίνεται στο Παράδειγμα 3-10

«Συνάρτηση» («Όρισμα 1»)

Παράδειγμα 3-10: Παράδειγμα Κλήσης Συνάρτησης

- Αναφορές σε διευθύνσεις και κριτήρια σχετικά με το περιεχόμενο και τη δομή των αναφερόμενων αντικειμένων. Η ακριβής σύνταξη για αυτή τη λειτουργικότητα δεν είναι ακόμα διαθέσιμη.
- Συνδυασμούς (Joins), που επιτρέπουν το συνδυασμό πληροφορίας από πολλαπλές πηγές με σκοπό τη δημιουργία μιας ενοποιημένης άποψης.
- Τελεστές Μετονομασίας (Renaming Operators), που επιτρέπουν την αναφορά στα αποτελέσματα μιας ερώτησης (ή υποερώτησης) με τη χρήση κάποιου ονόματος. Ο τελεστής μετονομασίας που υποστηρίζεται από την XQL είναι ο τελεστής «->».

Στο σημείο αυτό πρέπει να σημειωθεί ότι σε οποιοδήποτε σημείο των XQL εκφράσεων μπορούν να χρησιμοποιηθούν χαρακτήρες – μπαλαντέρ (wildcards), και συγκεκριμένα ο χαρακτήρας «*» για την αναπαράσταση οποιουδήποτε συνδυασμού χαρακτήρων και ο χαρακτήρας «?» για την αναπαράσταση οποιουδήποτε χαρακτήρα.

Τέλος, πρέπει να σημειωθεί ότι η σύνταξη της XQL, σε αντίθεση με αυτή της XML-QL, δε μοιάζει με SQL. Η ερώτηση που αντιστοιχεί, σε XQL μορφή, στην ερώτηση που φαίνεται στο Παράδειγμα 3-8, φαίνεται στο Παράδειγμα 3-11.

memo[memo/title contains WWW]

Παράδειγμα 3-11: XQL Ερώτηση για την επιλογή των σημειωμάτων που στον τίτλο τους περιέχεται η συμβολοσειρά WWW

3.6. Ανακεφαλαίωση

Στο κεφάλαιο αυτό έγινε μια αναφορά στη συσχετιζόμενη με την παρούσα διατριβή εργασία. Πιο συγκεκριμένα, περιγράφηκαν τόσο η *UML (Unified Modeling Language)*, η οποία χρησιμοποιήθηκε για την περιγραφή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» και των τμημάτων από τα οποία αυτό απαρτίζεται, όσο και τα πρότυπα που καθοδήγησαν το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» (*Πρότυπα Αναπαράστασης Εγγράφων, Πρότυπα Αναπαράστασης Μεταδεδομένων, Document Object Model, Γλώσσες Διαχείρισης & Αναζήτησης Εγγράφων*) και ενσωματώθηκαν σε αυτόν.

4^ο Κεφάλαιο

Αρχιτεκτονική του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»

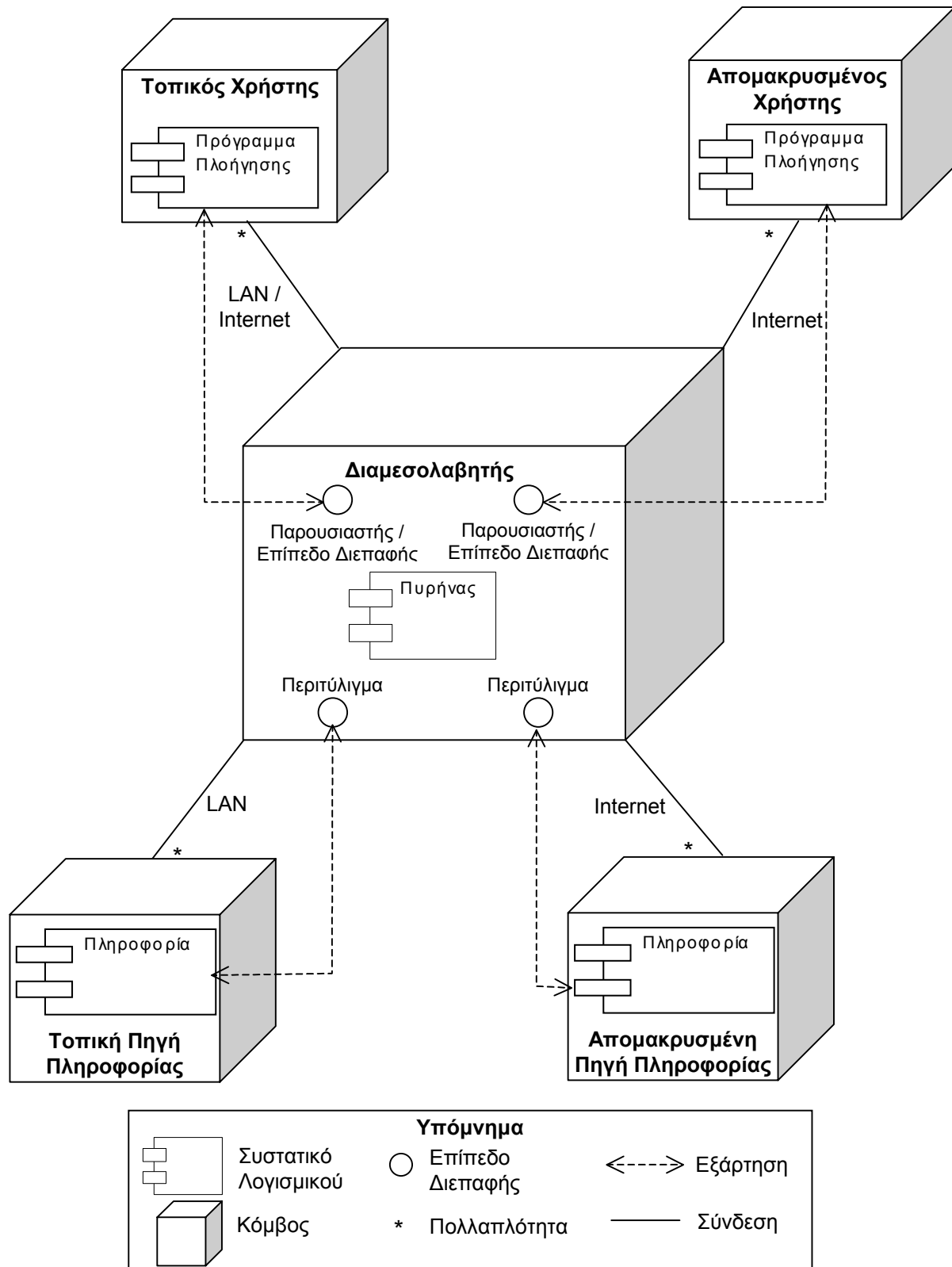
Το θέμα αυτού του κεφαλαίου είναι η αρχιτεκτονική του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Πιο συγκεκριμένα, στο τμήμα 4.1 δίνεται μια *Γενική Περιγραφή της Αρχιτεκτονικής του Συστήματος*. Τα *Τμήματα της Αρχιτεκτονικής του Συστήματος* περιγράφονται σύντομα στο τμήμα 4.2. Στο τμήμα 4.3 περιγράφεται η *Λειτουργικότητα του Συστήματος* και το κεφάλαιο κλείνει με μια σύντομη *Ανακεφαλαίωση* στο τμήμα 4.4.

4.1. Γενική Περιγραφή της Αρχιτεκτονικής του Συστήματος

Σ' αυτό το τμήμα γίνεται μια προσπάθεια γενικής περιγραφής της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Μια πρώτη απεικόνιση της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» δόθηκε στο *Σχήμα 1-1* και στο *Διάγραμμα 3-1*. Στο *Διάγραμμα 4-1* επιχειρείται μια λεπτομερέστερη απεικόνιση της αρχιτεκτονικής του συστήματος με τη χρήση ενός UML διαγράμματος παράταξης.

Στο *Διάγραμμα 4-1* φαίνεται ότι το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» απαρτίζεται από ένα αριθμό πηγών πληροφορίας και από ένα διαμεσολαβητή. Οι πηγές πληροφορίας μπορεί να είναι παραδοσιακές βάσεις δεδομένων, πληροφοριακά συστήματα

πολυμέσων, πηγές ημι-δομημένης ή αδόμητης πληροφορίας κ.α. Ο διαμεσολαβητής, ο οποίος αποτελεί το σύστημα διεπαφής χρήστη – πηγών πληροφορίας, είναι η «καρδιά» του συστήματος καθώς είναι υπεύθυνος για:



Διάγραμμα 4-1: UML Διάγραμμα Παράταξης του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»

- Την επικοινωνία με τους (τοπικούς και απομακρυσμένους) χρήστες.

- Τη μεταβίβαση ερωτήσεων και αιτήσεων από τους χρήστες στις πηγές πληροφορίας.
- Τη μεταφορά πληροφορίας από τις πηγές πληροφορίας στους χρήστες.

Ο σχεδιασμός τόσο της συνολικής αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» όσο και των τμημάτων από τα οποία αυτή απαρτίζεται, βασίστηκε στο συνδυασμό των τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό (XML, RDF, XQL κλπ.).

Μια εκτενέστερη αναφορά στα επιμέρους τμήματα που απαρτίζουν την αρχιτεκτονική του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» γίνεται στο τμήμα 4.2.

4.2. Τμήματα της Αρχιτεκτονικής του Συστήματος

Στο τμήμα αυτό θα γίνει μια σύντομη περιγραφή των τμημάτων της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Τα τμήματα που απαρτίζουν την αρχιτεκτονική του συστήματος είναι τα εξής:

- Ο διαμεσολαβητής, που αποτελεί το επίπεδο διεπαφής χρηστών-συστήματος. Ο διαμεσολαβητής είναι υπεύθυνος για:
 - Την υποστήριξη του γραφικού επιπέδου διεπαφής των χρηστών (graphical user interface) του συστήματος. Αυτό περιλαμβάνει:
 - Τη μετάφραση των ερωτήσεων και των αιτήσεων που εκφράζονται με γραφικό τρόπο από τους χρήστες στη γλώσσα του διαμεσολαβητή.
 - Την παρουσίαση στους χρήστες ενός μοντέλου πληροφορίας που καλύπτει τα ενδιαφέροντα και τις προτιμήσεις τους όπως, αυτά εκφράζονται στα διαγράμματα προτιμήσεών τους.
 - Την παρουσίαση στους χρήστες ενός επιπέδου διεπαφής προσαρμοσμένου στις προτιμήσεις τους και τις δυνατότητες των υπολογιστών τους (π.χ. παρουσίαση λιγότερων εικόνων στους χρήστες που η σύνδεσή τους με το Διαδίκτυο είναι αργή, μη παρουσίαση ήχων στους χρήστες που δε διαθέτουν κάρτα ήχου κλπ.).
 - Τη μετατροπή των αποτελεσμάτων των ερωτήσεων από μορφή συμβατή με τη γλώσσα του διαμεσολαβητή σε μορφή κατάλληλη για παρουσίαση από την εφαρμογή που αποτελεί το σύστημα διεπαφής του χρήστη.
 - Την αποστολή (μαζικά ή τμηματικά) των αποτελεσμάτων των ερωτήσεων και των αιτήσεων των χρηστών στις εφαρμογές που αυτοί χρησιμοποιούν για την επικοινωνία τους με το διαμεσολαβητή.
 - Τη μετάφραση των ερωτήσεων και των αιτήσεων από τη γλώσσα του διαμεσολαβητή στις γλώσσες των πηγών πληροφορίας.
 - Τη μετάφραση των απαντήσεων στις ερωτήσεις και τις αιτήσεις από τη μορφή που υποστηρίζει η κάθε πηγή πληροφορίας σε μια κοινή μορφή, συμβατή με τη γλώσσα του διαμεσολαβητή.

- Τη μετάφραση, αν είναι απαραίτητη, των όρων των ερωτήσεων από τις γλώσσες στις οποίες εκφράζονται από τους χρήστες στη βασική γλώσσα του συστήματος.
- Τη διαχείριση των διαγραμμάτων προτιμήσεων των χρηστών και την παροχή σ' αυτούς των υπηρεσιών που ζητούνται κάθε φορά με τρόπο σύμφωνο με τις προτιμήσεις τους.
- Τη διαχείριση της μεταπληροφορίας που σχετίζεται με τη δομή και το περιεχόμενο των πηγών πληροφορίας και την πληροφορία που αποθηκεύεται σ' αυτές.

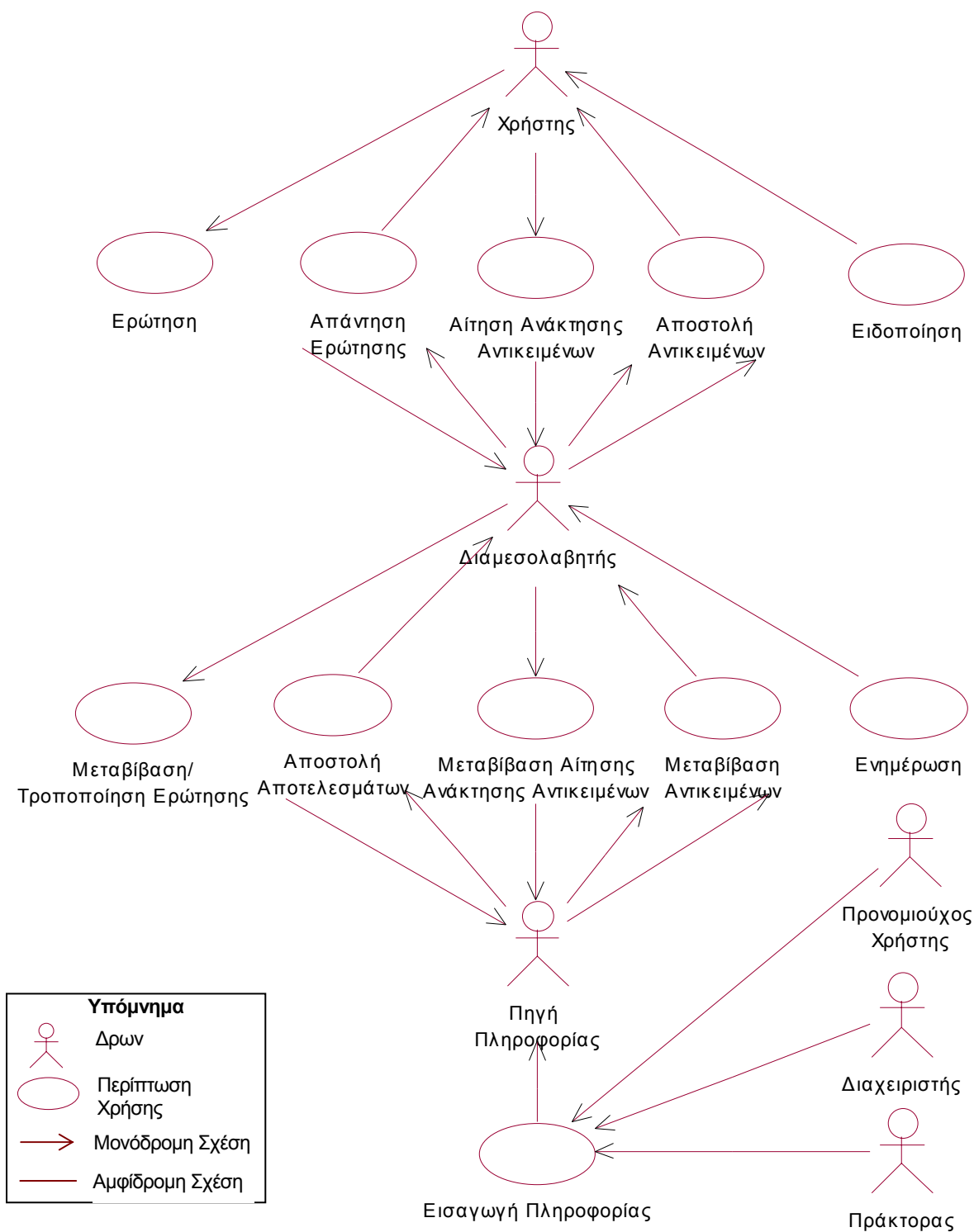
Στο σημείο αυτό πρέπει να αναφερθεί ότι η μετάφραση πληροφορίας που λαμβάνει χώρα στο διαμεσολαβητή από και προς τις πηγές πληροφορίας γίνεται μέσω των περιτυλιγμάτων (wrappers) τους, ενώ η αλληλεπίδραση του διαμεσολαβητή με τους χρήστες υποστηρίζεται από τους παρουσιαστές (presentators) και το επίπεδο διεπαφής (interface) του διαμεσολαβητή. Λεπτομερέστερη αναφορά στο διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» γίνεται στο 5^ο κεφάλαιο.

- Οι πηγές πληροφορίας, που είναι οργανωμένες συλλογές δεδομένων (εγγράφων, δομημένων δεδομένων που οργανώνονται σε βάσεις δεδομένων, αντικειμένων πολυμέσων, ημι-δομημένων ή αδόμητων αντικειμένων κλπ.). Πρόσβαση στα περιεχόμενα μιας πηγής πληροφορίας παρέχεται από τη γλώσσα της πηγής πληροφορίας. Έτσι, π.χ., η γλώσσα μιας πηγής πληροφορίας που υλοποιείται από μια σχεσιακή βάση δεδομένων είναι η SQL, μιας πηγής πληροφορίας που υλοποιείται από μια αντικειμενοστραφή βάση δεδομένων είναι η OQL κλπ. Μέσω της γλώσσας μιας πηγής πληροφορίας μπορούν να γίνουν οι εξής κατηγορίες λειτουργιών:
 - Λειτουργίες που δεν επηρεάζουν τη δομή και το περιεχόμενο των πηγών πληροφορίας και διατίθενται μέσω του διαμεσολαβητή στους χρήστες. Στην κατηγορία αυτή εντάσσονται οι εξής λειτουργίες:
 - Αναζήτηση στιγμιοτύπων των υπαρχουσών κλάσεων που πληρούν συγκεκριμένα κριτήρια.
 - Ανάκτηση στιγμιοτύπων που αποθηκεύονται στις πηγές πληροφορίας.
 - Λειτουργίες που επηρεάζουν τη δομή και το περιεχόμενο των πηγών πληροφορίας. Οι λειτουργίες αυτές εκτελούνται από «έξυπνα» εργαλεία που αυτοματοποιούν την εκτέλεσή τους (όπως είναι οι πράκτορες), από τους διαχειριστές των πηγών πληροφορίας και από προνομιούχους χρήστες, που έχουν το δικαίωμα μετατροπής της δομής και/ή του περιεχομένου των πηγών πληροφορίας. Στην κατηγορία αυτή εντάσσονται οι εξής λειτουργίες:
 - Ορισμός νέων κλάσεων, που υποστηρίζονται από την πηγή πληροφορίας (π.χ. νέοι τύποι εγγράφων για πηγές πληροφορίας που περιέχουν έγγραφα).
 - Εισαγωγή νέων στιγμιοτύπων των υπαρχουσών κλάσεων.
 - Δεικτοδότηση, μέσα στις πηγές πληροφορίας, των αντικειμένων που γίνονται διαθέσιμα και σχετίζονται με το περιεχόμενο των πηγών πληροφορίας.

Μια λεπτομερέστερη περιγραφή των πηγών πληροφορίας γίνεται στο 6^ο κεφάλαιο, ενώ τα εργαλεία διαχείρισης των πηγών πληροφορίας περιγράφονται στο 7^ο κεφάλαιο.

4.3. Λειτουργικότητα του Συστήματος

Το θέμα αυτού του τμήματος είναι η περιγραφή της λειτουργικότητας του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Καθώς για την περιγραφή του συστήματος χρησιμοποιείται η UML, γίνεται προσπάθεια η λειτουργικότητά του να αποδοθεί με τη χρήση ενός UML Διαγράμματος Περιπτώσεων Χρήσης στο Διάγραμμα 4-2.



Διάγραμμα 4-2: Διάγραμμα Περιπτώσεων Χρήσης του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»

Όπως φαίνεται στο Διάγραμμα 4-2, οι χρήστες χρησιμοποιούν το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» στις εξής περιπτώσεις:

1. Όταν εκφράζουν μια ερώτηση (συνήθως μέσω ενός γραφικού συστήματος διεπαφής). Όπως φαίνεται στο Διάγραμμα 4-2, στην περίπτωση χρήσης «Ερώτηση» ο χρήστης αλληλεπιδρά με το διαμεσολαβητή. Πιο συγκεκριμένα, η ερώτηση που όρισε ο χρήστης με γραφικό ή άλλο τρόπο εκφράζεται στη γλώσσα του διαμεσολαβητή. Στη συνέχεια, ο διαμεσολαβητής στέλνει την ερώτηση στις πηγές πληροφορίας μέσω των *Περιτυλιγμάτων* (*Wrappers*) τους, τα οποία είναι υπεύθυνα για τη μετάφραση της ερώτησης από τη γλώσσα του διαμεσολαβητή στη γλώσσα της πηγής πληροφορίας (περίπτωση χρήσης «Μεταβίβαση / Τροποποίηση Ερώτησης»).

Οι πηγές πληροφορίας πραγματοποιούν την αναζήτηση στα πληροφοριακά συστήματα που διαθέτουν και επιστρέφουν τα αποτελέσματα στο διαμεσολαβητή (περίπτωση χρήσης «Αποστολή Αποτελέσματος Ερώτησης»).

Όταν όλες οι πηγές πληροφορίας επιστρέψουν τα αποτελέσματα της ερώτησης, ο διαμεσολαβητής τα συνδυάζει, κανονικοποιεί τα βάρη τους (αν υπάρχουν) και τα στέλνει στο χρήστη μέσω των *Παρουσιαστών* (*Presentators*) διαμορφωμένα σύμφωνα με τον τρόπο που περιγράφεται στο διάγραμμα προτιμήσεων του τελευταίου (περίπτωση χρήσης «Απάντηση Ερώτησης»).

2. Όταν υποβάλλουν αίτηση ανάκτησης κάποιων αντικειμένων (που μπορεί να είναι, για παράδειγμα, τμήμα της απάντησης μιας ερώτησης). Όπως φαίνεται στο Διάγραμμα 4-2, στην περίπτωση χρήσης «Αίτηση Ανάκτησης Αντικειμένων», ο χρήστης αλληλεπιδρά με το διαμεσολαβητή (συνήθως μέσω της διάσχισης κάποιου συνδέσμου υπερκειμένου αν πρόκειται για ένα αντικείμενο ή του πατήματος ενός κουμπιού αν πρόκειται για περισσότερα από ένα αντικείμενα). Η αίτηση μεταφράζεται στη γλώσσα του διαμεσολαβητή και προκύπτουν οι πηγές πληροφορίας στις οποίες πρέπει να μεταβιβαστούν τμήματα της αίτησης¹⁷ μετά τις απαραίτητες τροποποιήσεις που θα γίνουν από τα περιτυλίγματά τους (περίπτωση χρήσης «Μεταβίβαση Αίτησης Ανάκτησης Αντικειμένων»).

Στις πηγές πληροφορίας λαμβάνει χώρα ο εντοπισμός των αντικειμένων, τα οποία στη συνέχεια στέλνονται στο διαμεσολαβητή, όπου μετατρέπονται σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή μέσω των περιτυλιγμάτων των πηγών πληροφορίας. Στη συνέχεια, ο διαμεσολαβητής στέλνει τα αντικείμενα στο χρήστη μορφοποιημένα από τον κατάλληλο παρουσιαστή σύμφωνα με τις προτιμήσεις που καταγράφονται στο διάγραμμα προτιμήσεων του χρήστη (περίπτωση χρήσης «Αποστολή Αντικειμένων»).

3. Όταν γίνει διαθέσιμο κάποιο αντικείμενο, το περιεχόμενο του οποίου ταιριάζει με τα ενδιαφέροντα του χρήστη. Στην περίπτωση αυτή το αντικείμενο εισάγεται σε κάποια πηγή πληροφορίας από το διαχειριστή (*administrator*) της, κάποιο προνομιούχο χρήστη που έχει το δικαίωμα εισαγωγής πληροφορίας στην πηγή πληροφορίας ή από κάποιο πράκτορα (*agent*) που είναι υπεύθυνος για την ενημέρωση της πηγής πληροφορίας για

¹⁷ Καθώς τα αιτούμενα αντικείμενα μπορεί να αποθηκεύονται σε περισσότερες από μια πηγές πληροφορίας, σε καθεμιά από αυτές αποστέλλεται το τμήμα της αίτησης που αντιστοιχεί στα αντικείμενα πληροφορίας που αποθηκεύονται σ' αυτήν.

νέα αντικείμενα που ταιριάζουν με το περιεχόμενο της πηγής πληροφορίας και γίνονται διαθέσιμα (περίπτωση χρήσης «Εισαγωγή Πληροφορίας»).

Στη συνέχεια, ο διαμεσολαβητής ενημερώνεται από την πηγή πληροφορίας (περίπτωση χρήσης «Ενημέρωση»), ο οποίος και ειδοποιεί τους ενδιαφερόμενους χρήστες με την αποστολή του κατάλληλου μηνύματος ειδοποίησης (περίπτωση χρήσης «Ειδοποίηση»).

Η περιγραφή της λειτουργικότητας του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» που έγινε σ' αυτό το τμήμα σκοπό είχε να δώσει μια γενική εικόνα του συστήματος. Εκτενέστερη και λεπτομερέστερη περιγραφή της λειτουργικότητας του διαμεσολαβητή και των πηγών πληροφορίας δίνεται, αντίστοιχα, στο 5^ο και το 6^ο κεφάλαιο.

4.4. Ανακεφαλαίωση

Στο κεφάλαιο αυτό έγινε μια γενική περιγραφή της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», δόθηκε μια σύντομη περιγραφή του διαμεσολαβητή και των πηγών πληροφορίας, που είναι τα βασικά τμήματα που την απαρτίζουν και περιγράφηκε η λειτουργικότητα του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Λεπτομερέστερη περιγραφή του διαμεσολαβητή και των πηγών πληροφορίας του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» δίνεται, αντίστοιχα, στο 5^ο και το 6^ο κεφάλαιο.

5^ο Κεφάλαιο

Διαμεσολαβητής

Στο κεφάλαιο αυτό περιγράφεται ο Διαμεσολαβητής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Όπως έχει αναφερθεί στα προηγούμενα κεφάλαια, ο Διαμεσολαβητής είναι η καρδιά του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»: Ο Διαμεσολαβητής αποτελεί το σύστημα διεπαφής των χρηστών με τις πηγές πληροφορίας και παρέχει σ' αυτούς ομοιόμορφη και διαφανή πρόσβαση στα περιεχόμενα των ετερογενών πηγών πληροφορίας των οποίων τα περιεχόμενα γίνονται διαθέσιμα μέσω αυτού.

Στο τμήμα 5.1 καθορίζεται η *Αρχιτεκτονική του Διαμεσολαβητή*, ενώ στο τμήμα 5.2 περιγράφονται τα *Τμήματα της Αρχιτεκτονικής του Διαμεσολαβητή*. Τα τμήματα της αρχιτεκτονικής του Διαμεσολαβητή είναι τα εξής:

1. Το επίπεδο διεπαφής του διαμεσολαβητή, που περιγράφεται στην ενότητα 5.2.1 και είναι υπεύθυνο για τη μετατροπή των ερωτήσεων και των αιτήσεων που υποβάλλουν οι χρήστες στη γλώσσα του διαμεσολαβητή μέσω των (συνήθως γραφικών) επιπέδων διεπαφής που χρησιμοποιούν.
 2. Οι παρουσιαστές, που περιγράφονται στην ενότητα 5.2.2 και είναι υπεύθυνοι για τη μορφοποίηση, με σκοπό τη μετέπειτα παρουσίαση στους χρήστες, της πληροφορίας που παίρνουν σαν είσοδο σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή.
 3. Ο πυρήνας του διαμεσολαβητή, που περιγράφεται στην ενότητα 5.2.3 και είναι υπεύθυνος για:
-

- Την προσαρμογή του περιβάλλοντος εργασίας των χρηστών ανάλογα με τις προτιμήσεις τους και την υπολογιστική πλατφόρμα που χρησιμοποιούν.
 - Την υποστήριξη πολύγλωσσου περιεχόμενου.
 - Την παροχή ταυτόχρονης, διαφανούς και ομοιόμορφης πρόσβασης σε ετερογενείς πηγές πληροφορίας.
4. Τα περιτυλίγματα των πηγών πληροφορίας, που περιγράφονται στην ενότητα 5.2.4 και είναι υπεύθυνα για τη μετάφραση των ερωτήσεων και των αιτήσεων που είναι εκφρασμένες στη γλώσσα του διαμεσολαβητή σε ερωτήσεις που εκφράζονται στις γλώσσες των αντίστοιχων πηγών πληροφορίας και τη μετατροπή της πληροφορίας που είναι εκφρασμένη σε μορφή συμβατή με τις γλώσσες των πηγών πληροφορίας σε μορφή που είναι συμβατή με τη γλώσσα του διαμεσολαβητή..

Ο σχεδιασμός της αρχιτεκτονικής του διαμεσολαβητή, όπως και όλων των τμημάτων του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», βασίστηκε στο συνδυασμό των τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό (XML, RDF, XQL κλπ.).

Στο τμήμα 5.3 περιγράφεται η *Λειτουργικότητα του Διαμεσολαβητή* που συνοψίζεται στην αλληλεπίδραση με τους χρήστες, την παροχή πολύγλωσσας υποστήριξης, την αλληλεπίδραση με τις πηγές πληροφορίας, την εξατομίκευση των παρεχόμενων στους χρήστες υπηρεσιών και τη διαχείριση μεταδεδομένων.

Βασικό ρόλο στη λειτουργία του Διαμεσολαβητή παίζει η *Γλώσσα του Διαμεσολαβητή*, που περιγράφεται στο τμήμα 5.4. Οι αιτήσεις και οι ερωτήσεις των χρηστών του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» μεταφράζονται στη γλώσσα του διαμεσολαβητή πριν σταλούν στις πηγές πληροφορίας μέσω των περιτυλιγμάτων τους. Επιπλέον, τα αποτελέσματα των ερωτήσεων και των αιτήσεων που επιστρέφονται από τις πηγές πληροφορίας, πριν μορφοποιηθούν για παρουσίαση από τους παρουσιαστές, μεταφράζονται από τα περιτυλίγματα των πηγών πληροφορίας σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή. Η γλώσσα του Διαμεσολαβητή, που αποτελεί συνεισφορά της παρούσας διατριβής, έχει τα εξής χαρακτηριστικά:

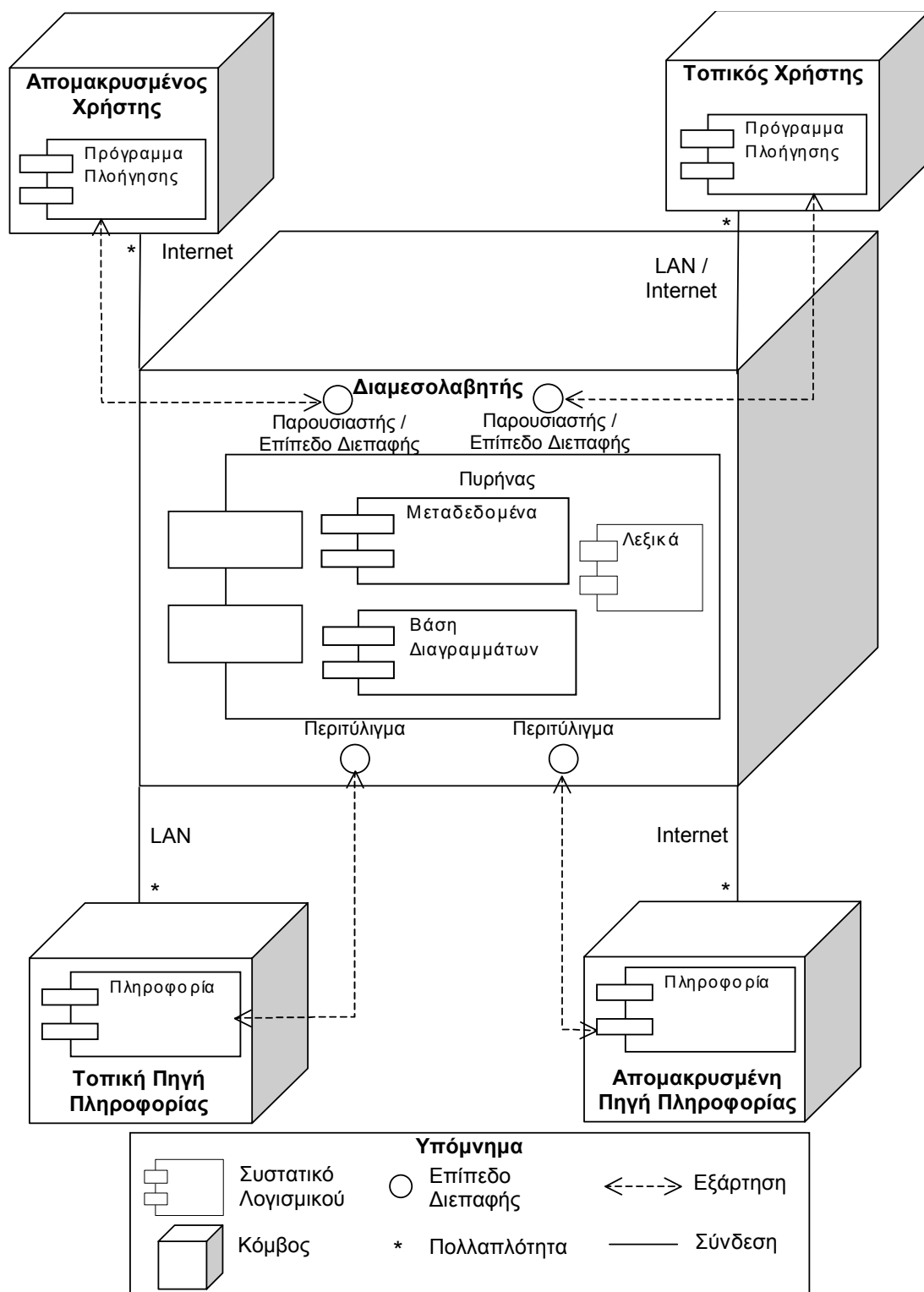
- Διαχειρίζεται ένα μοντέλο πληροφορίας που περιγράφεται από ένα RDF σχήμα, το οποίο είναι στην πραγματικότητα ένα επεκταμένο E-R διάγραμμα.
- Μοιάζει με την SQL (είναι δηλαδή SQL-like).
- Επιτρέπει την έκφραση ερωτήσεων ομοιότητας.
- Επιστρέφει τα αποτελέσματα των ερωτήσεων και των αιτήσεων σε XML μορφή.

Σημαντικό ρόλο στη λειτουργία του Διαμεσολαβητή παίζει και η *Βάση Δεδομένων του Διαμεσολαβητή*, που σχεδιάστηκε στα πλαίσια της παρούσας διατριβής. Η βάση δεδομένων του Διαμεσολαβητή, που περιγράφεται στο τμήμα 5.5, είναι απαραίτητη για την αποδοτική λειτουργία του πυρήνα του Διαμεσολαβητή, καθώς περιέχει τα διαγράμματα όπου καταγράφονται οι προτιμήσεις των χρηστών, τις δομές που χρησιμοποιούνται για την παροχή πολύγλωσσας υποστήριξης και τη σχετική με τις πηγές πληροφορίας μεταπληροφορία.

Τέλος, το κεφάλαιο κλείνει με μια σύντομη *Ανακεφαλαίωση* στο τμήμα 5.6.

5.1. Αρχιτεκτονική του Διαμεσολαβητή

Το θέμα αυτού του τμήματος είναι η αρχιτεκτονική του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Η αρχιτεκτονική του διαμεσολαβητή απεικονίζεται, με τη μορφή διαγράμματος παράταξης, στο *Διάγραμμα 5-1*.



Διάγραμμα 5-1: Διάγραμμα Παράταξης για την Αρχιτεκτονική του Διαμεσολαβητή

Όπως φαίνεται στο Διάγραμμα 5-1, ο διαμεσολαβητής απαρτίζεται από:

- Ένα επίπεδο διεπαφής με τους χρήστες (interface), όπου οι αιτήσεις και οι ερωτήσεις των χρηστών μετατρέπονται στη γλώσσα του διαμεσολαβητή.
- Ένα σύνολο παρουσιαστών (presentators), οι οποίοι είναι υπεύθυνοι για τη μορφοποίηση της πληροφορίας με σκοπό την παρουσίαση στους χρήστες, με τρόπο που καθορίζεται για κάθε χρήστη ξεχωριστά, ανάλογα με τις προτιμήσεις του, την υπολογιστική πλατφόρμα και το λογισμικό που χρησιμοποιεί. Επίσης, όταν κάποιοι χρήστες ειδοποιούνται για το ότι γίνεται διαθέσιμη πληροφορία που πιθανώς τους ενδιαφέρει, μέσω των παρουσιαστών στέλνεται το κατάλληλο μήνυμα στο (συνήθως γραφικό) επίπεδο διεπαφής των χρηστών.
- Ένα σύνολο περιτυλιγμάτων, που χρησιμοποιούνται για τη μετατροπή των αιτήσεων, των ερωτήσεων και των δεδομένων από μορφές συμβατές με τη γλώσσα του διαμεσολαβητή σε μορφές συμβατές με τις γλώσσες των πηγών πληροφορίας και αντιστρόφως.
- Ένα πυρήνα, που αποτελεί τη βάση στην οποία στηρίζονται και λειτουργούν τόσο οι παρουσιαστές όσο και τα περιτυλίγματα. Πιο συγκεκριμένα, ο πυρήνας του διαμεσολαβητή απαρτίζεται από:
 - Τη βάση που περιέχει τα διαγράμματα προτιμήσεων των χρηστών του συστήματος (user profiles).
 - Τα λεξικά και τους θησαυρούς που χρησιμοποιούνται για την υποστήριξη πολύγλωσσων αναζητήσεων.
 - Τα σχετικά με τις πηγές πληροφορίας μεταδεδομένα.

Τα τμήματα που απαρτίζουν το διαμεσολαβητή περιγράφονται εκτενέστερα στο τμήμα 5.2. Ο σχεδιασμός της αρχιτεκτονικής του διαμεσολαβητή, όπως και όλων των τμημάτων του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», βασίστηκε στο συνδυασμό των τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό.

5.2. Τμήματα της Αρχιτεκτονικής του Διαμεσολαβητή

Στο τμήμα αυτό περιγράφονται τα τμήματα που απαρτίζουν την αρχιτεκτονική του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Όπως αναφέρθηκε ήδη στο τμήμα 5.1, ο διαμεσολαβητής απαρτίζεται από τα εξής 4 τμήματα:

1. Το επίπεδο διεπαφής του διαμεσολαβητή, που περιγράφεται στην ενότητα 5.2.1.
2. Τους παρουσιαστές, που περιγράφονται στην ενότητα 5.2.2.
3. Τον πυρήνα του διαμεσολαβητή, που περιγράφεται στην ενότητα 5.2.3.
4. Τα περιτυλίγματα των πηγών πληροφορίας, που περιγράφονται στην ενότητα 5.2.4.

5.2.1. Επίπεδο Διεπαφής του Διαμεσολαβητή

Σ' αυτή την ενότητα γίνεται αναφορά στο επίπεδο διεπαφής του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» με τους χρήστες. Η λειτουργικότητα του επιπέδου διεπαφής συνοψίζεται στη μετατροπή των ερωτήσεων και των αιτήσεων που υποβάλλουν οι χρήστες στη γλώσσα του διαμεσολαβητή μέσω των (συνήθως γραφικών) επιπέδων διεπαφής που χρησιμοποιούν.

Πιο συγκεκριμένα, το επίπεδο διεπαφής του διαμεσολαβητή δέχεται ως είσοδο τις επιλογές του χρήστη (που είναι συνήθως μια URL κωδικοποιημένη συμβολοσειρά σε προσυμφωνημένη, ανάλογα με την εφαρμογή που χρησιμοποιεί ο χρήστης κάθε φορά, μορφή) και τη μετατρέπει σε ερώτηση σύμφωνη με τη σύνταξη της γλώσσας του διαμεσολαβητή.

Όπως θα φανεί και στη συνέχεια, το επίπεδο διεπαφής του διαμεσολαβητή, σε συνδυασμό με τους παρουσιαστές, κάνει δυνατή την αλληλεπίδραση των χρηστών με το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση».

5.2.2. Παρουσιαστές

Το θέμα αυτής της ενότητας είναι οι παρουσιαστές. Ο ρόλος των παρουσιαστών συνίσταται στη μορφοποίηση, με σκοπό την παρουσίαση, της πληροφορίας που παίρνουν σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή. Η πληροφορία αυτή μπορεί να είναι τα αποτελέσματα κάποιας ερώτησης, τα αντικείμενα που επιστρέφονται μετά την εκτέλεση κάποιας αίτησης ή η αποστολή σε μια ομάδα χρηστών ενός μηνύματος ειδοποίησης ότι γίνεται διαθέσιμη κάποια πληροφορία που πιθανώς τους ενδιαφέρει.

Ο τρόπος μορφοποίησης της πληροφορίας εξαρτάται από τις προτιμήσεις και την υπολογιστική πλατφόρμα του χρήστη στον οποίο θα παρουσιαστεί η πληροφορία. Αυτό επιτυγχάνεται με την αλληλεπίδραση του παρουσιαστή με τα διαγράμματα προτιμήσεων των χρηστών για την ανάκτηση από αυτά των παραμέτρων μορφοποίησης της πληροφορίας. Για το σκοπό αυτό, η συσχέτιση των περιεχομένων των διαγραμμάτων προτιμήσεων των χρηστών με τις ενέργειες που εκτελούνται από τους παρουσιαστές εκφράζεται μέσω της σύνταξης της XSL.

Οι παρουσιαστές στο διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», σε πρώτη φάση τουλάχιστον, δέχονται XML πληροφορία ως είσοδο και παράγουν ως έξοδο μορφοποιημένη για παρουσίαση πληροφορία σε μορφή που είναι δυνατό να διαχειριστούν τα προγράμματα που χρησιμοποιούν οι χρήστες. Καθώς η XSL προσφέρεται για τον προσδιορισμό της μορφοποίησης XML πληροφορίας, ανάλογα με τις προτιμήσεις του χρήστη δημιουργείται ένα ιδεατό XSL style sheet, που προσδιορίζει τη μορφοποίηση της πληροφορίας που πρόκειται να παρουσιαστεί. Έτσι, η έξοδος του παρουσιαστή είναι η XML πληροφορία που πρόκειται να παρουσιαστεί μαζί με την XSL πληροφορία που καθορίζει τη μορφοποίησή της. Στο σημείο αυτό πρέπει να σημειωθεί ότι προγράμματα που διαθέτουν XSL επεξεργαστές (όπως είναι τα προγράμματα πλοήγησης του Παγκοσμίου Ιστού) μπορούν να παρουσιάσουν XML πληροφορία που συνδυάζεται με XSL σε μη-XML μορφή (π.χ. HTML).

Εκτός από την παροχή της απαραίτητης πληροφορίας μορφοποίησης, οι παρουσιαστές είναι υπεύθυνοι και για την ενσωμάτωση στην προς παρουσίαση πληροφορία των μεταδεδομένων που περιγράφουν τα αντικείμενα που πρόκειται να παρουσιαστούν στους χρήστες. Έτσι, τα μεταδεδομένα που περιγράφουν τα αντικείμενα που θα παρουσιαστούν στους χρήστες παρέχονται από τους παρουσιαστές σε RDF μορφή, ενσωματωμένα στην XML μορφή των αντικειμένων. Τα μεταδεδομένα αυτά περιλαμβάνουν μεταπληροφορία που μπορεί να σχετίζεται με τα ίδια τα αντικείμενα (όπως είναι, π.χ., τα Dublin Core στοιχεία) ή/και μεταπληροφορία που σχετίζεται με τις πηγές πληροφορίας από τις οποίες τα αντικείμενα προέρχονται. Η διαδικασία ενσωμάτωσης των μεταδεδομένων είναι απλή: για κάθε στοιχείο μεταδεδομένων που δεν είναι ήδη σε RDF μορφή, το στοιχείο αυτό μετατρέπεται σε RDF μορφή. Στη συνέχεια, όλα μαζί τα στοιχεία μεταδεδομένων σε RDF μορφή συνθέτουν ένα ιδεατό RDF έγγραφο που συνδέεται με το ιδεατό XML έγγραφο που αναπαριστά το προς παρουσίαση αντικείμενο και στέλνονται στο χρήστη για παρουσίαση.

5.2.3. Ο Πυρήνας του Διαμεσολαβητή

Σ' αυτή την ενότητα περιγράφεται ο πυρήνας του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Όπως έχει ήδη αναφερθεί, ο πυρήνας του διαμεσολαβητή αποτελεί τη βάση στην οποία στηρίζονται και λειτουργούν τόσο οι παρουσιαστές όσο και τα περιτυλίγματα. Πιο συγκεκριμένα, ο πυρήνας παρέχει υποστήριξη σε ότι αφορά:

- Προσαρμογή του περιβάλλοντος εργασίας των χρηστών ανάλογα με τις προτιμήσεις τους και την υπολογιστική πλατφόρμα που χρησιμοποιούν. Η υποστήριξη αυτή παρέχεται μέσω της αποθήκευσης και της διαχείρισης διαγραμμάτων χρηστών (user profiles).
- Πολύγλωσσο περιεχόμενο. Ένας από τους στόχους του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» είναι η υποστήριξη τουλάχιστον 2 γλωσσών (σε πρώτη φάση Αγγλικών και Ελληνικών) και η σταδιακή υποστήριξη και άλλων γλωσσών, πράγμα που προϋποθέτει μια ανοικτή και εύκολα επεκτάσιμη αρχιτεκτονική. Η πολύγλωσση υποστήριξη που παρέχεται από το σύστημα περιλαμβάνει:
 - Την παροχή διαγλωσσικής ανάκτησης μέσω της υποστήριξης ερωτήσεων που επιστρέφουν όχι μόνο τα αντικείμενα που περιέχουν τους όρους των ερωτήσεων στη γλώσσα που αυτοί έχουν δοθεί ή συσχετίζονται με αυτούς, αλλά και τα αντικείμενα που περιέχουν ή συσχετίζονται με τους συνώνυμους τους όρους σε άλλες γλώσσες.
 - Την παροχή πρόσβασης σε θησαυρούς και λεξικά, που περιέχουν γενικεύσεις και εξειδικεύσεις όρων για τις υποστηριζόμενες γλώσσες. Οι θησαυροί και τα λεξικά χρησιμοποιούνται και κατά την επεξεργασία των ερωτήσεων.
- Ταυτόχρονη, διαφανή και ομοιόμορφη πρόσβαση σε ετερογενείς πηγές πληροφορίας. Αυτό επιτυγχάνεται με την αποθήκευση και τη διαχείριση μεταπληροφορίας σχετικής με καθεμιά από τις πηγές πληροφορίας που τα περιεχόμενά της γίνονται διαθέσιμα μέσω του διαμεσολαβητή.

Για να υποστηριχθεί η λειτουργικότητα που αναφέρθηκε παραπάνω, όπως φαίνεται στο Διάγραμμα 5-1, ο πυρήνας διαθέτει μια βάση διαγραμμάτων προτιμήσεων χρηστών, ένα σύνολο λεξικών και θησαυρών και μια βάση μεταπληροφορίας που σχετίζεται με τις πηγές πληροφορίας. Τα συστατικά αυτά αποθηκεύονται στη βάση δεδομένων του διαμεσολαβητή, η οποία περιγράφεται αναλυτικά στο τμήμα 5.5.

5.2.4. Περιτυλίγματα

Στην ενότητα αυτή συζητούνται τα περιτυλίγματα των πηγών πληροφορίας που προσπελαύνονται μέσω ενός διαμεσολαβητή. Όπως έχει ήδη αναφερθεί, τα περιτυλίγματα είναι υπεύθυνα για τη μετάφραση των ερωτήσεων και των αιτήσεων που είναι εκφρασμένες στη γλώσσα του διαμεσολαβητή σε ερωτήσεις που εκφράζονται στις γλώσσες των αντίστοιχων πηγών πληροφορίας και τη μετατροπή της πληροφορίας που είναι εκφρασμένη σε μορφή συμβατή με τις γλώσσες των πηγών πληροφορίας σε μορφή που είναι συμβατή με τη γλώσσα του διαμεσολαβητή.

Οι πηγές πληροφορίας, στα πλαίσια του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», εντάσσονται σε κλάσεις πηγών πληροφορίας. Οι πηγές πληροφορίας που ανήκουν στην ίδια κλάση διαθέτουν, εκτός των άλλων κοινών χαρακτηριστικών τους, κοινή γλώσσα και κοινό περιτύλιγμα. Έτσι, για κάθε κλάση πηγών πληροφορίας ορίζεται ένα σύνολο κανόνων για τη μετατροπή της πληροφορίας από σύνταξη συμβατή με τη γλώσσα του διαμεσολαβητή σε σύνταξη συμβατή με τη γλώσσα των στιγμιοτύπων της κλάσης της πηγών πληροφορίας και αντιστρόφως.

Καθώς στα πλαίσια της παρούσας διατριβής δίνεται έμφαση στην υποστήριξη πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα, ιδιαίτερη αναφορά στα περιτυλίγματα αυτών των κλάσεων πηγών πληροφορίας γίνεται στο 6^ο κεφάλαιο.

5.3. Λειτουργικότητα του Διαμεσολαβητή

Η λειτουργικότητα του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» είναι το θέμα αυτού του τμήματος. Η λειτουργικότητα αυτή μπορεί να συνοψιστεί στα εξής σημεία:

- *Αλληλεπίδραση με τους Χρήστες.* Η αλληλεπίδραση του διαμεσολαβητή με τους χρήστες υποστηρίζεται κυρίως από το επίπεδο διεπαφής του διαμεσολαβητή και τους παρουσιαστές. Η αλληλεπίδραση με τους χρήστες περιλαμβάνει, από την πλευρά των χρηστών, την έκφραση αιτήσεων και ερωτήσεων. Από την πλευρά του διαμεσολαβητή περιλαμβάνει την αποστολή στους χρήστες των απαντήσεων των αιτήσεων και των ερωτήσεων και την ειδοποίησή τους για αντικείμενα που ταιριάζουν με τα ενδιαφέροντά τους και γίνονται διαθέσιμα.
- *Παροχή Πολύγλωσσης Υποστήριξης.* Η πολύγλωσση υποστήριξη περιλαμβάνει την παροχή πρόσβασης σε θησαυρούς και λεξικά για τις υποστηριζόμενες γλώσσες, τα οποία περιέχουν επιπλέον εξειδικεύσεις και γενικεύσεις όρων. Τα λεξικά και οι θησαυροί χρησιμοποιούνται και κατά την επεξεργασία των ερωτήσεων, έτσι ώστε να

υποστηριχθεί η διαγλωσσική ανάκτηση των αντικειμένων που συσχετίζονται με μια ερώτηση.

Για την παροχή των παραπάνω υπηρεσιών, χρησιμοποιούνται ένα φίλτρο και ένας ευρέτης ρίζας (stemmer) για κάθε γλώσσα. Καθώς η παροχή πολύγλωσσης υποστήριξης δεν αποτελεί τμήμα της παρούσας διατριβής αλλά της διατριβής [Πουτ2000], στη συνέχεια δε θα δοθεί ιδιαίτερη έμφαση στην περιγραφή της.

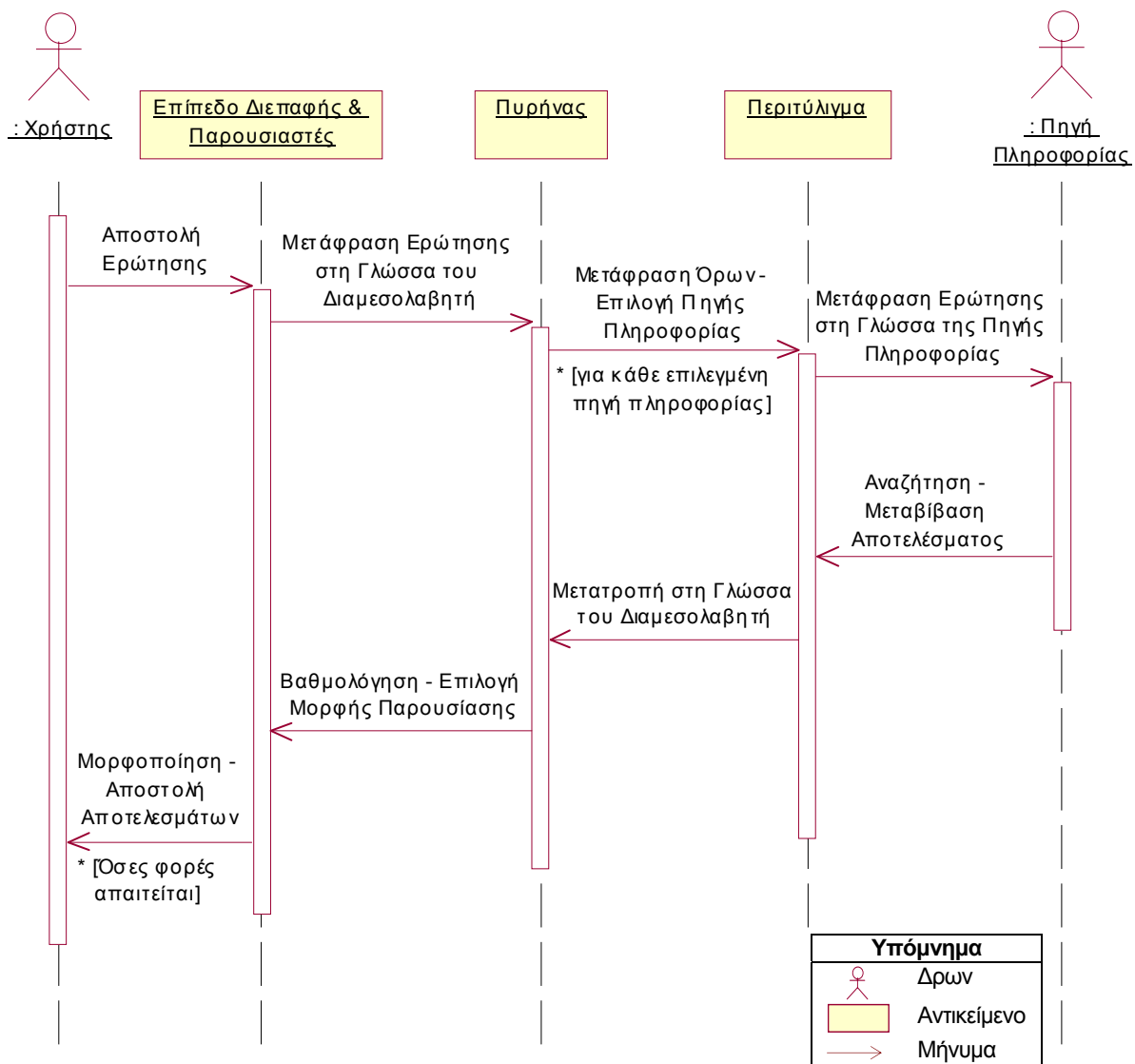
- *Αλληλεπίδραση με τις Πηγές Πληροφορίας.* Η αλληλεπίδραση του διαμεσολαβητή με τις πηγές πληροφορίας υποστηρίζεται από τα περιτυλίγματα των πηγών πληροφορίας. Η αλληλεπίδραση με τις πηγές πληροφορίας περιλαμβάνει, από την πλευρά του διαμεσολαβητή, τη μεταβίβαση αιτήσεων και ερωτήσεων στις πηγές πληροφορίας και την προώθηση των αποτελεσμάτων και των αιτήσεων, μετά την κατάλληλη επεξεργασία, στους χρήστες. Από την πλευρά των πηγών πληροφορίας, η αλληλεπίδραση περιλαμβάνει την επιστροφή των αποτελεσμάτων των ερωτήσεων και των αιτήσεων που μεταβιβάζονται σ' αυτές από το διαμεσολαβητή και στην αποστολή σε αυτόν ειδοποιήσεων για αλλαγές στο περιεχόμενο και/ή στη δομή τους (π.χ. νέα αντικείμενα που γίνονται διαθέσιμα, υποστήριξη νέων κλάσεων κλπ.).
- *Εξατομίκευση των παρεχόμενων στους χρήστες υπηρεσιών.* Η εξατομίκευση των παρεχόμενων στους χρήστες υπηρεσιών πραγματοποιείται μέσω της διαχείρισης και της χρήσης των διαγραμμάτων προτιμήσεων των χρηστών και περιλαμβάνει:
 - Την αποθήκευση των διαγραμμάτων προτιμήσεων των χρηστών σε μια βάση διαγραμμάτων μέσα στον πυρήνα του διαμεσολαβητή. Η βάση διαγραμμάτων αποτελεί τμήμα της βάσης δεδομένων του διαμεσολαβητή.
 - Το γεγονός ότι τα αποτελέσματα των ερωτήσεων και των αιτήσεων παρουσιάζονται στους χρήστες μορφοποιημένα με βάση τα διαγράμματα προτιμήσεών τους. Οι κανόνες μορφοποίησης τόσο των αποτελεσμάτων των ερωτήσεων όσο και των αντικειμένων καθορίζονται από:
 - Την υπολογιστική πλατφόρμα που χρησιμοποιεί ο κάθε χρήστης και τις δυνατότητες που αυτή παρέχει. Έτσι, για παράδειγμα, αν η σύνδεση κάποιου χρήστη με το Διαδίκτυο είναι αργή, στο χρήστη αυτό στέλνονται μόνο τα απαραίτητα κάθε φορά γραφικά. Επίσης, αν ο υπολογιστής κάποιου χρήστη δε διαθέτει κάρτα ήχου, δε στέλνονται στο χρήστη αυτό ηχητικά δεδομένα.
 - Τα χρώματα (κειμένου, φόντου κλπ.) και τις γραμματοσειρές που προτιμά ο χρήστης να χρησιμοποιούνται κατά την παρουσίαση πληροφορίας σ' αυτόν.
 - Τον τρόπο παρουσίασης της πληροφορίας που προτιμά ο χρήστης για τα αντικείμενα (πλήρη παρουσίαση, σύντομη, τμηματική κλπ.) και τα αποτελέσματα των ερωτήσεων (όλα μαζί, σε ομάδες ή στοιχείων κάθε φορά κλπ.).
 - Όταν κάποιο αντικείμενο γίνει διαθέσιμο, οι ενδιαφερόμενοι χρήστες ειδοποιούνται γι αυτό. Ένας χρήστης χαρακτηρίζεται ως «ενδιαφερόμενος» για κάποιο αντικείμενο αν στο διάγραμμα προτιμήσεών του, στον τομέα των ενδιαφερόντων, περιλαμβάνονται όροι που σχετίζονται με το περιεχόμενο του αντικειμένου αυτού.

Καθώς τα διαγράμματα προτιμήσεων των χρηστών δεν αποτελούν τμήμα της παρούσας διατριβής, στη συνέχεια δε θα δοθεί ιδιαίτερη έμφαση στην περιγραφή τους.

- *Διαχείριση Μεταδεδομένων.* Η διαχείριση μεταδεδομένων περιλαμβάνει την αποθήκευση και τη χρήση μεταδεδομένων για τη δομή και το περιεχόμενο των πηγών πληροφορίας. Επιπλέον, περιλαμβάνει την παροχή, σε μορφή RDF, της μεταπληροφορίας που σχετίζεται με τα αντικείμενα από τους παρουσιαστές.

Για να γίνει πιο εύκολα κατανοητή η παρουσίαση της λειτουργικότητας του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», ακολουθεί η προσπάθεια εντοπισμού του ρόλου του μέσα από ένα σύνολο UML διαγραμμάτων αλληλεπίδρασης (και πιο συγκεκριμένα σειριακών διαγραμμάτων). Τα διαγράμματα αυτά περιγράφουν τις επιμέρους αλληλεπιδράσεις των τμημάτων του διαμεσολαβητή, τόσο μεταξύ τους όσο και με τους χρήστες και τις πηγές πληροφορίας, σε κάθε περίπτωση χρήσης του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».

Το διάγραμμα αλληλεπίδρασης για την απάντηση μιας ερώτησης απεικονίζεται στο *Διάγραμμα 5-2*.



Διάγραμμα 5-2: Σειριακό Διάγραμμα για την Απάντηση Ερώτησης

Όπως φαίνεται στο *Διάγραμμα 5-2*, μόλις ο χρήστης συνθέσει την ερώτησή του, τη στέλνει μέσω του γραφικού επιπέδου διεπαφής που χρησιμοποιεί στο Επίπεδο Διεπαφής του διαμεσολαβητή. Εκεί, η ερώτηση μεταφράζεται στη γλώσσα του διαμεσολαβητή και στέλνεται στον πυρήνα του διαμεσολαβητή, όπου με βάση τα λεξικά μεταφράζονται οι όροι της ερώτησης στη βασική γλώσσα που υποστηρίζεται από το σύστημα. Επιπλέον, με τη βοήθεια των μεταδεδομένων που περιγράφουν τη δομή και το περιεχόμενο των πηγών πληροφορίας, επιλέγονται οι πηγές πληροφορίας στις οποίες θα μεταβιβαστεί η ερώτηση.

Στη συνέχεια, για κάθε μια από τις επιλεγμένες πηγές πληροφορίας, το περιτύλιγμά της κλάσης πηγών πληροφορίας στην οποία εντάσσεται μεταφράζει την ερώτηση στη γλώσσα της πηγής πληροφορίας και τη μεταβιβάζει στην πηγή πληροφορίας. Κάθε πηγή πληροφορίας πραγματοποιεί την αναζήτηση για αντικείμενα που ικανοποιούν τα κριτήρια που θέτει η ερώτηση μεταξύ των περιεχομένων της και μεταβιβάζει το αποτέλεσμα στο περιτύλιγμά της. Το περιτύλιγμα μεταφράζει το αποτέλεσμα στη γλώσσα του διαμεσολαβητή και το προωθεί στον πυρήνα του διαμεσολαβητή.

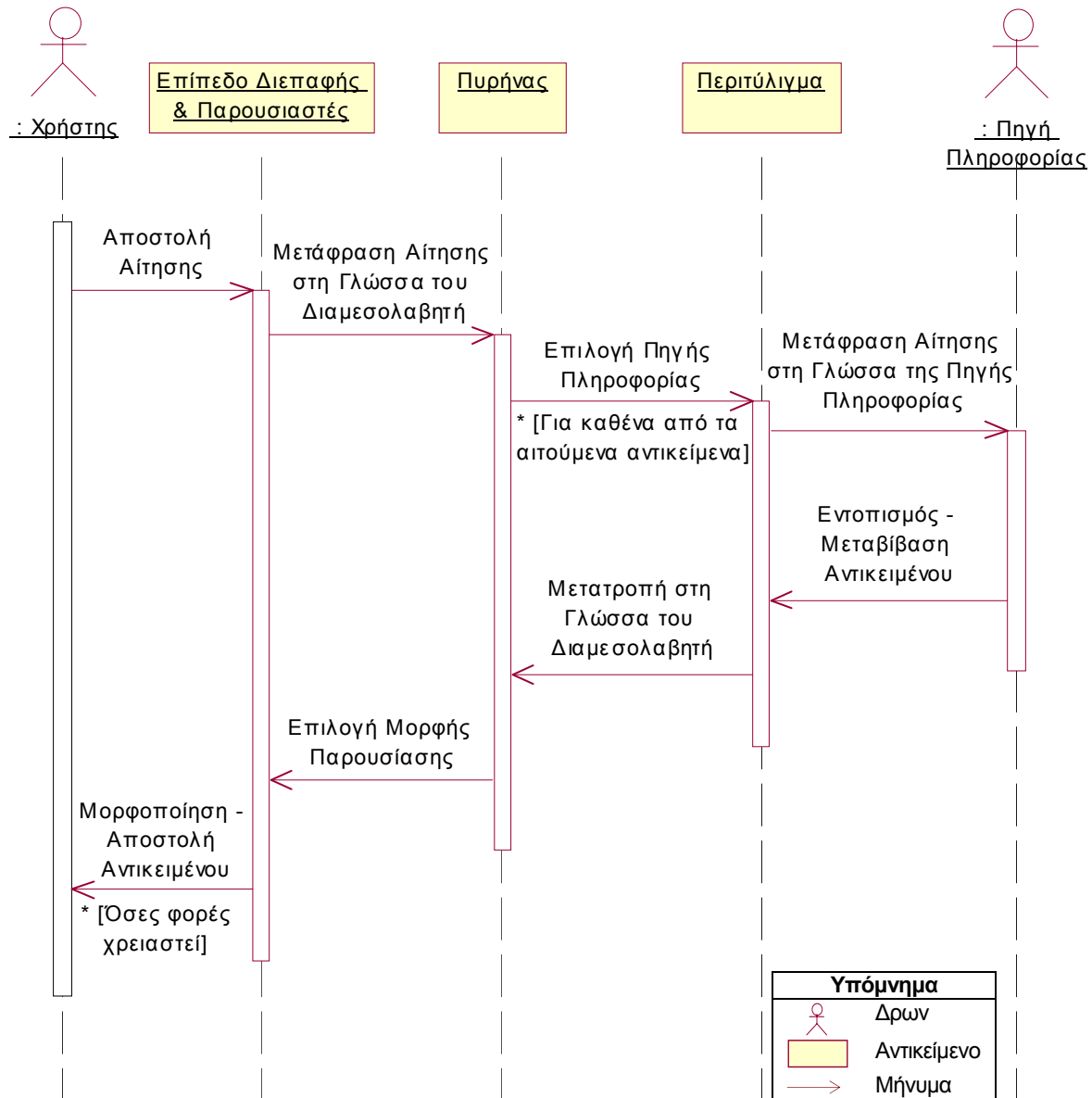
Όταν όλες οι πηγές πληροφορίας επιστρέψουν τα αποτελέσματα της αναζήτησης που πραγματοποίησαν, στον πυρήνα του διαμεσολαβητή κανονικοποιούνται οι βαθμολογήσεις των αποτελεσμάτων και επιλέγεται, με βάση το διάγραμμα προτιμήσεων του χρήστη, η μορφή παρουσίασης των αποτελεσμάτων. Τέλος, οι παρουσιαστές αναλαμβάνουν τη μορφοποίηση των αποτελεσμάτων με βάση την επιλεγμένη μορφή παρουσίασης και τα αποτελέσματα στέλνονται για παρουσίαση στο χρήστη. Κατά τη διάρκεια της μορφοποίησης γίνονται οι απαραίτητες ρυθμίσεις, ώστε οι σύνδεσμοι στα αντικείμενα που απαρτίζουν το αποτέλεσμα της ερώτησης να καθορίζουν και τη μορφή της ανακτώμενης πληροφορίας¹⁸. Επιπλέον, καθώς τα αποτελέσματα μπορεί να σταλούν στους χρήστες, ανάλογα και με τις προτιμήσεις των τελευταίων, τμηματικά (π.χ. σε ομάδες: τα η πρώτα αντικείμενα αρχικά, μετά τα η επόμενα κλπ.), η διαδικασία μορφοποίησης και αποστολής των αποτελεσμάτων μπορεί να επαναληφθεί περισσότερες από μία φορές.

Το διάγραμμα αλληλεπίδρασης για την ανάκτηση αντικειμένων φαίνεται στο *Διάγραμμα 5-3*.

Όπως φαίνεται στο *Διάγραμμα 5-3*, ο χρήστης στέλνει την αίτηση μέσω του γραφικού επιπέδου διεπαφής που χρησιμοποιεί στο Επίπεδο Διεπαφής του διαμεσολαβητή, όπου η αίτηση μεταφράζεται στη γλώσσα του διαμεσολαβητή. Στη συνέχεια, η αίτηση μεταβιβάζεται στον πυρήνα του διαμεσολαβητή, όπου, για καθένα από τα αιτούμενα αντικείμενα, επιλέγεται η πηγή πληροφορίας όπου βρίσκεται το συγκεκριμένο αντικείμενο. Μέσω του περιτύλιγματος της πηγής πληροφορίας, το αντίστοιχο τμήμα της αίτησης μεταφράζεται στη γλώσσα της πηγής πληροφορίας και μεταβιβάζεται στην πηγή πληροφορίας, όπου λαμβάνει χώρα ο εντοπισμός των αντικειμένων. Στη συνέχεια τα αντικείμενα μεταβιβάζονται στο περιτύλιγμα, όπου μετατρέπονται σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή και προωθούνται στον πυρήνα. Όταν όλα τα αντικείμενα μεταβιβαστούν στο

¹⁸ Έτσι, για παράδειγμα, αν ο υπολογιστής κάποιου χρήστη δε διαθέτει κάρτα ήχου και γίνει απ' αυτόν αίτηση αποθήκευσης κάποιων αντικειμένων που περιέχουν και ηχητικά δεδομένα στον υπολογιστή του, στέλνεται σε αυτόν όλη η πληροφορία από την οποία απαρτίζονται τα αντικείμενα (μαζί με τα ηχητικά δεδομένα). Αν όμως ο χρήστης θέλει να του παρουσιαστούν τα αντικείμενα αυτά, η διάσχιση του αντίστοιχου συνδέσμου έχει ως συνέπεια να σταλούν σε αυτόν όλα τα δεδομένα από τα οποία απαρτίζονται τα αντικείμενα αυτά εκτός από τα ηχητικά.

διαμεσολαβητή, στον πυρήνα του διαμεσολαβητή, με βάση το διάγραμμα προτιμήσεων του χρήστη, επιλέγεται η μορφή παρουσίασης¹⁹ των αντικειμένων, τα οποία μορφοποιούνται από τον κατάλληλο παρουσιαστή και στέλνονται για παρουσίαση στο χρήστη. Καθώς μπορεί να έχει επιλεγεί τα αντικείμενα να παρουσιάζονται τμηματικά, η διαδικασία μορφοποίησης των αντικειμένων (ή τμημάτων των αντικειμένων) και αποστολής των μορφοποιημένων αντικειμένων (ή τμημάτων αντικειμένων) μπορεί να επαναληφθεί περισσότερες από μία φορές.

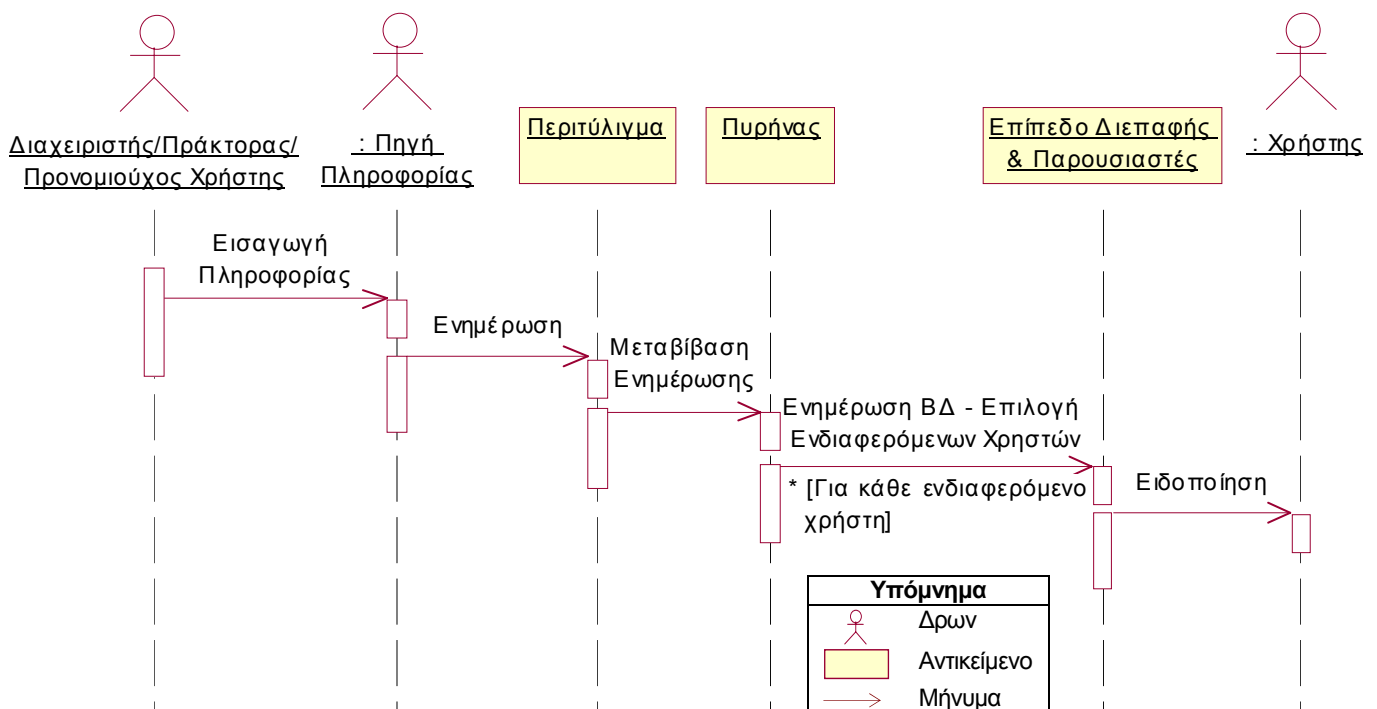


Διάγραμμα 5-3: Σειριακό Διάγραμμα για την Ανάκτηση Αντικειμένου

Το Διάγραμμα 5-4 είναι το διάγραμμα αλληλεπίδρασης που περιγράφει τη διαδικασία ειδοποίησης των πιθανά ενδιαφερόμενων χρηστών για το ότι κάποιο αντικείμενο, το περιεχόμενο του οποίου ταιριάζει με τα ενδιαφέροντά τους, γίνεται διαθέσιμο.

¹⁹ Στη μορφή παρουσίασης των αντικειμένων περιλαμβάνεται και η ενσωμάτωση ή μη μεταδεδομένων στα προς παρουσίαση αντικείμενα καθώς και η επιλογή, στην πρώτη περίπτωση, των μεταδεδομένων που θα ενσωματωθούν

Όπως φαίνεται στο *Διάγραμμα 5-4*, ο διαχειριστής της πηγής πληροφορίας, κάποιος πράκτορας υπεύθυνος για την ενημέρωσή της ή κάποιος προνομιούχος χρήστης που έχει το δικαίωμα εισαγωγής πληροφορίας, εισάγει νέα πληροφορία στην πηγή πληροφορίας. Τότε στέλνεται ένα μήνυμα ενημέρωσης του διαμεσολαβητή στο περιτύλιγμα της πηγής πληροφορίας. Το μήνυμα μεταβιβάζεται στον πυρήνα του διαμεσολαβητή, όπου ενημερώνεται, αν χρειάζεται, η βάση δεδομένων του διαμεσολαβητή για αλλαγές που πιθανώς έγιναν στο περιεχόμενο και/ή τη δομή μιας πηγής πληροφορίας. Στη συνέχεια επιλέγονται, με βάση τα ενδιαφέροντα που καταγράφονται στα διαγράμματα προτιμήσεών τους, οι πιθανά ενδιαφερόμενοι για την πληροφορία που έγινε διαθέσιμη χρήστες. Τέλος, από τους παρουσιαστές στέλνεται σε καθέναν από τους ενδιαφερόμενους χρήστες το κατάλληλο μήνυμα ειδοποίησης, μορφοποιημένο σύμφωνα με τις προτιμήσεις τους.



Διάγραμμα 5-4: Σειριακό Διάγραμμα για την Ειδοποίηση Χρηστών

5.4. Γλώσσα του Διαμεσολαβητή

Το θέμα αυτού του τμήματος είναι η γλώσσα του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Η χρησιμότητα και η σημασία της γλώσσας του διαμεσολαβητή είναι προφανής: Οι αιτήσεις και οι ερωτήσεις των χρηστών του συστήματος μεταφράζονται στη γλώσσα του διαμεσολαβητή πριν σταλούν στις πηγές πληροφορίας μέσω των περιτυλιγμάτων τους. Επιπλέον, τα αποτελέσματα των ερωτήσεων και των αιτήσεων που επιστρέφονται από τις πηγές πληροφορίας, πριν μορφοποιηθούν για παρουσίαση από τους παρουσιαστές, μεταφράζονται από τα περιτυλίγματα των πηγών πληροφορίας σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή.

Η ύπαρξη της γλώσσας του διαμεσολαβητή συνεπάγεται ότι η γλώσσα αυτή διαχειρίζεται κάποιο μοντέλο πληροφορίας. Το μοντέλο αυτό δεν είναι προφανές, καθώς αυτό θα πρέπει να καλύπτει τις ανάγκες όλων των κλάσεων πηγών πληροφορίας που γίνονται διαθέσιμες

μέσω του διαμεσολαβητή. Αυτό σημαίνει ότι το μοντέλο πληροφορίας που θα διαχειρίζεται η γλώσσα του διαμεσολαβητή πρέπει να έχει τη δυνατότητα να περιγράψει ικανοποιητικά πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα (που η γλώσσα τους διαχειρίζεται το σχεσιακό μοντέλο), πηγές πληροφορίας που περιέχουν δομημένα έγγραφα (που η γλώσσα τους διαχειρίζεται το ημιδομημένο ή το διατεταγμένο δεντρικό μοντέλο), πηγές πληροφορίας που περιέχουν ημιδομημένα αντικείμενα (που η γλώσσα τους διαχειρίζεται το ημιδομημένο μοντέλο) κλπ. Έτσι επιλέχθηκε το μοντέλο πληροφορίας που διαχειρίζεται η γλώσσα του διαμεσολαβητή να περιγράφεται από ένα RDF σχήμα, που προκύπτει από το συνδυασμό των RDF σχημάτων που περιγράφουν τις βάσεις δεδομένων των πηγών πληροφορίας. Η επιλογή αυτή έγινε καθώς τα RDF σχήματα αποτελούν ένα πρότυπο (standard) τρόπο για την κάλυψη των αναγκών αναπαράστασης πολλών μοντέλων πληροφορίας. Τα RDF σχήματα αρκούν για την πλήρη περιγραφή E-R διαγραμμάτων, μπορούν όμως επιπλέον να περιγράψουν και αντικειμενοστραφή μοντέλα και γράφους.

Από τα παραπάνω προκύπτει ότι το μοντέλο πληροφορίας που διαχειρίζεται η γλώσσα του διαμεσολαβητή περιγράφεται από ένα επεκταμένο με αντικειμενοστραφή χαρακτηριστικά και χαρακτηριστικά υποστήριξης γράφων E-R διάγραμμα. Οι οντότητες που γίνονται διαθέσιμες μέσω του διαμεσολαβητή, παρουσιάζεται στο χρήστη να διαθέτουν ένα επιπλέον χαρακτηριστικό, που αναπαριστά το όνομα της πηγής πληροφορίας από την οποία προέρχεται το κάθε στιγμιότυπο.

Καθώς το μοντέλο πληροφορίας που διαχειρίζεται η γλώσσα του διαμεσολαβητή περιγράφεται από ένα επεκταμένο E-R διάγραμμα, αντίστοιχα η σύνταξη της γλώσσας του διαμεσολαβητή αποτελεί μια επέκταση της SQL, που περιλαμβάνει τα απαραίτητα για την υποστήριξη της επιπλέον λειτουργικότητας χαρακτηριστικά. Τα χαρακτηριστικά αυτά περιλαμβάνουν:

- Την υποστήριξη ερωτήσεων που αναφέρονται σε κλάσεις αντικειμένων, που αποτελεί χαρακτηριστικό του αντικειμενοστραφούς μοντέλου. Για το σκοπό αυτό, ο σχεδιασμός της γλώσσας του διαμεσολαβητή επηρεάστηκε από την OQL [OQL] σε ότι αφορά την υποστήριξη κλάσεων.
- Την υποστήριξη ερωτήσεων που περιέχουν κριτήρια που βασίζονται σε ιεραρχικές σχέσεις μεταξύ αντικειμένων (πατέρα – παιδιού και προγόνου – απογόνου), που αποτελεί χαρακτηριστικό του διατεταγμένου δενδρικού μοντέλου. Για το σκοπό αυτό, ο σχεδιασμός της γλώσσας του διαμεσολαβητή επηρεάστηκε από την XQL σε ότι αφορά την υποστήριξη ιεραρχικών σχέσεων μεταξύ αντικειμένων.

Οι απαιτήσεις που τίθενται για τη γλώσσα του διαμεσολαβητή, από το μοντέλο πληροφορίας που αυτή διαχειρίζεται, τη λειτουργικότητά του και την αλληλεπίδρασή του με τις πηγές πληροφορίας και τους χρήστες είναι οι εξής:

- Η δυνατότητα έκφρασης αιτήσεων ανάκτησης πληροφορίας από τις πηγές πληροφορίας που το περιεχόμενό τους γίνεται διαθέσιμο μέσω του διαμεσολαβητή.
- Η δυνατότητα έκφρασης τόσο «κλασσικών» ερωτήσεων δίτιμης λογικής όσο και ερωτήσεων που μπορεί να περιέχουν όρους ομοιότητας και/ή όρους δίτιμης λογικής με

βάση το περιεχόμενο και/ή τη δομή των δομημένων αντικειμένων που αποθηκεύονται στις πηγές πληροφορίας.

- Η δυνατότητα περιγραφής δομημένων αντικειμένων από τη γλώσσα του διαμεσολαβητή, προκειμένου αυτά είτε να επιστραφούν ως αποτελέσματα ερωτήσεων είτε να χρησιμοποιηθούν για την έκφραση κριτηρίων ερωτήσεων.
- Η δυνατότητα έκφρασης ερωτήσεων που περιλαμβάνουν τόσο τα στιγμιότυπα μιας κλάσης όσο και αυτά των υποκλάσεων της.
- Η δυνατότητα έκφρασης ερωτήσεων που περιέχουν κριτήρια που βασίζονται σε ιεραρχικές σχέσεις.

Με βάση το μοντέλο πληροφορίας της γλώσσας του διαμεσολαβητή και τη λειτουργικότητα που πρέπει αυτή να παρέχει, επιλέχθηκε η γλώσσα του διαμεσολαβητή να έχει τα εξής χαρακτηριστικά:

- Να μοιάζει με την SQL (να είναι δηλαδή SQL-like). Η σημαντικότερη αιτία για την επιλογή να μοιάζει η γλώσσα του διαμεσολαβητή με SQL είναι το ότι το μοντέλο πληροφορίας που διαχειρίζεται η γλώσσα του διαμεσολαβητή περιγράφεται από ένα επεκταμένο E-R διάγραμμα. Έτσι, μια επεκταμένη με τα απαραίτητα χαρακτηριστικά έκδοση της SQL θα καλύπτει τις ανάγκες έκφρασης ερωτήσεων για το μοντέλο αυτό.

Επιπλέον, αναμένεται ότι οι γλώσσες των περισσότερων κλάσεων πηγών πληροφορίας είτε θα είναι SQL είτε θα μοιάζουν με SQL. Έτσι γίνονται πιο εύκολα τόσο ο σχεδιασμός όσο και η υλοποίηση των περιτυλιγμάτων των πηγών πληροφορίας. Επιπλέον, η διαδικασία της μετατροπής των αιτήσεων και των ερωτήσεων από τα περιτυλίγματα στη γλώσσα των πηγών πληροφορίας θα είναι λιγότερο χρονοβόρα.

- Να επιτρέπει την έκφραση ερωτήσεων ομοιότητας. Το μοντέλο ανάκτησης πληροφορίας που υποστηρίζεται από τη γλώσσα του διαμεσολαβητή επιτρέπει την έκφραση κριτηρίων που συσχετίζουν αντικείμενα ή χαρακτηριστικά αντικειμένων με σύνολα όρων που συνδέονται μεταξύ τους με AND και OR. Επιπλέον, επιτρέπει την έκφραση κριτηρίων που περιέχουν ασαφή (fuzzy) χαρακτηριστικά [Μουμ97].

Στην κατεύθυνση αυτή επιλέχθηκε να περιλαμβάνονται στη σύνταξη της γλώσσας του διαμεσολαβητή τμήματα XML πληροφορίας για την περιγραφή σύνθετων κριτηρίων που επιτρέπουν την υποστήριξη δυνατοτήτων ανάκτησης πληροφορίας που δεν υποστηρίζονται από την SQL.

- Να επιστρέφει τα αποτελέσματα των ερωτήσεων και των αιτήσεων σε XML μορφή. Η σημαντικότερη αιτία για τη χρήση σύνταξης βασισμένης στην XML για την περιγραφή των αποτελεσμάτων των ερωτήσεων και των αιτήσεων είναι το γεγονός ότι η πληροφορία που είναι σε μορφή XML υποστηρίζεται από τους παρουσιαστές, ενώ αναμένεται να υποστηρίζεται και από τις περισσότερες πηγές πληροφορίας. Πιο συγκεκριμένα, οι παρουσιαστές έχουν τη δυνατότητα να δέχονται για μορφοποίηση και να εξάγουν XML πληροφορία, ενώ αναμένεται οι πηγές πληροφορίας να εξάγουν την πληροφορία που επιστρέφουν σε μορφή XML. Αυτό σημαίνει ότι οι μετατροπές της μορφής της πληροφορίας, καθώς αυτή θα ανταλλάσσεται μεταξύ των τμημάτων που

απαρτίζουν την αρχιτεκτονική του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», θα είναι απλούστερες και λιγότερο χρονοβόρες.

Επιπλέον, μ' αυτό τον τρόπο, θα είναι ευκολότερη στη συνέχεια η ενσωμάτωση ή η διασύνδεση της πληροφορίας μορφοποίησης στην επιστρεφόμενη πληροφορία από τους παρουσιαστές (με τη μορφή στυλ).

Ο καθορισμός της σύνταξης της γλώσσας του διαμεσολαβητή χωρίζεται σε δυο τμήματα: στο πρώτο τμήμα, που ασχολείται με την *Υποστήριξη Ερωτήσεων και Αιτήσεων* (βλέπε ενότητα 5.4.1) και στο τμήμα που περιγράφεται η *Επιστρεφόμενη Πληροφορία* (βλέπε ενότητα 5.4.2).

5.4.1. Υποστήριξη Ερωτήσεων και Αιτήσεων

Το θέμα αυτής της ενότητας είναι η περιγραφή της υποστήριξης που παρέχεται για την έκφραση ερωτήσεων και αιτήσεων από τη γλώσσα του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Όπως έχει ήδη αναφερθεί, η γλώσσα του διαμεσολαβητή μοιάζει με SQL, παρέχει όμως αυξημένη λειτουργικότητα σε σχέση με την SQL. Έτσι, η λειτουργικότητα της γλώσσας του διαμεσολαβητή συνοψίζεται στα εξής σημεία:

- Στην παροχή της λειτουργικότητας που παρέχεται από την SQL για την υποστήριξη ερωτήσεων δίτιμης λογικής.
- Στην παροχή της δυνατότητας επιλογής των πηγών πληροφορίας στα περιεχόμενα των οποίων θα πραγματοποιηθεί κάθε φορά η αναζήτηση.
- Στην παροχή της δυνατότητας αναζήτησης τόσο μεταξύ των στιγμιοτύπων μιας κλάσης όσο και μεταξύ των στιγμιοτύπων των υποκλάσεων της κλάσης αυτής²⁰.
- Στην παροχή της δυνατότητας έκφρασης τόσο ερωτήσεων ομοιότητας όσο και ερωτήσεων που περιέχουν όρους δίτιμης λογικής και όρους ομοιότητας.
- Στην παροχή της δυνατότητας έκφρασης κριτηρίων σχετικών με τη δομή των αντικειμένων.
- Στην παροχή της δυνατότητας έκφρασης αιτήσεων για τα χαρακτηριστικά των αντικειμένων και για την ολική ή μερική ανάκτηση του περιεχομένου αντικειμένων και δομικών στοιχείων αντικειμένων.

Ο Πίνακας 5-1 περιέχει μια σύντομη περιγραφή της επιπλέον της SQL λειτουργικότητας που παρέχεται από τη γλώσσα του διαμεσολαβητή και την περιγραφή της σύνταξης που την υποστηρίζει.

Περιγραφή	Σύνταξη
Επιλογή/Επιστροφή της Πηγής Πληροφορίας όπου περιέχονται τα Αντικείμενα	Ιδεατό Χαρακτηριστικό Source

²⁰ Αυτό ισχύει φυσικά μόνο για της πηγές πληροφορίας όπου υπάρχουν κλάσεις και υποκλάσεις. Η πληροφορία που επιτρέπει την εξαγωγή συμπερασμάτων σχετικά με το μοντέλο πληροφορίας κάθε πηγής πληροφορίας, όπως αυτό εξάγεται προς το διαμεσολαβητή αποτελεί τμήμα της σχετικής με την κάθε πηγή πληροφορίας μεταπληροφορίας.

Επιλογή/Επιστροφή περιεχομένου των Αντικειμένων	Ιδεατό Χαρακτηριστικό Content
Εξαίρεση δομικών στοιχείων από το επιστρεφόμενο περιεχόμενο των Αντικειμένων	Τελεστής «-»
Σχέση πατέρα – παιδιού	Τελεστής «/»
Σχέση προγόνου – απογόνου	Τελεστής «//»
Αναφορά στο n-ιστό δομικό στοιχείο	Τελεστής «[]»
Σχέσεις Διάταξης	Τελεστές «Before» και «After»
Επιλογή στιγμιοτύπων κλάσης	<Όνομα Κλάσης>
Επιλογή στιγμιοτύπων μιας κλάσης και των υποκλάσεων της	<Όνομα Κλάσης>*
Συσχέτιση με όρους ομοιότητας	Τελεστής «RELATES TO», σύνολα όρων ((AND OR)TERMSETS), όροι (TERMS)
Ασαφή χαρακτηριστικά	<Όνομα Ασαφούς Χαρακτηριστικού>

Πίνακας 5-1: Η επιπλέον της SQL λειτουργικότητα που παρέχεται από τη Γλώσσα του Διαμεσολαβητή

Όπως φαίνεται στον πίνακα, η δυνατότητα επιλογής και/ή επιστροφής του ονόματος των πηγών πληροφορίας στα περιεχόμενα των οποίων θα πραγματοποιηθεί κάθε φορά η αναζήτηση για αντικείμενα που ικανοποιούν τα κριτήρια της ερώτησης που τίθεται, επιτυγχάνεται με το να τεθεί κάποιος περιορισμός στο ιδεατό χαρακτηριστικό Source όλων των κλάσεων. Έτσι, για την επιλογή του τίτλου και της ταυτότητας των στιγμιοτύπων της κλάσης «Άρθρο» που έχουν γραφτεί από τον «X Y» και αποθηκεύονται στην πηγή πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης» δημιουργείται η ερώτηση που φαίνεται στο Παράδειγμα 5-1.

```
SELECT Τίτλος, Ταυτότητα FROM Άρθρο
WHERE Συγγραφέας= "X Y" AND
Source="Βιβλιοθήκη Πολυτεχνείου Κρήτης"
```

Παράδειγμα 5-1: Ερώτηση όπου καθορίζεται η Πηγή Πληροφορίας στα περιεχόμενα της οποίας θα πραγματοποιηθεί η αναζήτηση

Στο σημείο αυτό πρέπει να διευκρινιστεί ότι, όταν μια ερώτηση μεταβιβάζεται από το διαμεσολαβητή στις πηγές πληροφορίας, οι συνθήκες που αναφέρονται στο ιδεατό χαρακτηριστικό Source παραλείπονται.

Η γλώσσα του διαμεσολαβητή, παρ' όλο που δε χρειάζεται συχνά η δυνατότητα να πραγματοποιείται αναζήτηση μεταξύ των στιγμιοτύπων μιας κλάσης αντικειμένων χωρίς να λαμβάνονται υπ' όψιν τα στιγμιότυπα των υποκλάσεων της, υποστηρίζει τη δυνατότητα αναζήτησης τόσο μόνο μεταξύ των στιγμιοτύπων μιας κλάσης όσο και μεταξύ των στιγμιοτύπων της κλάσης και των υποκλάσεων της κλάσης αυτής.

Έτσι, για την επιλογή μεταξύ των στιγμιοτύπων μιας κλάσης αντικειμένων χωρίς να λαμβάνονται υπ' όψιν τα στιγμιότυπα των υποκλάσεων της, όπως αναφέρει ο Πίνακας 5-1, στο FROM τμήμα των ερωτήσεων χρησιμοποιείται το όνομα της κλάσης. Αντίστοιχα, για

την επιλογή μεταξύ των στιγμιοτύπων μιας κλάσης αντικειμένων και των υποκλάσεων της κλάσης αυτής, στο FROM τμήμα των ερωτήσεων χρησιμοποιείται το όνομα της κλάσης ακολουθούμενο από ένα αστεράκι (*). Έτσι, η ερώτηση που φαίνεται στο Παράδειγμα 5-1 πραγματοποιεί αναζήτηση μεταξύ των στιγμιοτύπων της κλάσης «Άρθρο» χωρίς να λαμβάνονται υπ' όψιν τα στιγμιότυπα των υποκλάσεων της, ενώ η ερώτηση που φαίνεται στο Παράδειγμα 5-2 πραγματοποιεί αναζήτηση τόσο μεταξύ των στιγμιοτύπων της κλάσης «Άρθρο» όσο και μεταξύ των στιγμιοτύπων των υποκλάσεων της.

```
SELECT Τίτλος, Ταυτότητα FROM Άρθρο*
WHERE Συγγραφέας= "X Y" AND
Source="Βιβλιοθήκη Πολυτεχνείου Κρήτης"
```

Παράδειγμα 5-2: Ερώτηση που πραγματοποιεί αναζήτηση μεταξύ των στιγμιοτύπων της κλάσης «Άρθρο» και των υποκλάσεών της

Για την έκφραση ερωτήσεων ομοιότητας η γλώσσα του διαμεσολαβητή διαθέτει τον τελεστή RELATES TO, που δηλώνει ότι το περιεχόμενο κάποιου αντικειμένου ή τμήματος αντικειμένου σχετίζεται με κάποιο σύνολο όρων. Ένα σύνολο όρων περιέχει όρους (TERMS), καθένας από τους οποίους διαθέτει κάποιο βάρος (WEIGHT). Μεταξύ των όρων αυτών εφαρμόζεται είτε η λειτουργία της σύζευξης (AND), οπότε το σύνολο όρων είναι ένα ANDTERMSET, είτε η λειτουργία της διάζευξης (OR), οπότε το σύνολο όρων είναι ένα ORTERMSET. Τέλος, ένα σύνολο όρων μπορεί να περιέχει, εκτός από όρους, άλλα σύνολα όρων.

Η ερώτηση που φαίνεται στο Παράδειγμα 5-3 επιστρέφει τους τίτλους, τις ταυτότητες και του συγγραφείς των Εγγράφων που το περιεχόμενό τους σχετίζεται με τον όρο «Κνωσός» με βαθμό συσχέτισης 100% και με τον όρο «Ανακατασκευή» με βαθμό συσχέτισης 50%.

```
SELECT Τίτλος, Συγγραφέας, Ταυτότητα
FROM Έγγραφο* WHERE Έγγραφο RELATES TO
<ORTERMSET>
<TERM WEIGHT=100%>Κνωσός</TERM>
<TERM WEIGHT=50%>Ανακατασκευή</TERM>
</ORTERMSET>
```

Παράδειγμα 5-3: Ερώτηση Ομοιότητας για το περιεχόμενο Αντικειμένων

Επιπλέον, η γλώσσα του διαμεσολαβητή υποστηρίζει ασαφή χαρακτηριστικά (fuzzy attributes). Τα ασαφή χαρακτηριστικά είναι χαρακτηριστικά που η τιμή τους υπολογίζεται με βάση μια συνάρτηση αξιολόγησης πάνω στις τιμές ενός ή περισσότερων χαρακτηριστικών των αντίστοιχων αντικειμένων. Έτσι, για παράδειγμα, έστω ότι για όλα τα έγγραφα το χαρακτηριστικό «Παλαιότητα» ορίζεται με βάση την εφαρμογή της συνάρτησης

$$f(x) = \begin{cases} 0, & \forall x > 1950 \\ 1 - \frac{1}{10(y - x)}, & \forall x \leq 1950 \end{cases}, \text{ όπου } x \text{ η ημερομηνία συγγραφής του εγγράφου και}$$

y η τρέχουσα ημερομηνία. Μια ερώτηση που χρησιμοποιεί κριτήριο που βασίζεται στην

παλαιότητα των εγγράφων για την ανάκτηση των εγγράφων με βαθμό παλαιότητας μεγαλύτερο από 50% φαίνεται στο Παράδειγμα 5-4.

```
SELECT Τίτλος, Συγγραφέας, Ταυτότητα  
FROM Έγγραφο* WHERE Παλαιότητα>50%
```

Παράδειγμα 5-4: Ερώτηση με βάση Ασαφή Χαρακτηριστικά

Σε ότι αφορά τη δυνατότητα έκφρασης κριτηρίων σχετικών με τη δομή των αντικειμένων, η γλώσσα του διαμεσολαβητή επιτρέπει την έκφραση συνθηκών για τα δομικά στοιχεία που τα συνθέτουν. Πιο συγκεκριμένα, η γλώσσα του διαμεσολαβητή επιτρέπει:

- Την έκφραση κριτηρίων που χρησιμοποιούν τις σχέσεις διάταξης μεταξύ στοιχείων, μέσω των τελεστών «Before» και «After».
- Την έκφραση κριτηρίων για συγκεκριμένα δομικά στοιχεία, μέσω του τελεστή «[]», όπου περιέχεται ο αύξων αριθμός τους στο αντικείμενο στο οποίο αυτά περιέχονται.
- Την έκφραση κριτηρίων για τα δομικά στοιχεία – παιδιά των αντικειμένων (ή δομικών στοιχείων αντικειμένων) μέσω του τελεστή «/»²¹.
- Την έκφραση κριτηρίων για τα δομικά στοιχεία – απογόνους των αντικειμένων (ή δομικών στοιχείων αντικειμένων) μέσω του τελεστή «//».

Έτσι, για παράδειγμα, αν τα Άρθρα επιτρέπεται να διαθέτουν ενότητες, η ερώτηση που φαίνεται στο Παράδειγμα 5-5 επιστρέφει τους τίτλους, τις ταυτότητες και τους συγγραφείς των Άρθρων που κάποια ενότητά τους σχετίζεται με τον όρο «Κνωσός» με βαθμό συσχέτισης 100% και με τον όρο «Ανακατασκευή» με βαθμό συσχέτισης 50%.

```
SELECT Τίτλος, Συγγραφέας, Ταυτότητα  
FROM Άρθρο* WHERE Άρθρο/Ενότητα RELATES TO  
<ORTERMSET>  
    <TERM WEIGHT=100%>Κνωσός</TERM>  
    <TERM WEIGHT=50%>Ανακατασκευή</TERM>  
</ORTERMSET>
```

Παράδειγμα 5-5: Ερώτηση με βάση τη Δομή των Αντικειμένων, για «Παιδιά»Αντικειμένων

Σε συνέχεια του προηγούμενου παραδείγματος, έστω ότι οι ενότητες των άρθρων επιτρέπεται να απαρτίζονται από υποενότητες και παραγράφους, ενώ οι υποενότητες επιτρέπεται να απαρτίζονται από παραγράφους. Στην περίπτωση αυτή, η ερώτηση που φαίνεται στο Παράδειγμα 5-6 επιστρέφει τους τίτλους, τις ταυτότητες και τους συγγραφείς των Άρθρων που κάποια παράγραφός τους (είτε αυτή ανήκει σε κάποια ενότητα είτε σε κάποια υποενότητα) σχετίζεται με τον όρο «Κνωσός» με βαθμό συσχέτισης 100% και με τον όρο «Ανακατασκευή» με βαθμό συσχέτισης 50%.

²¹ Ο τελεστής «/» επιλέχθηκε αντί του τελεστή «.» για την έκφραση της σχέσης πατέρα–παιδιού για να μην υπάρξει σύγχυση με τα χαρακτηριστικά των αντικειμένων, στα οποία γίνεται αναφορά μέσω του τελεστή «.»

```

SELECT Τίτλος, Συγγραφέας, Ταυτότητα
FROM Άρθρο* WHERE Άρθρο//Παράγραφος RELATES TO
<ORTERMSET>
    <TERM WEIGHT=100%>Κνωσός</TERM>
    <TERM WEIGHT=50%>Ανακατασκευή</TERM>
</ORTERMSET>

```

Παράδειγμα 5-6: Ερώτηση με βάση τη Δομή των Αντικειμένων, για «Απογόνους» Αντικειμένων

Η ερώτηση που φαίνεται στο Παράδειγμα 5-7 επιστρέφει τους τίτλους, τις ταυτότητες και τους συγγραφείς των Άρθρων που μια από τις 3 πρώτες παραγράφους τους σχετίζεται με τον όρο «Κνωσός» με βαθμό συσχέτισης 100% και με τον όρο «Ανακατασκευή» με βαθμό συσχέτισης 50%.

```

SELECT Τίτλος, Συγγραφέας, Ταυτότητα
FROM Άρθρο* WHERE Άρθρο//Παράγραφος[1 to 3] RELATES TO
<ORTERMSET>
    <TERM WEIGHT=100%>Κνωσός</TERM>
    <TERM WEIGHT=50%>Ανακατασκευή</TERM>
</ORTERMSET>

```

Παράδειγμα 5-7: Ερώτηση με βάση τη Δομή των Αντικειμένων, για συγκεκριμένους «Απογόνους» Αντικειμένων

Αντίστοιχα, η ερώτηση που φαίνεται στο Παράδειγμα 5-8 επιστρέφει τους τίτλους, τις ταυτότητες και τους συγγραφείς των Άρθρων που η 1^η παράγραφος της πρώτης ενότητάς τους σχετίζεται με τον όρο «Κνωσός» με βαθμό συσχέτισης 100% και με τον όρο «Ανακατασκευή» με βαθμό συσχέτισης 50%.

```

SELECT Τίτλος, Συγγραφέας, Ταυτότητα
FROM Άρθρο* WHERE Άρθρο/Ενότητα[1]//Παράγραφος[1] RELATES TO
<ORTERMSET>
    <TERM WEIGHT=100%>Κνωσός</TERM>
    <TERM WEIGHT=50%>Ανακατασκευή</TERM>
</ORTERMSET>

```

Παράδειγμα 5-8: Ερώτηση με βάση τη Δομή των Αντικειμένων, για συγκεκριμένους «Απογόνους» Αντικειμένων

Τέλος, ένα παράδειγμα ερώτησης που χρησιμοποιείται σχέση διάταξης φαίνεται στο Παράδειγμα 5-9, όπου επιλέγονται οι τίτλοι, οι ταυτότητες και οι συγγραφείς των Άρθρων που η πρώτη παράγραφός τους εμφανίζεται πριν την πρώτη ενότητά τους.

```

SELECT Τίτλος, Συγγραφέας, Ταυτότητα
FROM Άρθρο* WHERE Άρθρο//Παράγραφος[1] BEFORE Άρθρο/Ενότητα[1]

```

Παράδειγμα 5-9: Ερώτηση με κριτήρια που περιέχουν Σχέσεις Διάταξης

Ο Πίνακας 5-1, περιέχει και την περιγραφή των αιτήσεων ανάκτησης περιεχομένου αντικειμένων που παρέχει η γλώσσα του διαμεσολαβητή: Για την υποστήριξη αιτήσεων για το περιεχόμενο των αντικειμένων υποστηρίζεται το ιδεατό χαρακτηριστικό Content. Έτσι, για την ανάκτηση του περιεχομένου ολόκληρου του Άρθρου με Ταυτότητα 125 που αποθηκεύεται στην πηγή πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης» εκφράζεται η ερώτηση που φαίνεται στο Παράδειγμα 5-10.

```
SELECT Content
FROM Άρθρο WHERE Άρθρο.Ταυτότητα="125" AND
Source="Βιβλιοθήκη Πολυτεχνείου Κρήτης"
```

Παράδειγμα 5-10: Αίτηση Ανάκτησης Περιεχομένου Αντικειμένου

Εκτός από τις αιτήσεις ανάκτησης ολόκληρου του περιεχομένου των αντικειμένων, η γλώσσα του διαμεσολαβητή υποστηρίζει και τη μερική ανάκτηση του περιεχομένου των αντικειμένων μέσω του τελεστή «-». Έτσι, για την ανάκτηση του περιεχομένου, εκτός από το περιεχόμενο των υποενοτήτων του²², του Άρθρου με Ταυτότητα 125 που αποθηκεύεται στην πηγή πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης», εκφράζεται η ερώτηση που φαίνεται στο Παράδειγμα 5-11.

```
SELECT Content-Υποενότητα
FROM Άρθρο WHERE Άρθρο.Ταυτότητα="125" AND
Source="Βιβλιοθήκη Πολυτεχνείου Κρήτης"
```

Παράδειγμα 5-11: Αίτηση Μερικής Ανάκτησης Περιεχομένου Αντικειμένου

Επιπλέον, η γλώσσα του διαμεσολαβητή υποστηρίζει την ανάκτηση συγκεκριμένων δομικών στοιχείων αντικειμένων. Έτσι, για την ανάκτηση του περιεχομένου της 1^{ης} ενότητας του Άρθρου με Ταυτότητα 125 που αποθηκεύεται στην πηγή πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης» εκφράζεται η ερώτηση που φαίνεται στο Παράδειγμα 5-12.

```
SELECT Content
FROM Άρθρο/Ενότητα[1] WHERE Άρθρο.Ταυτότητα="125" AND
Source="Βιβλιοθήκη Πολυτεχνείου Κρήτης"
```

Παράδειγμα 5-12: Αίτηση Ανάκτησης του Περιεχομένου Δομικού Στοιχείου Αντικειμένου

Στο σημείο αυτό πρέπει να σημειωθεί ότι, προφανώς, η γλώσσα του διαμεσολαβητή επιτρέπει μαζί με το περιεχόμενο των αντικειμένων να ανακτώνται και χαρακτηριστικά τους. Έτσι, στο Παράδειγμα 5-13 φαίνεται η αίτηση ανάκτησης του τίτλου και του περιεχομένου ολόκληρου του Άρθρου με Ταυτότητα 125 που αποθηκεύεται στην πηγή πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης».

²² Στο αποτέλεσμα της αίτησης, στις υποενοότητες του Άρθρου οδηγούν σύνδεσμοι υπερκειμένου, με επικέτες τους τίτλους των αντίστοιχων υποενοτήτων, που τοποθετούνται στη θέση των υποενοτήτων.

```
SELECT Τίτλος, Content
FROM Αρθρο WHERE Αρθρο.Ταυτότητα="125" AND
Source="Βιβλιοθήκη Πολυτεχνείου Κρήτης"
```

Παράδειγμα 5-13: Αίτηση Ανάκτησης Περιεχομένου και Χαρακτηριστικών Αντικειμένου

5.4.2. Επιστρεφόμενη Πληροφορία

Σ' αυτή την ενότητα περιγράφεται η μορφή στην οποία επιστρέφεται η πληροφορία από τη γλώσσα του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Όπως έχει αναφερθεί ήδη, πληροφορία επιστρέφεται από τη γλώσσα του διαμεσολαβητή στις εξής περιπτώσεις:

- Όταν επιστρέφονται τα αποτελέσματα κάποιας ερώτησης.
- Όταν επιστρέφεται το αποτέλεσμα κάποιας αίτησης ανάκτησης αντικειμένων.

Η πληροφορία που επιστρέφεται από τη γλώσσα του διαμεσολαβητή είναι, σε όλες τις περιπτώσεις που αναφέρθηκαν, σε XML μορφή. Έτσι, η πληροφορία που επιστρέφεται από τη γλώσσα του διαμεσολαβητή μπορεί σε κάθε περίπτωση να αντιμετωπιστεί σαν ένα ιδεατό XML έγγραφο που η δομή του είναι σύμφωνη με το DTD Result.dtd που φαίνεται στο DTD 5-1 και περιγράφεται από το URL "http://irida.ced.tuc.gr/XML/Result.dtd".

```
<!ELEMENT Result      (Item* )>
<!ELEMENT Item        (Value* )>
<!ATTLIST Item
  Source CDATA #REQUIRED
  Class  CDATA #REQUIRED
  Weight CDATA #IMPLIED >

<!ELEMENT Error       (#PCDATA )>
<!ELEMENT Warning     (#PCDATA )>
<!ELEMENT Value       (#PCDATA )>
<!ATTLIST Value
  Name CDATA #REQUIRED >
```

DTD 5-1: O DTD Result.dtd

Όπως φαίνεται και στο DTD 5-1, η πληροφορία που επιστρέφεται κάθε φορά από τη γλώσσα του διαμεσολαβητή αναπαρίσταται από ένα στιγμιότυπο του στοιχείου «Result».

Κάθε στιγμιότυπο του στοιχείου «Result» απαρτίζεται από μια λίστα στιγμιοτύπων του στοιχείου «Item», καθένα από τα οποία διαθέτει τρία χαρακτηριστικά:

- Το χαρακτηριστικό «Weight», που είναι μια συμβολοσειρά²³ που η τιμή της δηλώνει το βάρος του τρέχοντος αντικειμένου σε σχέση με την ερώτηση.
- Το χαρακτηριστικό «Class», που είναι μια συμβολοσειρά που η τιμή της δηλώνει το όνομα της κλάσης στην οποία ανήκει το τρέχον στιγμιότυπο του στοιχείου «Item».

²³ Η αναπαράσταση του βάρους ως συμβολοσειράς οφείλεται στο ότι οι DTDs δεν υποστηρίζουν τύπους δεδομένων. Αν, για παράδειγμα, είχε χρησιμοποιηθεί η XML Schema Language αντί των DTDs, ο τύπος του χαρακτηριστικού «Weight» θα ήταν «πραγματικός αριθμός».

- Το χαρακτηριστικό «Source», που είναι μια συμβολοσειρά που η τιμή της δηλώνει το όνομα της πηγής πληροφορίας από την οποία προέρχεται το τρέχον στιγμιότυπο του στοιχείου «Item».

Το περιεχόμενο κάθε στιγμιότυπου του στοιχείου «Item» απαρτίζεται από μια λίστα στιγμιοτύπων του στοιχείου «Value», που αναπαριστούν τις τιμές καθενός από τα πεδία που επιλέγονται με βάση την τρέχουσα ερώτηση για καθένα από τα αντικείμενα που πληρούν τα κριτήριά της. Κάθε στιγμιότυπο του στοιχείου «Value» διαθέτει το χαρακτηριστικό «Name» που είναι μια συμβολοσειρά που η τιμή της δηλώνει το όνομα του πεδίου στο οποίο αντιστοιχεί το τρέχον στιγμιότυπο. Η τιμή του πεδίου δίνεται στο περιεχόμενο του στοιχείου.

Τέλος, αν από τον επεξεργαστή της γλώσσας ερωτήσεων επιστραφεί κάποιο μήνυμα λάθους ή προειδοποίησης, η συμβολοσειρά που περιέχει το μήνυμα επιστρέφεται ως περιεχόμενο ενός στιγμιότυπου του στοιχείου Error ή Warning αντίστοιχα.

Ένα παράδειγμα πληροφορίας που επιστρέφεται ως αποτέλεσμα ερώτησης φαίνεται στο Παράδειγμα 5-14, όπου υποτίθεται ότι το αποτέλεσμα της ερώτησης που φαίνεται στο Παράδειγμα 5-3 απαρτίζεται από τα εξής έγγραφα:

- Το στιγμιότυπο της κλάσης «Άρθρο» με τίτλο «Ιδεατή Ανακατασκευή της Κνωσού», του Χ Υ, από την πηγή πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης», με βαθμό ομοιότητας 100% ως προς την ερώτηση.
- Το στιγμιότυπο της κλάσης «Βιβλίο» με τίτλο «Η Αρχαία Κνωσός», του Α Υ, από την πηγή πληροφορίας «Βιβλιοθήκη Δήμου Χανίων», με βαθμό ομοιότητας 75% ως προς την ερώτηση.
- Το στιγμιότυπο της κλάσης «Αναφορά» με τίτλο «Ιδεατές Ανακατασκευές Αρχαιολογικών Χώρων», του Υ Α, από την πηγή πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης», με βαθμό ομοιότητας 70% ως προς την ερώτηση.
- Το στιγμιότυπο της κλάσης «Άρθρο» με τίτλο «Περιήγηση στην Αρχαία Κνωσό», του Χ Α, από την πηγή πληροφορίας «Βιβλιοθήκη Δήμου Χανίων», με βαθμό ομοιότητας 50% ως προς την ερώτηση.

Στην περίπτωση αυτή, η πληροφορία που επιστρέφεται από τη γλώσσα του διαμεσολαβητή έχει τη μορφή που φαίνεται στο Παράδειγμα 5-14.

```
<Result>
  <Item Class="Άρθρο" Source="Βιβλιοθήκη Πολυτεχνείου Κρήτης"
    Weight="100%">
    <Value Name="Τίτλος"> Ιδεατή Ανακατασκευή της Κνωσού </Value>
    <Value Name="Συγγραφέας"> Χ Υ </Value>
    <Value Name="Ταυτότητα"> 125 </Value>
  </Item>
  <Item Class="Βιβλίο" Source="Βιβλιοθήκη Δήμου Χανίων" Weight="75%">
    <Value Name="Τίτλος"> Η Αρχαία Κνωσός </Value>
    <Value Name="Συγγραφέας"> Α Υ </Value>
    <Value Name="Ταυτότητα"> 145 </Value>
  </Item>
  <Item Class="Αναφορά" Source="Βιβλιοθήκη Πολυτεχνείου Κρήτης"
```

```

Weight="70%">
  <Value Name="Τίτλος">
    Ίδρυματικές Ανακατασκευές Αρχαιολογικών Χώρων
  </Value>
  <Value Name="Συγγραφέας"> Y A </Value>
  <Value Name="Ταυτότητα"> 115 </Value>
</Item>
<Item Class="Άρθρο" Source="Βιβλιοθήκη Δήμου Χανίων" Weight="50%">
  <Value Name="Τίτλος"> Περιήγηση στην Αρχαία Κνωσό </Value>
  <Value Name="Συγγραφέας"> X A </Value>
  <Value Name="Ταυτότητα"> 105 </Value>
</Item>
</Result>

```

Παράδειγμα 5-14: Παράδειγμα Πληροφορίας που επιστρέφεται ως Αποτέλεσμα Ερώτησης

Έστω ότι ο χρήστης που έκανε την ερώτηση και έλαβε τα αποτελέσματα που φαίνονται στο Παράδειγμα 5-14 διαλέγει να δει ολόκληρο το 1^ο αντικείμενο από τα αποτελέσματα. Η επιλογή αυτή γίνεται συνήθως μέσω του γραφικού επιπέδου διεπαφής που χρησιμοποιεί ο χρήστης, με τη διάσχιση ενός συνδέσμου ή το πάτημα ενός κουμπιού και έχει ως συνέπεια την εκτέλεση της ερώτησης που φαίνεται στο Παράδειγμα 5-10. Στην περίπτωση αυτή, η πληροφορία που επιστρέφεται από τη γλώσσα του διαμεσολαβητή έχει τη μορφή που φαίνεται στο Παράδειγμα 5-15.

```

<Result>
  <Item Class="Άρθρο" Source="Βιβλιοθήκη Πολυτεχνείου Κρήτης">
    <Value Name="Content">
      <Άρθρο .....>
        ... περιεχόμενο άρθρου ...
      </Άρθρο>
    </Value>
  </Item>
</Result>

```

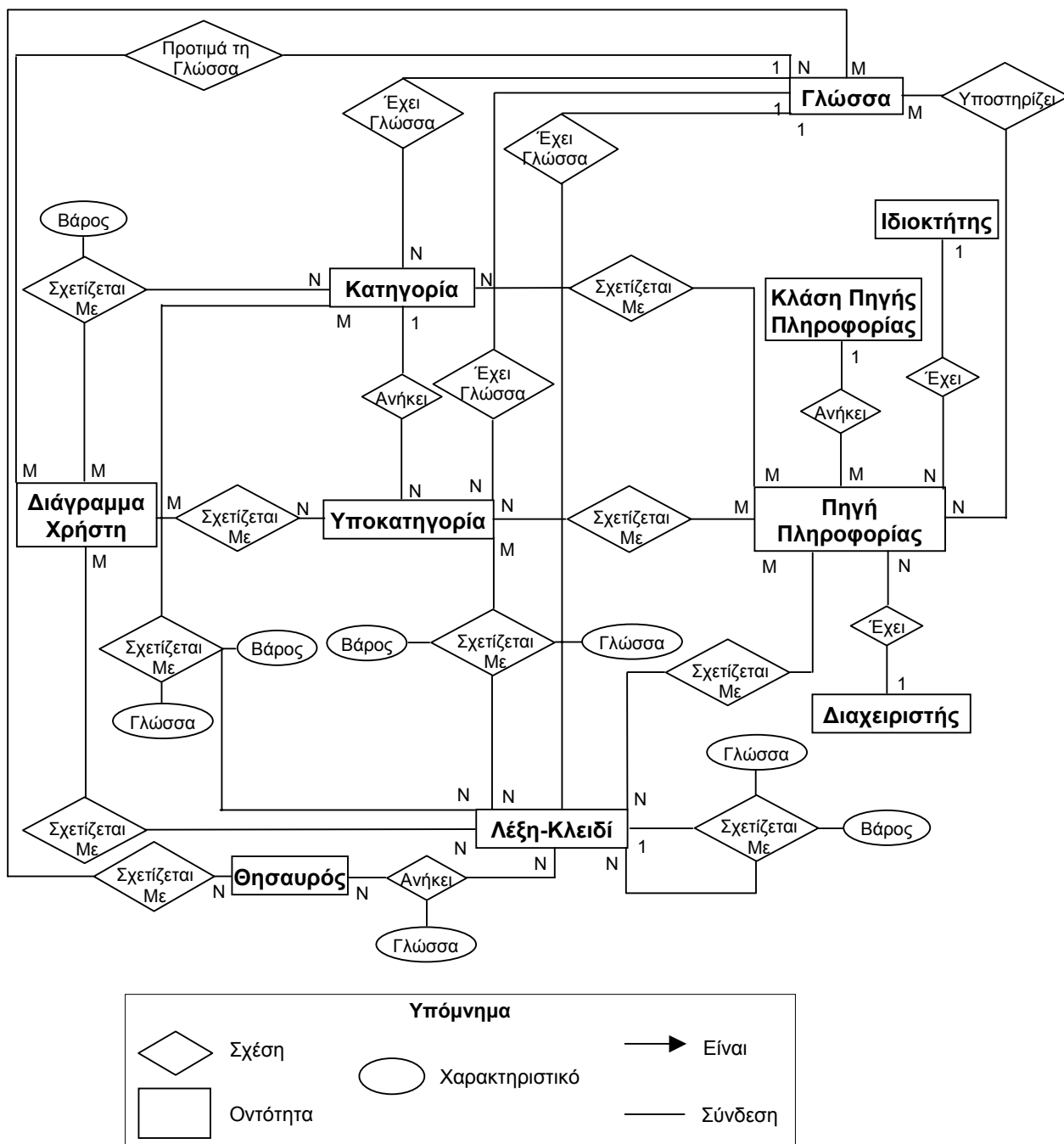
Παράδειγμα 5-15: Παράδειγμα Πληροφορίας που επιστρέφεται ως Αποτέλεσμα Αίτησης

5.5. Βάση Δεδομένων του Διαμεσολαβητή

Η βάση δεδομένων του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» είναι το θέμα αυτού του τμήματος. Η χρησιμότητα και η σημασία της βάσης δεδομένων του διαμεσολαβητή έχει γίνει προφανής, καθώς, όπως έχει ήδη αναφερθεί, αυτή περιέχει:

- Τα διαγράμματα όπου καταγράφονται οι προτιμήσεις των χρηστών.
- Τις δομές που χρησιμοποιούνται για την παροχή πολύγλωσσης υποστήριξης.
- Τη σχετική με τις πηγές πληροφορίας μεταπληροφορία.

Το σχήμα της βάσης δεδομένων του διαμεσολαβητή περιγράφεται από το διάγραμμα οντοτήτων-σχέσεων (E-R διάγραμμα) που απεικονίζεται στο *Διάγραμμα 5-5*. Στο σημείο αυτό πρέπει να αναφερθεί ότι η βάση δεδομένων του διαμεσολαβητή δεν περιέχει τις βάσεις δεδομένων των πηγών πληροφορίας, περιέχει όμως την απαραίτητη μεταπληροφορία για την εκτέλεση των λειτουργιών που σχετίζονται με αυτές. Όπως φαίνεται στο *Διάγραμμα 5-5*, η βάση δεδομένων του διαμεσολαβητή περιέχει τις εξής κατηγορίες οντοτήτων:



Διάγραμμα 5-5: E-R Διάγραμμα που περιγράφει τη Βάση Δεδομένων του Διαμεσολαβητή

- Οντότητες που σχετίζονται με τους θησαυρούς και τα λεξικά που χρησιμοποιούνται από τον πυρήνα του διαμεσολαβητή. Σ' αυτή την κατηγορία οντοτήτων εντάσσονται οι εξής οντότητες:

- *Γλώσσα*. Η οντότητα Γλώσσα αναπαριστά τις γλώσσες που υποστηρίζονται από το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση».
- *Θησαυρός*. Η οντότητα Θησαυρός αναπαριστά τα λεξικά και τους θησαυρούς που αποθηκεύονται στη βάση δεδομένων του διαμεσολαβητή και χρησιμοποιούνται από τον πυρήνα του. Κάθε Θησαυρός συσχετίζεται με μια τουλάχιστον γλώσσα. Οι θησαυροί που σχετίζονται με μια μόνο γλώσσα παρέχουν τους βαθμούς συσχέτισης μεταξύ των όρων της γλώσσας αυτής, ενώ αυτοί που συσχετίζονται με περισσότερες από μια γλώσσες παρέχουν τους βαθμούς συσχέτισης μεταξύ των όρων περισσότερων από μίας γλωσσών.
- Οντότητες που σχετίζονται με την κατηγοριοποίηση της πληροφορίας. Σ' αυτή την κατηγορία οντοτήτων εντάσσονται οι εξής οντότητες:
 - *Λέξη-Κλειδί*. Η οντότητα Λέξη-Κλειδί αναπαριστά τους όρους από τους οποίους απαρτίζονται οι θησαυροί. Κάθε λέξη-κλειδί συσχετίζεται, μέσω της συσχέτισης «Έχει Γλώσσα», με τη γλώσσα της οποίας αποτελεί μέρος. Επιπλέον, με βάση τη γλώσσα της οποίας αποτελεί μέρος, κάθε λέξη-κλειδί συσχετίζεται με ένα τουλάχιστον Θησαυρό μέσω της συσχέτισης «Ανήκει».
 - *Κατηγορία*. Η οντότητα Κατηγορία αναπαριστά τις κατηγορίες στις οποίες μπορεί να εντάσσεται το περιεχόμενο κάποιου αντικειμένου (π.χ. Αθλητικά, Πολιτικά, Επιστημονικά κλπ.). Κάθε κατηγορία συσχετίζεται, μέσω της συσχέτισης «Έχει Γλώσσα», με τη γλώσσα στην οποία ανήκει ο όρος που αποτελεί το όνομά της. Με τις κατηγορίες συσχετίζονται μέσω της συσχέτισης «Σχετίζεται Με», με βάση τη γλώσσα της οποίας αποτελούν μέρος και με διαφορετικό βαθμό συσχέτισης η καθεμιά, οι λέξεις-κλειδιά.
 - *Υποκατηγορία*. Η οντότητα Υποκατηγορία αναπαριστά τις υποκατηγορίες των κατηγοριών στις οποίες μπορεί να εντάσσεται το περιεχόμενο κάποιου αντικειμένου (έτσι, π.χ., υποκατηγορίες που μπορεί να εντάσσονται στην κατηγορία Αθλητικά είναι οι υποκατηγορίες Μπάσκετ, Βόλεϊ κλπ.). Κάθε υποκατηγορία συσχετίζεται, μέσω της συσχέτισης «Έχει Γλώσσα», με τη γλώσσα στην οποία ανήκει ο όρος που αποτελεί το όνομά της. Με τις υποκατηγορίες συσχετίζονται, συσχετίζονται μέσω της συσχέτισης «Σχετίζεται Με», με βάση τη γλώσσα της οποίας αποτελούν μέρος και με διαφορετικό βαθμό συσχέτισης η καθεμιά, οι λέξεις-κλειδιά. Επιπλέον, κάθε υποκατηγορία συσχετίζεται με μια ακριβώς κατηγορία μέσω της συσχέτισης «Ανήκει».
- Οντότητες που σχετίζονται με τα διαγράμματα προτιμήσεων των χρηστών. Σ' αυτή την κατηγορία οντοτήτων εντάσσεται η οντότητα *Διάγραμμα Χρήστη*. Η οντότητα Διάγραμμα Χρήστη αναπαριστά τα διαγράμματα προτιμήσεων των χρηστών. Κάθε διάγραμμα χρήστη συσχετίζεται, μέσω της συσχέτισης «Σχετίζεται Με», με κατηγορίες, υποκατηγορίες και λέξεις κλειδιά που χαρακτηρίζουν το περιεχόμενο των αντικειμένων που τους ενδιαφέρουν τον αντίστοιχο χρήστη. Χαρακτηριστικό στις συσχετίσεις αυτές είναι το βάρος, το οποίο δηλώνει το βαθμό συσχέτισης του διαγράμματος με κατηγορίες, υποκατηγορίες και λέξεις-κλειδιά.

Επιπλέον, τα διαγράμματα προτιμήσεων των χρηστών συσχετίζονται, μέσω της συσχέτισης «Προτιμά τη Γλώσσα», με γλώσσες, οι οποίες είναι οι γλώσσες που προτιμά ο χρήστης στον οποίο αντιστοιχεί το κάθε διάγραμμα.

- Οντότητες που σχετίζονται με τη μεταπληροφορία που αποθηκεύεται στον πυρήνα του διαμεσολαβητή, σχετικά με τις πηγές πληροφορίας. Σ' αυτή την κατηγορία οντοτήτων εντάσσονται οι εξής οντότητες:

- *Πηγή Πληροφορίας*. Η οντότητα Πηγή Πληροφορίας αναπαριστά τις πηγές πληροφορίας που το περιεχόμενό τους γίνεται διαθέσιμο μέσω του διαμεσολαβητή. Κάθε στιγμιότυπο της οντότητας Πηγή Πληροφορίας περιέχει μεταπληροφορία σχετική με την αντίστοιχη πηγή πληροφορίας. Η περιγραφή του περιεχομένου της πηγής πληροφορίας επιτυγχάνεται με τη συσχέτισή της, μέσω της συσχέτισης «Σχετίζεται Με», με κατηγορίες, υποκατηγορίες και λέξεις-κλειδιά.

Επιπλέον, κάθε πηγή πληροφορίας συσχετίζεται, μέσω της συσχέτισης «Υποστηρίζει», με έναν αριθμό γλωσσών, δηλώνοντας έτσι ότι υποστηρίζει τις γλώσσες αυτές.

- *Κλάση Πηγής Πληροφορίας*. Η οντότητα Κλάση Πηγής Πληροφορίας αναπαριστά τις κλάσεις στις οποίες ανήκουν οι πηγές πληροφορίας. Κάθε πηγή πληροφορίας συσχετίζεται με μια ακριβώς κλάση πηγής πληροφορίας μέσω της συσχέτισης «Ανήκει». Παραδείγματα κλάσεων πηγών πληροφορίας είναι η κλάση των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, η κλάση των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα κλπ.
- *Ιδιοκτήτης*. Η οντότητα Ιδιοκτήτης αναπαριστά τους ιδιοκτήτες των πηγών πληροφορίας, με την έννοια ότι αυτά τα φυσικά ή νομικά πρόσωπα έχουν τα πνευματικά δικαιώματα για το περιεχόμενο των αντίστοιχων πηγών πληροφορίας (έτσι, π.χ., ιδιοκτήτης της πηγής πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης» είναι το Πολυτεχνείο Κρήτης). Κάθε πηγή πληροφορίας συσχετίζεται με ένα ακριβώς ιδιοκτήτη μέσω της συσχέτισης «Έχει».
- *Διαχειριστής*. Η οντότητα Διαχειριστής αναπαριστά τους διαχειριστές των πηγών πληροφορίας. Κάθε πηγή πληροφορίας συσχετίζεται με ένα ακριβώς διαχειριστή μέσω της συσχέτισης «Έχει».

Καθώς η πολύγλωσση υποστήριξη και τα διαγράμματα προτιμήσεων των χρηστών δεν αποτελούν τμήμα της παρούσας διατριβής, δε θα γίνει περαιτέρω συζήτηση για τα θέματα αυτά. Αντίθετα, εκτενέστερη αναφορά στο κομμάτι εκείνο της βάσης δεδομένων του διαμεσολαβητή που χρησιμοποιείται για την υποστήριξη της σχετικής με τις πηγές πληροφορίας μεταπληροφορίας θα γίνει στην ενότητα 5.5.1. Το τρέχον τμήμα θα κλείσει στην ενότητα 5.5.2 με την περιγραφή της λειτουργικότητας που παρέχεται από τη βάση δεδομένων του διαμεσολαβητή

5.5.1. Μεταπληροφορία Σχετική με τις Πηγές Πληροφορίας

Αυτή η ενότητα θα επικεντρώσει στο κομμάτι εκείνο της βάσης δεδομένων του διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» που

χρησιμοποιείται για την υποστήριξη της σχετικής με τις πηγές πληροφορίας μεταπληροφορίας. Η μεταπληροφορία αυτή χρησιμοποιείται για την περιγραφή της δομής και της οργάνωσης των πηγών πληροφορίας που γίνονται διαθέσιμες μέσω του διαμεσολαβητή, καθώς και του περιεχομένου της πληροφορίας που αποθηκεύεται σε καθεμιά απ' αυτές.

Η μεταπληροφορία που σχετίζεται με τις πηγές πληροφορίας αποθηκεύεται στη βάση δεδομένων του διαμεσολαβητή και ενημερώνεται κάθε φορά που αλλάζει κάποιο χαρακτηριστικό πηγής πληροφορίας, το περιεχόμενο ή η δομή της.

Η σχετική με κάποια πηγή πληροφορίας μεταπληροφορία αποθηκεύεται στο αντίστοιχο στιγμιότυπο Πηγής Πληροφορίας και στις σχέσεις του με άλλα στιγμιότυπα (κατηγορίες, ιδιοκτήτη, διαχειριστή κλπ.). Έτσι, η μεταπληροφορία που είναι διαθέσιμη για κάθε πηγή πληροφορίας περιλαμβάνει:

- Μεταπληροφορία σχετική με την κλάση πηγών πληροφορίας στην οποία ανήκει η τρέχουσα πηγή πληροφορίας. Στη μεταπληροφορία αυτή περιλαμβάνονται το όνομα της κλάσης, η βάση δεδομένων, η γλώσσα και το περιτύλιγμα των στιγμιότυπων της κλάσης.
- Μεταπληροφορία σχετική με την τρέχουσα πηγή πληροφορίας. Στη μεταπληροφορία αυτή περιλαμβάνονται το όνομα της πηγής πληροφορίας, η διεύθυνσή της στο Διαδίκτυο, η δομή της, ο ιδιοκτήτης της και ο διαχειριστής της.
- Μεταπληροφορία σχετική με το περιεχόμενο της τρέχουσας πηγής πληροφορίας. Στη μεταπληροφορία αυτή περιλαμβάνονται οι κατηγορίες, οι υποκατηγορίες και οι λέξεις-κλειδιά που σχετίζονται με την τρέχουσα πηγή πληροφορίας, καθώς και οι γλώσσες στις οποίες υπάρχει διαθέσιμη πληροφορία στην τρέχουσα πηγή πληροφορίας.

Έτσι, το διάγραμμα κλάσεων που απεικονίζεται στο *Διάγραμμα 5-6* παρουσιάζει την οργάνωση της βάσης δεδομένων του διαμεσολαβητή με ιδιαίτερη έμφαση στις κλάσεις που περιγράφουν τη σχετική με τις πηγές πληροφορίας μεταπληροφορία. Όπως φαίνεται στο *Διάγραμμα 5-6*, κάθε πηγή πληροφορίας ανήκει σε μια συγκεκριμένη κλάση πηγής πληροφορίας (π.χ. πηγή πληροφορίας που περιέχει δομημένα έγγραφα). Για κάθε κλάση πηγής πληροφορίας καθορίζονται τα εξής:

- Η ταυτότητα της τρέχουσας κλάσης πηγών πληροφορίας, που αναπαρίσταται από το χαρακτηριστικό «Ταυτότητα».
- Το όνομα της τρέχουσας κλάσης πηγών πληροφορίας, που αναπαρίσταται από το χαρακτηριστικό «Όνομα». Το όνομα αυτό είναι μοναδικό για κάθε κλάση πηγής πληροφορίας (π.χ. πηγή πληροφορίας που περιέχει δομημένα έγγραφα, πηγή πληροφορίας που περιέχει παραδοσιακά δεδομένα κλπ.).
- Ο τύπος βάσης δεδομένων που υποστηρίζει τα στιγμιότυπα της τρέχουσας κλάσης πηγών πληροφορίας (σχεσιακή, αντικειμενοστραφής κλπ.) , που αναπαρίσταται από το χαρακτηριστικό «Βάση Δεδομένων».
- Το περιτύλιγμα που είναι διαθέσιμο για τα στιγμιότυπα της τρέχουσας κλάσης πηγών πληροφορίας, που αναπαρίσταται από το χαρακτηριστικό «Περιτύλιγμα».

- Ο ιδιοκτήτης της πηγής πληροφορίας, που αναπαρίσταται από ένα στιγμιότυπο της κλάσης «Ιδιοκτήτης», το οποίο συσχετίζεται με την πηγή πληροφορίας.
- Ο διαχειριστής της πηγής πληροφορίας, που αναπαρίσταται από ένα στιγμιότυπο της κλάσης «Διαχειριστής», το οποίο συσχετίζεται με την πηγή πληροφορίας.
- Η δομή της βάσης δεδομένων της τρέχουσας πηγής πληροφορίας, που αναπαρίσταται από το χαρακτηριστικό «Δομή» και αναπαριστά, το μοντέλο πληροφορίας της πηγής πληροφορίας όπως αυτό γίνεται διαθέσιμο από τη γλώσσα και το περιτύλιγμά της. Πιο συγκεκριμένα, το χαρακτηριστικό δομή περιέχει το URL του αρχείου όπου περιέχεται, με τη μορφή RDF σχήματος, το μοντέλο πληροφορίας της πηγής πληροφορίας.

Όταν δημιουργείται μια πηγή πληροφορίας, στη βάση δεδομένων της υπάρχουν μόνο οι βασικές οντότητες που επιτρέπουν την παραπέρα εξέλιξη του σχήματός της, οπότε η δομή της περιγράφεται από το RDF σχήμα που αντιστοιχεί στις κενές βάσεις δεδομένων της κλάσης πηγών πληροφορίας στην οποία αυτή εντάσσεται. Στη συνέχεια, το μοντέλο πληροφορίας της πηγής πληροφορίας μπορεί να εξελιχθεί μέσω της υποστήριξης νέων κλάσεων αντικειμένων, οπότε και η δομή της στη βάση δεδομένων του διαμεσολαβητή ενημερώνεται. Αυτό συμβαίνει καθώς κάθε φορά που υπάρχει κάποια αλλαγή στο αρχείο όπου περιέχεται με τη μορφή RDF σχήματος, το μοντέλο πληροφορίας της πηγής πληροφορίας, η αλλαγή γίνεται ορατή μέσω του URL του, ο οποίος δε μεταβάλλεται και αποτελεί την τιμή του χαρακτηριστικού «Δομή» της πηγής πληροφορίας.

Στο σημείο αυτό πρέπει να σημειωθεί ότι, καθώς τα μοντέλα δυο πηγών πληροφορίας που ανήκουν στην ίδια κλάση πηγής πληροφορίας εξελίσσονται, η δομές τους μπορεί να είναι διαφορετικές. Για παράδειγμα, έστω ότι από κάποιο διαμεσολαβητή υποστηρίζονται δυο πηγές πληροφορίας που περιέχουν έγγραφα: Η πηγή πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης» και η πηγή πληροφορίας «Βιβλιοθήκη Δήμου Χανίων». Αρχικά και οι δύο πηγές πληροφορίας έχουν την ίδια δομή, που περιγράφεται από το RDF σχήμα που αντιστοιχεί στις κενές βάσεις δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Στη συνέχεια, έστω ότι η πρώτη κλάση εγγράφων που υποστηρίζεται από την πηγή πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης» είναι η κλάση εγγράφων «Μεταπτυχιακή Διατριβή», ενώ η πρώτη κλάση εγγράφων που υποστηρίζεται από την πηγή πληροφορίας «Βιβλιοθήκη Δήμου Χανίων» είναι η κλάση εγγράφων «Βιβλίο». Ως συνέπεια της υποστήριξης των νέων κλάσεων, τα μοντέλα πληροφορίας των δύο πηγών πληροφορίας εξελίσσονται και είναι διαφορετικά μεταξύ τους.

Η συσχέτιση των πηγών πληροφορίας με κατηγορίες, υποκατηγορίες και λέξεις κλειδιά αποτελεί τη μεταπληροφορία που σχετίζεται με το περιεχόμενό τους. Επιπλέον, η συσχέτιση των πηγών πληροφορίας με έναν αριθμό γλωσσών δηλώνει σε ποιες γλώσσες περιέχει πληροφορία κάθε πηγή πληροφορίας.

Η μεταπληροφορία που σχετίζεται με τη δομή των πηγών πληροφορίας στην πραγματικότητα ενημερώνεται κάθε φορά που υπάρχει ειδοποίηση από κάποια πηγή πληροφορίας για νέα αντικείμενα που γίνονται διαθέσιμα μέσω αυτής, καθώς η εξέλιξη του σχήματος της βάσης δεδομένων μιας πηγής πληροφορίας είναι συνήθως συνέπεια της

εισαγωγής σε αυτήν αντικειμένων που ανήκουν σε κλάση που δεν υποστηρίζεται ήδη από την πηγή πληροφορίας.

Η μεταπληροφορία που σχετίζεται με τις πηγές πληροφορίας χρησιμοποιείται τόσο κατά τη διάρκεια της επεξεργασίας ερωτήσεων όσο και κατά την παρουσίαση αντικειμένων που προέρχονται από τις πηγές πληροφορίας. Πιο συγκεκριμένα:

- Κατά την επεξεργασία ερωτήσεων, αρχικά χρησιμοποιείται η σχετική με τη δομή και το περιεχόμενο των πηγών πληροφορίας μεταπληροφορία, ώστε να αποφασιστεί αν η ερώτηση θα μεταβιβαστεί στην πηγή πληροφορίας.

Έτσι, για παράδειγμα, αν γίνει μια ερώτηση για έγγραφα, η ερώτηση αυτή δε θα μεταβιβαστεί στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Αντίστοιχα, αν γίνει μια ερώτηση σχετική αποκλειστικά με τον Αθλητισμό, η ερώτηση αυτή δε θα μεταβιβαστεί στις πηγές πληροφορίας που το περιεχόμενό τους σχετίζεται αποκλειστικά με τη Γεωπονική.

Στη συνέχεια, για τις πηγές πληροφορίας στις οποίες θα μεταβιβαστεί κάθε φορά η ερώτηση, χρησιμοποιούνται τα αντίστοιχα περιτυλίγματα, τα οποία εντοπίζονται από τη σχετική με τις κλάσεις πηγών πληροφορίας μεταπληροφορία.

Τέλος, όταν τα αποτελέσματα της ερώτησης επιστραφούν από τις πηγές πληροφορίας, με βάση τη συσχέτιση κάθε πηγής πληροφορίας με τους όρους της ερώτησης βαθμολογούνται με ενιαίο τρόπο τα αντικείμενα που απαρτίζουν το αποτέλεσμα και κανονικοποιούνται τα βάρη τους.

Στο σημείο αυτό πρέπει να σημειωθεί ότι οι τεχνικές που σχετίζονται με την επιλογή των πηγών πληροφορίας στις οποίες θα μεταβιβαστεί μια ερώτηση και την αξιολόγηση των πηγών πληροφορίας βάσει της σχέσης του περιεχομένου της κάθε μιας με τους όρους της ερώτησης δεν αποτελεί τμήμα αυτής της διατριβής αλλά της διατριβής [Πουτ2000]. Επίσης, η ενιαία βαθμολόγηση και η κανονικοποίηση των βαρών των αντικειμένων που επιστρέφει η κάθε πηγή πληροφορίας σε μια ερώτηση που περιέχει όρους ομοιότητας αποτελεί τμήμα της διατριβής [Πουτ2000].

- Κατά την παρουσίαση των αντικειμένων που αποθηκεύονται στις πηγές πληροφορίας, ανάλογα με τις προτιμήσεις του χρήστη στον οποίο θα γίνει η παρουσίαση, είναι πιθανό να παρουσιαστούν σε αυτόν τα ονόματα των αντίστοιχων πηγών πληροφορίας και/ή η σχετική με αυτές μεταπληροφορία.

5.5.2. Λειτουργικότητα της Βάσης Δεδομένων του Διαμεσολαβητή

Το θέμα αυτής της ενότητας είναι η περιγραφή της λειτουργικότητας της βάσης δεδομένων του διαμεσολαβητή. Καθώς ο διαμεσολαβητής αποτελεί το σύστημα διεπαφής των χρηστών με τις πηγές πληροφορίας, η λειτουργικότητα της βάσης δεδομένων του διαμεσολαβητή εξυπηρετεί την αλληλεπίδραση των χρηστών του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» με το διαμεσολαβητή. Όπως φάνηκε στο τμήμα 5.3, η αλληλεπίδραση των χρηστών με το διαμεσολαβητή συνοψίζεται στα εξής σημεία:

- Στην αλληλεπίδραση κατά την υποβολή και εκτέλεση ερωτήσεων και την επιστροφή των αποτελεσμάτων τους.

- Στην αλληλεπίδραση κατά την υποβολή αιτήσεων και την επιστροφή των αποτελεσμάτων τους.
- Στην αλληλεπίδραση κατά την ειδοποίηση των χρηστών για πληροφορία που γίνεται διαθέσιμη και σχετίζεται με τα θέματα που τους ενδιαφέρουν.

Για να γίνει περισσότερο κατανοητή η λειτουργικότητα που παρέχει η βάση δεδομένων του διαμεσολαβητή κατά τη διάρκεια των αλληλεπιδράσεων αυτών, στη συνέχεια παρατίθενται διαγράμματα περιπτώσεων χρήσης για καθένα από τα σημεία που αναφέρθηκαν.

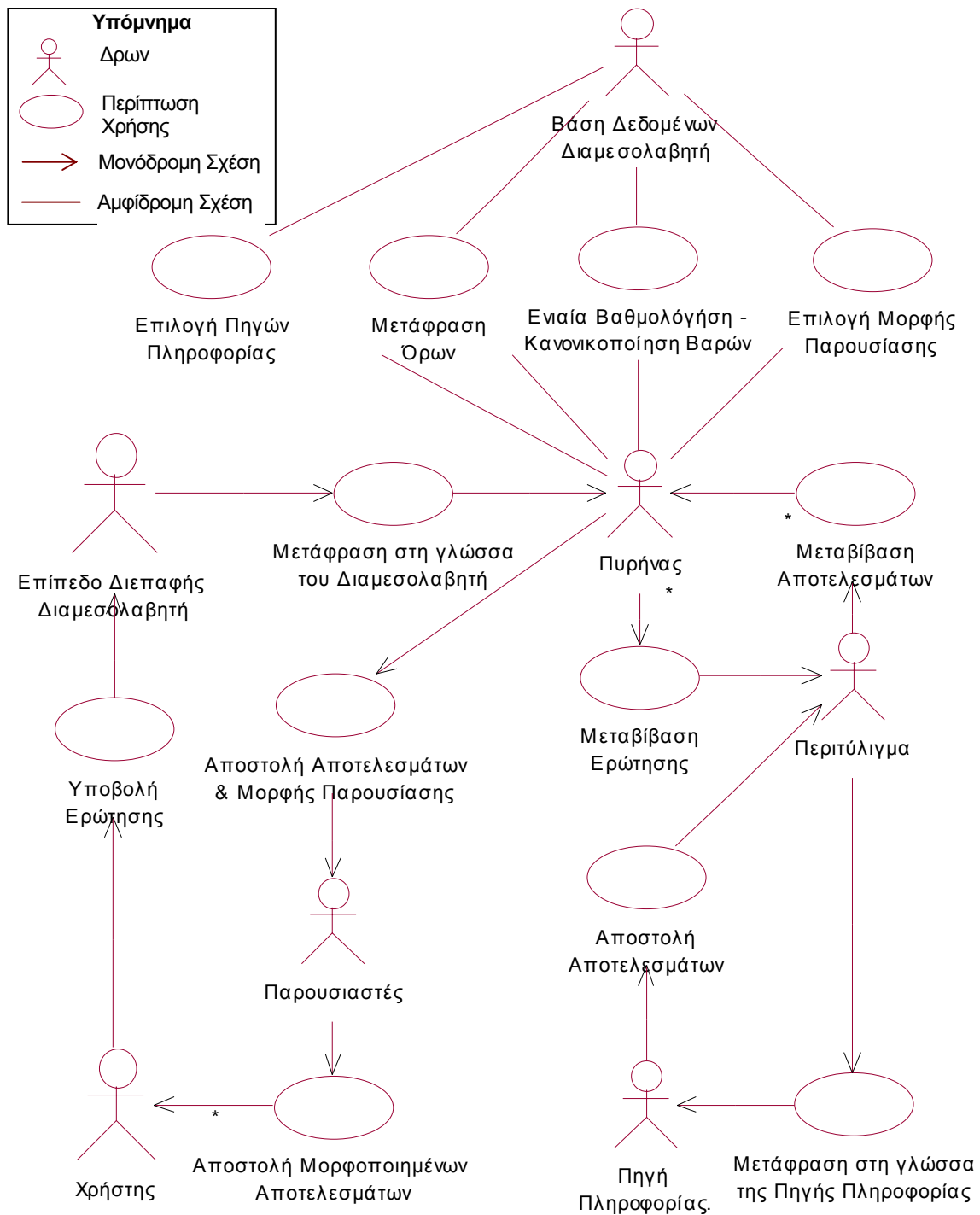
Η λειτουργικότητα της βάσης δεδομένων του διαμεσολαβητή κατά τη διάρκεια της υποβολής και επεξεργασίας ερωτήσεων και της επιστροφής των αποτελεσμάτων τους φαίνεται στο *Διάγραμμα 5-7*, που απεικονίζει ένα διάγραμμα περιπτώσεων χρήσης για τη λειτουργικότητα της βάσης δεδομένων του διαμεσολαβητή κατά την επεξεργασία ερωτήσεων.

Όπως φαίνεται στο *Διάγραμμα 5-7*, οι χρήστες (συνήθως μέσω γραφικών επιπέδων διεπαφής) υποβάλλουν ερωτήσεις, οι οποίες αποστέλλονται στο επίπεδο διεπαφής του διαμεσολαβητή, όπου και μετατρέπονται, αν χρειάζεται, σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή πριν μεταβιβαστούν στον πυρήνα του διαμεσολαβητή. Ο πυρήνας χρησιμοποιεί τη βάση δεδομένων του διαμεσολαβητή (η οποία αποτελεί τμήμα του πυρήνα) για τις εξής λειτουργίες:

- Τη μετάφραση στη βασική γλώσσα που υποστηρίζεται από το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση», αν αυτό είναι απαραίτητο, των όρων που χρησιμοποιούνται στην ερώτηση. Στην περίπτωση αυτή χρησιμοποιούνται οι θησαυροί που αποθηκεύονται στη βάση δεδομένων του διαμεσολαβητή.
- Την επιλογή των πηγών πληροφορίας στις οποίες θα μεταβιβαστεί η ερώτηση. Στην περίπτωση αυτή χρησιμοποιείται η μεταπληροφορία που περιγράφει τις πηγές πληροφορίας: Η ερώτηση δε μεταβιβάζεται στις πηγές πληροφορίας που το περιεχόμενό τους δε σχετίζεται με τους όρους που περιέχονται στην ερώτηση, ούτε σε αυτές που δεν υποστηρίζουν τις κλάσεις που προσδιορίζονται στην ερώτηση. Επιπλέον, τα περιτυλίγματα των πηγών πληροφορίας στις οποίες θα μεταβιβαστεί η ερώτηση εντοπίζονται από τη σχετική με τις κλάσεις πηγών πληροφορίας μεταπληροφορία.

Στη συνέχεια, η ερώτηση προωθείται στα περιτυλίγματα των πηγών πληροφορίας που επελέγησαν, απ' όπου η ερώτηση μεταβιβάζεται στις πηγές πληροφορίας αφού πρώτα μετατραπεί σε μορφή συμβατή με τη γλώσσα των αντίστοιχων πηγών πληροφορίας.

Κάθε πηγή πληροφορίας, αφού εκτελέσει την ερώτηση, μεταβιβάζει το αποτέλεσμα, μέσω του περιτυλίσματός της, όπου το αποτέλεσμα μετατρέπεται σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή, στον πυρήνα του διαμεσολαβητή. Εκεί λαμβάνει χώρα ένας δεύτερος κύκλος αλληλεπίδρασης με τη βάση δεδομένων με σκοπό:



Διάγραμμα 5-7: Διάγραμμα Περιπτώσεων Χρήσης για τη Λειτουργικότητα της Βάσης Δεδομένων του Διαμεσολαβητή κατά την Επεξεργασία Ερωτήσεων

- Την ενιαία βαθμολόγηση και την κανονικοποίηση των βαρών των αντικειμένων που απαρτίζουν το συνολικό αποτέλεσμα της ερώτησης, ανάλογα και με το βαθμό συσχέτισης των όρων της ερώτησης με τις πηγές πληροφορίας από τις οποίες προέρχονται τα αντικείμενα που απαρτίζουν το αποτέλεσμα. Στην περίπτωση αυτή χρησιμοποιείται η σχετική με το περιεχόμενο των πηγών πληροφορίας μεταπληροφορία και πιο συγκεκριμένα οι βαθμοί συσχέτισης των όρων της ερώτησης με το περιεχόμενο των πηγών πληροφορίας.

- Την επιλογή της μορφής παρουσίασης των αποτελεσμάτων της ερώτησης. Στην περίπτωση αυτή ανακτώνται οι προτιμήσεις του χρήστη που σχετίζονται με την παρουσίαση αποτελεσμάτων ερωτήσεων από το διάγραμμα προτιμήσεών του.

Τα ενιαία βαθμολογημένα αποτελέσματα μαζί με τους κανόνες μορφοποίησής τους αποστέλλονται στους παρουσιαστές. Εκεί πραγματοποιείται η μορφοποίηση των αποτελεσμάτων και η μορφοποιημένη πληροφορία αποστέλλεται στους χρήστες. Η αποστολή των αποτελεσμάτων στους χρήστες μπορεί, ανάλογα με τις προτιμήσεις των τελευταίων, να γίνει σε έναν ή περισσότερους κύκλους αλληλεπίδρασης.

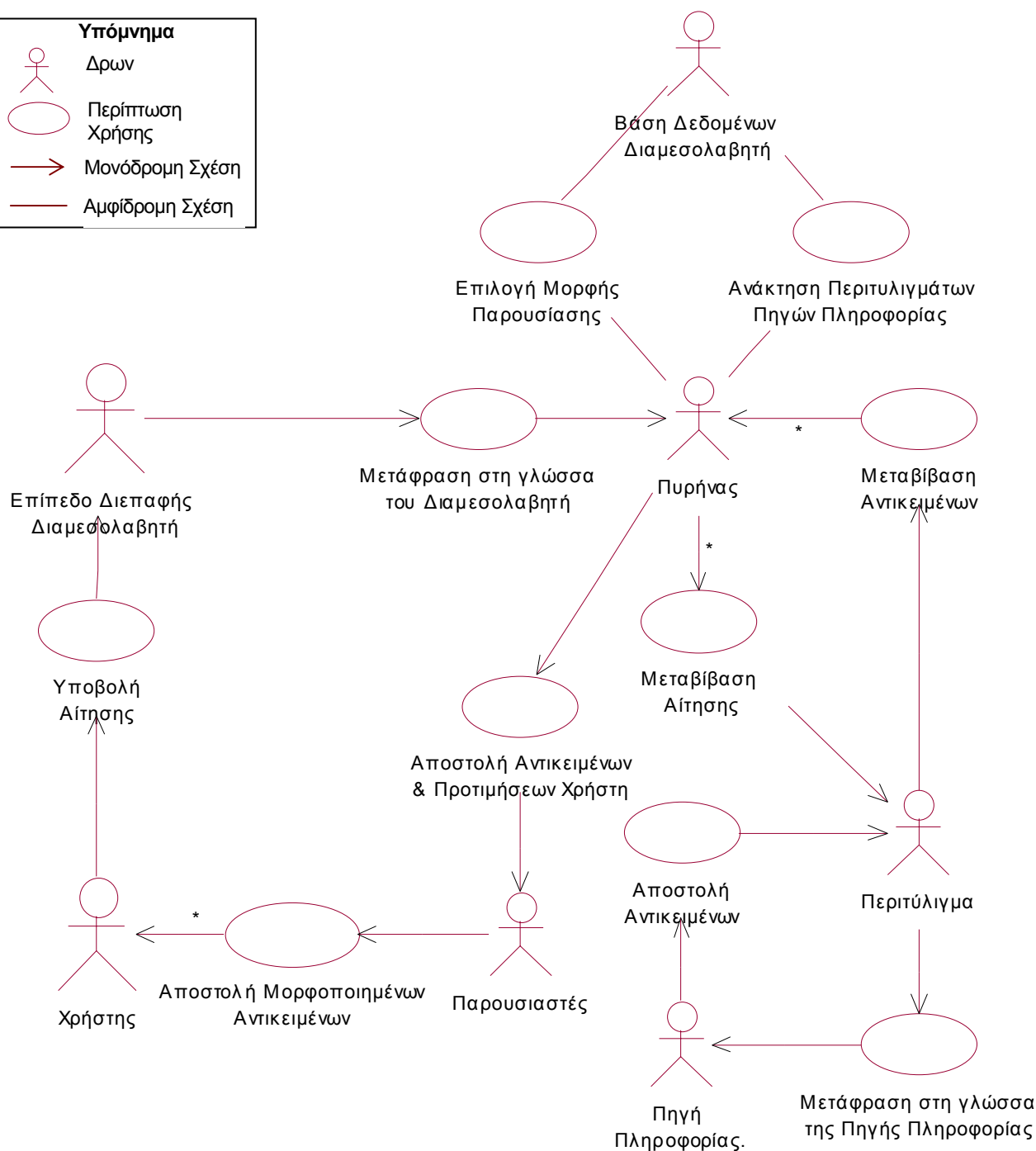
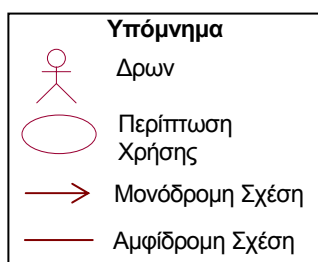
Η λειτουργικότητα της βάσης δεδομένων του διαμεσολαβητή κατά τη διάρκεια της υποβολής αιτήσεων και της επιστροφής των αποτελεσμάτων τους απεικονίζεται στο *Διάγραμμα 5-8*, που είναι ένα διάγραμμα περιπτώσεων χρήσης για τη λειτουργικότητα της βάσης δεδομένων του διαμεσολαβητή κατά την εξυπηρέτηση αιτήσεων.

Όπως φαίνεται στο *Διάγραμμα 5-8*, οι χρήστες (συνήθως μέσω γραφικών επιπέδων διεπαφής) υποβάλλουν αιτήσεις, οι οποίες αποστέλλονται στο επίπεδο διεπαφής του διαμεσολαβητή, όπου και μετατρέπονται, αν χρειάζεται, σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή πριν μεταβιβαστούν στον πυρήνα του διαμεσολαβητή. Ο πυρήνας χρησιμοποιεί τη βάση δεδομένων του διαμεσολαβητή για την ανάκτηση των περιτυλιγμάτων των πηγών πληροφορίας στις οποίες θα προωθηθούν τμήματα της αίτησης: Η πηγή πληροφορίας στην οποία αποθηκεύεται καθένα από τα αιτούμενα αντικείμενα περιέχεται στην αίτηση, και το περιτύλιγμά της προσδιορίζεται μέσα στη μεταπληροφορία που αποθηκεύεται για την κλάση πηγών πληροφορίας στην οποία εντάσσεται.

Στη συνέχεια, τα τμήματα της αίτησης, καθένα από τα οποία αναφέρεται στην ανάκτηση ενός ή περισσότερων αντικειμένων από την ίδια πηγή πληροφορίας, προωθούνται στα περιτυλίγματα των αντίστοιχων πηγών πληροφορίας. Εκεί κάθε τμήμα της αίτησης μεταβιβάζεται στην αντίστοιχη πηγή πληροφορίας αφού πρώτα μετατραπεί σε μορφή συμβατή με τη γλώσσα της.

Κάθε πηγή πληροφορίας μεταβιβάζει το αποτέλεσμα του τμήματος της αίτησης που εκτέλεσε μέσω του περιτυλίγματός της, όπου το αποτέλεσμα μετατρέπεται σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή, στον πυρήνα του διαμεσολαβητή. Εκεί, μετά από αλληλεπίδραση με τη βάση δεδομένων επιλέγεται η μορφή παρουσίασης των αντικειμένων που πρόκειται να επιστραφούν, καθώς ανακτώνται από το διάγραμμα προτιμήσεών του οι προτιμήσεις του χρήστη που σχετίζονται με την παρουσίαση των αντικειμένων που ανήκουν σε κάθε μια από τις κλάσεις που στιγμιότυπά της περιλαμβάνονται στα αποτελέσματα της ερώτησης.

Τα αντικείμενα μαζί με τους κανόνες μορφοποίησής τους αποστέλλονται στους παρουσιαστές. Εκεί πραγματοποιείται η μορφοποίηση των αντικειμένων και η μορφοποιημένη πληροφορία αποστέλλεται στους χρήστες. Η αποστολή των αντικειμένων στους χρήστες μπορεί, ανάλογα με τις προτιμήσεις των τελευταίων, να γίνει σε έναν ή περισσότερους κύκλους αλληλεπίδρασης (ειδικά αν οι χρήστες έχουν επιλέξει να μην τους παρουσιάζεται ολόκληρο το περιεχόμενο των αντικειμένων).

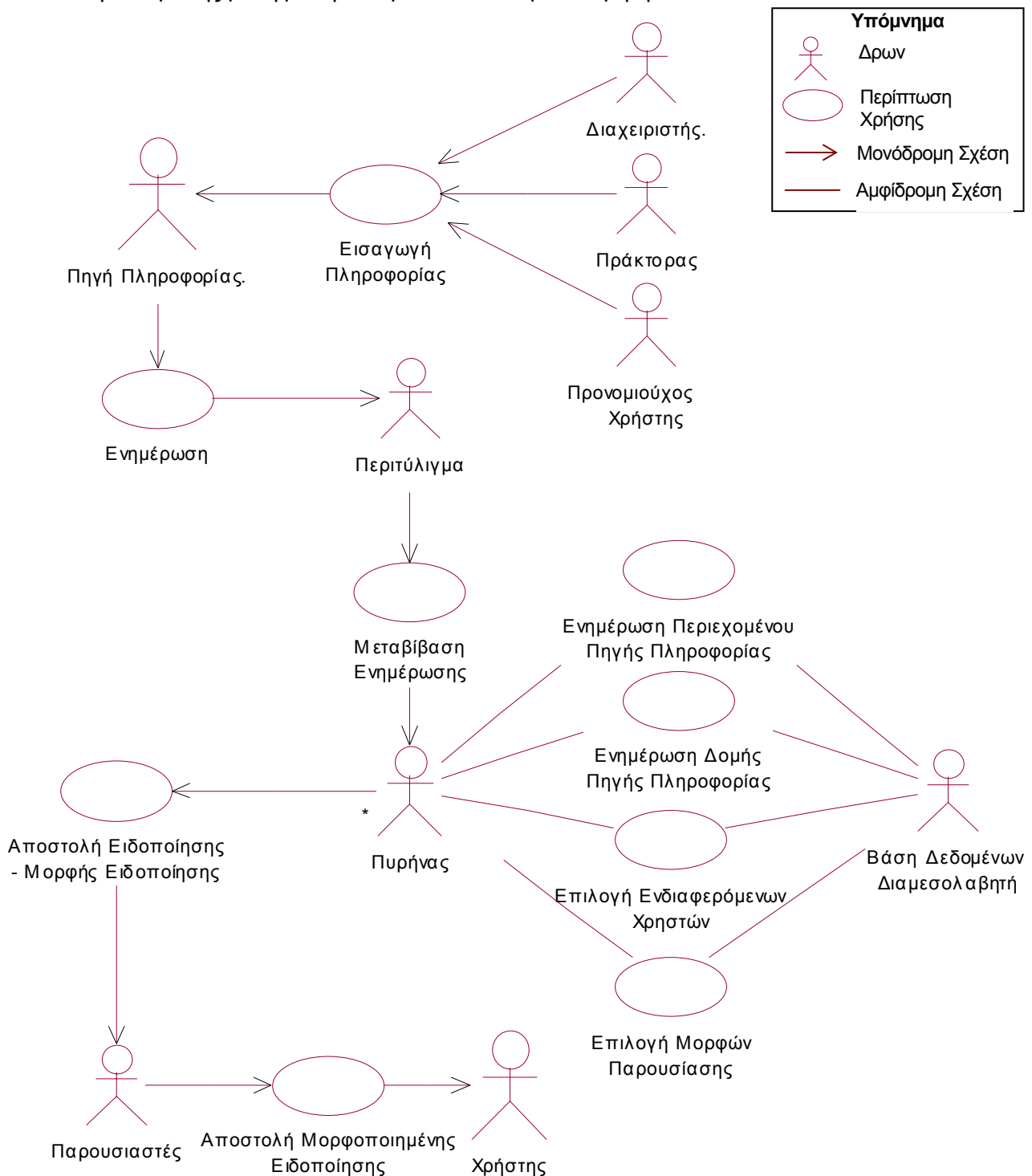


Διάγραμμα 5-8: Διάγραμμα Περιπτώσεων Χρήσης για τη Λειτουργικότητα της Βάσης Δεδομένων του Διαμεσολαβητή κατά την Εξυπηρέτηση Αιτήσεων

Η λειτουργικότητα της βάσης δεδομένων του διαμεσολαβητή κατά τη διάρκεια της ειδοποίησης των χρηστών για πληροφορία που γίνεται διαθέσιμη και σχετίζεται με τα θέματα που τους ενδιαφέρουν απεικονίζεται στο Διάγραμμα 5-9, που είναι ένα διάγραμμα περιπτώσεων χρήσης για τη λειτουργικότητα της βάσης δεδομένων του διαμεσολαβητή κατά την εισαγωγή πληροφορίας στις πηγές πληροφορίας.

Όπως φαίνεται στο Διάγραμμα 5-9, ο διαχειριστής μιας πηγής πληροφορίας, κάποιος προνομιούχος χρήστης που έχει το δικαίωμα να εισάγει πληροφορία στην πηγή πληροφορίας ή κάποιος πράκτορας υπεύθυνος για την ενημέρωση της πηγής

πληροφορίας πραγματοποιεί την εισαγωγή πληροφορίας στην πηγή πληροφορίας. Η πηγή πληροφορίας στέλνει, μέσω του περιτύλιγμάς της, ένα μήνυμα ενημέρωσης του διαμεσολαβητή, το οποίο προωθείται στον πυρήνα. Εκεί λαμβάνει χώρα ένας κύκλος αλληλεπίδρασης με τη βάση δεδομένων του διαμεσολαβητή έτσι ώστε:



Διάγραμμα 5-9: Διάγραμμα Περιπτώσεων Χρήσης για τη Λειτουργικότητα της Βάσης Δεδομένων του Διαμεσολαβητή κατά την Εισαγωγή Πληροφορίας στις Πηγές Πληροφορίας

- Να ενημερωθεί, αν χρειάζεται, η μεταπληροφορία που σχετίζεται με το περιεχόμενο και/ή τη δομή της πηγής πληροφορίας στην οποία πραγματοποιήθηκε η εισαγωγή:

- Αν η κλάση στην οποία ανήκει το αντικείμενο που εισήχθη ξεκινά να υποστηρίζεται από την πηγή πληροφορίας με την εισαγωγή του²⁴, ενημερώνεται η σχετική με τη δομή της πηγής πληροφορίας μεταπληροφορία. Η ενημέρωση αυτή γίνεται αυτόματα, καθώς στη βάση δεδομένων του διαμεσολαβητή αποθηκεύεται ο URL του RDF σχήματος που περιγράφει τη δομή της πηγής πληροφορίας
- Αν το περιεχόμενο του αντικειμένου τροποποιεί τη σχέση της πηγής πληροφορίας με κάποια κατηγορία, υποκατηγορία ή λέξη-κλειδί, πραγματοποιούνται οι αντίστοιχες ενημερώσεις στη βάση δεδομένων του διαμεσολαβητή.
- Να επιλεγούν οι χρήστες που θα ειδοποιηθούν για το ότι το αντικείμενο που εισήχθη έγινε διαθέσιμο. Οι χρήστες αυτοί επιλέγονται με βάση την ομοιότητα των ενδιαφερόντων τους με το περιεχόμενο του αντικειμένου. Η σχετική με τα ενδιαφέροντά τους πληροφορία ανακτάται από το διάγραμμα προτιμήσεών τους. Το θέμα της αξιολόγησης της σχέσης του περιεχομένου του αντικειμένου με τα ενδιαφέροντα των χρηστών περιγράφεται στη διατριβή [Πουτ2000].
- Να ανακτηθεί από το διάγραμμα προτιμήσεών του, για καθένα από τους χρήστες που θα ειδοποιηθούν για το ότι το αντικείμενο που εισήχθη έγινε διαθέσιμο, η επιθυμητή απ' αυτόν μορφή της ειδοποίησης που θα του σταλεί.

Στη συνέχεια, ο πυρήνας του διαμεσολαβητή αποστέλλει στους παρουσιαστές το περιεχόμενο της ειδοποίησης και το σύνολο των επιθυμητών μορφών της. Οι παρουσιαστές μορφοποιούν κατάλληλα για τον κάθε χρήστη την ειδοποίηση και την αποστέλλουν σ' αυτόν.

5.6. Ανακεφαλαίωση

Στο κεφάλαιο αυτό περιγράφηκε ο Διαμεσολαβητής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Πιο συγκεκριμένα, καθορίστηκε η *Αρχιτεκτονική του Διαμεσολαβητή* και περιγράφηκαν τα *Τμήματα της Αρχιτεκτονικής του Διαμεσολαβητή* και η *Λειτουργικότητα του Διαμεσολαβητή*. Ο σχεδιασμός τόσο της συνολικής αρχιτεκτονικής του διαμεσολαβητή όσο και των τμημάτων από τα οποία αυτή απαρτίζεται, βασίστηκε στο συνδυασμό των τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό.

Επιπλέον, περιγράφηκαν αναλυτικότερα η *Γλώσσα του Διαμεσολαβητή* και η *Βάση Δεδομένων του Διαμεσολαβητή*, καθώς αποτελούν, μαζί με την αρχιτεκτονική του διαμεσολαβητή και την περιγραφή της λειτουργικότητάς του, συνεισφορά της παρούσας διατριβής.

²⁴ Αυτό συμβαίνει συχνά όταν εισάγεται σε μια πηγή πληροφορίας κάποιο αντικείμενο που εντοπίστηκε από πράκτορα.

6^ο Κεφάλαιο

Πηγές Πληροφορίας

Το κεφάλαιο αυτό επικεντρώνει στην περιγραφή των πηγών πληροφορίας που αποτελούν τμήμα του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», με ιδιαίτερη έμφαση στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Οι πηγές πληροφορίας είναι ανεξάρτητα πληροφοριακά συστήματα που μπορεί να διαφέρουν μεταξύ τους τόσο στη μορφή της πληροφορίας που περιέχουν (π.χ. δομημένα έγγραφα, παραδοσιακά δεδομένα, αντικείμενα πολυμέσων κλπ.) όσο και στα τεχνικά χαρακτηριστικά τους όπως είναι το είδος της βάσης δεδομένων τους (π.χ. σχεσιακή, αντικειμενοστραφής κλπ.), η οργάνωση της βάσης δεδομένων τους σε επίπεδο σχήματος και η γλώσσα μέσω της οποίας εξάγουν τη λειτουργικότητά τους. Επιπλέον, οι πηγές πληροφορίας που αποτελούν μέρος του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» μπορούν να χαρακτηριστούν ως συγχρονιζόμενες, καθώς το περιεχόμενό τους ενημερώνεται από πράκτορες για αντικείμενα που οι τελευταίοι εντοπίζουν στο Διαδίκτυο και το περιεχόμενό τους ταιριάζει με αυτό των πηγών πληροφορίας.

Στο τμήμα 6.1 περιγράφεται η *Αρχιτεκτονική των Πηγών Πληροφορίας*, όπου φαίνεται ότι όλες οι πηγές πληροφορίας διαθέτουν δυο βασικά συστατικά λογισμικού:

- Τη *Γλώσσα της Πηγής Πληροφορίας*, που παρέχει το επίπεδο διεπαφής της πηγής πληροφορίας τόσο με το Διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ'

Απαίτηση» όσο και με τα εργαλεία και τους μηχανισμούς που χρησιμοποιούνται για τη διαχείρισή της.

- Τη *Βάση Δεδομένων της Πηγής Πληροφορίας*, που είναι μια (συνήθως σχεσιακή) βάση δεδομένων όπου αποθηκεύονται αντικείμενα, σύνδεσμοι σε απομακρυσμένα (ως προς την πηγή πληροφορίας) αντικείμενα ή τμήματα αντικειμένων και ένα σύνολο μεταπληροφορίας που σχετίζεται με τα αντικείμενα (τοπικά και απομακρυσμένα ως προς την πηγή πληροφορίας).

Στο τμήμα 6.2 περιγράφεται η *Λειτουργικότητα των Πηγών Πληροφορίας*, η οποία συνοψίζεται στα εξής σημεία:

- Αρχικοποίηση της βάσης δεδομένων μιας πηγής πληροφορίας.
- Υποστήριξη νέων κλάσεων στη βάση δεδομένων μιας πηγής πληροφορίας.
- Εισαγωγή νέων αντικειμένων στη βάση δεδομένων μιας πηγής πληροφορίας.
- Δεικτοδότηση, μέσα σε μια πηγή πληροφορίας, απομακρυσμένων αντικειμένων που το περιεχόμενό τους σχετίζεται με αυτό της πηγής πληροφορίας.
- Αναζήτηση (τοπικών και απομακρυσμένων ως προς μια πηγή πληροφορίας) αντικειμένων και/ή τμημάτων αντικειμένων που ικανοποιούν συγκεκριμένα κριτήρια σχετικά με το περιεχόμενο και/ή τη δομή τους.
- Ανάκτηση αντικειμένων και/ή τμημάτων αντικειμένων που αποθηκεύονται στη βάση δεδομένων μιας πηγής πληροφορίας.

Στο τμήμα 6.3 περιγράφονται οι *Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα*, οι οποίες είναι ανεξάρτητα συστήματα αποθήκευσης και διαχείρισης εγγράφων και συνδέσμων.

Στο τμήμα 6.4 περιγράφονται οι *Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα*, οι οποίες βασίζονται σε σχεσιακές βάσεις δεδομένων και το περιεχόμενό τους είναι, σε ότι αφορά τη δομή και τη μορφή του, αντίστοιχο με αυτό που αποθηκεύεται παραδοσιακά στα πληροφοριακά συστήματα που βασίζονται σε σχεσιακές βάσεις δεδομένων.

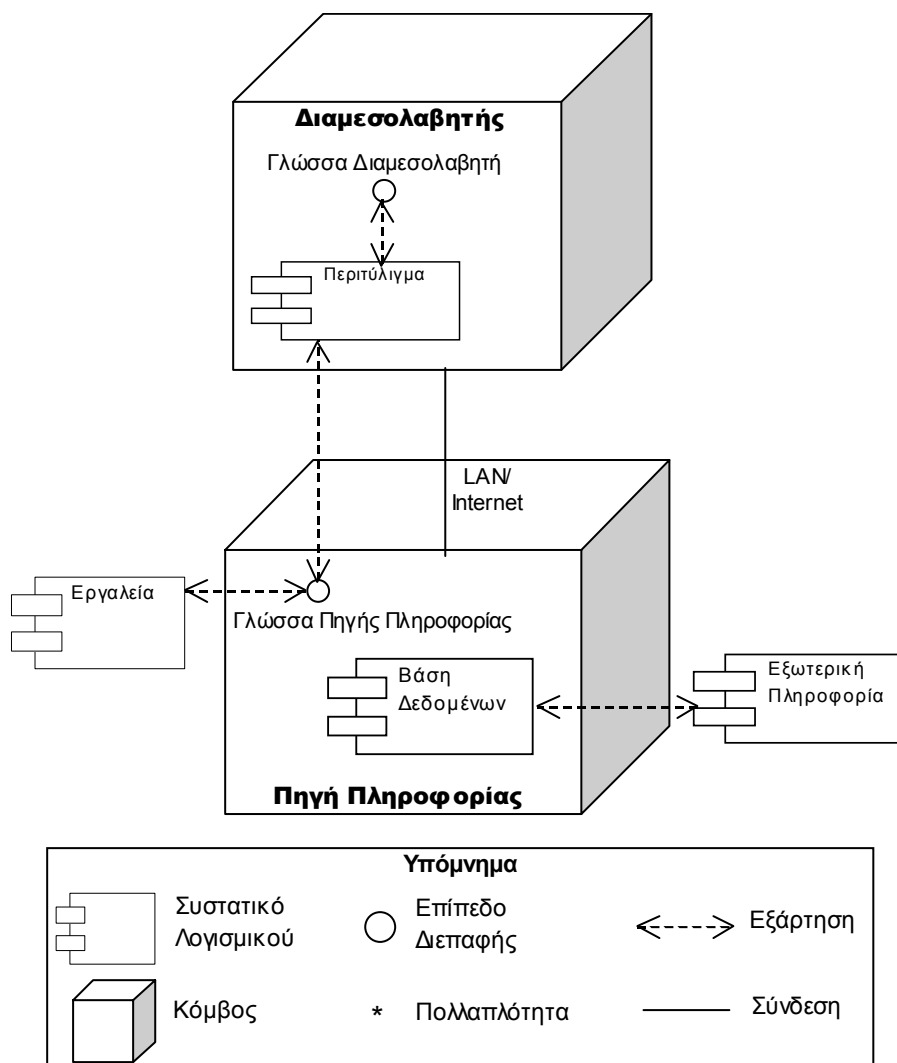
Τέλος, το κεφάλαιο κλείνει με μια σύντομη *Ανακεφαλαίωση* στο τμήμα 6.5.

6.1. Αρχιτεκτονική των Πηγών Πληροφορίας

Οι Πηγές Πληροφορίας είναι αυτόνομα πληροφοριακά συστήματα που μέσω της γλώσσας τους επιτρέπουν τόσο την πρόσβαση στο περιεχόμενό τους όσο και την ενημέρωση του περιεχομένου τους. Στα πλαίσια του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» οι πηγές πληροφορίας επικοινωνούν με τους χρήστες μέσω του διαμεσολαβητή, ο οποίος «εκδίδει» τα περιεχόμενά τους στο Διαδίκτυο. Ο διαμεσολαβητής επικοινωνεί με τις πηγές πληροφορίας που ανήκουν στην ίδια κλάση μέσω του κοινού περιτυλίγμάτος τους.

Σ' αυτό το τμήμα γίνεται μια σύντομη περιγραφή της γενικευμένης αρχιτεκτονικής των πηγών πληροφορίας στα πλαίσια συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Η

γενικευμένη αρχιτεκτονική των πηγών πληροφορίας δίνεται μέσω ενός διαγράμματος παράταξης στο Διάγραμμα 6-1.



Διάγραμμα 6-1: Διάγραμμα Παράταξης για τη Γενικευμένη Αρχιτεκτονική των Πηγών Πληροφορίας

Όπως φαίνεται στο Διάγραμμα 6-1, όλες οι πηγές πληροφορίας διαθέτουν τα εξής βασικά συστατικά λογισμικού:

- Τη *Γλώσσα της Πηγής Πληροφορίας*, που παρέχει το επίπεδο διεπαφής της πηγής πληροφορίας τόσο με το Διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» όσο και με τα εργαλεία που χρησιμοποιούνται για τη δημιουργία, τη διαχείριση και την ενημέρωση των πηγών πληροφορίας, τα οποία περιγράφονται αναλυτικά στο 7^ο Κεφάλαιο. Οι πηγές πληροφορίας που ανήκουν στην ίδια κλάση έχουν κοινή γλώσσα.
- Τη *Βάση Δεδομένων της Πηγής Πληροφορίας*, που είναι μια (συνήθως σχεσιακή) βάση δεδομένων όπου αποθηκεύονται αντικείμενα, σύνδεσμοι σε απομακρυσμένα (ως προς την πηγή πληροφορίας) αντικείμενα ή τμήματα αντικείμενων και ένα σύνολο μεταπληροφορίας που σχετίζεται με τα αντικείμενα (τοπικά και απομακρυσμένα ως προς την πηγή πληροφορίας). Παρ' όλο που όλες οι πηγές πληροφορίας που ανήκουν στην ίδια κλάση έχουν βάσεις δεδομένων του ίδιου τύπου, οι οποίες αρχικά έχουν όλες

το ίδιο σχήμα και το ίδιο μοντέλο πληροφορίας, στην πορεία τα σχήματα και τα μοντέλα πληροφορίας των διαφορετικών πηγών πληροφορίας που ανήκουν στην ίδια κλάση μπορεί να εξελιχθούν διαφορετικά, μέσω της υποστήριξης διαφορετικών κλάσεων.

Λεπτομερέστερη περιγραφή της αρχιτεκτονικής των πηγών πληροφορίας και των τμημάτων από τα οποία αυτή απαρτίζεται έχει νόημα να γίνει ξεχωριστά για κάθε κλάση πηγών πληροφορίας. Μ' αυτό τον τρόπο θα είναι δυνατό να εντοπισθούν οι ιδιαιτερότητες κάθε κλάσης πηγών πληροφορίας και να γίνει λεπτομερέστερη περιγραφή των στιγμιοτύπων της. Καθώς η παρούσα διατριβή επικεντρώνει στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, ιδιαίτερη αναφορά σ' αυτές τις κλάσεις πηγών πληροφορίας γίνεται στα τμήματα 6.3 και 6.4 αντίστοιχα.

6.2. Λειτουργικότητα των Πηγών Πληροφορίας

Το θέμα αυτού του τμήματος είναι η περιγραφή της βασικής λειτουργικότητας που παρέχεται από όλες τις πηγές πληροφορίας (μέσω των γλωσσών τους), ανεξάρτητα από την κλάση πηγών πληροφορίας στην οποία εντάσσονται. Στο σημείο αυτό πρέπει να σημειωθεί ότι ο σχεδιασμός και η υλοποίηση των εργαλείων και των μηχανισμών που υποστηρίζουν τη λειτουργικότητα αυτή περιγράφεται στο 7^ο κεφάλαιο, θέμα του οποίου είναι η *Διαχείριση των Πηγών Πληροφορίας*. Η λειτουργικότητα των πηγών πληροφορίας απεικονίζεται στο Διάγραμμα 6-2, που είναι ένα γενικευμένο διάγραμμα περιπτώσεων χρήσης, και συνοψίζεται στα εξής σημεία:

- Αρχικοποίηση της βάσης δεδομένων μιας πηγής πληροφορίας. Μ' αυτό τον τρόπο ο διαχειριστής μιας πηγής πληροφορίας μπορεί να δημιουργήσει μια κενή βάση δεδομένων, που περιέχει μόνο τις οντότητες που είναι απαραίτητες για τη λειτουργία της πηγής πληροφορίας και την εξέλιξη του μοντέλου πληροφορίας και του σχήματός της. Ο μηχανισμός που υποστηρίζει τη λειτουργία της αρχικοποίησης της βάσης δεδομένων μιας πηγής πληροφορίας περιγράφεται στο τμήμα 7.1.
- Υποστήριξη νέων κλάσεων στη βάση δεδομένων μιας πηγής πληροφορίας. Αυτό σημαίνει ότι παρέχεται η δυνατότητα να εξελιχθεί το σχήμα της βάσης δεδομένων των πηγών πληροφορίας έτσι ώστε να υποστηριχθούν νέες κλάσεις. Με τον τρόπο αυτό δίνεται η δυνατότητα υποστήριξης νέων κλάσεων εγγράφων από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και η δυνατότητα υποστήριξης νέων οντοτήτων και συσχετίσεων από τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της υποστήριξης νέων κλάσεων στη βάση δεδομένων μιας πηγής πληροφορίας περιγράφονται στο τμήμα 7.2.
- Εισαγωγή νέων αντικειμένων στη βάση δεδομένων μιας πηγής πληροφορίας. Με τον τρόπο αυτό δίνεται η δυνατότητα εισαγωγής νέων στιγμιοτύπων στις βάσεις δεδομένων των πηγών πληροφορίας από:
 - Το διαχειριστή κάθε πηγής πληροφορίας.
 - Κάποιο προνομιούχο χρήστη (τοπικό ή απομακρυσμένο), που ο κωδικός πρόσβασής του του παρέχει τη δυνατότητα εισαγωγής δεδομένων.

- Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της εισαγωγής νέων αντικειμένων στη βάση δεδομένων μιας πηγής πληροφορίας περιγράφονται στο τμήμα 7.3.



- Δεικτοδότηση, μέσα σε μια πηγή πληροφορίας, απομακρυσμένων αντικειμένων που το περιεχόμενό τους σχετίζεται με αυτό της πηγής πληροφορίας, τα οποία όμως δεν εισάγονται (τουλάχιστον όχι εξ' ολοκλήρου) στη βάση δεδομένων της πηγής πληροφορίας. Αυτό σημαίνει ότι μπορεί να εντοπιστούν κάποια αντικείμενα που το περιεχόμενό τους είναι σχετικό με αυτό της πηγής πληροφορίας, τα οποία όμως δεν εισάγονται εξ' ολοκλήρου στη βάση δεδομένων²⁵ της. Στην περίπτωση αυτή ενημερώνεται η πηγή πληροφορίας (και ειδικότερα η βάση δεδομένων, οι δομές δεικτοδότησης και οι μέθοδοι πρόσβασης που η πηγή πληροφορίας διαθέτει) για το περιεχόμενο των αντικειμένων αυτών και για την τοποθεσία στην οποία αυτά βρίσκονται. Έτσι γίνεται δυνατή η παραπομπή των χρηστών στα αντικείμενα αυτά όταν οι τελευταίοι εκφράσουν κάποια ερώτηση, τα κριτήρια της οποίας πληρούνται και από αυτά τα αντικείμενα. Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της δεικτοδότησης σε μια πηγή πληροφορίας των αντικειμένων που το περιεχόμενό τους σχετίζεται με αυτό της πηγής πληροφορίας, τα οποία όμως δεν εισάγονται (τουλάχιστον όχι εξ' ολοκλήρου) στη βάση δεδομένων της πηγής πληροφορίας, περιγράφονται στο τμήμα 7.5.
- Αναζήτηση (τοπικών και απομακρυσμένων ως προς μια πηγή πληροφορίας) αντικειμένων και/ή τμημάτων αντικειμένων που ικανοποιούν συγκεκριμένα κριτήρια σχετικά με το περιεχόμενο και/ή τη δομή τους. Έτσι γίνεται δυνατή η υποστήριξη από τις πηγές πληροφορίας τόσο ερωτήσεων δίτιμης λογικής όσο και ερωτήσεων ομοιότητας για τη δομή και το περιεχόμενο των αντικειμένων. Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της αναζήτησης αντικειμένων και/ή τμημάτων αντικειμένων περιγράφονται στο τμήμα 7.4.
- Ανάκτηση αντικειμένων και/ή τμημάτων αντικειμένων που αποθηκεύονται στη βάση δεδομένων μιας πηγής πληροφορίας. Με τον τρόπο αυτό γίνεται δυνατή η πρόσβαση των χρηστών στην πληροφορία που τους ενδιαφέρει και γίνεται διαθέσιμη από κάποια πηγή πληροφορίας. Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της ανάκτησης αντικειμένων και/ή τμημάτων αντικειμένων περιγράφονται στο τμήμα 7.4.

6.3. Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Το θέμα αυτού του τμήματος είναι η περιγραφή των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και η καταγραφή της λειτουργικότητάς τους στα πλαίσια του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».

Οι πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, στα πλαίσια του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», είναι ανεξάρτητα συστήματα αποθήκευσης και διαχείρισης εγγράφων και συνδέσμων. Οι πηγές πληροφορίας που περιέχουν δομημένα έγγραφα διαθέτουν *Μεθόδους Πρόσβασης (Access Methods)* που επιτρέπουν την ανάκτηση τόσο τοπικών όσο και απομακρυσμένων εγγράφων. Για να εξασφαλιστούν η ικανοποιητική κάλυψη σε ότι αφορά απομακρυσμένα έγγραφα που σχετίζονται με τα

²⁵ Αυτό μπορεί να συμβαίνει για λόγους πνευματικών δικαιωμάτων, για λόγους οικονομίας χώρου αποθήκευσης, επειδή ο βαθμός συσχέτισης των αντικειμένων με το περιεχόμενο της πηγής πληροφορίας είναι μικρός κλπ.

θέματα που καλύπτονται από το περιεχόμενο μιας πηγής πληροφορίας και η ενημέρωσή της για νέα έγγραφα που γίνονται διαθέσιμα, χρησιμοποιείται ένας *Πράκτορας (Agent)* που ενημερώνει την πηγή πληροφορίας όταν ανακαλύψει κάποιο έγγραφο που το περιεχόμενό του σχετίζεται με αυτό της πηγής πληροφορίας. Έτσι, στην πραγματικότητα, οι πηγές πληροφορίας που περιέχουν δομημένα έγγραφα αποτελούν ανεξάρτητες Ψηφιακές Βιβλιοθήκες.

Στην παρούσα φάση του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», οι πηγές πληροφορίας που περιέχουν δομημένα έγγραφα υποστηρίζουν XML έγγραφα.

Οι πηγές πληροφορίας που περιέχουν δομημένα έγγραφα επικοινωνούν με το διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» μέσω του κοινού περιτύλιγμά τους: Το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα είναι υπεύθυνο για τη μετάφραση ερωτήσεων και αιτήσεων από τη γλώσσα του διαμεσολαβητή στη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και για τη μετατροπή των αποτελεσμάτων των ερωτήσεων και των αιτήσεων από μορφή συμβατή με τη γλώσσα των πηγών πληροφορίας σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή.

Για την αποθήκευση των δομημένων εγγράφων στις πηγές πληροφορίας χρησιμοποιείται μια σχεσιακή βάση δεδομένων, ενώ για την παροχή δυνατοτήτων αποδοτικής αναζήτησης εγγράφων με βάση το περιεχόμενο και τη δομή τους έχουν υλοποιηθεί ειδικές μέθοδοι πρόσβασης στα πλαίσια της διατριβής [Πουτ2000].

Η γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, εκτός από την επικοινωνία με τους διαμεσολαβητές, παρέχει και την απαιτούμενη λειτουργικότητα για την επικοινωνία των πηγών πληροφορίας με τα εργαλεία που χρησιμοποιούνται για τη δημιουργία τους ή συνεργάζονται με αυτές για τη διαχείριση εγγράφων.

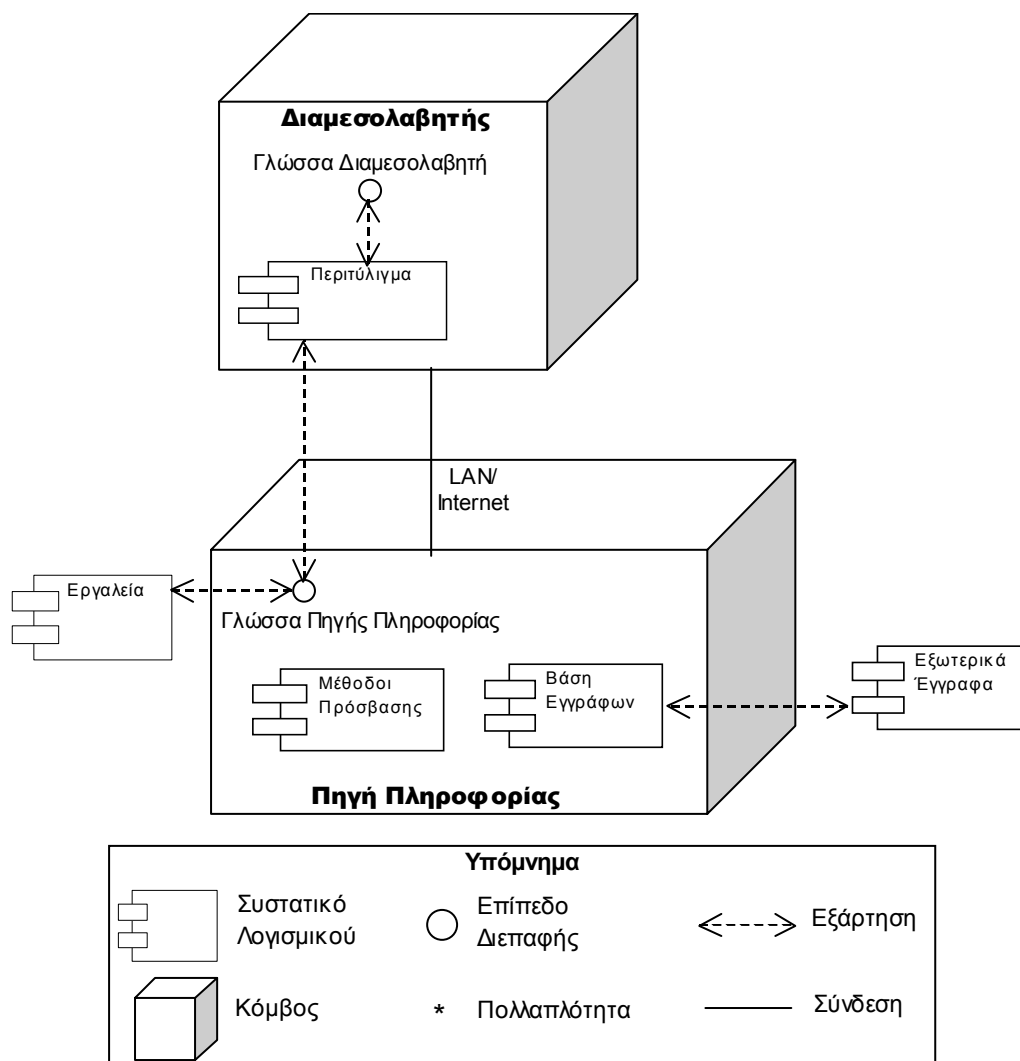
Στη συνέχεια αυτού του τμήματος γίνεται εκτενέστερη αναφορά στην *Αρχιτεκτονική των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα* (ενότητα 6.3.1) και στα σημαντικότερα από τα συστατικά της (*Γλώσσα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα* στην ενότητα 6.3.2, *Μέθοδοι Πρόσβασης* στην ενότητα 6.3.3, *Βάση Δεδομένων των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα* στην ενότητα 6.3.4), στη *Λειτουργικότητα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα* (ενότητα 6.3.5) και στην *Αναπαράσταση Δομημένων Εγγράφων* (ενότητα 6.3.6).

6.3.1. Αρχιτεκτονική των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Στο τμήμα αυτό συζητείται η αρχιτεκτονική των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Η αρχιτεκτονική μιας πηγής πληροφορίας που περιέχει δομημένα έγγραφα απεικονίζεται στο Διάγραμμα 6-3.

Όπως φαίνεται και στο Διάγραμμα 6-3, μια πηγή πληροφορίας που περιέχει δομημένα έγγραφα γίνεται διαθέσιμη σε ένα διαμεσολαβητή μέσω του περιτύλιγματος της κλάσης των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, καθώς το περιτύλιγμα μεταφράζει τις αιτήσεις από τη γλώσσα του διαμεσολαβητή στη γλώσσα των στιγμιοτύπων της συγκεκριμένης κλάσης πηγών πληροφορίας και αντίστροφα.

Από το Διάγραμμα 6-3 φαίνεται ότι η αρχιτεκτονική μιας πηγής πληροφορίας που περιέχει δομημένα έγγραφα απαρτίζεται από τα εξής βασικά τμήματα:



Διάγραμμα 6-3: Διάγραμμα Παράταξης για την Αρχιτεκτονική των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

- Τη *Γλώσσα της Πηγής Πληροφορίας*, που παρέχει το επίπεδο διεπαφής της πηγής πληροφορίας τόσο με το διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» όσο και με τα εργαλεία που χρησιμοποιούνται για τη δημιουργία, τη διαχείριση και την ενημέρωση των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα²⁶.
- Τις *Μεθόδους Πρόσβασης*, που έχουν υλοποιηθεί ειδικά για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, έτσι ώστε να παρέχεται γρήγορη και αποτελεσματική πρόσβαση στα (τοπικά και απομακρυσμένα ως προς μια πηγή πληροφορίας) έγγραφα. Επιπλέον, από τις μεθόδους πρόσβασης υποστηρίζονται ερωτήσεις που μπορεί να περιέχουν όρους δίτημης λογικής και/ή όρους ομοιότητας, με βάση τόσο τη δομή όσο και το περιεχόμενο των εγγράφων [Πουτ2000].

²⁶ Τα εργαλεία που χρησιμοποιούνται για τη δημιουργία, τη διαχείριση και την ενημέρωση των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα περιγράφονται αναλυτικά στο 7^ο Κεφάλαιο.

- Τη *Βάση Δεδομένων της Πηγής Πληροφορίας*, που είναι μια σχεσιακή βάση δεδομένων όπου αποθηκεύονται έγγραφα, σύνδεσμοι σε απομακρυσμένα (ως προς την πηγή πληροφορίας) έγγραφα ή τμήματα εγγράφων και ένα σύνολο μεταπληροφορίας που σχετίζεται με τα αντικείμενα αυτά (τοπικά και απομακρυσμένα).

Μια λεπτομερέστερη αναφορά στα τμήματα από τα οποία απαρτίζεται η αρχιτεκτονική των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα γίνεται στις ενότητες 6.3.2 (όπου περιγράφεται η *Γλώσσα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα*), 6.3.3 (όπου περιγράφονται οι *Μέθοδοι Πρόσβασης*) και 6.3.4 (όπου περιγράφεται η *Βάση Δεδομένων των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα*).

6.3.2. Γλώσσα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Η γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα είναι το θέμα αυτής της ενότητας. Όπως έχει ήδη αναφερθεί, η γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα παρέχει το επίπεδο διεπαφής των πηγών πληροφορίας που εντάσσονται σ' αυτές τόσο με το διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» όσο και με τα εργαλεία που χρησιμοποιούνται για τη δημιουργία, τη διαχείριση και την ενημέρωση των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Με δεδομένο το ρόλο της, οι απαιτήσεις που έπρεπε να καλύπτονται από τη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα ήταν οι εξής:

- Η παροχή της δυνατότητας έκφρασης ερωτήσεων που υποστηρίζονται από τις Μεθόδους Πρόσβασης. Οι ερωτήσεις αυτές μπορεί να περιέχουν όρους δίτιμης λογικής και/ή όρους ομοιότητας, να αναφέρονται τόσο σε τοπικά όσο και σε απομακρυσμένα ως προς μια πηγή πληροφορίας έγγραφα και να επιτρέπουν την έκφραση συνθηκών σχετικών με το περιεχόμενο και/ή τη δομή των εγγράφων και των τμημάτων από τα οποία αυτά απαρτίζονται.
- Η παροχή της απαραίτητης για την υποστήριξη του DOM λειτουργικότητας που αφορά την πρόσβαση σε XML έγγραφα, σε τμήματα XML εγγράφων και στη δομή των XML εγγράφων.
- Η υποστήριξη των βασικών λειτουργιών που είναι απαραίτητες για κάθε πηγή πληροφορίας: υποστήριξη νέων κατηγοριών εγγράφων, εισαγωγή νέων εγγράφων, ανάκτηση εγγράφων και τμημάτων εγγράφων και ενημέρωση των πηγών πληροφορίας για έγγραφα που δεν εισάγονται στη σχεσιακή βάση δεδομένων της πηγής πληροφορίας.

Με βάση τα παραπάνω, η γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα επιλέχθηκε να απαρτίζεται από 3 τμήματα:

- Το τμήμα που καθορίζει την *Υποστήριξη Ερωτήσεων και Αιτήσεων* και περιγράφεται στην υποενότητα 6.3.2.1.
- Το τμήμα που καθορίζει την *Υποστήριξη Εργαλείων Διαχείρισης των Πηγών Πληροφορίας* και περιγράφεται στην υποενότητα 6.3.2.2.

- Το τμήμα που καθορίζει τη μορφή που θα έχει η *Επιστρεφόμενη Πληροφορία*, το οποίο περιγράφεται στην υποενότητα 6.3.2.3.

Καθώς ο καθορισμός των περιτυλιγμάτων μιας κλάσης πηγών πληροφορίας βασίζεται σε μεγάλο ποσοστό στη γλώσσα αυτής της κλάσης πηγών πληροφορίας, η ενότητα αυτή κλείνει με μια αναφορά στα *Περιτυλίγματα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα*, στην υποενότητα 6.3.2.4.

Στο σημείο αυτό και πριν προχωρήσει η περιγραφή της σύνταξης της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, πρέπει να αναφερθεί ότι το μοντέλο πληροφορίας που διαχειρίζεται η γλώσσα αυτή είναι (όπως και το μοντέλο πληροφορίας που διαχειρίζεται η XQL) το διατεταγμένο δένδρικό μοντέλο. Η επιλογή αυτή έγινε καθώς θεωρήθηκε ότι το μοντέλο του διατεταγμένου δέντρου περιγράφει ικανοποιητικά τη δομή των XML εγγράφων.

Για το λόγο αυτό και με δεδομένη τη δυναμική της XQL, που είναι η επικρατέστερη υποψήφια για να γίνει πρότυπο έκφρασης ερωτήσεων για XML έγγραφα, επιλέχθηκε η γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα να ακολουθήσει τη σύνταξη της XQL με μικρές μόνο διαφοροποιήσεις που έχουν να κάνουν με τις ανάγκες των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Οι διαφοροποιήσεις αυτές αναφέρονται στη μορφή της επιστρεφόμενης πληροφορίας (βλέπε υποενότητα 6.3.2.3) και στη λειτουργικότητα επιπλέον της XQL που απαιτείται από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα. Πρέπει να σημειωθεί πάντως ότι, με σκοπό τη διασφάλιση της συμβατότητας της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα με την XQL, η επιπλέον της XQL λειτουργικότητα της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα παρέχεται από ένα σύνολο συναρτήσεων σε μορφή υποστηριζόμενη από την XQL.

Πρέπει επίσης να σημειωθεί ότι για την αλληλεπίδραση με τη σχεσιακή βάση δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα χρησιμοποιείται από τον επεξεργαστή της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα ένα επίπεδο διεπαφής με XML έγγραφα, που επιτρέπει την πρόσβαση στα έγγραφα και τα δομικά στοιχεία των εγγράφων που αποθηκεύονται στη βάση δεδομένων. Το *Επίπεδο Διεπαφής με XML Έγγραφα* περιγράφεται στην υποενότητα 6.3.6.4.

6.3.2.1. Υποστήριξη Ερωτήσεων και Αιτήσεων

Η υποστήριξη που παρέχεται από τη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα για την έκφραση ερωτήσεων και αιτήσεων σε αυτήν περιγράφεται σ' αυτή την υποενότητα. Όπως έχει ήδη αναφερθεί, η γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα υποστηρίζει, σε ότι αφορά τις ερωτήσεις, μέσω XQL σύνταξης τόσο «κλασσικές» XQL ερωτήσεις για έγγραφα και τμήματα εγγράφων όσο και ερωτήσεις που μπορεί να περιέχουν όρους δίτιμης λογικής και/ή όρους ομοιότητας.

Η λειτουργικότητα της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, σε ότι αφορά την υποστήριξη ερωτήσεων και αιτήσεων, συνοψίζεται στα εξής σημεία:

- Λειτουργικότητα, αντίστοιχη με αυτήν που παρέχεται από την XQL, για ερωτήσεις δίτιμης λογικής. Η λειτουργικότητα αυτή έχει περιγραφεί στο 3^ο κεφάλαιο (βλέπε ενότητα 3.5.2).
- Υποστήριξη ερωτήσεων ομοιότητας και ερωτήσεων που μπορεί να περιέχουν όρους δίτιμης λογικής και όρους ομοιότητας.
- Υποστήριξη της δυνατότητας αναζήτησης τόσο μεταξύ των στιγμιοτύπων και μόνο μιας ή περισσότερων κλάσεων εγγράφων ή δομικών στοιχείων εγγράφων όσο και μεταξύ των στιγμιοτύπων των ίδιων των κλάσεων και των υποκλάσεών τους.
- Υποστήριξη της ολικής ή μερικής ανάκτησης του περιεχομένου των εγγράφων και των δομικών στοιχείων των εγγράφων.

Η επιπλέον της XQL λειτουργικότητα, σε ότι αφορά την υποστήριξη ερωτήσεων και αιτήσεων, υλοποιείται σε επίπεδο σύνταξης με τη χρήση συναρτήσεων²⁷. Ο Πίνακας 6-1 απεικονίζει τη σύνταξη των συναρτήσεων που υλοποιούν την επιπλέον της XQL λειτουργικότητα που παρέχεται από τη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα για την υποστήριξη ερωτήσεων και αιτήσεων.

Λειτουργικότητα	Σύνταξη
Έκφραση κριτηρίου ομοιότητας για το περιεχόμενο στοιχείων με σύνολα όρων που σχετίζονται μεταξύ τους με τους τελεστές AND και OR	<code>relates(<σύνολο όρων και φωλιασμένων συνόλων όρων>)</code>
Έκφραση κριτηρίου για τα στιγμιότυπα μιας κλάσης και των υποκλάσεών της	<code>include_subclasses(<όνομα κλάσης>)</code>
Αίτηση ανάκτησης του περιεχομένου στιγμιοτύπων της κλάσης «Κλάση Στοιχείου» όπου το περιεχόμενο των υποστοιχείων της που είναι στιγμιότυπα της κλάσης «Κλάση Υποστοιχείου» αντικαθίσταται από συνδέσμους προς αυτά	<code>only_link_to(<Κλάση Στοιχείου>, <Κλάση Υποστοιχείου>)</code>

Πίνακας 6-1: Οι συναρτήσεις που παρέχουν την επιπλέον της XQL λειτουργικότητα για την υποστήριξη ερωτήσεων και αιτήσεων που παρέχεται από τη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα

6.3.2.2. Υποστήριξη Εργαλείων Διαχείρισης των Πηγών Πληροφορίας

Το θέμα αυτής της υποενότητας είναι η περιγραφή της υποστήριξης που παρέχεται από τη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα στα εργαλεία διαχείρισης των πηγών πληροφορίας αυτής της κλάσης. Η υποστήριξη αυτή παρέχεται από μηχανισμούς που καλούνται από συναρτήσεις που ενσωματώνονται στην XQL σύνταξη της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Έτσι, εκτός από την υποστήριξη ερωτήσεων και αιτήσεων, η γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα υποστηρίζει, μέσω XQL σύνταξής, τις εξής λειτουργίες:

- Αρχικοποίηση της βάσης δεδομένων της τρέχουσας πηγής πληροφορίας.

²⁷ Υπενθυμίζεται ότι η σύνταξη της XQL υποστηρίζει τη χρήση συναρτήσεων οριζόμενων από τους χρήστες.

- Δημιουργία κλάσεων δομημένων εγγράφων και δομικών στοιχείων δομημένων εγγράφων. Στην περίπτωση αυτή, η εισαγωγή και η υποστήριξη μιας κλάσης εγγράφων έχει ως συνέπεια τη δημιουργία και όσων κλάσεων δομικών στοιχείων εγγράφων είναι απαραίτητες για την αναπαράσταση των στιγμιοτύπων της κλάσης εγγράφων. Για να γίνει δυνατή η εισαγωγή μιας κλάσης εγγράφων, πρέπει να δοθούν ως ορίσματα:
 - Ο URL του DTD ή του XML σχήματος που περιγράφει τη δομή και τη σύνταξη των στιγμιοτύπων της κλάσης
 - Ο URL του αρχείου όπου περιέχονται οι Κανόνες Αναπαράστασης των στιγμιοτύπων της κλάσης στη βάση δεδομένων της πηγής πληροφορίας
 - Ο URL του αρχείου όπου περιέχεται το RDF Σχήμα που αναπαριστά το μοντέλο της συγκεκριμένης κλάσης εγγράφων, όπως αυτό γίνεται διαθέσιμο από τη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
- Εισαγωγή στιγμιοτύπων δομημένων εγγράφων και δομικών στοιχείων δομημένων εγγράφων. Για να γίνει δυνατή η εισαγωγή ενός εγγράφου, πρέπει να δοθούν ως ορίσματα το όνομα της κλάσης εγγράφων στην οποία το έγγραφο ανήκει, ο URL του εγγράφου και οι τιμές των χαρακτηριστικών του. Αντίστοιχα, για να γίνει δυνατή η εισαγωγή ενός δομικού στοιχείου εγγράφου, πρέπει να δοθούν ως ορίσματα το όνομα της κλάσης στην οποία ανήκει το δομικό στοιχείο και οι τιμές των χαρακτηριστικών του δομικού στοιχείου.
- Δεικτοδότηση απομακρυσμένων δομημένων εγγράφων και δομικών στοιχείων δομημένων εγγράφων. Για να γίνει δυνατή η δεικτοδότηση κάποιου απομακρυσμένου εγγράφου ή δομικού στοιχείου εγγράφου πρέπει να δοθεί ως όρισμα ο URL του εγγράφου ή του δομικού στοιχείου εγγράφου που πρόκειται να δεικτοδοτηθεί.

Οι λειτουργίες αυτές υλοποιούνται από τους κατάλληλους για κάθε περίπτωση μηχανισμούς (βλέπε 7^ο κεφάλαιο), οι οποίοι ουσιαστικά υποστηρίζουν και τα αντίστοιχα εργαλεία. Ο Πίνακας 6-2 περιέχει τη σύνταξη των συναρτήσεων που υποστηρίζουν τις λειτουργίες αυτές.

Λειτουργία	Συνάρτηση
Αρχικοποίηση της Βάσης Δεδομένων	<code>initdb</code>
Εισαγωγή Κλάσης Εγγράφων	<code>insert_doc_class(<DTD ή XML Σχήμα>, <Αρχείο Κανόνων Αναπαράστασης>, <RDF Σχήμα>)</code>
Εισαγωγή Εγγράφου	<code>insert_doc(<Όνομα Κλάσης>, <URL>, <Τιμές Χαρακτηριστικών>)</code>
Εισαγωγή Δομικού Στοιχείου Εγγράφου	<code>insert_doc_element(<Όνομα Κλάσης>, <Τιμές Χαρακτηριστικών>)</code>
Δεικτοδότηση Απομακρυσμένου Εγγράφου ή Δομικού Στοιχείου Εγγράφου	<code>index(<URL>)</code>

Πίνακας 6-2: Οι συναρτήσεις που υλοποιούν τις λειτουργίες υποστήριξης των εργαλείων διαχείρισης των πηγών πληροφορίας για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και ενσωματώνονται στη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα

6.3.2.3. Επιστρεφόμενη Πληροφορία

Η πληροφορία που επιστρέφεται από τη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα είναι το θέμα αυτής της υποενότητας. Καθώς στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα τα έγγραφα αποθηκεύονται σε XML μορφή, είναι προφανές ότι το περιεχόμενο της επιστρεφόμενης πληροφορίας είναι σε μορφή XML. Για το λόγο αυτό και από τη στιγμή που και η XQL επιστρέφει πληροφορία σε XML μορφή, επιλέχθηκε η επιστρεφόμενη πληροφορία, είτε είναι αποτέλεσμα ερώτησης είτε είναι αποτέλεσμα αίτησης, να δομείται με τρόπο συμβατό με την XML.

Η XQL επιστρέφει την XML πληροφορία από την οποία απαρτίζεται το αποτέλεσμα ερωτήσεων και αιτήσεων χωρίς να την οργανώνει με κάποιο συγκεκριμένο τρόπο. Καθώς όμως από τις πηγές πληροφορίας υποστηρίζονται και ερωτήσεις ομοιότητας, είναι απαραίτητο τα αποτελέσματα των ερωτήσεων να επιστρέφονται δομημένα, έτσι ώστε τα βάρη των αντικειμένων από τα οποία απαρτίζονται να μην αναμειγνύονται με το περιεχόμενο των αντικειμένων.

Έτσι επιλέχθηκε η επιστρεφόμενη πληροφορία να υπακούει στο DTD SourceResult.dtd που παρατίθεται στο DTD 6-1.

```
<!ELEMENT Result      (Item* )>
<!ELEMENT Item        (Value* )>
<!ATTLIST Item        Class CDATA  #REQUIRED
                        Weight CDATA #IMPLIED >
<!ELEMENT Error       (#PCDATA )>
<!ELEMENT Warning     (#PCDATA )>
<!ELEMENT Value       (#PCDATA )>
<!ATTLIST Value       Name CDATA  #REQUIRED >
```

DTD 6-1: O DTD SourceResult.dtd

Ο DTD SourceResult.dtd είναι ταυτόσημος με το DTD 5-1 (Result.dtd) που ορίζει τη δόμηση της πληροφορίας που επιστρέφει η γλώσσα του διαμεσολαβητή (βλέπε ενότητα 5.4.2)²⁸, με μόνη διαφορά ότι τα Items δεν διαθέτουν πια το χαρακτηριστικό Source, που αναπαριστά το όνομα της πηγής πληροφορίας από την οποία προέρχεται καθένα από τα αντικείμενα από τα οποία απαρτίζεται η επιστρεφόμενη πληροφορία σε μια συγκεκριμένη ερώτηση ή αίτηση. Η παράλειψη αυτή έγινε καθώς είναι περιττή η πληροφορία για την πηγή πληροφορίας απ' όπου προέρχεται κάποιο αντικείμενο όταν όλα τα αντικείμενα προέρχονται από την ίδια πηγή πληροφορίας.

6.3.2.4. Περιτυλίγματα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Σ' αυτή την υποενότητα περιγράφονται τα περιτυλίγματα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Ο ρόλος των περιτυλιγμάτων είναι η μετάφραση ερωτήσεων και αιτήσεων από τη γλώσσα του διαμεσολαβητή στη γλώσσα των πηγών πληροφορίας

²⁸ Η επιλογή αυτή έγινε καθώς οι δύο DTDs καλύπτουν αντίστοιχες ανάγκες, οπότε δε θεωρήθηκε σκόπιμη η εισαγωγή διπλής ορολογίας για την περιγραφή των ίδιων εννοιών

που περιέχουν δομημένα έγγραφα. Έτσι, τα περιτυλίγματα είναι στην πραγματικότητα μεταφραστές που βασίζονται σε ένα σύνολο κανόνων αντιστοίχισης. Ο Πίνακας 6-3 περιέχει τους κανόνες αντιστοίχισης που αντιστοιχίζουν τη δομή της γλώσσας του διαμεσολαβητή στη δομή της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και αντιστρόφως.

Λειτουργία	Γλώσσα Διαμεσολαβητή	Γλώσσα Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα
Επιλογή χαρακτηριστικών από στιγμιότυπα κλάσεων	SELECT <Όνομα Κλάσης 1>.<Όνομα Χαρακτηριστικού 1>, ..., <Όνομα Χαρακτηριστικού 1>, ..., Κλάσης M>.<Όνομα Χαρακτηριστικού<Όνομα Κλάσης M>@<Όνομα N> FROM <Όνομα Κλάσης 1>, ..., Χαρακτηριστικού N> <Όνομα Κλάσης M>	<Όνομα Κλάσης 1>@<Όνομα Χαρακτηριστικού 1>, ..., <Όνομα Χαρακτηριστικού 1>, ..., Κλάσης M>@<Όνομα N>
Συνθήκη	WHERE <Συνθήκη>	[<Συνθήκη>]
Σχέση Πατέρα – Παιδιού	<Όνομα Κλάσης Πατέρα>/<Όνομα Κλάσης Παιδιού>	<Όνομα Κλάσης Πατέρα>/<Όνομα Κλάσης Παιδιού>
Σχέση Προγόνου – Απογόνου	<Όνομα Κλάσης Προγόνου>//<Όνομα Κλάσης Απογόνου>	<Όνομα Κλάσης Προγόνου>//<Όνομα Κλάσης Απογόνου>
N-ιοστό παιδί	<Όνομα Κλάσης Πατέρα>/<Όνομα Κλάσης Παιδιού>[N]	<Όνομα Κλάσης Πατέρα>/<Όνομα Κλάσης Παιδιού>[N]
N-ιοστός απόγονος	<Όνομα Κλάσης Προγόνου>//<Όνομα Κλάσης Απογόνου>[N]	<Όνομα Κλάσης Προγόνου>//<Όνομα Κλάσης Απογόνου>[N]
Πριν	Before	Before
Μετά	After	After
Μετονομασία	As	->
Ομαδοποίηση με βάση το χαρακτηριστικό «Χαρακτηριστικό 1»	<Ερώτηση> Group By «Χαρακτηριστικό 1»	«Χαρακτηριστικό 1»{<Ερώτηση>}
Σύγκριση για την ισότητα των εκφράσεων «Έκφραση 1» και «Έκφραση 2»	«Έκφραση 1» = «Έκφραση 2»	«Έκφραση 1» = «Έκφραση 2» και «Έκφραση 1» eq «Έκφραση 2»
Ανισότητα	«Έκφραση 1» != «Έκφραση 2»	«Έκφραση 1» != «Έκφραση 2» και «Έκφραση 1» ne «Έκφραση 2»
Σύγκριση των εκφράσεων «Έκφραση 1» και «Έκφραση 2» (ανεξάρτητη κεφαλαίων – πεζών)	lower («Έκφραση 1») = lower («Έκφραση 2»)	«Έκφραση 1» ieq «Έκφραση 2»
Υπαρξη του προτύπου κειμένου «Κείμενο 1»	Like («Κείμενο 1»)	contains («Κείμενο 1»)

Υπαρξη του προτύπου κειμένου «Κείμενο 1» (ανεξάρτητη κεφαλαίων – πεζών)	Like (lower («Κείμενο 1»))	icontains («Κείμενο 1»)
Μικρότερο	<	< και lt
Μικρότερο ή ίσο	<=	<= και lte
Μεγαλύτερο	>	> και gt
Μεγαλύτερο ή ίσο	>=	>= και gte
Στοιχεία με όνομα «Όνομα Στοιχείου», ανεξάρτητα από τη namespace στην οποία το «Όνομα Στοιχείου» ορίζεται	<Όνομα Στοιχείου>	<Όνομα Στοιχείου>
Στοιχεία με όνομα «Όνομα Στοιχείου», με namespace με όνομα «Όνομα Namespace»	<Όνομα Namespace>:<Όνομα Στοιχείου>	<Όνομα Namespace>:<Όνομα Στοιχείου>
Τα στοιχεία με οποιοδήποτε * όνομα, ανεξάρτητα από namespaces	*	*
Τα στοιχεία με όνομα «Όνομα Στοιχείου», όπου το «Όνομα Στοιχείου» δεν ορίζεται σε κάποια namespace	:<Όνομα Στοιχείου>	:<Όνομα Στοιχείου>
Τα στοιχεία με όνομα «Όνομα Στοιχείου», όπου το «Όνομα Στοιχείου» ορίζεται σε κάποια namespace	*:<Όνομα Στοιχείου>	*:<Όνομα Στοιχείου>
Τα στοιχεία με οποιοδήποτε όνομα με namespace με όνομα «Όνομα Namespace»	<Όνομα Namespace>:*	<Όνομα Namespace>:*
Τα στοιχεία με οποιοδήποτε : * όνομα που δεν ορίζεται σε κάποια namespace	:*	:*
Τα στοιχεία με οποιοδήποτε * : * όνομα που ορίζεται σε κάποια namespace	*:*	*:*
Άρνηση της έκφρασης «Έκφραση 1»	not («Έκφραση 1»)	not («Έκφραση 1»)
Ένωση (Union) των εκφράσεων «Έκφραση 1»	«Έκφραση 1» OR «Έκφραση 2»	«Έκφραση 1» union «Έκφραση 2» και «Έκφραση 1» «Έκφραση 2»

και «Έκφραση 2»		
Τομή (Intersection) των εκφράσεων «Έκφραση 1» και «Έκφραση 2»	«Έκφραση 1» AND «Έκφραση 2»	«Έκφραση 1» intersect «Έκφραση 2»
Και οι δύο (Both) εκφράσεις «Έκφραση 1» και «Έκφραση 2»	«Έκφραση 1» AND «Έκφραση 2»	«Έκφραση 1» ~ «Έκφραση 2»
Ή (Or) των εκφράσεων «Έκφραση 1» και «Έκφραση 2»	«Έκφραση 1» OR «Έκφραση 2»	«Έκφραση 1» or «Έκφραση 2»
Και (And) των εκφράσεων «Έκφραση 1» και «Έκφραση 2»	«Έκφραση 1» AND «Έκφραση 2»	«Έκφραση 1» and «Έκφραση 2»
Κριτήριο ομοιότητας για το περιεχόμενο στοιχείων με σύνολο όρων που σχετίζονται μεταξύ τους με τους τελεστές AND και OR	RELATES TO < (OR AND) TERMSET> <TERM WEIGHT=Βάρος> Όρος</TERM> . . . </ (OR AND) TERMSET>	relates(<όνομα στοιχείων>, <σύνολο όρων και φωλιασμένων συνόλων όρων>)
Κριτήριο για τα στιγμιότυπα μιας κλάσης και των υποκλάσεών της	<όνομα κλάσης>*	include_subclasses(<όνομα κλάσης>)
Περιεχόμενο στοιχείων κάποιας κλάσης, όπου το περιεχόμενο των υποστοιχείων της κλάσης «Κλάση Υποστοιχείου» αντικαθίσταται από συνδέσμους προς αυτά	Content - <Κλάση Υποστοιχείου>	only_link_to(<Κλάση Στοιχείου>, <Κλάση Υποστοιχείου>)

Πίνακας 6-3: Κανόνες Αντιστοίχισης Ερωτήσεων και Αιτήσεων μεταξύ της Γλώσσας του Διαμεσολαβητή και της Γλώσσας των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Στο σημείο αυτό πρέπει να σημειωθεί ότι, με δεδομένη τη μορφή της πληροφορίας στην οποία επιστρέφουν τα αποτελέσματα ερωτήσεων και αιτήσεων, η γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και η γλώσσα του διαμεσολαβητή, κατά τη μετατροπή πληροφορίας μεταξύ αυτών των μορφών η μόνη αλλαγή είναι η πρόσθεση κατά τη μετατροπή στη γλώσσα του διαμεσολαβητή του χαρακτηριστικού Source στα Items και η αφαίρεσή του κατά τη μετατροπή στη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.

Για να γίνει περισσότερο κατανοητή η λειτουργία των περιτυλιγμάτων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, στη συνέχεια παρατίθεται μια σειρά παραδειγμάτων μετατροπής αιτήσεων και ερωτήσεων μεταξύ της γλώσσας του διαμεσολαβητή και της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.

Με βάση τους κανόνες μετατροπής ερωτήσεων και αιτήσεων μεταξύ της γλώσσας του διαμεσολαβητή και της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και το γεγονός ότι κατά τη μεταβίβαση των ερωτήσεων από τον πυρήνα του διαμεσολαβητή στα περιτύλιγματα των πηγών πληροφορίας παραλείπονται οι συνθήκες για το ιδεατό χαρακτηριστικό Source, η ερώτηση που φαίνεται στο Παράδειγμα 5-1 μετατρέπεται στην έξοδο του κοινού περιτύλιγματος των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, για λογαριασμό της πηγής πληροφορίας «Βιβλιοθήκη Πολυτεχνείου Κρήτης», στην ερώτηση που φαίνεται στο Παράδειγμα 6-1.

Άρθρο@Τίτλος, Άρθρο@Ταυτότητα [Άρθρο@Συγγραφέας="X Y"]

Παράδειγμα 6-1: Η Ερώτηση που φαίνεται στο Παράδειγμα 5-1 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα

Αντίστοιχα, η ερώτηση ομοιότητας που φαίνεται στο Παράδειγμα 5-3 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα παίρνει τη μορφή της ερώτησης που φαίνεται στο Παράδειγμα 6-2.

Έγγραφο@Τίτλος, Έγγραφο@Ταυτότητα, Έγγραφο@Συγγραφέας
[relates(include_subclasses(Έγγραφο), OR(Κνωσός, 100%), (Ανακατασκευή, 75%))]

Παράδειγμα 6-2: Η Ερώτηση που φαίνεται στο Παράδειγμα 5-3 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα

Τα ασαφή χαρακτηριστικά, τόσο στη γλώσσα του διαμεσολαβητή όσο και στη γλώσσα των πηγών πληροφορίας αντιμετωπίζονται όπως και τα απλά χαρακτηριστικά. Έτσι, η ερώτηση που φαίνεται στο Παράδειγμα 5-4 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα φαίνεται στο Παράδειγμα 6-3.

Έγγραφο@Τίτλος, Έγγραφο@Ταυτότητα, Έγγραφο@Συγγραφέας
[Έγγραφο@Παλαιότητα>50%]

Παράδειγμα 6-3: Η Ερώτηση που φαίνεται στο Παράδειγμα 5-4 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα

Τέλος, η αίτηση μερικής ανάκτησης του περιεχομένου αντικειμένου που περιγράφηκε στο Παράδειγμα 5-11, μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, φαίνεται στο Παράδειγμα 6-4, ενώ η αίτηση ανάκτησης του περιεχομένου συγκεκριμένου δομικού στοιχείου που περιγράφηκε στο Παράδειγμα 5-12 φαίνεται στο Παράδειγμα 6-5.

only_link_to(<Άρθρο>, <Υποενότητα>) [Άρθρο.Ταυτότητα="125"]

Παράδειγμα 6-4: Η Ερώτηση που φαίνεται στο Παράδειγμα 5-11 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα

Παράδειγμα 6-5 Η Ερώτηση που φαίνεται στο Παράδειγμα 5-12 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα

6.3.3. Μέθοδοι Πρόσβασης

Όπως έχει ήδη αναφερθεί, οι μέθοδοι πρόσβασης είναι ένα από τα βασικά δομικά στοιχεία που απαρτίζουν την αρχιτεκτονική των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Μια σύντομη αναφορά στις μεθόδους πρόσβασης των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, που σχεδιάστηκαν και υλοποιήθηκαν στα πλαίσια της διατριβής [Πουτ2000], γίνεται σ' αυτή την ενότητα.

Στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα χρησιμοποιείται, με τις απαιτούμενες τροποποιήσεις ώστε να ανταποκρίνεται στις απαιτήσεις και την πλατφόρμα υλοποίησης του συστήματός «Παροχής Πληροφοριών κατ' Απαίτηση», η μέθοδος πρόσβασης BOSI η οποία έχει χρησιμοποιηθεί και στο σύστημα KYDONIA [Καπετ95].

Η μέθοδος πρόσβασης BOSI υποστηρίζει ερωτήσεις που βασίζονται στη δομή και/ή το περιεχόμενο των εγγράφων. Οι ερωτήσεις αυτές επιτρέπεται να περιέχουν όρους δίτιμης λογικής και/ή όρους ομοιότητας. Επιπλέον, οι ερωτήσεις μπορεί να αφορούν είτε ολόκληρα έγγραφα είτε δομικά στοιχεία εγγράφων, και να εφαρμόζονται είτε σε όλα τα έγγραφα της πηγής πληροφορίας είτε σε συγκεκριμένες κλάσεις εγγράφων. Οι όροι των ερωτήσεων μπορεί να είναι είτε λέξεις, είτε πρότυπα (patterns), ενώ υποστηρίζονται και ερωτήσεις γειννίας λέξεων και προτύπων.

Η λειτουργικότητα αυτή υποστηρίζεται τόσο για έγγραφα αποθηκευμένα στις πηγές πληροφορίας όσο και για απομακρυσμένα έγγραφα, για τη δομή και το περιεχόμενο των οποίων έχουν ενημερωθεί οι μέθοδοι πρόσβασης.

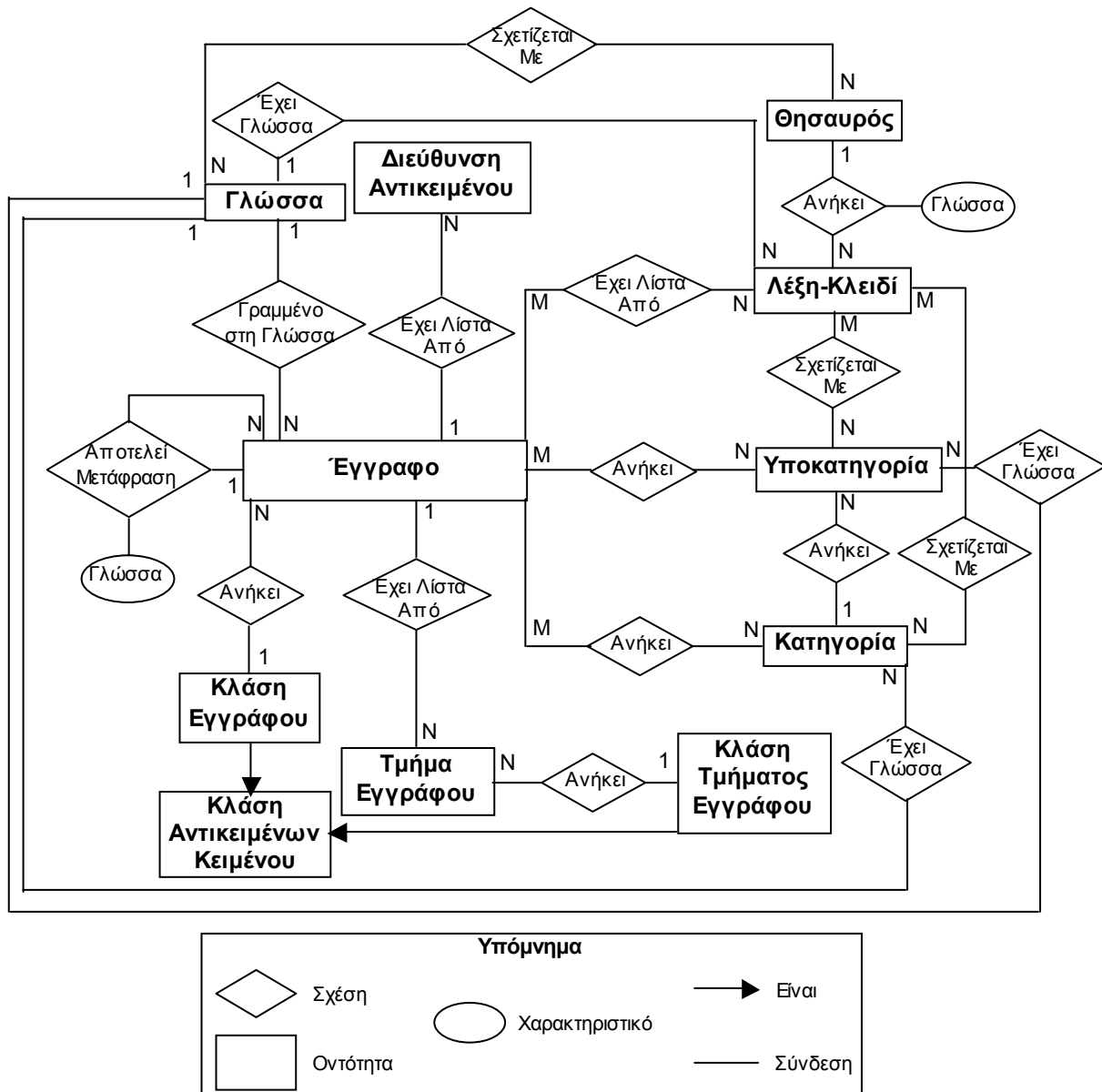
Η μέθοδος πρόσβασης BOSI βασίζεται σε ανεστραμμένα αρχεία και υπογραφές. Πιο συγκεκριμένα, αποτελείται από ένα φίλτρο και ένα σύνολο υπογραφών. Το φίλτρο, που ονομάζεται LEXMAN, χρησιμοποιείται για να προσδιοριστεί σε ποια έγγραφα περιέχεται κάθε όρος της ερώτησης και πόσες φορές περιέχεται κάθε όρος στο κάθε έγγραφο. Οι υπογραφές παρέχουν τη δυνατότητα να βρεθεί ο αριθμός των εμφανίσεων ενός ή περισσοτέρων όρων στα δομικά στοιχεία των εγγράφων χωρίς να είναι απαραίτητη η ανάκτηση των ίδιων των εγγράφων.

Ο LEXMAN αποτελείται από ένα λεξιλόγιο για κάθε γλώσσα, ένα σύνολο επεκτάσιμων λιστών για καθεμιά από τις υποστηριζόμενες γλώσσες και έναν πίνακα αντιστοίχησης λέξεων - επεκτάσιμων λιστών για κάθε γλώσσα.

6.3.4. Βάση Δεδομένων των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Το θέμα αυτής της ενότητας είναι η Βάση Δεδομένων των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα. Το E-R διάγραμμα του σχήματος της βάσης δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα φαίνεται στο Διάγραμμα 6-4.

Με βάση το ρόλο που παίζουν οι διαφορετικές οντότητες που απαρτίζουν το E-R διάγραμμα, οι οντότητες κατηγοριοποιούνται στις εξής ομάδες:



Διάγραμμα 6-4: Αφαιρετικό E-R Διάγραμμα για το Σχήμα της Βάσης Δεδομένων των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

- Οντότητες που αναπαριστούν έγγραφα και τμήματα εγγράφων. Οι οντότητες που ανήκουν σε αυτή την ομάδα είναι οι εξής:
 - Η οντότητα *Έγγραφο*, που αναπαριστά τα δομημένα έγγραφα.
 - Η οντότητα *Τμήμα Εγγράφου*, που αναπαριστά τα δομικά στοιχεία των δομημένων εγγράφων που επιτρέπεται να αποθηκεύονται ανεξάρτητα από τα έγγραφα στα οποία περιέχονται. Τα στιγμιότυπα της οντότητας *Τμήμα Εγγράφου* συσχετίζονται, μέσω της συσχέτισης «Ανήκει», με τα στιγμιότυπα της οντότητας *Έγγραφο* στα οποία περιέχονται.

- Η οντότητα *Διεύθυνση Αντικειμένου*, που αναπαριστά τις διευθύνσεις των αντικειμένων στο Διαδίκτυο. Με τον όρο διευθύνσεις γίνεται αναφορά τόσο σε URLs (Uniform Resource Locators) και URIs (Uniform Resource Identifiers) όσο και σε XPointers και XLinks.

Στο σημείο αυτό πρέπει να αναφερθεί ότι οι οντότητες Έγγραφο και Τμήμα Εγγράφου είναι οι βασικές οντότητες για την υποστήριξη δομημένων εγγράφων από το σύστημα. Το τμήμα του σχήματος όπου αναπαρίστανται οι σχετικές με τα δομημένα έγγραφα οντότητες θα εξελίσσεται και θα επεκτείνεται καθώς θα υποστηρίζονται νέες κλάσεις εγγράφων και τμημάτων εγγράφων από το σύστημα (για περισσότερες λεπτομέρειες βλέπε ενότητα 6.3.6).

- *Οντότητες που υποστηρίζουν την κατηγοριοποίηση των εγγράφων με βάση το περιεχόμενό τους.* Οι οντότητες που ανήκουν σε αυτή την ομάδα είναι οι εξής:
 - Η οντότητα *Κατηγορία*, που αναπαριστά τις κατηγορίες περιεχομένου στις οποίες εντάσσεται το περιεχόμενο των εγγράφων (π.χ. Αθλητικά, Πολιτικά, Επιστημονικά κλπ.). Κάθε στιγμιότυπο της οντότητας Έγγραφο συσχετίζεται, μέσω της συσχέτισης «Ανήκει», με ένα αριθμό κατηγοριών.
 - Η οντότητα *Υποκατηγορία*, που αναπαριστά τις υποκατηγορίες περιεχομένου στις οποίες εντάσσεται το περιεχόμενο των εγγράφων. Οι υποκατηγορίες αποτελούν το δεύτερο επίπεδο κατηγοριοποίησης των εγγράφων (έτσι, π.χ., υποκατηγορίες που μπορεί να εντάσσονται στην κατηγορία Αθλητικά είναι οι υποκατηγορίες Μπάσκετ, Βόλεϊ κλπ.). Κάθε Υποκατηγορία συσχετίζεται με μια ακριβώς Κατηγορία, την Κατηγορία στην οποία εντάσσεται, μέσω της συσχέτισης «Ανήκει». Κάθε στιγμιότυπο της οντότητας Έγγραφο συσχετίζεται, μέσω της συσχέτισης «Ανήκει», με ένα αριθμό υποκατηγοριών.
 - Η οντότητα *Λέξη-Κλειδί*, που αναπαριστά τις λέξεις-κλειδιά που χρησιμοποιούνται για το χαρακτηρισμό του περιεχομένου των εγγράφων. Κάθε Λέξη-Κλειδί συσχετίζεται, μέσω της συσχέτισης «Σχετίζεται Με», με τις Κατηγορίες και τις Υποκατηγορίες, που το περιεχόμενο των αντικειμένων που εντάσσονται σ' αυτές σχετίζεται με τη λέξη-κλειδί. Κάθε στιγμιότυπο της οντότητας Έγγραφο συσχετίζεται, μέσω της συσχέτισης «Σχετίζεται Με», με ένα αριθμό λέξεων-κλειδιών.
 - Η οντότητα *Θησαυρός*, που αναπαριστά τους θησαυρούς που χρησιμοποιούνται από την πηγή πληροφορίας. Κάθε Θησαυρός συσχετίζεται με ένα αριθμό Λέξεων-Κλειδιών που περιέχονται στο Θησαυρό αυτό μέσω της συσχέτισης «Ανήκει» και με μία τουλάχιστον γλώσσα μέσω της συσχέτισης «Σχετίζεται Με». Οι θησαυροί που σχετίζονται με μια μόνο γλώσσα παρέχουν τους βαθμούς συσχέτισης μεταξύ των όρων της γλώσσας αυτής, ενώ αυτοί που συσχετίζονται με περισσότερες από μια γλώσσες παρέχουν τους βαθμούς συσχέτισης μεταξύ των όρων περισσότερων από μίας γλωσσών.
- *Οντότητες που αναπαριστούν τη σχετική με τα έγγραφα μεταπληροφορία.* Οι οντότητες που ανήκουν σε αυτή την ομάδα είναι οι εξής:

- Η οντότητα *Κλάση Εγγράφων*, που αναπαριστά τις κλάσεις των εγγράφων που αποθηκεύονται στο σύστημα. Κάθε στιγμιότυπο της οντότητας Έγγραφο συσχετίζεται, μέσω της συσχέτισης «Ανήκει», με μια ακριβώς κλάση εγγράφων.
- Η οντότητα *Κλάση Τμημάτων Εγγράφων*, που αναπαριστά τις κλάσεις των τμημάτων των εγγράφων που επιτρέπεται να αποθηκεύονται στο σύστημα ανεξάρτητα από τα έγγραφα στα οποία περιέχονται. Κάθε στιγμιότυπο της οντότητας Τμήμα Εγγράφου συσχετίζεται, μέσω της συσχέτισης «Ανήκει», με μια ακριβώς κλάση τμημάτων εγγράφων.
- Η οντότητα *Γλώσσα*, που αναπαριστά τις γλώσσες στις οποίες είναι γραμμένα τα έγγραφα που αποθηκεύονται στο σύστημα. Η οντότητα Γλώσσα συσχετίζεται, μέσω της συσχέτισης «Έχει Γλώσσα» με τα έγγραφα που είναι γραμμένα σε αυτήν, με τους θησαυρούς που περιέχουν όρους της γλώσσας αυτής, με τις λέξεις-κλειδιά της τρέχουσας γλώσσας και με τις κατηγορίες και τις υποκατηγορίες που το όνομά τους είναι εκφρασμένο στην τρέχουσα γλώσσα.

6.3.5. Λειτουργικότητα των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Η λειτουργικότητα που παρέχεται από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα είναι, σε γενικές γραμμές, αντίστοιχη με αυτή που παρέχεται γενικά από τις πηγές πληροφορίας και απεικονίζεται στο Διάγραμμα 6-2. Ειδικότερα τώρα, για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, τα σημεία λειτουργικότητας που περιγράφηκαν στην ενότητα 6.2 παίρνουν την ακόλουθη μορφή:

- Αρχικοποίηση της βάσης δεδομένων μιας πηγής πληροφορίας. Μ' αυτό τον τρόπο ο διαχειριστής μιας πηγής πληροφορίας μπορεί να δημιουργήσει μια κενή βάση δεδομένων, που περιέχει μόνο τις οντότητες που είναι απαραίτητες για τη λειτουργία της πηγής πληροφορίας και την παραπέρα εξέλιξη του σχήματός της. Έτσι, στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα περιέχονται μόνο οι οντότητες που είναι απαραίτητες για την αναπαράσταση των κλάσεων Έγγραφο, Τμήμα Εγγράφου και Διεύθυνση Αντικειμένου, οι οποίες είναι απαραίτητες για την υλοποίηση του μοντέλου αναπαράστασης εγγράφων της ενότητας 6.3.6 και των κλάσεων Κατηγορία, Υποκατηγορία, Λέξη – Κλειδί, Θησαυρός και Γλώσσα που χρησιμοποιούνται για την περιγραφή της μεταπληροφορίας που σχετίζεται με το περιεχόμενο των εγγράφων.

Ο μηχανισμός που υποστηρίζει τη λειτουργία της αρχικοποίησης της βάσης δεδομένων μιας πηγής πληροφορίας που περιέχει δομημένα έγγραφα περιγράφεται στο τμήμα 7.1.

- Υποστήριξη νέων κλάσεων εγγράφων από τη σχεσιακή βάση δεδομένων μιας πηγής πληροφορίας. Καθώς από το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» υποστηρίζονται XML έγγραφα, είναι προφανές ότι για να γίνει δυνατή η αποτελεσματική αναπαράσταση και διαχείρισή τους, είναι απαραίτητο από το σύστημα να υποστηρίζονται διαφορετικές κλάσεις εγγράφων²⁹, το σύνολο των οποίων θα είναι

²⁹ Οι κλάσεις εγγράφων που υποστηρίζονται από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα περιγράφονται από XML DTDs ή XML σχήματα.

δυνατό να εμπλουτίζεται. Φαίνεται λοιπόν ότι είναι απαραίτητο να δίνεται η δυνατότητα εξέλιξης του σχήματος της βάσης δεδομένων των πηγών πληροφορίας, καθώς η ανάγκη υποστήριξης νέων κλάσεων εγγράφων επιβάλλεται από την XML.

Για να γίνει δυνατή η υποστήριξη μιας νέας κλάσης εγγράφων, υποστηρίζονται ταυτόχρονα και οι κλάσεις δομικών στοιχείων εγγράφων που είναι απαραίτητες για την αναπαράσταση των εγγράφων που εντάσσονται στη νέα κλάση εγγράφων.

Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της υποστήριξης νέων κλάσεων εγγράφων στη βάση δεδομένων μιας πηγής πληροφορίας που περιέχει δομημένα έγγραφα περιγράφονται στο τμήμα 7.2.

- Εισαγωγή νέων εγγράφων στη σχεσιακή βάση δεδομένων μιας πηγής πληροφορίας. Με τον τρόπο αυτό δίνεται η δυνατότητα εισαγωγής νέων εγγράφων στη βάση δεδομένων μιας πηγής πληροφορίας από:
 - Το διαχειριστή της πηγής πληροφορίας.
 - Κάποιο προνομιούχο χρήστη (τοπικό ή απομακρυσμένο) που ο κωδικός πρόσβασής του του παρέχει τη δυνατότητα πραγματοποίησης εισαγωγής εγγράφων.
 - Κάποιο πράκτορα, που εντοπίζει έγγραφα και ενημερώνει τις πηγές πληροφορίας το περιεχόμενο των οποίων σχετίζεται με το περιεχόμενο των εγγράφων αυτών.

Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της εισαγωγής νέων εγγράφων στη βάση δεδομένων μιας πηγής πληροφορίας που περιέχει δομημένα έγγραφα περιγράφονται στο τμήμα 7.3.

- Ενημέρωση της πηγής πληροφορίας (και πιο συγκεκριμένα της βάσης δεδομένων και των μεθόδων πρόσβασης) για έγγραφα που δεν εισάγονται (ή τουλάχιστον δεν εισάγονται εξ' ολοκλήρου) στη σχεσιακή βάση δεδομένων της πηγής πληροφορίας, αλλά δεικτοδοτούνται σ' αυτήν. Αυτό σημαίνει ότι μπορεί να εντοπιστούν κάποια έγγραφα που το περιεχόμενό τους είναι σχετικό με αυτό της πηγής πληροφορίας, τα οποία όμως δεν εισάγονται εξ' ολοκλήρου στη βάση δεδομένων της.

Στην περίπτωση αυτή ενημερώνεται η πηγή πληροφορίας (και ειδικότερα η βάση δεδομένων, οι δομές δεικτοδότησης και οι μέθοδοι πρόσβασης) για το περιεχόμενο των εγγράφων αυτών και για την τοποθεσία στην οποία αυτά βρίσκονται. Επιπλέον, για κάποιο έγγραφο τμήματα του οποίου συσχετίζονται σε μεγάλο βαθμό με το περιεχόμενο της πηγής πληροφορίας μπορούν να αποθηκεύονται στη βάση δεδομένων μόνο τα σε μεγάλο βαθμό συσχετιζόμενα με το περιεχόμενο της πηγής πληροφορίας τμήματα του εγγράφου και να υπάρχουν σύνδεσμοι προς τα υπόλοιπα τμήματα του εγγράφου.

Έτσι γίνεται δυνατή η παραπομπή των χρηστών στα έγγραφα που δεν εισάγονται εξ' ολοκλήρου στη βάση δεδομένων της πηγής πληροφορίας όταν οι τελευταίοι εκφράσουν κάποια ερώτηση, τα κριτήρια της οποίας πληρούνται και από αυτά τα έγγραφα.

Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της δεικτοδότησης σε μια πηγή πληροφορίας που περιέχει δομημένα έγγραφα εγγράφων που το περιεχόμενό τους σχετίζεται με αυτό της πηγής πληροφορίας, τα οποία όμως δεν εισάγονται (τουλάχιστον

όχι εξ' ολοκλήρου) στη βάση δεδομένων της πηγής πληροφορίας, περιγράφονται στο τμήμα 7.5.

- Αναζήτηση (τοπικών και απομακρυσμένων ως προς μια πηγή πληροφορίας) εγγράφων ή τμημάτων εγγράφων που ικανοποιούν συγκεκριμένα κριτήρια σχετικά με το περιεχόμενο και/ή τη δομή τους. Έτσι γίνεται δυνατή η υποστήριξη ερωτήσεων που περιέχουν όρους δίτιμης λογικής και/ή όρους ομοιότητας από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα. Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της αναζήτησης δομημένων εγγράφων και/ή τμημάτων εγγράφων περιγράφονται στο τμήμα 7.4.
- Ανάκτηση εγγράφων και τμημάτων εγγράφων που αποθηκεύονται στη σχεσιακή βάση δεδομένων μιας πηγής πληροφορίας. Με τον τρόπο αυτό γίνεται δυνατή η πρόσβαση των χρηστών στα έγγραφα που τους ενδιαφέρουν και παρέχεται η δυνατότητα εξ' ολοκλήρου και τμηματικής παρουσιάσής τους. Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της ανάκτησης δομημένων εγγράφων και/ή τμημάτων εγγράφων περιγράφονται στο τμήμα 7.4.

6.3.6. Αναπαράσταση Δομημένων Εγγράφων

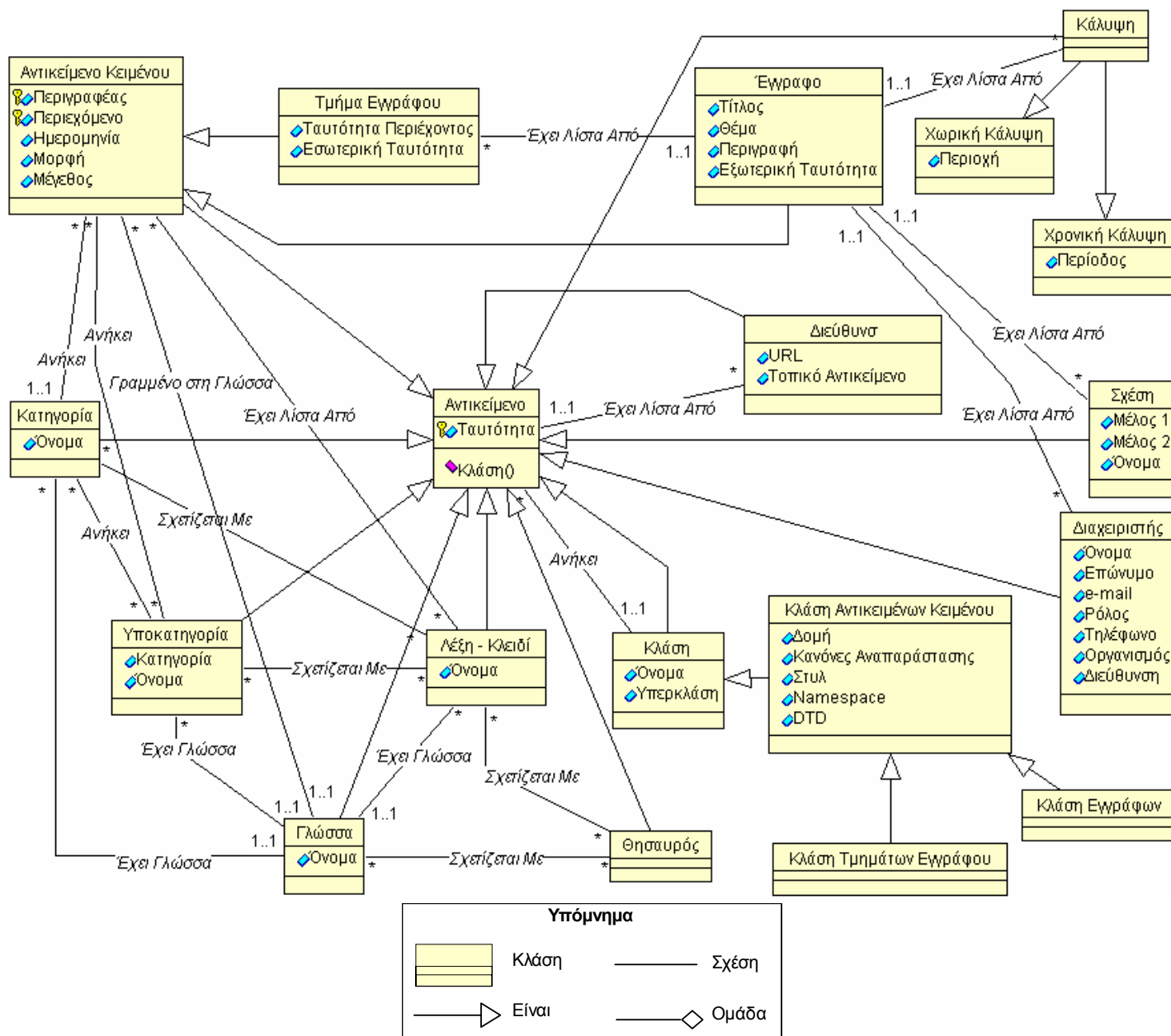
Στην ενότητα αυτή περιγράφεται η αναπαράσταση δομημένων εγγράφων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα. Πιο συγκεκριμένα, στην υποενότητα 6.3.6.1 περιγράφεται το *Μοντέλο Αναπαράστασης Δομημένων Εγγράφων*. Καθώς το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» επικεντρώνει σε XML έγγραφα, στην υποενότητα 6.3.6.2 περιγράφεται η αναπαράσταση XML εγγράφων με βάση το μοντέλο αναπαράστασης δομημένων εγγράφων. Στην υποενότητα 6.3.6.3 περιγράφεται το *Σχήμα Βάσης Δεδομένων* που χρησιμοποιείται για την υλοποίηση του μοντέλου αναπαράστασης XML εγγράφων σε σχεσιακές βάσεις δεδομένων και στην υποενότητα 6.3.6.4 γίνεται αναφορά στο *Επίπεδο Διεπαφής με XML Έγγραφα*.

6.3.6.1. Μοντέλο Αναπαράστασης Δομημένων Εγγράφων

Το μοντέλο αναπαράστασης δομημένων εγγράφων στη σχεσιακή βάση δεδομένων μιας πηγής πληροφορίας που περιέχει δομημένα έγγραφα περιγράφεται σ' αυτή την υποενότητα. Ανάμεσα στις απαιτήσεις του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» ήταν η ύπαρξη αποδοτικών μηχανισμών διαχείρισης εγγράφων. Για να επιτευχθεί αυτό, έπρεπε η αναπαράσταση των εγγράφων στις σχεσιακές βάσεις δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα να είναι ευέλικτη και να εκμεταλλεύεται όσο είναι δυνατόν καλύτερα τις δυνατότητες του ΣΣΔΒΔ (Σχεσιακού Συστήματος Διαχείρισης Βάσεων Δεδομένων) χωρίς να υποφέρει από πλευράς απόδοσης. Έπρεπε επίσης να επιτρέπει την αναπαράσταση απομακρυσμένων εγγράφων, τα οποία δεν αποθηκεύονται (ή τουλάχιστον δεν αποθηκεύονται εξ' ολοκλήρου) στη βάση δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Επιπλέον, η αναπαράσταση των δομημένων εγγράφων έπρεπε να παρέχει δυνατότητες αναφοράς και ανάκτησης των επιμέρους δομικών στοιχείων από τα οποία απαρτίζονται τα έγγραφα.

Λόγω των προβλημάτων που παρουσιάζουν οι προσεγγίσεις αναπαράστασης δομημένων εγγράφων που περιγράφηκαν στο 2^ο κεφάλαιο, στα πλαίσια του συστήματος «Παροχής

Πληροφοριών κατ' Απαίτηση» προτιμήθηκε η αναπαράσταση που απεικονίζεται στο Διάγραμμα 6-5. Το Διάγραμμα 6-5 είναι ένα UML διάγραμμα κλάσεων, το οποίο απεικονίζει ένα αφαιρετικό μοντέλο αναπαράστασης εγγράφων που ικανοποιεί τις απαιτήσεις που αναφέρθηκαν πριν χωρίς να υποφέρει από τα προβλήματα των προσεγγίσεων που περιγράφηκαν στο 2^ο κεφάλαιο.



Διάγραμμα 6-5: Αφαιρετικό Μοντέλο Αναπαράστασης Εγγράφων στις Βάσεις Δεδομένων των Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Με βάση την προσέγγιση αυτή, στη βάση δεδομένων αποθηκεύονται Έγγραφα (Documents) και Τμήματα Εγγράφων (Document Fragments), που κατηγοριοποιούνται ως Αντικείμενα Κειμένου (Text Objects). Επιπλέον, στη βάση δεδομένων αποθηκεύονται Διευθύνσεις Αντικειμένων (Object Addresses) καθώς και οντότητες που αναπαριστούν στοιχεία μεταπληροφορίας που χρησιμοποιούνται για την περιγραφή των εγγράφων.

Τα έγγραφα μικρού και μεσαίου μεγέθους αποθηκεύονται ως ενιαία στιγμιότυπα της κλάσης Έγγραφο. Αντίθετα, τα μεγαλύτερα έγγραφα αποθηκεύονται τμηματικά: Για κάθε μεγάλο έγγραφο υπάρχει ένα στιγμιότυπο της κλάσης Έγγραφο και ένα σύνολο στιγμιότυπων της κλάσης Τμήμα Εγγράφου, που αναπαριστούν δομικά στοιχεία του εγγράφου. Στη λογική δομή των εγγράφων αυτών περιέχονται δείκτες προς τα στιγμιότυπα εκείνα της κλάσης Τμήμα Εγγράφου που αναπαριστούν το πρώτο επίπεδο των δομικών στοιχείων των εγγράφων³⁰.

Στην αποθήκευση των δομικών στοιχείων των εγγράφων ακολουθείται η ίδια λογική που ακολουθείται για τα έγγραφα: τα μικρά και μεσαίου μεγέθους στοιχεία αποθηκεύονται ενιαία, ενώ τα μεγάλα τμηματικά³¹. Τα στιγμιότυπα της κλάσης Τμήμα Εγγράφου διαθέτουν λογική δομή αντίστοιχη με αυτή των στιγμιότυπων της κλάσης Έγγραφο: η λογική δομή των Τμημάτων Εγγράφου περιέχει δείκτες στα υποστοιχεία τους, είτε αυτά είναι ανεξάρτητα Τμήματα Εγγράφου, είτε Τμήματα Εγγράφου που αποθηκεύονται μαζί με το Τμήμα Εγγράφου στο οποίο περιέχονται, είτε διευθύνσεις.

Το μοντέλο αναπαράστασης εγγράφων που περιγράφηκε παραπάνω είναι κατάλληλο και για την περιγραφή εγγράφων που δεν αποθηκεύονται στην τοπική βάση δεδομένων (ή αποθηκεύονται τμηματικά σ' αυτήν). Για τα έγγραφα αυτά αποθηκεύονται τα εξής:

- Η μεταπληροφορία που σχετίζεται με το περιεχόμενο και τη δομή τους.
- Τα χαρακτηριστικά τους.
- Τα τμήματα των εγγράφων που είναι διαθέσιμα τοπικά.
- Η λογική δομή των εγγράφων, η οποία περιέχει αναφορές σε διευθύνσεις εγγράφων ή τμημάτων εγγράφων στο Διαδίκτυο.

Όλα τα αντικείμενα που αποθηκεύονται στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα εντάσσονται σε κάποια *Κλάση (Class)*, η οποία περιγράφει το τρέχον κάθε φορά αντικείμενο και τα χαρακτηριστικά του (π.χ. πιθανές κλάσεις εγγράφων είναι το βιβλίο, το άρθρο κ.α. ενώ πιθανές κλάσεις δομικών στοιχείων εγγράφων η παράγραφος, το κεφάλαιο κ.α.). Η κλάση αποτελεί μέθοδο για όλα τα αντικείμενα και επιστρέφει το όνομα της κλάσης με την οποία συσχετίζεται το τρέχον στιγμιότυπο μέσω της συσχέτισης «Ανήκει». Το όνομα αυτό, ανάλογα με την κλάση στην οποία εντάσσεται το τρέχον στιγμιότυπο, παίρνει τις εξής τιμές:

- Για όλα τα στιγμιότυπα της κλάσης «Τμήμα Εγγράφου» και των υποκλάσεων της, το όνομα της Κλάσης Τμημάτων Εγγράφου με την οποία συσχετίζεται το τρέχον στιγμιότυπο μέσω της συσχέτισης «Ανήκει».

³⁰ Στο σημείο αυτό υποστηρίζεται η δυνατότητα να γίνει η εξής παραλλαγή: κάποια δομικά στοιχεία μικρού μεγέθους αποθηκεύονται μαζί με το έγγραφο ενώ τα υπόλοιπα χωριστά. Έτσι π.χ. για ένα βιβλίο, τα κεφάλαια και τα παραρτήματα που το απαρτίζουν αποθηκεύονται ως ανεξάρτητες οντότητες, ενώ η σύνοψη, η αφιέρωση κ.α. μαζί με το στιγμιότυπο του βιβλίου.

³¹ Στο σημείο αυτό υποστηρίζεται η δυνατότητα να γίνει η εξής παραλλαγή: κάποια δομικά στοιχεία, όπως π.χ. τα κεφάλαια και τα υποκεφάλαια να μπορούν να αποθηκευτούν, ανάλογα με το μέγεθός τους, ενιαία ή τμηματικά, ενώ κάποια άλλα όπως π.χ. οι λίστες να αποθηκεύονται πάντα, ανεξαρτήτως μεγέθους, ενιαία.

- Για όλα τα στιγμιότυπα της κλάσης «Έγγραφο» και των υποκλάσεών της, το όνομα της Κλάσης Εγγράφων με την οποία συσχετίζεται το τρέχον στιγμιότυπο μέσω της συσχέτισης «Ανήκει».
- Για όλα τα στιγμιότυπα της κλάσης «Διαχειριστής» τη συμβολοσειρά «Διαχειριστής».
- Για όλα τα στιγμιότυπα της κλάσης «Διεύθυνση Αντικειμένου» τη συμβολοσειρά «Διεύθυνση».
- Για όλα τα στιγμιότυπα της κλάσης «Κατηγορία» τη συμβολοσειρά «Κατηγορία».
- Για όλα τα στιγμιότυπα της κλάσης «Υποκατηγορία» τη συμβολοσειρά «Υποκατηγορία».
- Για όλα τα στιγμιότυπα της κλάσης «Λέξη – Κλειδί» τη συμβολοσειρά «Λέξη – Κλειδί».
- Για όλα τα στιγμιότυπα της κλάσης «Θησαυρός» τη συμβολοσειρά «Θησαυρός».
- Για όλα τα στιγμιότυπα της κλάσης «Γλώσσα» τη συμβολοσειρά «Γλώσσα».

Επιπλέον, κάθε Αντικείμενο Κειμένου (Έγγραφο ή Τμήμα Εγγράφου) διαθέτει:

- Μέγεθος (Size), που αναπαριστά το μέγεθος του αντικειμένου κειμένου σε bytes.
- Ημερομηνία (Date), που αναπαριστά την ημερομηνία κατά την οποία δημιουργήθηκε το αντικείμενο κειμένου.
- Περιεχόμενο (Content), που αναπαριστά το περιεχόμενο του αντικειμένου.
- Περιγραφέα (Descriptor), ο οποίος περιέχει τη λογική δομή του αντικειμένου. Η μορφή με την οποία αποθηκεύονται στον περιγραφέα ενός εγγράφου οι αναφορές στα δομικά στοιχεία από τα οποία αυτό απαρτίζεται φαίνεται στον Πίνακα 6-4. Όπως φαίνεται στον πίνακα, στους περιγραφείς των εγγράφων και των τμημάτων εγγράφων αποθηκεύονται:
 - Για τα δομικά στοιχεία των εγγράφων που αποθηκεύονται μαζί με τα έγγραφα:
 - Η εσωτερική ταυτότητα (ObjectID) του δομικού στοιχείου. Οι εσωτερικές ταυτότητες των δομικών στοιχείων χαρακτηρίζουν τα δομικά στοιχεία των εγγράφων και είναι μοναδικές στα όρια κάθε εγγράφου.
 - Η Κλάση (Class) του δομικού στοιχείου.
 - Τα χαρακτηριστικά (Attributes) του δομικού στοιχείου.
 - Η απόσταση της αρχής του δομικού στοιχείου από την αρχή του εγγράφου σε bytes (Offset).
 - Το μέγεθος του δομικού στοιχείου σε bytes (Size).
 - Για τα δομικά στοιχεία των εγγράφων που αποθηκεύονται ανεξάρτητα από τα έγγραφα:
 - Το αναγνωριστικό (ID) του δομικού στοιχείου μέσα στην οντότητα στην οποία αποθηκεύεται.
 - Η εσωτερική ταυτότητα (ObjectID) του δομικού στοιχείου.
 - Η Κλάση (Class) του δομικού στοιχείου.

- Το όνομα της οντότητας (EntityName) στην οποία αποθηκεύεται το δομικό στοιχείο.
- Για τα δομικά στοιχεία των εγγράφων που περιέχονται σε δομικά στοιχεία εγγράφων που αποθηκεύονται ανεξάρτητα από τα έγγραφα:
 - Η εσωτερική ταυτότητα (ObjectID) του δομικού στοιχείου.
 - Η Κλάση (Class) του δομικού στοιχείου.
 - Το όνομα της οντότητας (EntityName) στην οποία αποθηκεύεται το δομικό στοιχείο που περιέχει το τρέχον στοιχείο.
 - Το αναγνωριστικό (ID) του δομικού στοιχείου που περιέχει το τρέχον στοιχείο μέσα στην οντότητα στην οποία αποθηκεύεται.
 - Η εσωτερική ταυτότητα (FObjectID) του τρέχοντος δομικού στοιχείου μέσα στο δομικό στοιχείο που το περιέχει.
- Για τις διευθύνσεις:
 - Η εσωτερική ταυτότητα (ObjectID) της διεύθυνσης.
 - Η Κλάση (Class) της διεύθυνσης.
 - Το αναγνωριστικό (ID) της διεύθυνσης μέσα στην οντότητα στην οποία αποθηκεύεται.

Τύπος Αντικειμένου	Μορφή
Στοιχείο Εγγράφου	ObjectID.Class.Attributes.Offset.Size
Ανεξάρτητο Στοιχείο (ή Τμήμα) Εγγράφου	ObjectID.Class.EntityName.ID
Στοιχείο Εγγράφου που αποθηκεύεται σε Ανεξάρτητο Στοιχείο Εγγράφου	ObjectID.Class.EntityName.ID.FObjectID
Διεύθυνση	ObjectID.Class.ID

Πίνακας 6-4: Περιγραφή Στοιχείων Εγγράφου ή Ανεξάρτητων Στοιχείων Εγγράφου στον Περιγραφέα του Εγγράφου στο οποίο περιέχονται

Τα έγγραφα διαθέτουν μια σειρά χαρακτηριστικών (Attributes). Εκτός από τα εξειδικευμένα χαρακτηριστικά, που μπορεί να διαφέρουν από έγγραφο σε έγγραφο, ανάλογα με την κλάση στην οποία κάθε έγγραφο εντάσσεται, όλα τα έγγραφα διαθέτουν ένα σύνολο χαρακτηριστικών που αναπαριστούν τη σχετική με τα έγγραφα μεταπληροφορία. Στην πραγματικότητα, τα χαρακτηριστικά αυτά αντιστοιχούν στα στοιχεία μεταπληροφορίας που προτείνονται από την κοινότητα Dublin Core για την περιγραφή πόρων στο Διαδίκτυο. Τα χαρακτηριστικά αυτά είναι τα εξής:

- *Τίτλος*, που αποτελεί τον τίτλο του τρέχοντος εγγράφου.
- *Θέμα*, που αναπαριστά το θέμα του τρέχοντος εγγράφου.
- *Γλώσσα*, που αναπαριστά τη γλώσσα στην οποία είναι γραμμένο το τρέχον έγγραφο.
- *Τύπος*, που αντιστοιχεί στην κλάση του εγγράφου (και για το λόγο αυτό δε φαίνεται στο Διάγραμμα 6-5).

- *Ημερομηνία*, που δηλώνει τη χρονική στιγμή κατά την οποία δημιουργήθηκε το έγγραφο.
- *Περιγραφή*, που αποτελεί μια περιγραφή του περιεχομένου του τρέχοντος εγγράφου. Η περιγραφή ενός εγγράφου συμπίπτει με τη Σύνοψη (Abstract) του εγγράφου, αν το έγγραφο διαθέτει σύνοψη.
- *Σχέση*, που αναπαριστά τη σχέση του τρέχοντος εγγράφου με άλλα αντικείμενα (τοπικά ή απομακρυσμένα).
- *Διαχειριστής*, που αναπαριστά κάποιο πρόσωπο που σχετίζεται με την παραγωγή του τρέχοντος εγγράφου (π.χ. συγγραφέας, συνεργάτης, εκδότης κλπ.).
- *Ταυτότητα*, που αναπαριστά ένα μοναδικό αναγνωριστικό (εκτός βάσης δεδομένων) για το έγγραφο (π.χ. URL, αριθμός ISBN κλπ.) και αντιστοιχεί στο χαρακτηριστικό Εξωτερική Ταυτότητα.
- *Κάλυψη*, που αναπαριστά τη χωρική περιοχή και τη χρονική περίοδο που καλύπτει το τρέχον έγγραφο. Πιο συγκεκριμένα, η *Χρονική Κάλυψη* αναπαριστά τη χρονική περίοδο που καλύπτει το τρέχον έγγραφο και η *Χωρική Κάλυψη* αναπαριστά την αντίστοιχη χωρική περιοχή.
- *Μορφή*, που περιγράφει τη μορφή στην οποία ήταν αποθηκευμένο το έγγραφο πριν εισαχθεί στη βάση δεδομένων (π.χ. για XML έγγραφα το χαρακτηριστικό αυτό παίρνει την τιμή XML).

Επιπλέον, η σχετική με το περιεχόμενο των εγγράφων και των δομικών στοιχείων των εγγράφων μεταπληροφορία απαρτίζεται από τις συσχετίσεις τους με κατηγορίες, υποκατηγορίες και λέξεις – κλειδιά.

Όπως φαίνεται στο Διάγραμμα 6-5, τα στιγμιότυπα των οντοτήτων «Κλάση Εγγράφων» και «Κλάση Τμημάτων Εγγράφων» περιέχουν μεταπληροφορία που σχετίζεται με τις κλάσεις που αναπαριστούν και πιο συγκεκριμένα:

- Το όνομα της κλάσης που αναπαρίσταται από το τρέχον στιγμιότυπο.
- Την υπερκλάση της κλάσης που αναπαρίσταται από το τρέχον στιγμιότυπο.
- Το προκαθορισμένο, για τα στιγμιότυπα της κλάσης που αναπαρίσταται από το τρέχον στιγμιότυπο, στυλ μορφοποίησης.
- Τη namespace της κλάσης που αναπαρίσταται από το τρέχον στιγμιότυπο.
- Το DTD της κλάσης που αναπαρίσταται από το τρέχον στιγμιότυπο.
- Τους Κανόνες Αναπαράστασης της κλάσης που αναπαρίσταται από το τρέχον στιγμιότυπο. Για κάθε κλάση εγγράφων και τμημάτων εγγράφων καθορίζεται, πριν αυτές υποστηριχθούν από μια πηγή πληροφορίας, το σύνολο των Κανόνων Αναπαράστασης των στιγμιότυπων της. Οι κανόνες αυτοί αποθηκεύονται σε RDF αρχεία, των οποίων οι URLs αποτελούν την τιμή του πεδίου «Κανόνες Αναπαράστασης» των στιγμιότυπων των αντίστοιχων κλάσεων. Τα αρχεία αυτά είναι δομημένα με βάση τη γλώσσα ΚΑΔΔΑ (βλέπε 7^ο κεφάλαιο).

- Τη Δομή της κλάσης που αναπαρίσταται από το τρέχον στιγμιότυπο. Για κάθε κλάση εγγράφων αποθηκεύεται στο πεδίο «Δομή» του στιγμιότυπου από το οποίο αυτή αναπαρίσταται, ο URL του αρχείου όπου αποθηκεύεται, με τη μορφή RDF σχήματος, η δομή και τα χαρακτηριστικά των στιγμιότυπων της τρέχουσας κλάσης εγγράφων και των δομικών στοιχείων από τα οποία απαρτίζονται. Το σύνολο αυτών των RDF σχημάτων, ενσωματωμένο στο RDF σχήμα που περιγράφει το αρχικό μοντέλο πληροφορίας της βάσης δεδομένων της πηγής πληροφορίας, αποτελεί τη δομή της πηγής πληροφορίας όπως αυτή αποθηκεύεται στη βάση δεδομένων του διαμεσολαβητή. Το RDF σχήμα που περιγράφει τη δομή της βάσης δεδομένων μιας πηγής πληροφορίας που περιέχει δομημένα έγγραφα πριν την υποστήριξη κάποιας κλάσης εγγράφων φαίνεται στο RDF Σχήμα 6-1 και αντιστοιχεί στο URL "http://irida.ced.tuc.gr/RDF/ΠηγήΠληροφορίαςΕγγράφων.rdfs".

```
<?xml version='1.0'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema"
  xmlns:ΠηγήΠληροφορίαςΕγγράφων="">
  <rdfs:Class ID="αντικείμενο">
    <rdfs:label>αντικείμενο</rdfs:label>
    <rdfs:SubclassOf
      rdf:resource="http://www.w3.org/TR/PR-rdf-schema#Resource" />
  </rdfs:Class>
  <rdf:Property ID="ταυτότητα">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#Number">
    <rdfs:domain rdf:resource="#αντικείμενο" />
  </rdf:Property>
  <rdfs:Class ID="κλάση">
    <rdfs:label>κλάση</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
    <rdfs:domain rdf:resource="#αντικείμενο" />
  </rdfs:Class>
  <rdf:Property ID="υπερκλάση">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#κλάση" />
  </rdf:Property>
  <rdf:Property ID="όνομα">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#κλάση" />
    <rdfs:domain rdf:resource="#κατηγορία" />
    <rdfs:domain rdf:resource="#υποκατηγορία" />
    <rdfs:domain rdf:resource="#λέξη-κλειδί" />
    <rdfs:domain rdf:resource="#σχέση" />
    <rdfs:domain rdf:resource="#διαχειριστής" />
  </rdf:Property>
  <rdfs:Class ID="κλάση_αντικειμένων_κειμένου">
    <rdfs:label>κλάση_αντικειμένων_κειμένου</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#κλάση" />
  </rdfs:Class>
  <rdfs:Class ID="κλάση_εγγράφων">
    <rdfs:label>κλάση_εγγράφων</rdfs:label>
```

```

    <rdfs:SubclassOf rdf:resource="#κλάση_αντικειμένων_κειμένου" />
</rdfs:Class>
<rdfs:Class ID="κλάση_τμημάτων_εγγράφου">
    <rdfs:label>κλάση_τμημάτων_εγγράφου</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#κλάση_αντικειμένων_κειμένου" />
</rdfs:Class>
<rdf:Property ID="δομή">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση_αντικειμένων_κειμένου" />
</rdf:Property>
<rdf:Property ID="κανόνες_αναπαράστασης">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση_αντικειμένων_κειμένου" />
</rdf:Property>
<rdf:Property ID="στιλ">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση_αντικειμένων_κειμένου" />
</rdf:Property>
<rdf:Property ID="namespace">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση_αντικειμένων_κειμένου" />
</rdf:Property>
<rdfs:Class ID="γλώσσα">
    <rdfs:label>γλώσσα</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdfs:Class ID="κατηγορία">
    <rdfs:label>κατηγορία</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdfs:Class ID="υποκατηγορία">
    <rdfs:label>υποκατηγορία</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdf:Property ID="κατηγορία">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#υποκατηγορία" />
</rdf:Property>
<rdfs:Class ID="λέξη-κλειδί">
    <rdfs:label>λέξη-κλειδί</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdfs:Class ID="θησαυρός">
    <rdfs:label>θησαυρός</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdfs:Class ID="διεύθυνση_αντικειμένου">
    <rdfs:label>διεύθυνση_αντικειμένου</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdf:Property ID="URL">

```

```

    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#διεύθυνση_αντικειμένου" />
</rdf:Property>
<rdf:Property ID="τοπικό_αντικείμενο">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#Number">
    <rdf:domain rdf:resource="#διεύθυνση_αντικειμένου" />
</rdf:Property>
<rdfs:Class ID="διαχειριστής">
    <rdfs:label>διαχειριστής</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdf:Property ID="επώνυμο">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
    <rdf:domain rdf:resource="#διαχειριστής" />
</rdf:Property>
<rdf:Property ID="e-mail">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
    <rdf:domain rdf:resource="#διαχειριστής" />
</rdf:Property>
<rdf:Property ID="ρόλος">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
    <rdf:domain rdf:resource="#διαχειριστής" />
</rdf:Property>
<rdf:Property ID="τηλέφωνο">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
    <rdf:domain rdf:resource="#διαχειριστής" />
</rdf:Property>
<rdf:Property ID="οργανισμός">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
    <rdf:domain rdf:resource="#διαχειριστής" />
</rdf:Property>
<rdf:Property ID="διεύθυνση">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
    <rdf:domain rdf:resource="#διαχειριστής" />
</rdf:Property>
<rdfs:Class ID="σχέση">
    <rdfs:label>σχέση</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdf:Property ID="μέλος_1">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
    <rdf:domain rdf:resource="#σχέση" />
</rdf:Property>
<rdf:Property ID="μέλος_2">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
    <rdf:domain rdf:resource="#σχέση" />
</rdf:Property>
<rdfs:Class ID="κάλυψη">
    <rdfs:label>κάλυψη</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>

```

```

<rdfs:Class ID="χωρική_κόλυψη">
  <rdfs:label>χωρική_κόλυψη</rdfs:label>
  <rdfs:SubclassOf rdf:resource="#κόλυψη" />
</rdfs:Class>
<rdf:Property ID="περιοχή">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
  <rdf:domain rdf:resource="#χωρική_κόλυψη" />
</rdf:Property>
<rdfs:Class ID="χρονική_κόλυψη">
  <rdfs:label>χρονική_κόλυψη</rdfs:label>
  <rdfs:SubclassOf rdf:resource="#κόλυψη" />
</rdfs:Class>
<rdf:Property ID="περίοδος">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
  <rdf:domain rdf:resource="#χρονική_κόλυψη" />
</rdf:Property>
<rdfs:Class ID="αντικείμενο_κειμένου">
  <rdfs:label>αντικείμενο_κειμένου</rdfs:label>
  <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdf:Property ID="περιγραφέας">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
  <rdf:domain rdf:resource="#αντικείμενο_κειμένου" />
</rdf:Property>
<rdf:Property ID="περιεχόμενο">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#Text">
  <rdf:domain rdf:resource="#αντικείμενο_κειμένου" />
</rdf:Property>
<rdf:Property ID="ημερομηνία">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#Date">
  <rdf:domain rdf:resource="#αντικείμενο_κειμένου" />
</rdf:Property>
<rdf:Property ID="μορφή">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
  <rdf:domain rdf:resource="#αντικείμενο_κειμένου" />
</rdf:Property>
<rdf:Property ID="μέγεθος">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
  <rdf:domain rdf:resource="#αντικείμενο_κειμένου" />
</rdf:Property>
<rdfs:Class ID="τιμήμα_εγγράφου">
  <rdfs:label>τιμήμα_εγγράφου</rdfs:label>
  <rdfs:SubclassOf rdf:resource="#αντικείμενο_κειμένου" />
</rdfs:Class>
<rdf:Property ID="ταυτότητα_περιέχοντος">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#Number">
  <rdf:domain rdf:resource="#τιμήμα_εγγράφου" />
</rdf:Property>
<rdf:Property ID="εσωτερική_ταυτότητα">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#Number">
  <rdf:domain rdf:resource="#τιμήμα_εγγράφου" />

```

```

</rdf:Property>
<rdfs:Class ID="#έγγραφο">
  <rdfs:label>έγγραφο</rdfs:label>
  <rdfs:SubclassOf rdf:resource="#αντικείμενο_κειμένου" />
</rdfs:Class>
<rdf:Property ID="τίτλος">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
  <rdf:domain rdf:resource="#έγγραφο" />
</rdf:Property>
<rdf:Property ID="θέμα">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
  <rdf:domain rdf:resource="#έγγραφο" />
</rdf:Property>
<rdf:Property ID="περιγραφή">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String">
  <rdf:domain rdf:resource="#έγγραφο" />
</rdf:Property>
<rdf:Property ID="εξωτερική_ταυτότητα">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#Number">
  <rdf:domain rdf:resource="#έγγραφο" />
</rdf:Property>
</rdf:RDF>

```

RDF Σχήμα 6-1: RDF Σχήμα που περιγράφει τη Δομή της Βάσης Δεδομένων μιας Πηγής Πληροφορίας που περιέχει Δομημένα Έγγραφα πριν την υποστήριξη κάποιας Κλάσης Εγγράφων

Στο σημείο αυτό πρέπει να αναφερθεί ότι μαζί με κάθε Τμήμα Εγγράφου αποθηκεύονται με τη μορφή χαρακτηριστικών πληροφορία και μεταπληροφορία (όπως είναι τα στυλ παρουσίασης) ικανές να εγγυηθούν ότι το τρέχον τμήμα μπορεί να παρουσιαστεί αυτόνομα, χωρίς να χαθούν τα χαρακτηριστικά που καθόρισε ο συγγραφέας του εγγράφου του οποίου αποτελεί τμήμα. Πρέπει ακόμα να σημειωθεί ότι κάθε Τμήμα Εγγράφου διαθέτει το πεδίο Ταυτότητα Περιέχοντος, που περιέχει το αναγνωριστικό του εγγράφου στο οποίο το Τμήμα Εγγράφου περιέχεται και ένα πεδίο που δηλώνει την εσωτερική ταυτότητα του τμήματος εγγράφου μέσα στο έγγραφο στο οποίο περιέχεται.

Τέλος, πρέπει να σημειωθεί ότι ο συνδυασμός της Δομής και των Κανόνων Αναπαράστασης για κάθε κλάση εγγράφων χρησιμοποιούνται από το *Επίπεδο Διεπαφής με XML Έγγραφα* (βλέπε υποενότητα 6.3.6.4) ώστε να είναι δυνατή η υποστήριξη ενός ενιαίου τρόπου πρόσβασης στα έγγραφα και τα δομικά στοιχεία τους, ανεξάρτητα από το πως τα τελευταία αποθηκεύονται.

6.3.6.2. XML Έγγραφα

Όπως αναφέρθηκε παραπάνω, στα πλαίσια του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» στην παρούσα φάση οι πηγές πληροφορίας που περιέχουν δομημένα έγγραφα περιέχουν και διαχειρίζονται XML έγγραφα. Έτσι, σ' αυτή την υποενότητα περιγράφεται ο τρόπος αναπαράστασης XML εγγράφων σύμφωνα με το μοντέλο αναπαράστασης δομημένων εγγράφων που περιγράφηκε παραπάνω.

Ένα βασικό χαρακτηριστικό των XML εγγράφων είναι ότι τα έγγραφα αυτά δεν υπακούουν σε μια προκαθορισμένη σύνταξη: Τα XML έγγραφα μπορεί να υπακούουν σε κάποιο (ή σε κάποιους) DTD(s) ή XML Σχήμα(τα) (οι οποίοι είναι γλώσσες περιγραφής εγγράφων ορισμένες σύμφωνα με την XML) ή να μην υπακούουν σε κανένα DTD ή XML Σχήμα. Έτσι λοιπόν μπορεί να θεωρηθεί ότι κάθε DTD ή XML Σχήμα περιγράφει μια κλάση εγγράφων, ενώ τα έγγραφα που δεν υπακούουν σε κάποιο DTD ή XML Σχήμα ανήκουν στην ίδια κλάση, την κλάση των μη κατηγοριοποιημένων εγγράφων.

Από τα παραπάνω φαίνεται ότι το σχήμα και το μοντέλο πληροφορίας της βάσης δεδομένων όπου θα αποθηκεύονται XML έγγραφα θα εξελίσσονται κάθε φορά που θα υποστηρίζεται ένας νέος DTD ή XML Σχήμα από το σύστημα.

Για την κάλυψη των απαιτήσεων που προκύπτουν από τις παραπάνω διαπιστώσεις ακολουθήθηκε η εξής μεθοδολογία:

- Τα έγγραφα που δε δομούνται με βάση κάποιο DTD ή κάποιο XML Σχήμα αποθηκεύονται ως στιγμιότυπα της κλάσης Έγγραφο.
- Για κάθε νέο DTD ή XML Σχήμα που υποστηρίζεται από το σύστημα, ορίζεται μια υποκλάση της κλάσης Έγγραφο ή κάποιας υποκλάσης της κλάσης Έγγραφο με όνομα το όνομα του DTD ή του XML Σχήματος.
- Τα έγγραφα που υπακούουν σε κάποιο DTD ή το XML Σχήμα αποθηκεύονται ως στιγμιότυπα της ομώνυμης κλάσης. Πιο συγκεκριμένα, εισάγεται ένα στιγμιότυπο της κλάσης Κλάση Εγγράφων με όνομα <όνομα DTD ή XML Σχήματος> και δημιουργείται η ομώνυμη οντότητα.
- Για κάθε κλάση εγγράφων καθορίζεται αν θα επιτρέπεται ή όχι η χωριστή αποθήκευση κάποιων από τα δομικά στοιχεία που απαρτίζουν τα στιγμιότυπά της. Καθορίζεται επίσης αν θα επιτρέπεται η χωριστή αποθήκευση των στοιχείων αυτών ανεξαρτήτως του μεγέθους του εγγράφου στο οποίο περιέχονται ή αν το μέγεθος του περιέχοντος εγγράφου πρέπει να ξεπερνά μια συγκεκριμένη τιμή. Αν το τελευταίο ισχύει, καθορίζεται και η ελάχιστη τιμή που πρέπει να έχει το μέγεθος του εγγράφου για να αποθηκευτούν χωριστά κάποια από τα δομικά στοιχεία του. Επιπλέον, αν θα πρέπει να ισχύουν αντίστοιχα κριτήρια για το μέγεθος των στοιχείων αυτών, καθορίζονται τα κριτήρια αυτά.
- Για καθένα από τα δομικά στοιχεία εγγράφων που ορίζονται σε κάποιο DTD ή XML Σχήμα καθορίζεται αν τα στιγμιότυπά του θα επιτρέπεται να αποθηκεύονται ανεξάρτητα. Για καθένα από τα δομικά στοιχεία των εγγράφων που επιτρέπεται να αποθηκεύεται ανεξάρτητα δημιουργείται μια υποκλάση της κλάσης Τμήμα Εγγράφου με το όνομα <όνομα DTD ή XML Σχήματος>:<όνομα δομικού στοιχείου>. Πιο συγκεκριμένα, εισάγεται ένα στιγμιότυπο της κλάσης Κλάση Τμημάτων Εγγράφων με όνομα <όνομα DTD ή XML Σχήματος>:<όνομα δομικού στοιχείου> και δημιουργείται η ομώνυμη οντότητα.

Για τα στιγμιότυπα της κλάσης αυτής προσδιορίζεται αν επιτρέπεται κάποια από τα στοιχεία που τα απαρτίζουν να αποθηκεύονται χωριστά. Καθορίζεται επίσης ποιο είναι το ελάχιστο μέγεθος που πρέπει να έχουν τα στιγμιότυπα καθεμιάς από αυτές τις κλάσεις για να αποθηκευτούν χωριστά τα υποστοιχεία τους. Επιπλέον, αν πρέπει να

ισχύουν αντίστοιχα κριτήρια για το μέγεθος των υποστοιχείων, καθορίζονται τα κριτήρια αυτά.

Αν τα στιγμιότυπα κάποιου δομικού στοιχείου δεν επιτρέπεται να αποθηκεύονται ανεξάρτητα, τότε εισάγεται ένα στιγμιότυπο της κλάσης Κλάση Τμημάτων Εγγράφων με όνομα <όνομα DTD ή XML Σχήματος>:<όνομα δομικού στοιχείου>, αλλά δε δημιουργείται η ομώνυμη οντότητα.

- Αν υπάρχουν κάποια στοιχεία ή χαρακτηριστικά στοιχείων στο DTD ή το XML Σχήμα που αναπαριστούν τα χαρακτηριστικά που περιγράφουν τη μεταπληροφορία των εγγράφων, ορίζονται οι αντιστοιχίσεις που ισχύουν. Έτσι, π.χ., αν για τα έγγραφα που υπακούουν σε κάποιο DTD ο συγγραφέας αναπαρίσταται από το στοιχείο Συγγραφέας, το χαρακτηριστικό διαχειριστής για τα έγγραφα αυτά παίρνει την τιμή {Διαχειριστής=<τιμή του στοιχείου Συγγραφέας>, Ρόλος="Συγγραφέας"}.

Η υλοποίηση της μεθοδολογίας αυτής γίνεται από το εργαλείο SupportDTD που περιγράφεται στο 7^ο κεφάλαιο.

6.3.6.3. Σχήμα Βάσης Δεδομένων

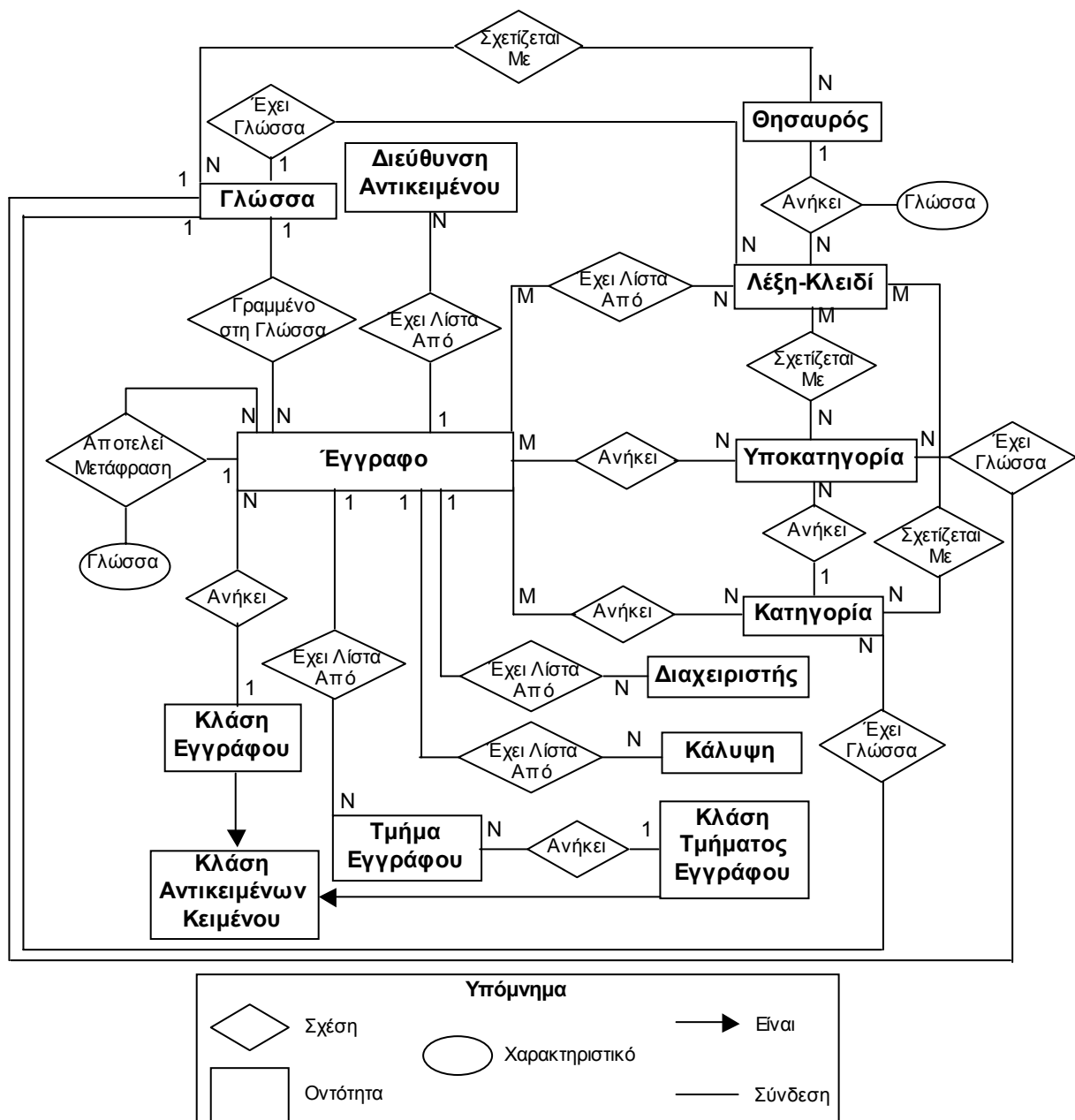
Στο τμήμα αυτό περιγράφεται το σχήμα που υλοποιεί το μοντέλο που περιγράφηκε στα τμήματα 6.3.6.1 και 6.3.6.2 για την αποθήκευση XML εγγράφων στις σχεσιακές βάσεις δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Το E-R διάγραμμα (στο οποίο φαίνονται και όλα τα στοιχεία που περιγράφουν τη σχετική με τα έγγραφα μεταπληροφορία που δεν αποτελούν πεδία των πινάκων που αναπαριστούν τις αντίστοιχες σχέσεις) της βάσης δεδομένων όπου αποθηκεύονται τα δομημένα έγγραφα φαίνεται στο Διάγραμμα 6-6.

Το διάγραμμα αυτό είναι γενικευμένο, καθώς απ' αυτό απουσιάζουν οι σχέσεις που αναπαριστούν συγκεκριμένες κλάσεις εγγράφων και τμημάτων εγγράφων, καθώς οι σχέσεις αυτές δημιουργούνται κάθε φορά που υποστηρίζεται από το σύστημα ένας νέος DTD ή ένα νέο XML Σχήμα. Έτσι, το Διάγραμμα 6-6 περιγράφει πλήρως τη βάση δεδομένων μιας πηγής πληροφορίας που περιέχει δομημένα έγγραφα πριν από την υποστήριξη του πρώτου DTD ή XML Σχήματος από την πηγή πληροφορίας.

Έστω τώρα ότι ο πρώτος DTD που υποστηρίζεται από το σύστημα είναι ο ακόλουθος DTD (που έχει το όνομα Άρθρο.dtd), ο οποίος περιγράφει τη δομή και τη σύνταξη Άρθρων.

```
<!ELEMENT ΑΡΘΡΟ (ΤΙΤΛΟΣ,ΣΥΓΓΡΑΦΕΑΣ,ΣΥΝΟΨΗ,ΕΝΟΤΗΤΑ*)>
<!ELEMENT ΣΥΓΓΡΑΦΕΑΣ EMPTY>
<!ATTLIST ΣΥΓΓΡΑΦΕΑΣ ΟΝΟΜΑ CDATA #IMPLIED
           ΕΠΩΝΥΜΟ CDATA #IMPLIED>
<!ELEMENT ΕΝΟΤΗΤΑ (ΠΑΡΑΓΡΑΦΟΣ*, ΥΠΟΕΝΟΤΗΤΑ*)>
<!ELEMENT ΥΠΟΕΝΟΤΗΤΑ (ΠΑΡΑΓΡΑΦΟΣ*)>
<!ELEMENT ΠΑΡΑΓΡΑΦΟΣ (#PCDATA|EM)*>
<!ELEMENT EM (#PCDATA)>
<!ELEMENT ΣΥΝΟΨΗ (#PCDATA)>
<!ELEMENT ΤΙΤΛΟΣ (#PCDATA)>
```

DTD 6-2: Ο DTD Άρθρο.dtd που περιγράφει τη δομή και τη σύνταξη Άρθρων



Διάγραμμα 6-6: E-R Διάγραμμα της Βάσης Δεδομένων πριν υποστηριχθεί κάποιος DTD ή κάποιο XML Σχήμα

Έστω ακόμα ότι οι κανόνες βάσει των οποίων θα αποθηκεύονται τα έγγραφα που υπακούουν στο DTD αυτό, είναι οι εξής³²:

- Η κλάση Άρθρο αποτελεί υποκλάση της κλάσης Έγγραφο (που είναι η μοναδική διαθέσιμη κλάση εγγράφων).
- Επιτρέπεται η τμηματική αποθήκευση των εγγράφων που υπακούουν στο DTD Άρθρο.dtd, τα οποία είναι και στιγμιότυπα της κλάσης Άρθρο, ανεξαρτήτως του μεγέθους τους.

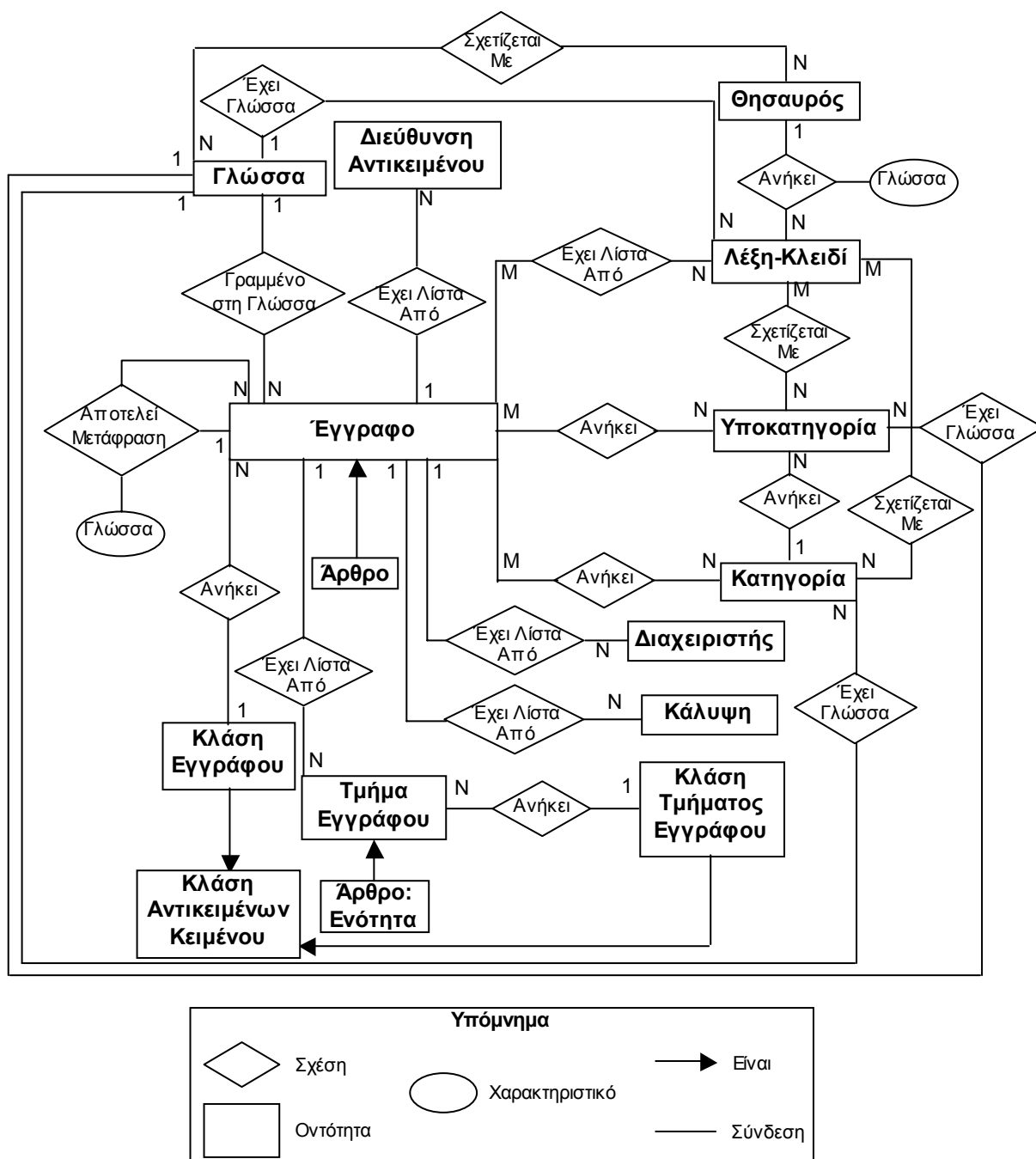
³² Η σύνταξη των κανόνων, η οποία είναι συμβατή με τη γλώσσα KΑΔΔΑ, περιγράφεται στο τμήμα 7.2. Στο τμήμα 7.2 περιγράφονται και οι μηχανισμοί που υλοποιούν την υποστήριξη των κανόνων αυτών.

- Από τα στιγμιότυπα της κλάσης Ενότητα επιτρέπεται να αποθηκεύονται τμηματικά μόνο αυτά που το μέγεθός τους υπερβαίνει τα 50 KB και μόνο στην περίπτωση που το μέγεθός του εγγράφου στο οποίο περιέχονται υπερβαίνει τα 200 KB.
- Τα στιγμιότυπα των κλάσεων Τίτλος, Συγγραφέας, Παράγραφος, Υποενότητα και Σύνοψη αποθηκεύονται πάντα μαζί με το αντικείμενο στο οποίο περιέχονται.
- Τα στιγμιότυπα της κλάσης Τίτλος, που αποτελούν τους τίτλους των αντίστοιχων άρθρων, αντιστοιχούν και στο ομώνυμο στοιχείο μεταπληροφορίας και αποθηκεύονται ως χαρακτηριστικά.
- Τα στιγμιότυπα της κλάσης Σύνοψη, που αποτελούν τις συνόψεις των αντίστοιχων άρθρων, αντιστοιχούν και στο στοιχείο μεταπληροφορίας «Περιγραφή» και αποθηκεύονται ως χαρακτηριστικά.
- Τα στιγμιότυπα της κλάσης Συγγραφέας, που αποτελούν τους συγγραφείς των αντίστοιχων άρθρων, αντιστοιχούν και στο στοιχείο μεταπληροφορίας «Διαχειριστής» με Ρόλο «Συγγραφέας».
- Η βαρύτητα των όρων που περιέχονται στο δομικό στοιχείο Τίτλος κάθε εγγράφου είναι 50% μεγαλύτερη από αυτή των όρων που περιέχονται στα υπόλοιπα δομικά στοιχεία που απαρτίζουν το έγγραφο.
- Η βαρύτητα των όρων που περιέχονται στο δομικό στοιχείο Σύνοψη κάθε εγγράφου είναι 25% μεγαλύτερη από αυτή των όρων που περιέχονται στα υπόλοιπα δομικά στοιχεία που απαρτίζουν το έγγραφο (εκτός το δομικό στοιχείο Τίτλος).
- Η βαρύτητα των όρων που περιέχονται και είναι τονισμένοι (περιέχονται σε κάποιο στιγμιότυπο του στοιχείου em) είναι κατά 5% μεγαλύτερη από αυτή των όρων που περιέχονται στις αντίστοιχες παραγράφους.
- Οι κλάσεις em και Τίτλος δεν εξάγονται μέσω του μοντέλου πληροφορίας της πηγής πληροφορίας.

Καθώς ο Άρθρο.dtd είναι ο πρώτος DTD που υποστηρίζεται από το σύστημα, το E-R διάγραμμα που περιγράφει τη βάση δεδομένων μετά την υποστήριξή του φαίνεται στο Διάγραμμα 6-7. Η βάση δεδομένων που περιγράφεται από αυτό το E-R διάγραμμα υλοποιεί το μοντέλο αναπαράστασης εγγράφων που περιγράφηκε στο τμήμα 6.3.6.1, καθώς:

- Τα έγγραφα που υπακούουν στο DTD Άρθρο.dtd ανήκουν στην κλάση εγγράφων Άρθρο και αναπαρίστανται από την ομώνυμη οντότητα.
- Οι ενότητες των άρθρων, τα οποία αποθηκεύονται ξεχωριστά από τα άρθρα, ανήκουν στην κλάση τμημάτων εγγράφων Άρθρο:Ενότητα και αναπαρίστανται από την ομώνυμη οντότητα.

Στο σημείο αυτό πρέπει να σημειωθεί ότι οι οντότητες που αποτελούν εξειδικεύσεις οντοτήτων κληρονομούν τις συσχετίσεις που έχουν οι οντότητες-γενικεύσεις τους.



Διάγραμμα 6-7: E-R Διάγραμμα της Βάσης Δεδομένων, μετά την υποστήριξη του DTD Άρθρο.dtd

6.3.6.4. Επίπεδο Διεπαφής με XML Έγγραφα

Το θέμα αυτής της ενότητας είναι η περιγραφή του επιπέδου διεπαφής με XML έγγραφα που παρέχεται από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα. Το επίπεδο διεπαφής έχει βασικό στόχο την παροχή πρόσβασης με ενιαίο τρόπο τόσο στα έγγραφα όσο και στα δομικά στοιχεία των εγγράφων, άσχετα από το αν τα τελευταία αποθηκεύονται μαζί με τα έγγραφα στα οποία περιέχονται ή ανεξάρτητα από αυτά. Με τον τρόπο αυτό παρέχεται η δυνατότητα χρήσης ενιαίας σύνταξης από τη γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα για όλα τα έγγραφα και τα δομικά στοιχεία εγγράφων, ανεξάρτητα από την πολιτική διαχείρισης και αποθήκευσής τους.

Το επίπεδο διεπαφής με XML έγγραφα βασίζεται στη Δομή και στους Κανόνες Αναπαράστασης κάθε κλάσης εγγράφων που υποστηρίζεται από μια πηγή πληροφορίας που περιέχει δομημένα έγγραφα. Η Δομή μιας κλάσης, όπως έχει ήδη αναφερθεί, είναι ένα RDF σχήμα, που αναπαριστά το τμήμα εκείνο του μοντέλου πληροφορίας της πηγής πληροφορίας, το οποίο σχετίζεται με την κλάση αυτή. Αντίστοιχα, οι Κανόνες Αναπαράστασης μιας κλάσης περιγράφουν την πολιτική αναπαράστασης και διαχείρισης των στιγμιοτύπων της κλάσης αυτής.

Το επίπεδο διεπαφής με XML έγγραφα αποτελεί τμήμα του επεξεργαστή της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και αναλαμβάνει τη μετατροπή των τμημάτων των ερωτήσεων και των αιτήσεων που έχουν να κάνουν με δομημένα έγγραφα σε SQL σύνταξη. Με βάση τα παραπάνω, η είσοδος που δέχεται το επίπεδο διεπαφής είναι σύμφωνη με τη Δομή της κλάσης, ενώ η έξοδος του είναι σύμφωνη με τους Κανόνες Αναπαράστασης για την κλάση αυτή.

Πιο συγκεκριμένα, η έξοδος που προκύπτει από το επίπεδο διεπαφής παίρνει, κατά περίπτωση, τις μορφές που φαίνονται στον πίνακα που ακολουθεί.

Αντικείμενο	Σύνταξη Γλώσσας Πηγής Πληροφορίας	Σύνταξη μετά την Επεξεργασία
Έγγραφα (μιας ή περισσότερων κλάσεων) που αποθηκεύονται, πλήρως ή μερικά, τοπικά	<Όνομα Κλάσης Εγγράφων 1>, ..., <Όνομα Κλάσης Εγγράφων N>	SELECT * FROM <Όνομα Κλάσης Εγγράφων 1>, ..., <Όνομα Κλάσης Εγγράφων N>
Τμήματα Εγγράφου που αποθηκεύονται <i>πάντα</i> ανεξάρτητα	<Όνομα Κλάσης Τμήματος 1>, ..., <Όνομα Κλάσης Τμήματος N>	SELECT * FROM <Όνομα Κλάσης Τμήματος 1>, ..., <Όνομα Κλάσης Τμήματος N>
Τμήματα Εγγράφου που αποθηκεύονται <i>πάντα</i> μαζί με το αντικείμενο στο οποίο περιέχεται	<Όνομα Κλάσης Τμήματος 1>, ..., <Όνομα Κλάσης Τμήματος N>	SELECT * FROM <Όνομα Κλάσης Εγγράφου 1/Όνομα Τμήματος 1>, ..., <Όνομα Κλάσης Εγγράφου N/Όνομα Τμήματος N>
Τμήματα Εγγράφου που αποθηκεύονται <i>πάντα</i> απομακρυσμένα	<Όνομα Κλάσης Τμήματος 1>, ..., <Όνομα Κλάσης Τμήματος N>	SELECT URL FROM <Όνομα Κλάσης Εγγράφου 1/Όνομα Τμήματος 1>, ..., <Όνομα Κλάσης Εγγράφου N/Όνομα Τμήματος N>
Τμήματα Εγγράφου που αποθηκεύονται <i>κατά περίπτωση</i> ανεξάρτητα, μαζί με το αντικείμενο στο οποίο περιέχεται ή απομακρυσμένα	<Όνομα Κλάσης Ανεξάρτητου Τμήματος 1>, ..., <Όνομα Κλάσης Ανεξάρτητου Τμήματος N>, <Όνομα Κλάσης Ιδεατού Τμήματος 1>, ..., <Όνομα Κλάσης Ιδεατού Τμήματος M>, <Όνομα Κλάσης Απομακρυσμένου Τμήματος 1>, ..., <Όνομα Απομακρυσμένου Τμήματος O>	SELECT * FROM <Όνομα Κλάσης Ανεξάρτητου Τμήματος 1>, ..., <Όνομα Κλάσης Ανεξάρτητου Τμήματος N>, <Όνομα Κλάσης Ιδεατού Τμήματος 1>, ..., <Όνομα Κλάσης Ιδεατού Τμήματος M>, <Όνομα Κλάσης Απομακρυσμένου Τμήματος 1>, ..., <Όνομα Κλάσης Απομακρυσμένου Τμήματος M+1/Όνομα Κλάσης Απομακρυσμένου Τμήματος M+O/Όνομα Απομακρυσμένου Τμήματος O>

		Τμήματος Ο>
Χαρακτηριστικό Εγγράφου ή Τμήματος Εγγράφου που αποθηκεύεται ανεξάρτητα	<Όνομα Κλάσης>@<Όνομα Χαρακτηριστικού>	<Όνομα Κλάσης>.<Όνομα Χαρακτηριστικού>
Χαρακτηριστικό Τμήματος Εγγράφου που αποθηκεύεται μαζί με το περιέχον Έγγραφο ή Τμήμα Εγγράφου	<Όνομα Κλάσης>@<Όνομα Χαρακτηριστικού>	<Όνομα Κλάσης Περιέχοντος/Όνομα Κλάσης>@<Όνομα Χαρακτηριστικού>

Πίνακας 6-5: Η Έξοδος του Επιπέδου Διεπαφής με XML Έγγραφα

Έτσι, για παράδειγμα, η αίτηση που φαίνεται στο Παράδειγμα 6-5 μετατρέπεται από το επίπεδο διεπαφής, αν υποθεθεί ότι οι ενότητες των άρθρων επιτρέπεται να αποθηκεύονται ανεξάρτητα από τα άρθρα στην περίπτωση που το μέγεθος των άρθρων όπου περιέχονται υπερβαίνει τα 200KB και μαζί με τα άρθρα σε κάθε άλλη περίπτωση, στην αίτηση που φαίνεται στο Παράδειγμα 6-6.

```
SELECT *
FROM Άρθρο/Ενότητα[3], Άρθρο:Ενότητα[3]
WHERE Άρθρο.Ταυτότητα="125"
```

Παράδειγμα 6-6: Αίτηση επεξεργασμένη από το Επίπεδο Διεπαφής με XML Έγγραφα

Στο σημείο αυτό πρέπει να αναφερθεί ότι ο μηχανισμός ανάκτησης αντικειμένων και δομικών στοιχείων αντικειμένων, που χρησιμοποιείται από τον επεξεργαστή της γλώσσας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, περιγράφεται στο 7^ο κεφάλαιο.

6.4. Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

Το θέμα αυτού του τμήματος είναι η περιγραφή των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα. Αυτή η κλάση πηγών πληροφορίας βασίζεται σε σχεσιακές βάσεις δεδομένων και το περιεχόμενό τους είναι, σε ό,τι αφορά τη δομή και τη μορφή του, αντίστοιχο με αυτό που αποθηκεύεται παραδοσιακά στα πληροφοριακά συστήματα που βασίζονται σε σχεσιακές βάσεις δεδομένων. Τα σημεία που διαφοροποιούν τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα από τις συνηθισμένες σχεσιακές βάσεις δεδομένων είναι τα εξής:

- Τα δεδομένα που αποθηκεύονται στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα μπορούν να εξαχθούν σε XML μορφή.
- Είναι δυνατή η εισαγωγή πληροφορίας που είναι σε μορφή XML στις βάσεις δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
- Υποστηρίζονται ερωτήσεις, οι οποίες αναφέρονται στο περιεχόμενο και/ή τη δομή των αντικειμένων και μπορεί να περιέχουν όρους δίτημης λογικής και/ή όρους ομοιότητας.

- Το σχήμα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα μπορεί να επεκταθεί για να υποστηρίξει κλάσεις που περιγράφονται από XML DTDs ή XML σχήματα.
- Το σχήμα των βάσεων δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα μπορεί να περιγραφεί μέσω ενός συνόλου XML DTDs ή XML σχημάτων.
- Το μοντέλο πληροφορίας των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα μπορεί να περιγραφεί από ένα RDF σχήμα. Έτσι υποστηρίζονται κάποιες επεκτάσεις του σχεσιακού μοντέλου πληροφορίας, όπως οι κλάσεις αντικειμένων και η κληρονομικότητα.

Από τα παραπάνω φαίνεται ότι η γενικευμένη αρχιτεκτονική των πηγών πληροφορίας που απεικονίζεται στο Διάγραμμα 6-1 καλύπτει τις απαιτήσεις των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Έτσι τα σημεία στα οποία θα επικεντρωθεί το υπόλοιπο αυτού του τμήματος είναι η αναλυτικότερη περιγραφή των τμημάτων που απαρτίζουν την αρχιτεκτονική των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα (αναφορά στη *Γλώσσα των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα* γίνεται στην ενότητα 6.4.1, ενώ στην ενότητα 6.4.3 περιγράφεται η *Βάση Δεδομένων των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα*) και η αναφορά στη *Λειτουργικότητα των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα* (που λαμβάνει χώρα στην ενότητα 6.4.3).

6.4.1. Γλώσσα των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

Σ' αυτή την ενότητα περιγράφεται η γλώσσα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα, η οποία έρχεται να καλύψει τις ανάγκες του μοντέλου πληροφορίας των πηγών πληροφορίας αυτής της κλάσης. Το μοντέλο πληροφορίας των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αποτελεί, όπως έχει ήδη αναφερθεί, μια επέκταση του σχεσιακού μοντέλου με αντικειμενοστραφή χαρακτηριστικά και μπορεί να περιγραφεί με τη χρήση ενός RDF σχήματος. Το γεγονός αυτό σημαίνει ότι το μοντέλο πληροφορίας των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα είναι διαχειρίσιμο για μια γλώσσα που μοιάζει με την SQL, καθώς το μοντέλο πληροφορίας που διαχειρίζεται η SQL μοιάζει με αυτό των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Η σύνταξη της γλώσσας των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα πρέπει να αποτελεί επέκταση της σύνταξης της SQL, καθώς το μοντέλο πληροφορίας που διαχειρίζεται η γλώσσα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα είναι πιο πλούσιο από πλευράς σημασιολογίας από το μοντέλο πληροφορίας που διαχειρίζεται η SQL.

Τέλος, υπάρχει η ανάγκη υποστήριξης από τη γλώσσα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα ερωτήσεων που περιέχουν τόσο όρους δίτημης λογικής όσο και όρους ομοιότητας.

Οι ανάγκες ικανοποίησης των απαιτήσεων αυτών, σε ότι αφορά την υποστήριξη ερωτήσεων και αιτήσεων υπήρχε και για τη γλώσσα του διαμεσολαβητή του συστήματος «Παροχής

Πληροφοριών κατ' Απαίτηση». Για το λόγο αυτό επιλέχθηκε η σύνταξη της γλώσσας των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα να αποτελεί υποσύνολο της σύνταξης της γλώσσας του διαμεσολαβητή. Πιο συγκεκριμένα, η λειτουργικότητα που παρέχει η γλώσσα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα σε ότι αφορά την υποστήριξη ερωτήσεων και αιτήσεων συνοψίζεται στα εξής σημεία:

- Στην παροχή της λειτουργικότητας που παρέχεται από την SQL για την υποστήριξη ερωτήσεων δίτιμης λογικής.
- Στην παροχή της δυνατότητας αναζήτησης τόσο μεταξύ των στιγμιοτύπων μιας κλάσης όσο και μεταξύ των στιγμιοτύπων των υποκλάσεων της κλάσης αυτής.
- Στην παροχή της δυνατότητας έκφρασης τόσο ερωτήσεων ομοιότητας όσο και ερωτήσεων που περιέχουν όρους δίτιμης λογικής και όρους ομοιότητας.
- Στην παροχή της δυνατότητας έκφρασης αιτήσεων για τα χαρακτηριστικά των αντικειμένων και για την ολική ή μερική ανάκτηση του περιεχομένου των αντικειμένων.

Η σύνταξη της γλώσσας των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα είναι ταυτόσημη, σε ότι αφορά την υποστήριξη της προαναφερθείσας λειτουργικότητας, με τη γλώσσα του διαμεσολαβητή. Ο Πίνακας 6-6 περιέχει το υποσύνολο της γλώσσας του διαμεσολαβητή που υποστηρίζει τη λειτουργικότητα αυτή.

Περιγραφή	Σύνταξη
Επιλογή στιγμιοτύπων κλάσης	<Όνομα Κλάσης>
Επιλογή στιγμιοτύπων μιας κλάσης και των υποκλάσεων της	<Όνομα Κλάσης>*
Συσχέτιση με όρους ομοιότητας	Τελεστής «RELATES TO», σύνολα όρων ((AND OR) TERMSETS), όροι (TERMS)
Ασαφή χαρακτηριστικά	<Όνομα Ασαφούς Χαρακτηριστικού>

Πίνακας 6-6: Η επιπλέον της SQL Λειτουργικότητα που παρέχεται από τη Γλώσσα των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

Σε ότι αφορά την πληροφορία που επιστρέφεται από τη γλώσσα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα, επιλέχθηκε αυτή να είναι σε XML μορφή και να δομείται σύμφωνα με το DTD 6-1. Η επιλογή αυτή έγινε με το ίδιο σκεπτικό που χρησιμοποιήθηκε για τη μορφή της πληροφορίας που επιστρέφει η γλώσσα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.

Το γεγονός ότι η γλώσσα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αποτελεί, σε ότι αφορά τη λειτουργικότητα και τη σύνταξη, γνήσιο υποσύνολο της γλώσσας του διαμεσολαβητή, έχει ως συνέπεια τα περιτυλίγματα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα να είναι, στην πραγματικότητα, απλοί διακομιστές των ερωτήσεων και των αιτήσεων από το διαμεσολαβητή προς τις πηγές πληροφορίας και των αντίστοιχων αποτελεσμάτων από τις πηγές πληροφορίας στο διαμεσολαβητή.

Εκτός όμως από την υποστήριξη ερωτήσεων και αιτήσεων, η γλώσσα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα υποστηρίζει, μέσω της σύνταξής της,

τις εξής λειτουργίες, που χρησιμοποιούνται από τα εργαλεία διαχείρισης των πηγών πληροφορίας και είναι απαραίτητες για κάθε πηγή πληροφορίας:

- Αρχικοποίηση της βάσης δεδομένων της τρέχουσας πηγής πληροφορίας.
- Δημιουργία νέων κλάσεων αντικειμένων. Για να γίνει δυνατή η εισαγωγή μιας κλάσης αντικειμένων, πρέπει να δοθούν τα εξής ορίσματα:
 - Ο URL του DTD ή του XML Σχήματος που περιγράφει τη δομή και τη σύνταξη των στιγμιотύπων της κλάσης.
 - Το όνομα του αρχείου που περιέχονται οι Κανόνες Αναπαράστασης των στιγμιотύπων της κλάσης στη βάση δεδομένων της πηγής πληροφορίας.
 - Το RDF Σχήμα που αναπαριστά το μοντέλο της συγκεκριμένης κλάσης, όπως αυτό γίνεται διαθέσιμο από τη γλώσσα της πηγής πληροφορίας.
- Εισαγωγή στιγμιотύπων αντικειμένων. Για να γίνει δυνατή η εισαγωγή ενός αντικειμένου, πρέπει να δοθούν ως ορίσματα το όνομα της κλάσης στην οποία το αντικείμενο ανήκει και οι τιμές των χαρακτηριστικών του ή, εναλλακτικά, ο URL του .
- Δεικτοδότηση απομακρυσμένων αντικειμένων. Για να γίνει δυνατή η δεικτοδότηση κάποιου απομακρυσμένου αντικειμένου πρέπει να δοθεί ως όρισμα ο URL του αντικειμένου που πρόκειται να δεικτοδοτηθεί.

Οι λειτουργίες αυτές υλοποιούνται από τους κατάλληλους για κάθε περίπτωση μηχανισμούς (βλέπε 7^ο κεφάλαιο), οι οποίοι ουσιαστικά υποστηρίζουν και τα αντίστοιχα εργαλεία. Ο Πίνακας 6-2 περιέχει τη σύνταξη που υποστηρίζει καθεμιά από τις λειτουργίες αυτές.

Λειτουργία	Σύνταξη
Αρχικοποίηση της Βάσης Δεδομένων	INITDB
Εισαγωγή Κλάσης	INSERTCLASS <DTD ή XML Σχήμα> <Αρχείο Κανόνων Αναπαράστασης> <RDF Schema>
Εισαγωγή Αντικειμένου	INSERT <Όνομα Κλάσης><Τιμές Χαρακτηριστικών> INSERT <URL>
Δεικτοδότηση Αντικειμένου	INDEX <URL>

Πίνακας 6-7: Οι λειτουργίες υποστήριξης των εργαλείων διαχείρισης των πηγών πληροφορίας που παρέχονται από τη γλώσσα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα

6.4.2. Βάση Δεδομένων των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

Το θέμα αυτής της ενότητας είναι η περιγραφή της βάσης δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Η βάση δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα είναι, στην πραγματικότητα, αντίστοιχη με τις παραδοσιακές βάσεις δεδομένων, με μόνη διαφορά την υποστήριξη των εξής επιπλέον χαρακτηριστικών:

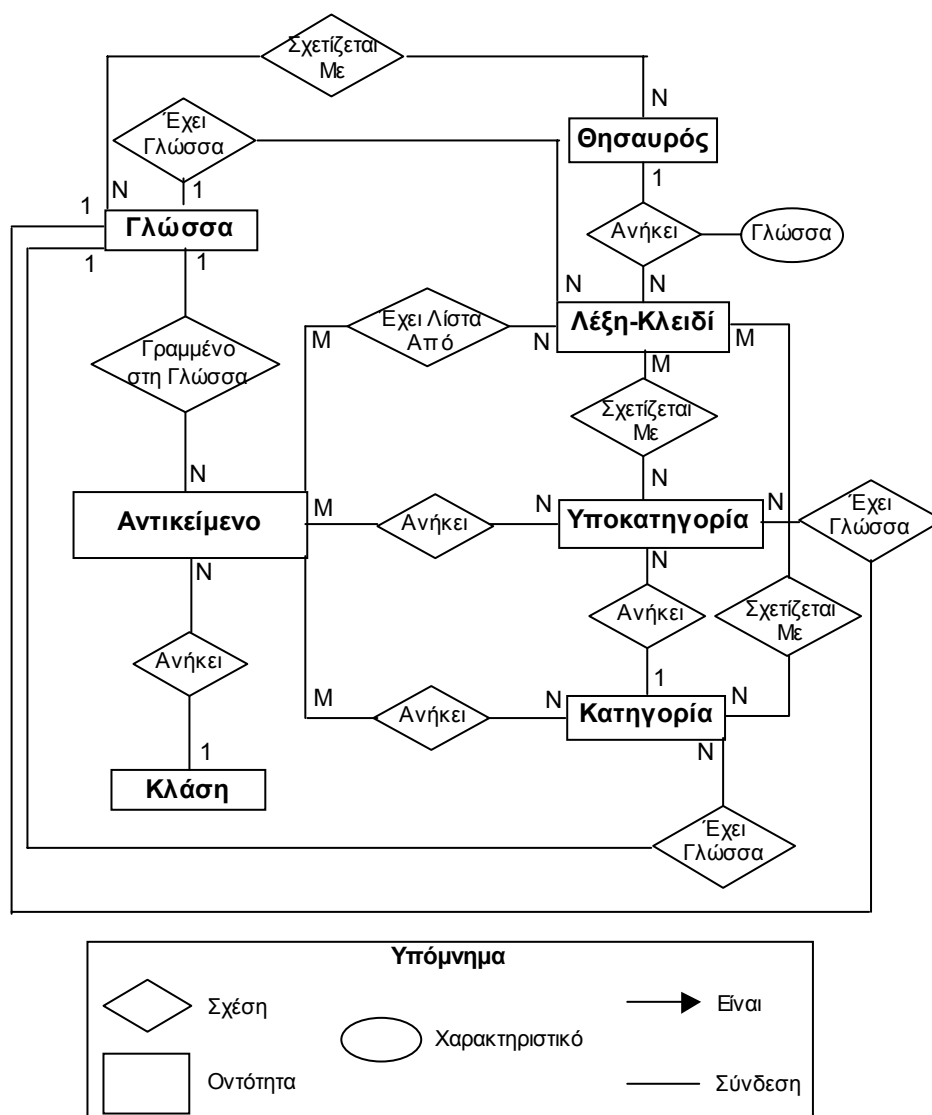
- Υποστήριξη κληρονομικότητας και κλάσεων αντικειμένων. Το χαρακτηριστικό αυτό απαιτεί την αρχικοποίηση του σχήματος της βάσης δεδομένων με τρόπο που θα κάνει δυνατή την υποστήριξη κλάσεων και κληρονομικότητας και θα επιτρέπει την επέκταση του σχήματος της βάσης δεδομένων με την υποστήριξη νέων κλάσεων και συσχετίσεων μεταξύ κλάσεων.
- Υποστήριξη ερωτήσεων ομοιότητας και ερωτήσεων που περιέχουν όρους δίτιμης λογικής και όρους ομοιότητας. Το χαρακτηριστικό αυτό απαιτεί την ύπαρξη κατηγοριών, υποκατηγοριών και λέξεων – κλειδιών που μπορούν να συσχετιστούν με το περιεχόμενο των αντικειμένων που αποθηκεύονται στη βάση δεδομένων και απαιτεί την κατάλληλη αρχικοποίηση της βάσης δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Για να γίνει δυνατή η υποστήριξη των χαρακτηριστικών αυτών, κατά την αρχικοποίηση της βάσης δεδομένων δημιουργείται το σχήμα που περιγράφεται από το E-R διάγραμμα που απεικονίζεται στο Διάγραμμα 6-8. Όπως φαίνεται στο Διάγραμμα 6-8, οι οντότητες που απαρτίζουν το E-R διάγραμμα χωρίζονται σε 3 ομάδες:

- *Οντότητες που αναπαριστούν αντικείμενα.* Σ' αυτή την ομάδα ανήκει αρχικά μόνο η οντότητα Αντικείμενο, η οποία αναπαριστά τα αντικείμενα που υπάρχουν στη βάση δεδομένων της πηγής πληροφορίας. Καθώς όμως το σχήμα της βάσης δεδομένων επεκτείνεται με την υποστήριξη νέων κλάσεων αντικειμένων, θα δημιουργούνται οντότητες που θα αναπαριστούν υποκλάσεις της κλάσης Αντικείμενο³³ (την οποία αναπαριστά η οντότητα Αντικείμενο).
- *Οντότητες που υποστηρίζουν την κατηγοριοποίηση των αντικειμένων με βάση το περιεχόμενό τους.* Οι οντότητες που ανήκουν σε αυτή την ομάδα είναι οι εξής:
 - Η οντότητα *Κατηγορία*, που αναπαριστά τις κατηγορίες περιεχομένου στις οποίες εντάσσεται το περιεχόμενο των αντικειμένων (π.χ. Αθλητικά, Πολιτικά, Επιστημονικά κλπ.).
 - Η οντότητα *Υποκατηγορία*, που αναπαριστά τις υποκατηγορίες περιεχομένου στις οποίες εντάσσεται το περιεχόμενο των αντικειμένων. Οι υποκατηγορίες αποτελούν το δεύτερο επίπεδο κατηγοριοποίησης των αντικειμένων (έτσι, π.χ., υποκατηγορίες που μπορεί να εντάσσονται στην κατηγορία Αθλητικά είναι οι υποκατηγορίες Μπάσκετ, Βόλεϊ κλπ.). Κάθε Υποκατηγορία συσχετίζεται με μια ακριβώς Κατηγορία, την Κατηγορία στην οποία εντάσσεται, μέσω της συσχέτισης «Ανήκει».
 - Η οντότητα *Λέξη-Κλειδί*, που αναπαριστά τις λέξεις-κλειδιά που χρησιμοποιούνται για το χαρακτηρισμό του περιεχομένου των αντικειμένων. Κάθε Λέξη-Κλειδί συσχετίζεται, μέσω της συσχέτισης «Σχετίζεται Με», με τις Κατηγορίες και τις Υποκατηγορίες που το περιεχόμενο των αντικειμένων που εντάσσονται σ' αυτές σχετίζεται με τη λέξη-κλειδί.
 - Η οντότητα *Θησαυρός*, που αναπαριστά τους θησαυρούς που χρησιμοποιούνται από την πηγή πληροφορίας. Κάθε Θησαυρός συσχετίζεται με ένα αριθμό Λέξεων-

³³ Οι οποίες μπορεί να είναι άμεσοι ή έμμεσοι απόγονοι της κλάσης Αντικείμενο στην «Είναι» ιεραρχία κλάσεων.

Κλειδιών που περιέχονται στο Θησαυρό αυτό μέσω της συσχέτισης «Ανήκει». Επιπλέον, κάθε Θησαυρός συσχετίζεται μέσω της συσχέτισης «Σχετίζεται Με» με μια τουλάχιστον γλώσσα. Οι θησαυροί που σχετίζονται με μια μόνο γλώσσα παρέχουν τους βαθμούς συσχέτισης μεταξύ των όρων της γλώσσας αυτής, ενώ αυτοί που συσχετίζονται με περισσότερες από μια γλώσσες παρέχουν τους βαθμούς συσχέτισης μεταξύ των όρων περισσότερων από μίας γλωσσών.



Διάγραμμα 6-8: E-R διάγραμμα της Βάσης Δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αμέσως μετά την αρχικοποίησή τους

- **Οντότητες που αναπαριστούν τη σχετική με τα αντικείμενα μεταπληροφορία.** Οι οντότητες που ανήκουν σε αυτή την ομάδα είναι οι εξής:
 - Η οντότητα *Κλάση*, που αναπαριστά τις κλάσεις των αντικειμένων που αποθηκεύονται στο σύστημα. Κάθε στιγμιότυπο της οντότητας Αντικείμενο συσχετίζεται, μέσω της συσχέτισης «Ανήκει», με μια Κλάση.
 - Η οντότητα *Γλώσσα*, που αναπαριστά τις γλώσσες στις οποίες είναι γραμμένα τα αντικείμενα που αποθηκεύονται στο σύστημα. Η οντότητα Γλώσσα συσχετίζεται,

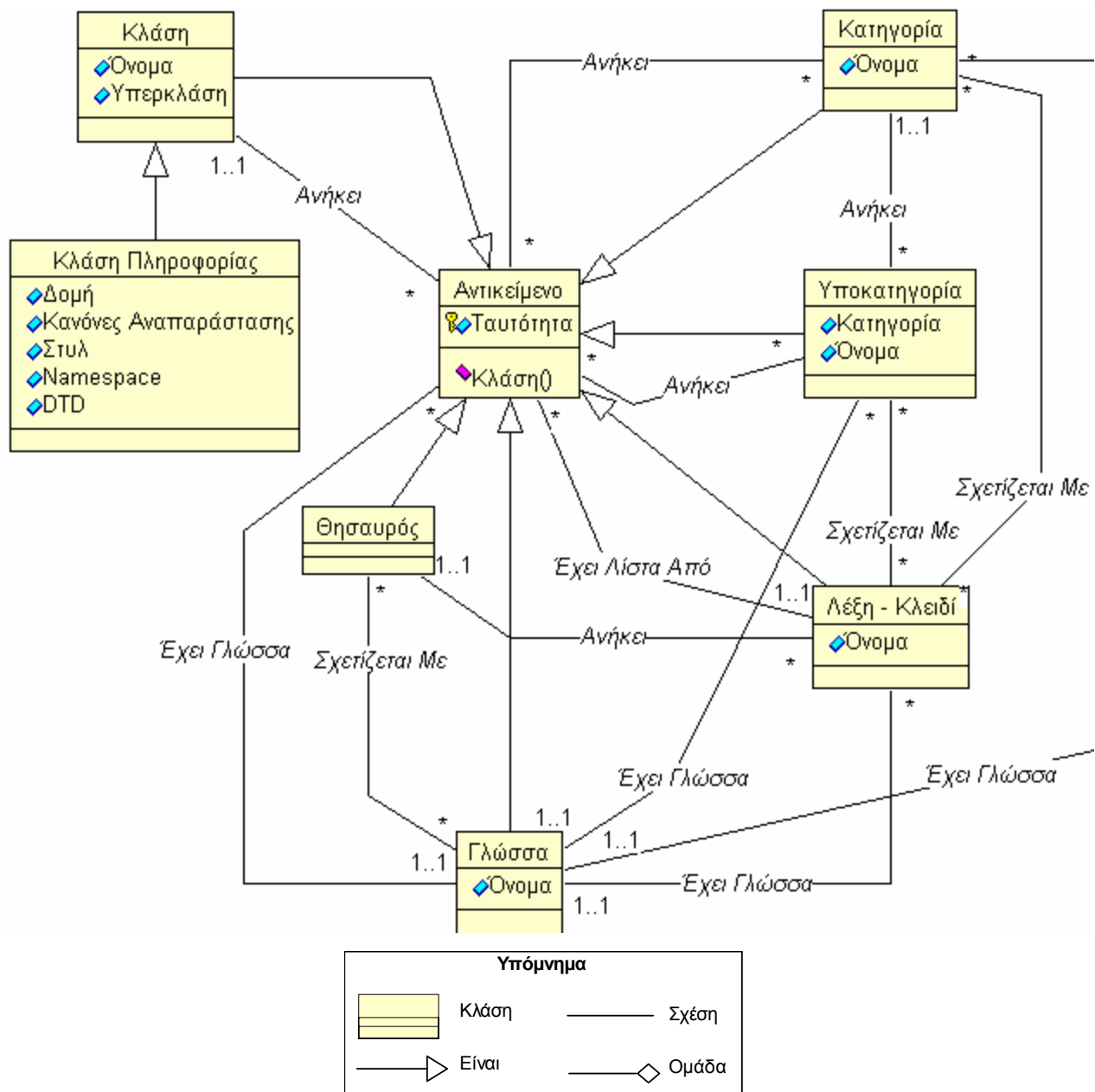
μέσω της συσχέτισης «Έχει Γλώσσα» με τα αντικείμενα που είναι το περιεχόμενό τους εκφράζεται σε αυτήν, με τις λέξεις-κλειδιά που ανήκουν σε αυτήν και με τις κατηγορίες και τις υποκατηγορίες που το όνομά τους είναι εκφρασμένο σε αυτή τη γλώσσα.

Καθώς η υποστήριξη δυνατοτήτων ανάκτησης πληροφορίας σε σχεσιακές βάσεις δεδομένων αποτέλεσε το θέμα της διατριβής [Μουμ98] και της διπλωματικής εργασίας [Ανεσ97] και η παροχή πολύγλωσσης υποστήριξης αποτελεί θέμα της διατριβής [Πουτ2000] και όχι της παρούσας διατριβής, στη συνέχεια θα δοθεί έμφαση μόνο στην αναπαράσταση κλάσεων αντικειμένων στη βάση δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Το Διάγραμμα 6-9 απεικονίζει τη βάση δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αμέσως μετά την αρχικοποίησή της, με έμφαση στην αναπαράσταση των αντικειμένων.

Όπως φαίνεται στο Διάγραμμα 6-9, κάθε κλάση που αναπαριστά αντικείμενα πληροφορίας και είναι στιγμιότυπο της κλάσης Κλάση Πληροφορίας διαθέτει:

- Το χαρακτηριστικό Όνομα, που αναπαριστά το όνομα της κλάσης.
- Το χαρακτηριστικό DTD, που αναπαριστά το DTD όπου περιγράφεται η κλάση.
- Το χαρακτηριστικό Κανόνες Αναπαράστασης, που αναπαριστά τους κανόνες αναπαράστασης των στιγμιότυπων της κλάσης στη βάση δεδομένων της πηγής πληροφορίας.
- Το χαρακτηριστικό Στυλ, που αναπαριστά το προκαθορισμένο στυλ μορφοποίησης που χρησιμοποιείται για τα στιγμιότυπα της κλάσης.
- Το χαρακτηριστικό Namespace, που αναπαριστά το όνομα της namespace όπου ορίζονται η δομή και τα χαρακτηριστικά των στιγμιότυπων της κλάσης.
- Το χαρακτηριστικό Υπερκλάση, που αναπαριστά την κλάση που είναι άμεσος πρόγονος της τρέχουσας κλάσης στην «Είναι» ιεραρχία κλάσεων.
- Το χαρακτηριστικό Δομή, που είναι ο URL του RDF Σχήματος που αναπαριστά το μοντέλο της συγκεκριμένης κλάσης, όπως αυτό γίνεται διαθέσιμο από τη γλώσσα της πηγής πληροφορίας. Το σύνολο των RDF σχημάτων των κλάσεων, ενσωματωμένο στο RDF σχήμα που περιγράφει το αρχικό μοντέλο πληροφορίας της βάσης δεδομένων της πηγής πληροφορίας, αποτελεί τη δομή της πηγής πληροφορίας, που αποθηκεύεται σε ένα αρχείο ο URL του οποίου αποθηκεύεται στη βάση δεδομένων του διαμεσολαβητή. Το RDF σχήμα που περιγράφει τη δομή της βάσης δεδομένων μιας πηγής πληροφορίας που περιέχει παραδοσιακά δεδομένα πριν την υποστήριξη κάποιας κλάσης φαίνεται στο RDF Σχήμα 6-2 και αντιστοιχεί στο URL "http://irida.ced.tuc.gr/RDF/ΠηγήΠληροφορίαςΔεδομένων.rdfs".



Διάγραμμα 6-9: Διάγραμμα Κλάσεων της Βάσης Δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αμέσως μετά την αρχικοποίησή τους, με έμφαση στην αναπαράσταση των αντικειμένων

```

<?xml version='1.0' ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema"
  xmlns:ΠηγήΠληροφορίαςΔεδομένων="">
  <rdfs:Class ID="αντικείμενο">
    <rdfs:label>αντικείμενο</rdfs:label>
    <rdfs:SubclassOf
      rdf:resource="http://www.w3.org/TR/PR-rdf-schema#Resource" />
  </rdfs:Class>
  <rdf:Property ID="ταυτότητα">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#Number">

```

```

    <rdf:domain rdf:resource="#αντικείμενο" />
</rdf:Property>
<rdfs:Class ID="κλάση">
    <rdfs:label>κλάση</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
    <rdf:domain rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdf:Property ID="υπερκλάση">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση" />
</rdf:Property>
<rdf:Property ID="όνομα">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση" />
    <rdf:domain rdf:resource="#κατηγορία" />
    <rdf:domain rdf:resource="#υποκατηγορία" />
    <rdf:domain rdf:resource="#λέξη-κλειδί" />
</rdf:Property>
<rdfs:Class ID="κλάση_πληροφορίας">
    <rdfs:label>κλάση</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#κλάση" />
</rdfs:Class>
<rdf:Property ID="DTD">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση_πληροφορίας" />
</rdf:Property>
<rdf:Property ID="δομή">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση_πληροφορίας" />
</rdf:Property>
<rdf:Property ID="κανόνες_αναπαράστασης">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση_πληροφορίας" />
</rdf:Property>
<rdf:Property ID="στυλ">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση_πληροφορίας" />
</rdf:Property>
<rdf:Property ID="namespace">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#κλάση_πληροφορίας" />
</rdf:Property>
<rdfs:Class ID="γλώσσα">
    <rdfs:label>γλώσσα</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdfs:Class ID="κατηγορία">
    <rdfs:label>κατηγορία</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdfs:Class ID="υποκατηγορία">

```

```

    <rdfs:label>υποκατηγορία</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdf:Property ID="κατηγορία">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#υποκατηγορία" />
</rdf:Property>
<rdfs:Class ID="λέξη-κλειδί">
    <rdfs:label>λέξη-κλειδί</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
<rdfs:Class ID="θησαυρός">
    <rdfs:label>θησαυρός</rdfs:label>
    <rdfs:SubclassOf rdf:resource="#αντικείμενο" />
</rdfs:Class>
</rdf:RDF>

```

RDF Σχήμα 6-2: RDF Σχήμα που περιγράφει τη Δομή της Βάσης Δεδομένων μιας Πηγής Πληροφορίας που περιέχει Παραδοσιακά Δεδομένα πριν την υποστήριξη κάποιας κλάσης

Τα αντικείμενα αναπαρίστανται ως στιγμιότυπα της κλάσης «Αντικείμενο» και των υποκλάσεων της. Η πληροφορία που σχετίζεται με κάποιο αντικείμενο αποθηκεύεται στις τιμές των χαρακτηριστικών του και στις συσχετίσεις του με άλλα αντικείμενα. Έτσι, στη συσχέτιση «Έχει Γλώσσα» μεταξύ αντικειμένων και γλωσσών καθορίζεται η γλώσσα στην οποία είναι εκφρασμένο το περιεχόμενο κάθε αντικειμένου. Αντίστοιχα, το περιεχόμενο κάθε αντικειμένου καθορίζεται από τις συσχετίσεις του αντικειμένου με κατηγορίες, υποκατηγορίες και λέξεις-κλειδιά.

Όπως έχει ήδη αναφερθεί, η επέκταση του σχήματος των βάσεων δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα μπορεί να περιγραφεί από XML DTDs ή XML σχήματα³⁴.

Έστω τώρα ότι πρόκειται να υποστηριχθεί από κάποια πηγή πληροφορίας που περιέχει παραδοσιακά δεδομένα ο DTD Πρόσωπο.dtd που φαίνεται στο DTD 6-3.

```

<!ELEMENT Πρόσωπο (Όνομα, Επώνυμο, Τηλέφωνο, Διεύθυνση, E-Mail, Οργανισμός)>
<!ATTLIST Πρόσωπο Ηλικία CDATA #IMPLIED >
<!ELEMENT Όνομα (#PCDATA )>
<!ELEMENT Επώνυμο (#PCDATA )>
<!ELEMENT Τηλέφωνο (#PCDATA )>
<!ELEMENT Διεύθυνση (#PCDATA )>
<!ELEMENT E-Mail (#PCDATA )>
<!ELEMENT Οργανισμός (#PCDATA )>

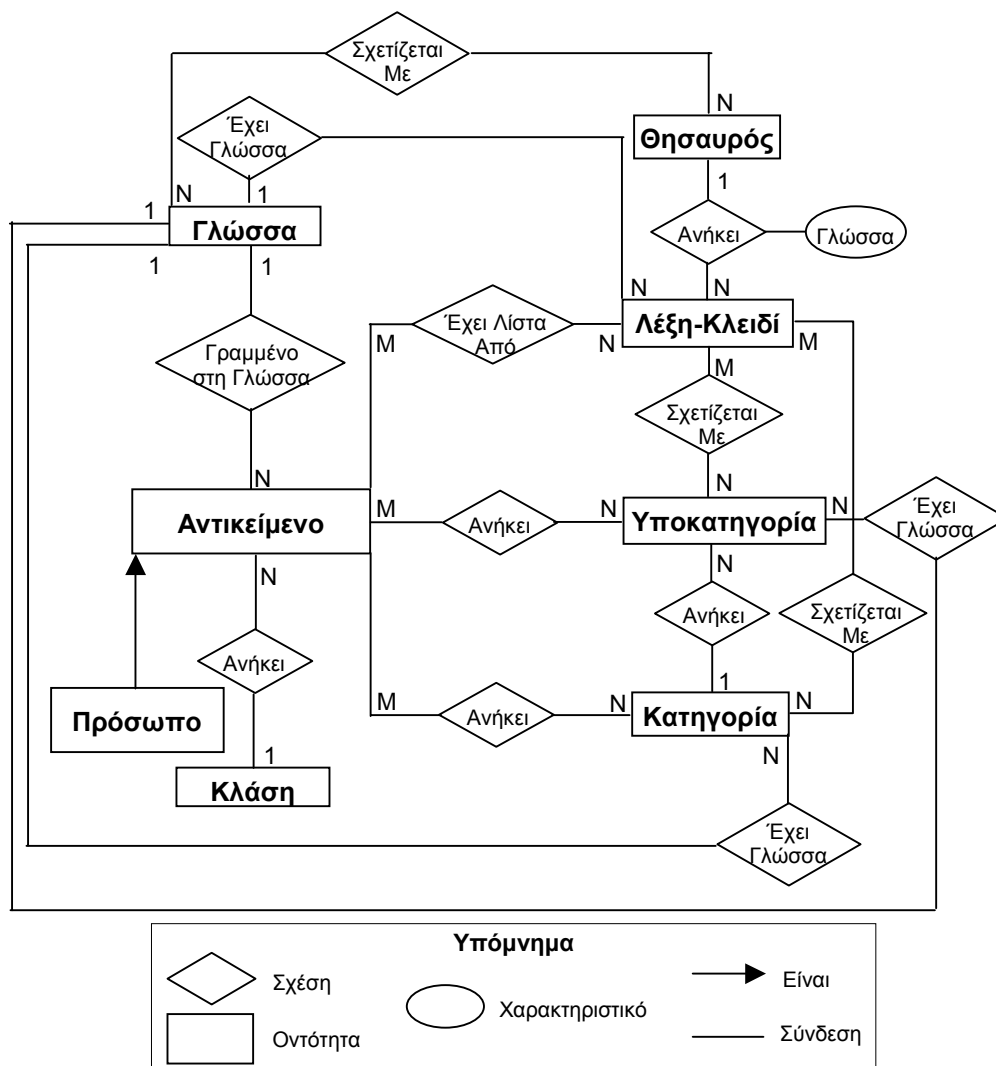
```

DTD 6-3: Ο DTD Πρόσωπο.dtd

Η υποστήριξη του DTD Πρόσωπο.dtd έχει ως συνέπεια:

³⁴ Για την περιγραφή του μηχανισμού που υλοποιεί την επέκταση του σχήματος της βάσης δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα βλέπε τμήμα 7.2.

- Τη δημιουργία ενός στιγμιότυπου της οντότητας «Κλάση», το οποίο έχει όνομα «Πρόσωπο» και αναπαριστά την κλάση με όνομα «Πρόσωπο».
- Τη δημιουργία της οντότητας «Πρόσωπο», τα στιγμιότυπα της οποίας θα αναπαριστούν τα πρόσωπα για τα οποία θα αποθηκεύεται πληροφορία στη βάση δεδομένων της πηγής πληροφορίας. Τα χαρακτηριστικά των στιγμιότυπων της οντότητας αυτής είναι το Όνομα, το Επώνυμο, το Τηλέφωνο, η Διεύθυνση, το E-Mail, ο Οργανισμός και η Ηλικία³⁵.



Διάγραμμα 6-10: E-R διάγραμμα της Βάσης Δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αμέσως μετά την υποστήριξη του DTD Πρόσωπο.dtd

Το E-R διάγραμμα της βάσης δεδομένων της πηγής πληροφορίας αμέσως μετά την υποστήριξη του DTD Πρόσωπο.dtd φαίνεται στο Διάγραμμα 6-10.

³⁵ Εναλλακτικά, το Όνομα, το Επώνυμο, το Τηλέφωνο, η Διεύθυνση, το E-Mail και ο Οργανισμός ή κάποια από αυτά θα μπορούσαν να αποτελούν ανεξάρτητες οντότητες. Για τον προσδιορισμό της αναπαράστασης των στιγμιότυπων των υποστοιχείων βλέπε τμήμα 7.2.

6.4.3. Λειτουργικότητα των Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

Η λειτουργικότητα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα αποτελεί το θέμα αυτής της ενότητας. Η λειτουργικότητα αυτή είναι αντίστοιχη με αυτή που παρέχεται από όλες τις πηγές πληροφορίας, απεικονίζεται στο Διάγραμμα 6-2 και περιγράφηκε σε γενικές γραμμές στην ενότητα 6.2. Οι μηχανισμοί που υλοποιούν τη λειτουργικότητα αυτή περιγράφονται στο 7^ο κεφάλαιο.

Σε ότι αφορά τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, η λειτουργικότητα που περιγράφηκε στην ενότητα 6.2 εξειδικεύεται ως εξής:

- Αρχικοποίηση της βάσης δεδομένων μιας πηγής πληροφορίας. Μ' αυτό τον τρόπο ο διαχειριστής μιας πηγής πληροφορίας μπορεί να δημιουργήσει μια κενή βάση δεδομένων, που περιέχει μόνο τις οντότητες που είναι απαραίτητες για τη λειτουργία της πηγής πληροφορίας και την παραπέρα εξέλιξη του σχήματός της. Έτσι, στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα περιέχονται μόνο οι σχέσεις που είναι απαραίτητες για την αναπαράσταση των κλάσεων Αντικείμενο, Κατηγορία, Υποκατηγορία, Λέξη – Κλειδί, Θησαυρός και Γλώσσα.

Ο μηχανισμός που υποστηρίζει τη λειτουργία της αρχικοποίησης της βάσης δεδομένων μιας πηγής πληροφορίας που περιέχει παραδοσιακά δεδομένα περιγράφεται στο τμήμα 7.1.

- Η υποστήριξη νέων κλάσεων στη βάση δεδομένων μιας πηγής πληροφορίας σημαίνει, για τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, τη δυνατότητα εξέλιξης του σχήματος της βάσης δεδομένων της πηγής πληροφορίας, έτσι ώστε να υποστηριχθούν από το μοντέλο πληροφορίας της πηγής πληροφορίας νέες οντότητες και/ή συσχετίσεις μεταξύ οντοτήτων.

Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της υποστήριξης νέων κλάσεων στη βάση δεδομένων μιας πηγής πληροφορίας που περιέχει παραδοσιακά δεδομένα περιγράφονται στο τμήμα 7.2

- Η εισαγωγή νέων αντικειμένων στη βάση δεδομένων μιας πηγής πληροφορίας σημαίνει, για τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, ότι παρέχεται η δυνατότητα σε όσους έχουν το δικαίωμα εισαγωγής πληροφορίας (διαχειριστές των πηγών πληροφορίας και τοπικούς ή απομακρυσμένους προνομιούχους χρήστες) να πραγματοποιούν εισαγωγές.

Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της εισαγωγής νέων αντικειμένων στη βάση δεδομένων μιας πηγής πληροφορίας που περιέχει παραδοσιακά δεδομένα περιγράφονται στο τμήμα 7.3.

- Η δεικτοδότηση στην πηγή πληροφορίας απομακρυσμένων αντικειμένων που το περιεχόμενό τους σχετίζεται με αυτό της πηγής πληροφορίας, τα οποία όμως δεν εισάγονται (τουλάχιστον όχι εξ' ολοκλήρου) στη βάση δεδομένων της πηγής πληροφορίας. Στην περίπτωση αυτή δημιουργείται μια *Αποψη (View)* των αντικειμένων στη βάση δεδομένων της πηγής πληροφορίας, η οποία απαρτίζεται από τις τιμές

επιλεγμένων χαρακτηριστικών των αντικειμένων, ανάλογα με την κλάση στην οποία εντάσσεται κάθε αντικείμενο.

Οι μηχανισμοί που υποστηρίζουν τη λειτουργία δεικτοδότηση σε μια πηγή πληροφορίας απομακρυσμένων αντικειμένων που το περιεχόμενό τους σχετίζεται με αυτό της πηγής πληροφορίας, τα οποία όμως δεν εισάγονται (τουλάχιστον όχι εξ' ολοκλήρου) στη βάση δεδομένων της πηγής πληροφορίας, περιγράφονται στο τμήμα 7.5.

- Η αναζήτηση αντικειμένων που ικανοποιούν συγκεκριμένα κριτήρια σχετικά με το περιεχόμενό τους σημαίνει, για τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, ότι παρέχεται από αυτές υποστήριξη ερωτήσεων. Οι ερωτήσεις που υποστηρίζονται επιτρέπεται να περιέχουν τόσο όρους ομοιότητας όσο και όρους δίτιμης λογικής. Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της αναζήτησης αντικειμένων περιγράφονται στο τμήμα 7.4.
- Η ανάκτηση αντικειμένων και/ή τμημάτων αντικειμένων που αποθηκεύονται στη βάση δεδομένων μιας πηγής πληροφορίας σημαίνει, για τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, ότι επιτρέπεται η πρόσβαση των χρηστών στην πληροφορία που τους ενδιαφέρει. Οι μηχανισμοί που υποστηρίζουν τη λειτουργία της ανάκτησης αντικειμένων και/ή τμημάτων αντικειμένων περιγράφονται στο τμήμα 7.4.

6.5. Ανακεφαλαίωση

Στο κεφάλαιο αυτό περιγράφηκαν οι πηγές πληροφορίας. Πιο συγκεκριμένα, αρχικά περιγράφηκε η *Αρχιτεκτονική των Πηγών Πληροφορίας* (τμήμα 6.1), ενώ στη συνέχεια (τμήμα 6.2) περιγράφηκε η *Λειτουργικότητα των Πηγών Πληροφορίας*. Τέλος, περιγράφηκαν οι *Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα* (τμήμα 6.3) και οι *Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα* (τμήμα 6.4). Ειδικότερα, έμφαση δόθηκε στις βάσεις δεδομένων, τις γλώσσες και τα περιτυλίγματα των πηγών πληροφορίας και στην αναπαράσταση δομημένων εγγράφων σε σχεσιακές βάσεις δεδομένων, καθώς τα σημεία αυτά αποτελούν συνεισφορά της παρούσας διατριβής.

Στο 7^ο κεφάλαιο, που ακολουθεί, περιγράφονται η διαχείριση των πηγών πληροφορίας και ο σχεδιασμός των εργαλείων και των μηχανισμών που απαιτούνται γι αυτήν.

7^ο Κεφάλαιο

Διαχείριση των Πηγών Πληροφορίας

Στο κεφάλαιο αυτό θα γίνει αναφορά στη διαχείριση των πηγών πληροφορίας. Κατ' επέκταση, αναφορά θα γίνει και στα εργαλεία που απαιτούνται, πάνω από τη βάση δεδομένων που υποστηρίζει κάθε πηγή πληροφορίας, για τη δημιουργία και τη διαχείριση των πηγών πληροφορίας. Καθώς η παρούσα διατριβή επικεντρώνει στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, το κεφάλαιο αυτό επικεντρώνει στη διαχείριση των πηγών πληροφορίας που ανήκουν στις κλάσεις αυτές.

Στο τμήμα 7.1 περιγράφονται η *Δημιουργία Πηγών Πληροφορίας* και η λειτουργία των προγραμμάτων εγκατάστασης (setup) των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα και των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Στο τμήμα 7.2 περιγράφεται η δυνατότητα που παρέχεται για την *Υποστήριξη Νέων Κλάσεων από τις Πηγές Πληροφορίας*, με τη χρήση του εργαλείου SupportDTD. Το εργαλείο SupportDTD επιτρέπει την εξέλιξη του σχήματος των βάσεων δεδομένων των πηγών πληροφορίας έτσι ώστε να υποστηριχθούν από αυτές κλάσεις που περιγράφονται από XML DTDs. Τόσο οι κανόνες αναπαράστασης, διαχείρισης και δεικτοδότησης των στιγμιotypών των κλάσεων που υποστηρίζονται από κάποια πηγή πληροφορίας, όσο και η δομή των κλάσεων αυτών, έτσι όπως εξάγεται από τη γλώσσα της πηγής πληροφορίας, παράγονται από το εργαλείο SupportDTD. Η δομή των κλάσεων περιγράφεται από RDF σχήματα, ενώ οι κανόνες αναπαράστασης, διαχείρισης και δεικτοδότησης των

στιγμιότυπων των κλάσεων περιγράφονται με βάση τη γλώσσα ΚΑΔΔΑ (Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Αντικειμένων), που είναι κατ' ουσίαν ένα RDF σχήμα.

Στο τμήμα 7.3 περιγράφεται η *Εισαγωγή Αντικειμένων* στις βάσεις δεδομένων που υποστηρίζουν τις πηγές πληροφορίας. Πιο συγκεκριμένα, γίνεται αναφορά στις κατηγορίες εφαρμογών εισαγωγής αντικειμένων που μπορούν να χρησιμοποιηθούν και περιγράφεται η λειτουργία των μηχανισμών που υποστηρίζουν την εισαγωγή αντικειμένων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Θέματα που σχετίζονται με την *Αναζήτηση & Ανάκτηση Αντικειμένων* στις πηγές πληροφορίας συζητούνται στο τμήμα 7.4.

Η *Δεικτοδότηση Απομακρυσμένων Αντικειμένων* συζητείται στο τμήμα 7.5. Η δεικτοδότηση απομακρυσμένων αντικειμένων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα περιλαμβάνει:

- Την ενημέρωση των δομών δεικτοδότησης των πηγών πληροφορίας για το περιεχόμενο των δεικτοδοτούμενων αντικειμένων.
- Την αντιγραφή στις βάσεις δεδομένων των πηγών πληροφορίας των μεταδεδομένων που χρησιμοποιούνται για την περιγραφή των δεικτοδοτούμενων αντικειμένων.
- Την αντιγραφή στις βάσεις δεδομένων των πηγών πληροφορίας των δομικών στοιχείων των δεικτοδοτούμενων αντικειμένων που καθορίζεται από τους κανόνες δεικτοδότησης ότι θα αποθηκεύονται στις βάσεις δεδομένων.

Αντίστοιχα, η δεικτοδότηση απομακρυσμένων αντικειμένων στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα περιλαμβάνει την ενημέρωση των δομών δεικτοδότησης των πηγών πληροφορίας για το περιεχόμενο των δεικτοδοτούμενων αντικειμένων και τη δημιουργία απόψεων των δεικτοδοτούμενων αντικειμένων με την αντιγραφή των χαρακτηριστικών εκείνων για τα οποία καθορίζεται από τους κανόνες δεικτοδότησης ότι θα αντιγράφονται.

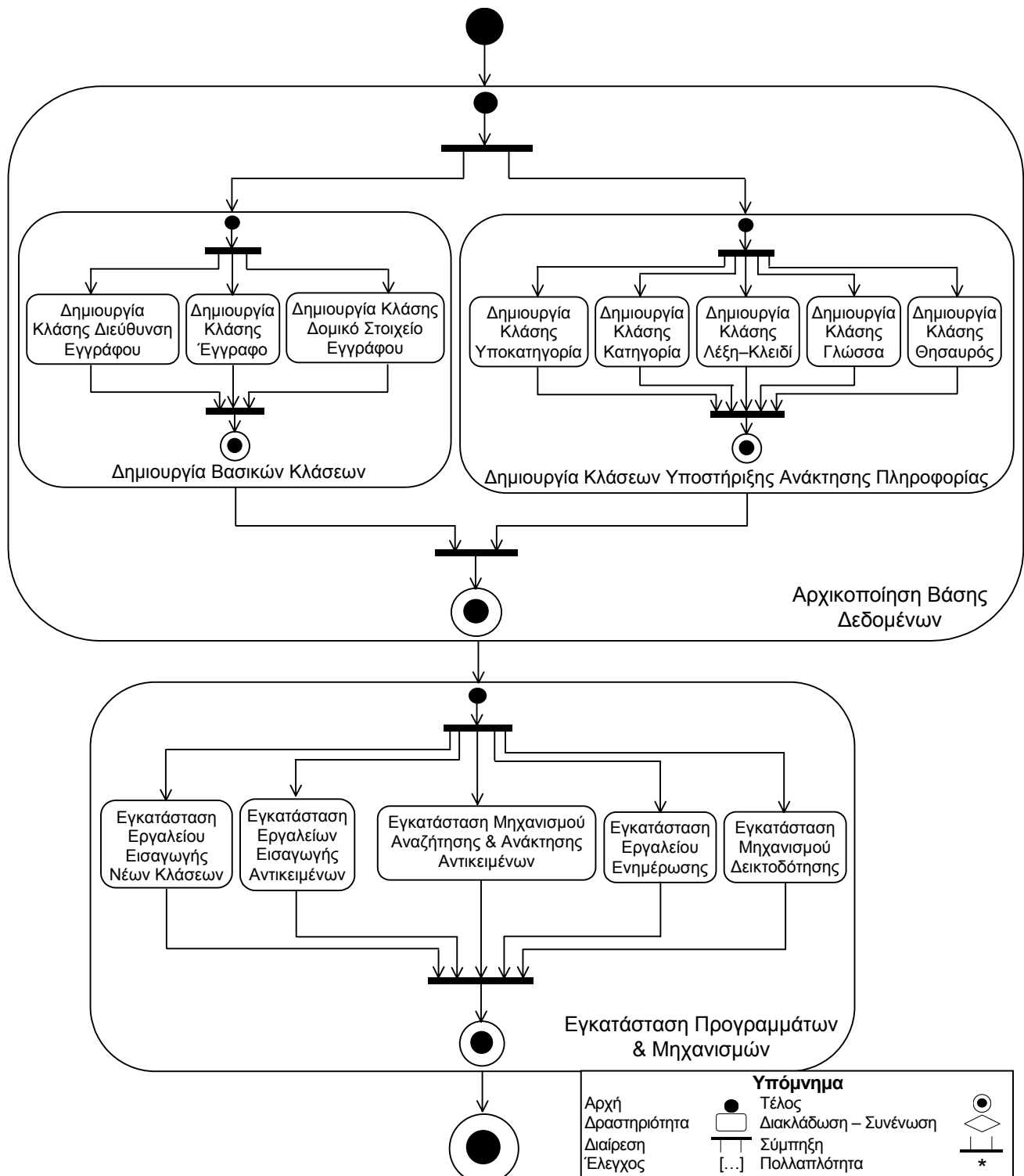
Το θέμα του τμήματος 7.6 είναι η *Ενημέρωση των Πηγών Πληροφορίας* για πληροφορία που γίνεται διαθέσιμη και σχετίζεται με το περιεχόμενο των συγκεκριμένων πηγών πληροφορίας. Η ενημέρωση των πηγών πληροφορίας γίνεται από πράκτορες που διασχίζουν το Διαδίκτυο για τον εντοπισμό αντικειμένων που το περιεχόμενό τους σχετίζεται με αυτό των πηγών πληροφορίας για την ενημέρωση των οποίων είναι υπεύθυνοι.

Τέλος, το κεφάλαιο κλείνει με μια σύντομη *Ανακεφαλαίωση* στο τμήμα 7.7.

7.1. Δημιουργία Πηγών Πληροφορίας

Στο τμήμα αυτό περιγράφεται η διαδικασία δημιουργίας μιας πηγής πληροφορίας. Η διαδικασία αυτή, που είναι σε ό,τι αφορά τη λογική της κοινή τόσο για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα όσο και για τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, γίνεται με τη χρήση του προγράμματος εγκατάστασης της πηγής πληροφορίας (setup). Το πρόγραμμα εγκατάστασης της πηγής πληροφορίας

είναι υπεύθυνο για την αρχικοποίηση της βάσης δεδομένων που υποστηρίζει την πηγή πληροφορίας και για την εγκατάσταση των απαραίτητων για τη διαχείριση της πηγής πληροφορίας εργαλείων και μηχανισμών.



Διάγραμμα 7-1: Διάγραμμα Δραστηριοτήτων για το Πρόγραμμα Εγκατάστασης Πηγών Πληροφορίας που περιέχουν Δομημένα Έγγραφα

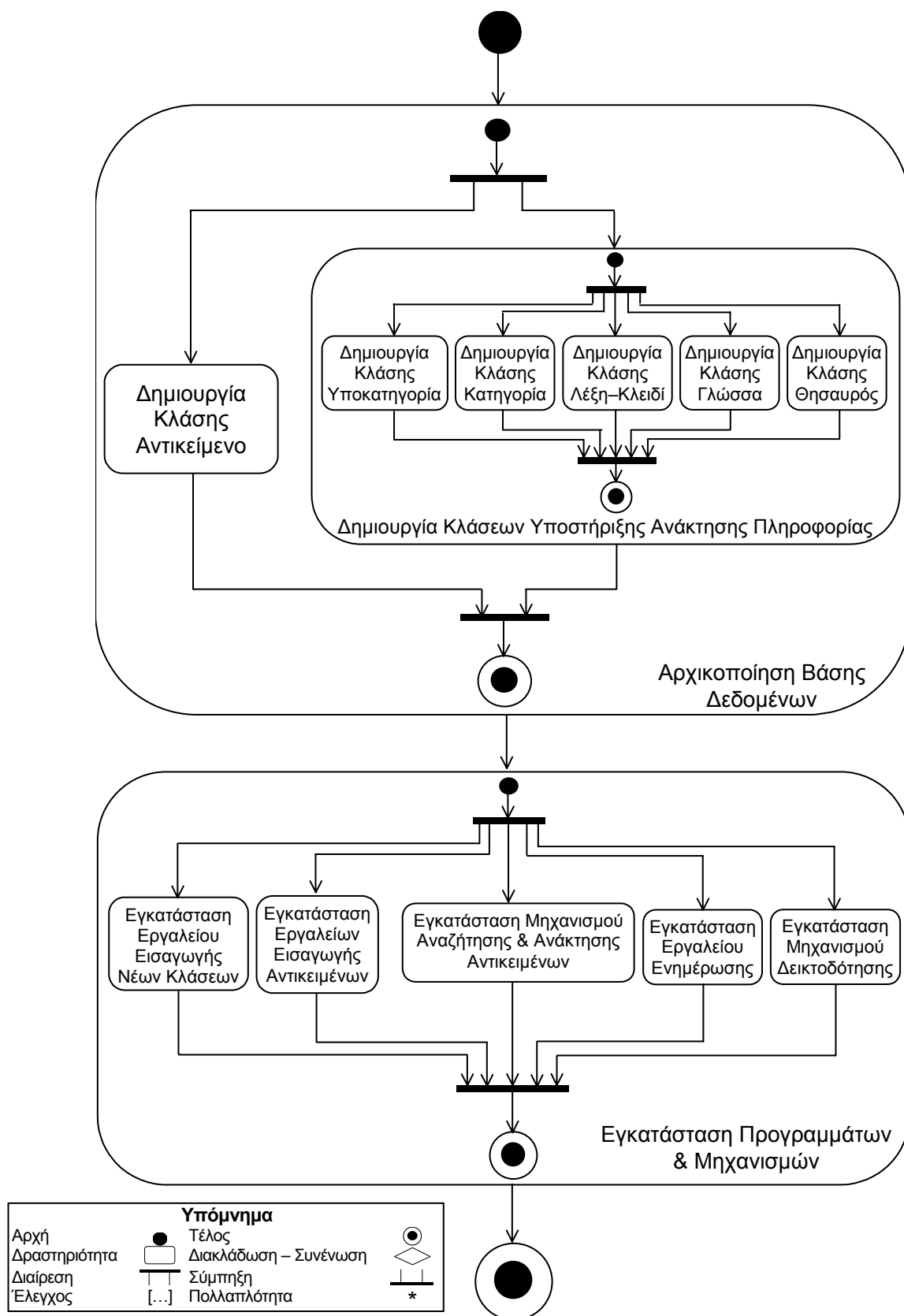
Η λειτουργία του προγράμματος εγκατάστασης των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα απεικονίζεται στο Διάγραμμα 7-1, που είναι ένα UML διάγραμμα

δραστηριοτήτων. Όπως φαίνεται στο Διάγραμμα 7-1, το πρόγραμμα εγκατάστασης των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα λειτουργεί με βάση τον εξής αλγόριθμο:

1. Αρχικοποιείται η βάση δεδομένων της υπό κατασκευή πηγής πληροφορίας:
 - 1.1. Δημιουργούνται οι οντότητες που είναι απαραίτητες για την παραπέρα εξέλιξη του σχήματός της βάσης δεδομένων και συγκεκριμένα:
 - 1.1.1. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Έγγραφο.
 - 1.1.2. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Τμήμα Εγγράφου.
 - 1.1.3. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Διεύθυνση Αντικειμένου.
 - 1.2. Δημιουργούνται οι οντότητες που χρησιμοποιούνται για την περιγραφή της μεταπληροφορίας που σχετίζεται με το περιεχόμενο των εγγράφων και συγκεκριμένα:
 - 1.2.1. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Κατηγορία.
 - 1.2.2. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Υποκατηγορία.
 - 1.2.3. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Λέξη – Κλειδί.
 - 1.2.4. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Θησαυρός.
 - 1.2.5. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Γλώσσα.
2. Εγκαθίστανται τα εργαλεία και οι μηχανισμοί που απαιτούνται για τη διαχείριση της πηγής πληροφορίας και συγκεκριμένα:
 - 2.1. Το εργαλείο που είναι υπεύθυνο για την υποστήριξη νέων κλάσεων δομημένων εγγράφων από την υπό κατασκευή πηγή πληροφορίας.
 - 2.2. Τα εργαλεία που είναι υπεύθυνα για την εισαγωγή δομημένων εγγράφων στην πηγή πληροφορίας.
 - 2.3. Ο μηχανισμός που είναι υπεύθυνος για την αναζήτηση δομημένων εγγράφων και δομικών στοιχείων εγγράφων μεταξύ των περιεχομένων της πηγής πληροφορίας και την ανάκτηση δομημένων εγγράφων και δομικών στοιχείων εγγράφων από την πηγή πληροφορίας.
 - 2.4. Ο μηχανισμός που είναι υπεύθυνος για τη δεικτοδότηση, μέσα στην πηγή πληροφορίας, απομακρυσμένων ως προς αυτήν εγγράφων.
 - 2.5. Το εργαλείο που είναι υπεύθυνο για την ενημέρωση της πηγής πληροφορίας για έγγραφα που γίνονται διαθέσιμα και σχετίζονται με το περιεχόμενό της.

Αντίστοιχα με τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, η λειτουργία του προγράμματος εγκατάστασης των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα απεικονίζεται στο Διάγραμμα 7-2, που επίσης είναι UML διάγραμμα δραστηριοτήτων.

Όπως φαίνεται στο Διάγραμμα 7-2, το πρόγραμμα εγκατάστασης πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα λειτουργεί με βάση τον εξής αλγόριθμο:



Διάγραμμα 7-2: Διάγραμμα Δραστηριοτήτων για το Πρόγραμμα Εγκατάστασης Πηγών Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

1. Αρχικοποιείται η βάση δεδομένων της υπό κατασκευή πηγής πληροφορίας:

- 1.1. Δημιουργούνται οι οντότητες που είναι απαραίτητες για την παραπέρα εξέλιξη του σχήματός της βάσης δεδομένων και συγκεκριμένα οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Αντικείμενο.
- 1.2. Δημιουργούνται οι οντότητες που χρησιμοποιούνται για την περιγραφή της μεταπληροφορίας που σχετίζεται με το περιεχόμενο των αντικειμένων και συγκεκριμένα:
 - 1.2.1. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Κατηγορία.
 - 1.2.2. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Υποκατηγορία.
 - 1.2.3. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Λέξη – Κλειδί.
 - 1.2.4. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Θησαυρός.
 - 1.2.5. Οι οντότητες που είναι απαραίτητες για την αναπαράσταση της κλάσης Γλώσσα.
2. Εγκαθίστανται τα εργαλεία και οι μηχανισμοί που απαιτούνται για τη διαχείριση της πηγής πληροφορίας και συγκεκριμένα:
 - 2.1. Το εργαλείο που είναι υπεύθυνο για την υποστήριξη νέων κλάσεων αντικειμένων από τη συγκεκριμένη πηγή πληροφορίας.
 - 2.2. Τα εργαλεία που είναι υπεύθυνα για την εισαγωγή αντικειμένων στην πηγή πληροφορίας.
 - 2.3. Ο μηχανισμός που είναι υπεύθυνος για την αναζήτηση αντικειμένων μεταξύ των περιεχομένων της πηγής πληροφορίας και την ανάκτηση αντικειμένων από την πηγή πληροφορίας.
 - 2.4. Ο μηχανισμός που είναι υπεύθυνος για τη δεικτοδότηση, μέσα στην πηγή πληροφορίας, απομακρυσμένων ως προς αυτήν αντικειμένων.
 - 2.5. Το εργαλείο που είναι υπεύθυνο για την ενημέρωση της πηγής πληροφορίας για αντικείμενα που γίνονται διαθέσιμα και σχετίζονται με το περιεχόμενό της.

7.2. Υποστήριξη Νέων Κλάσεων από τις Πηγές Πληροφορίας

Το θέμα αυτού του τμήματος είναι η περιγραφή της δυνατότητας που παρέχεται για την υποστήριξη νέων κλάσεων αντικειμένων από τις πηγές πληροφορίας. Η περιγραφή αυτή επικεντρώνει στην υποστήριξη νέων κλάσεων από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Όπως έχει αναφερθεί στο 6^ο κεφάλαιο, τα αντικείμενα που εισάγονται τόσο στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα όσο και στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα δίνονται σε XML μορφή. Αυτό σημαίνει ότι η δομή των κλάσεων στις οποίες εντάσσονται τα αντικείμενα αυτά περιγράφεται είτε από XML DTDs είτε από σχήματα. Προς το παρόν, παρ' όλο που έχουν γίνει αρκετές προτάσεις για γλώσσες περιγραφής κλάσεων XML πληροφορίας που είναι ισχυρότερες από τους DTDs, καμιά απ' αυτές δεν έχει υιοθετηθεί από το W3C³⁶. Έτσι, σε πρώτη φάση, από τις πηγές πληροφορίας του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» υποστηρίχθηκαν κλάσεις XML πληροφορίας που περιγράφονται από DTDs.

³⁶ Ο προσδιορισμός της XML Schema Language, που είναι η γλώσσα περιγραφής XML πληροφορίας που αναπτύσσεται από το W3C, δεν είναι ακόμα σταθερός.

Καθώς η εκφραστικότητα των DTDs είναι, κατά γενική ομολογία, περιορισμένη, κάθε φορά που πρόκειται να υποστηριχθεί μια νέα κλάση XML πληροφορίας από κάποια πηγή πληροφορίας, είναι απαραίτητη η παρέμβαση του διαχειριστή της πηγής πληροφορίας για να εξασφαλιστεί η σωστή υποστήριξη κλάσεων XML πληροφορίας.

Με βάση τις παραπάνω απαιτήσεις, αναπτύχθηκε ένα αλληλεπιδραστικό εργαλείο, το εργαλείο SupportDTD, το οποίο επιτρέπει την επέκταση του σχήματος της βάσης δεδομένων που υποστηρίζει την πηγή πληροφορίας, ώστε να υποστηριχθούν η εισαγωγή και η διαχείριση κλάσεων XML πληροφορίας που περιγράφονται από DTDs τόσο σε πηγές πληροφορίας πηγές πληροφορίας που περιέχουν δομημένα έγγραφα όσο και σε πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Για κάθε κλάση που υποστηρίζεται, εκτός από την εξέλιξη της βάσης δεδομένων της πηγής πληροφορίας, το εργαλείο SupportDTD εξάγει τη δομή των στιγμιοτύπων της με τη μορφή RDF σχήματος και ένα σύνολο κανόνων αναπαράστασης, διαχείρισης και δεικτοδότησης σε RDF μορφή.

Η *Λειτουργία του Εργαλείου SupportDTD* περιγράφεται στην ενότητα 7.2.1. Οι κανόνες αναπαράστασης, διαχείρισης και δεικτοδότησης τόσο των στιγμιοτύπων των κλάσεων δομημένων εγγράφων όσο και των στιγμιοτύπων των κλάσεων παραδοσιακών δεδομένων, οι οποίοι προκύπτουν από το εργαλείο SupportDTD, εκφράζονται με τη συμβατή με την RDF γλώσσα *ΚΑΔΔΑ (Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Αντικειμένων)*, που περιγράφεται στην ενότητα 7.2.2. Η δομή των αντικειμένων, που επίσης προκύπτει από το εργαλείο SupportDTD, όπως αυτή εξάγεται από τη γλώσσα των πηγών πληροφορίας στις οποίες αυτά αποθηκεύονται, δίνεται με τη μορφή RDF σχημάτων και περιγράφεται στην υποενότητα 7.2.3..

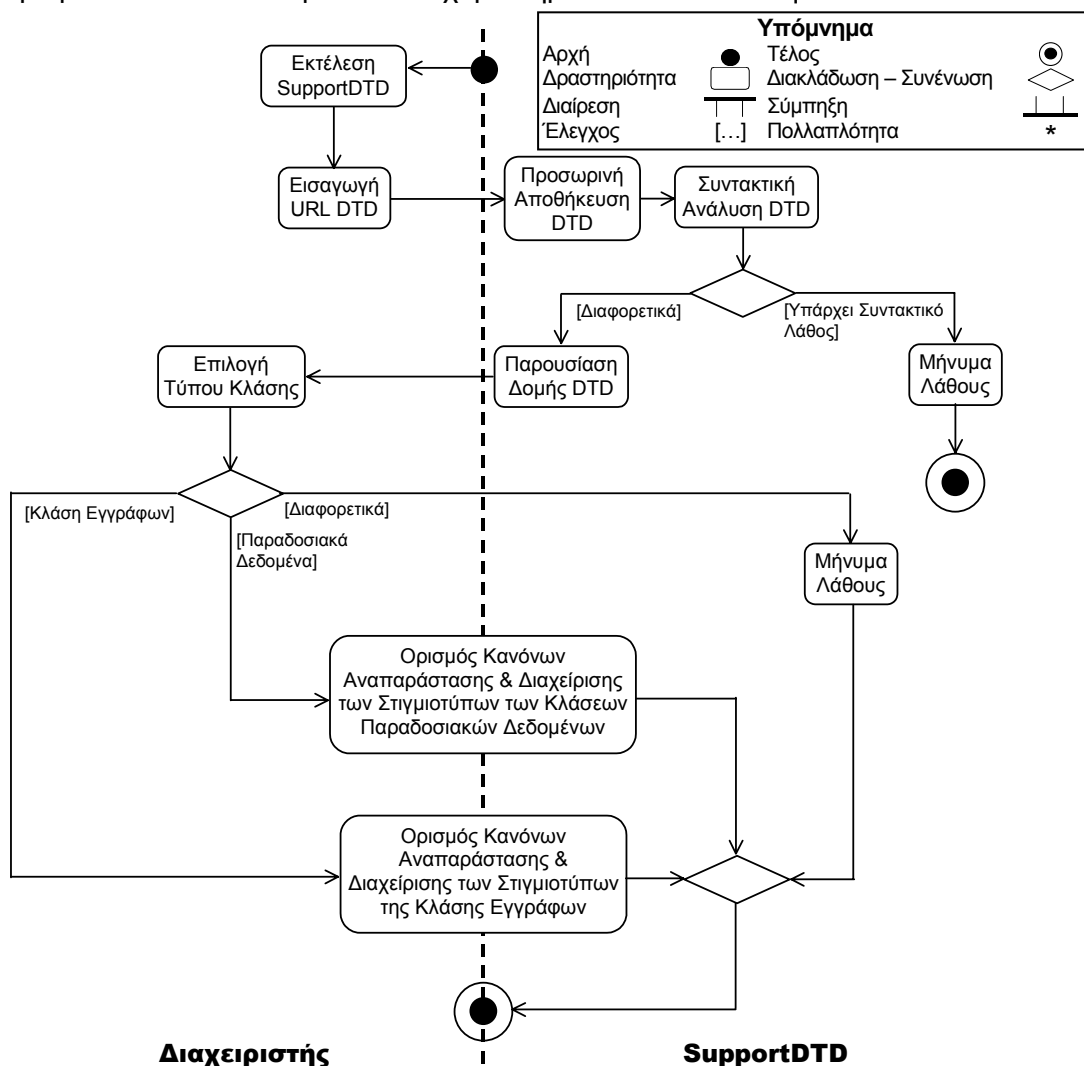
Στο σημείο αυτό πρέπει να σημειωθεί ότι, καθώς η XML χρησιμοποιείται για την περιγραφή τόσο εγγράφων όσο και παραδοσιακών δεδομένων, το εργαλείο SupportDTD έχει σχεδιαστεί έτσι ώστε να αντιμετωπίζει με διαφορετικό τρόπο τους DTDs που περιγράφουν κλάσεις δομημένων εγγράφων και με διαφορετικό τρόπο τους DTDs που περιγράφουν τη δομή κλάσεων παραδοσιακών δεδομένων. Έτσι, σε κάθε περίπτωση η αναπαράσταση στη σχεσιακή βάση δεδομένων της πηγής πληροφορίας είναι η πιο αποδοτική.

7.2.1. Λειτουργία του Εργαλείου SupportDTD

Το θέμα αυτής της ενότητας είναι η λειτουργία του εργαλείου SupportDTD, το οποίο επιτρέπει την επέκταση του σχήματος της σχεσιακής βάσης δεδομένων που υποστηρίζει είτε μια πηγή πληροφορίας που περιέχει δομημένα έγγραφα, είτε μια πηγή πληροφορίας που περιέχει παραδοσιακά δεδομένα με σκοπό την υποστήριξη νέων κλάσεων. Επιπλέον, καθορίζονται τόσο ο τρόπος αποθήκευσης και διαχείρισης των στιγμιοτύπων της κλάσης στην πηγή πληροφορίας όσο και το αν και ποια πληροφορία αποθηκεύεται στη βάση δεδομένων της πηγής πληροφορίας για τα στιγμιότυπα της κλάσης που δεν εισάγονται στην πηγή πληροφορίας αλλά δεικτοδοτούνται σε αυτήν. Μ' αυτό τον τρόπο γίνονται δυνατές στην πηγή πληροφορίας η εισαγωγή, η δεικτοδότηση και η διαχείριση κλάσεων XML εγγράφων ή κλάσεων παραδοσιακών δεδομένων που αναπαρίστανται με τη χρήση XML σύνταξης. Η δομή της XML πληροφορίας, είτε πρόκειται για έγγραφα είτε για παραδοσιακά δεδομένα, στην παρούσα φάση τουλάχιστον, περιγράφεται από XML DTDs.

Όπως έχει ήδη αναφερθεί, για τη σωστή και σύμφωνη με το μοντέλο που περιγράφηκε στο 6^ο κεφάλαιο υποστήριξη κλάσεων XML εγγράφων, κάθε φορά που πρόκειται να υποστηριχθεί μια νέα κλάση XML εγγράφων από κάποια πηγή πληροφορίας που περιέχει δομημένα έγγραφα, είναι απαραίτητη η παρέμβαση του διαχειριστή της πηγής πληροφορίας: Ο διαχειριστής της πηγής πληροφορίας είναι αυτός που καθορίζει τους κανόνες αποθήκευσης, διαχείρισης και δεικτοδότησης των στιγμιοτύπων της νέας κλάσης εγγράφων.

Επιπλέον, καθώς οι DTDs διαθέτουν περιορισμένη εκφραστικότητα και δεν υποστηρίζουν τύπους δεδομένων, απαραίτητη είναι η παρέμβαση του διαχειριστή και για τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα: Ο διαχειριστής της πηγής πληροφορίας είναι υπεύθυνος για τον ορισμό συσχετίσεων μεταξύ οντοτήτων και για τον καθορισμό των τύπων δεδομένων των χαρακτηριστικών των οντοτήτων.



Διάγραμμα 7-3: Διάγραμμα Δραστηριοτήτων για το εργαλείο SupportDTD

Με βάση τα παραπάνω, για να υποστηριχθούν από μια πηγή πληροφορίας η εισαγωγή και η δεικτοδότηση σε αυτήν XML πληροφορίας που η δομή της περιγράφεται από ένα DTD που δεν υποστηρίζεται ήδη από την πηγή πληροφορίας πρέπει να γίνουν τα εξής:

1. Ο διαχειριστής της πηγής πληροφορίας τρέχει το εργαλείο SupportDTD.

- 2.Ο διαχειριστής της πηγής πληροφορίας εισάγει στο εργαλείο, με τη μορφή URL, τη διεύθυνση του DTD που πρόκειται να υποστηριχθεί.
- 3.Ο DTD αποθηκεύεται προσωρινά στο δίσκο του υπολογιστή όπου έχει στηθεί η πηγή πληροφορίας.
- 4.Ο DTD αναλύεται συντακτικά.
 - 4.1.Αν υπάρχει κάποιο συντακτικό λάθος, τυπώνεται το κατάλληλο μήνυμα και η διαδικασία σταματά.
 - 4.2.Αν δεν υπάρχει συντακτικό λάθος, παρουσιάζεται στο διαχειριστή της πηγής πληροφορίας η ιεραρχία (με τη μορφή δέντρου) των στοιχείων που επιτρέπεται να αποτελούν συστατικά των αντικειμένων που δομούνται με βάση τον τρέχοντα DTD.
- 5.Ο διαχειριστής της πηγής πληροφορίας καθορίζει τον τύπο της κλάσης που περιγράφεται από το DTD.
 - 5.1.Αν ο DTD περιγράφει μια κλάση δομημένων εγγράφων, ο διαχειριστής της πηγής πληροφορίας κάνει τα εξής:
 - 5.1.1.Επιλέγει ποια από τις διαθέσιμες κλάσεις εγγράφων θα είναι η υπερκλάση της κλάσης εγγράφων³⁷ που θα δημιουργηθεί.
 - 5.1.2.Δημιουργείται, ως στιγμιότυπο της κλάσης «Κλάση Εγγράφων», μια υποκλάση της κλάσης που έχει επιλεγεί ως υπερκλάση της κλάσης εγγράφων που υποκούνται στον τρέχοντα DTD, η οποία έχει όνομα το όνομα του DTD.
 - 5.1.3.Καθορίζει τις δομές της κλάσης εγγράφων και των κλάσεων των δομικών στοιχείων εγγράφων, όπως αυτές θα εξαχθούν στο μοντέλο πληροφορίας της πηγής πληροφορίας.
 - 5.1.4.Καθορίζει ποια από τα δομικά στοιχεία από τα οποία απαρτίζονται τα στιγμιότυπα της κλάσης εγγράφων θα αποθηκεύονται στην πηγή πληροφορίας για τα δεικτοδοτούμενα στιγμιότυπα της κλάσης εγγράφων.
 - 5.1.5.Επιλέγει τα στοιχεία (αν υπάρχουν φυσικά) που επιτρέπεται να περιέχονται σε έγγραφα που υποκούνται στο DTD και χρησιμοποιούνται ως δεδομένα για την εξαγωγή στοιχείων της Dublin Core που περιγράφουν τα έγγραφα και επιλέγει τη μέθοδο εξαγωγής μεταδεδομένων που θα εφαρμοστεί σ' αυτά.
 - 5.1.6.Επιλέγει τις κλάσεις των δομικών στοιχείων που επιτρέπεται να περιέχονται σε έγγραφα που υποκούνται στο DTD και θα γίνονται διαθέσιμες μέσω του μοντέλου της πηγής πληροφορίας.
 - 5.1.7.Για κάθε κλάση δομικών στοιχείων που θα γίνεται διαθέσιμη μέσω του μοντέλου της πηγής πληροφορίας γίνονται τα εξής:
 - 5.1.7.1.Ο διαχειριστής της πηγής πληροφορίας επιλέγει ποια από τις διαθέσιμες κλάσεις τμημάτων εγγράφων³⁸ θα είναι η υπερκλάση της κλάσης τμημάτων εγγράφων που θα δημιουργηθεί.
 - 5.1.7.2.Δημιουργείται, ως στιγμιότυπο της κλάσης «Κλάση Τμημάτων Εγγράφων», μια υποκλάση της κλάσης που επιλέχθηκε ως υπερκλάση της τρέχουσας κλάσης τμημάτων εγγράφων, με όνομα <όνομα DTD>:<όνομα δομικού στοιχείου>.
 - 5.1.8.Καθορίζει αν τα έγγραφα που υποκούνται στον τρέχοντα DTD θα επιτρέπεται να αποθηκεύονται τμηματικά ή αν θα αποθηκεύονται πάντα ενιαία.
 - 5.1.8.1.Αν θα επιτρέπεται τα έγγραφα να αποθηκεύονται τμηματικά, γίνονται τα εξής:

³⁷ Η υπερκλάση της κλάσης εγγράφων που θα δημιουργηθεί είναι η κλάση Έγγραφο ή κάποια υποκλάση της.

³⁸ Η υπερκλάση της κλάσης τμημάτων εγγράφων που θα δημιουργηθεί είναι η κλάση «Τμήμα Εγγράφου» ή κάποια υποκλάση της.

- 5.1.8.1.1. Ο διαχειριστής της πηγής πληροφορίας καθορίζει αν η τμηματική αποθήκευση θα επιτρέπεται πάντα ή μόνο στην περίπτωση που το μέγεθος του εγγράφου υπερβαίνει κάποια συγκεκριμένη τιμή.
- 5.1.8.1.1.1. Αν η τμηματική αποθήκευση θα επιτρέπεται μόνο στην περίπτωση που το μέγεθος του εγγράφου υπερβαίνει κάποια συγκεκριμένη τιμή, ο διαχειριστής της πηγής πληροφορίας ορίζει την τιμή αυτή.
- 5.1.8.1.2. Ο διαχειριστής της πηγής πληροφορίας καθορίζει για ποιες από τις κλάσεις δομικών στοιχείων του εγγράφου επιτρέπεται τα στιγμιότυπά της να αποθηκεύονται ανεξάρτητα από το έγγραφο στο οποίο περιέχονται.
- 5.1.8.1.3. Για κάθε κλάση στοιχείων εγγράφου που επιτρέπεται στιγμιότυπά της να αποθηκεύονται ανεξάρτητα από το έγγραφο ή το δομικό στοιχείο εγγράφου στο οποίο περιέχονται, γίνονται τα εξής:
- 5.1.8.1.3.1. Δημιουργείται μια οντότητα που αναπαριστά τα στιγμιότυπα της κλάσης.
- 5.1.8.1.3.2. Αν η ανεξάρτητη αποθήκευση των στιγμιότυπων της κλάσης θα επιτρέπεται μόνο στην περίπτωση που το μέγεθος του εγγράφου υπερβαίνει κάποια συγκεκριμένη τιμή, ο διαχειριστής της πηγής πληροφορίας ορίζει την τιμή αυτή.
- 5.1.8.2. Για κάθε κλάση δομικών στοιχείων που θα γίνεται διαθέσιμη μέσω του μοντέλου της πηγής πληροφορίας γίνονται τα εξής:
- 5.1.8.2.1. Στο πεδίο DTD της κλάσης που δημιουργήθηκε αποθηκεύεται ο URL του DTD που περιγράφει τη δομή των στιγμιότυπων της.
- 5.1.8.2.2. Στο πεδίο Δομή της κλάσης που δημιουργήθηκε αποθηκεύεται ο URL του RDF σχήματος που περιγράφει τη δομή της όπως αυτή εξάγεται από τη γλώσσα της πηγής πληροφορίας.
- 5.1.8.2.3. Στο πεδίο Κανόνες Αναπαράστασης της κλάσης που δημιουργήθηκε αποθηκεύεται ο URL μέσω του οποίου δίνεται πρόσβαση στους κανόνες αναπαράστασης και διαχείρισης των στιγμιότυπων της κλάσης.
- 5.1.8.2.4. Στο πεδίο Στυλ της κλάσης που δημιουργήθηκε αποθηκεύεται (αν υπάρχει) ο URL του προκαθορισμένου από το διαχειριστή της πηγής πληροφορίας στυλ μορφοποίησης για τα στιγμιότυπα της κλάσης.
- 5.1.8.2.5. Στο πεδίο Namespace της κλάσης που δημιουργήθηκε αποθηκεύεται (αν υπάρχει) η namespace του τρέχοντα DTD.
- 5.1.8.3. Για κάθε κλάση στοιχείων εγγράφων, ανεξάρτητα από την πολιτική αποθήκευσης των στιγμιότυπων της, ο διαχειριστής της πηγής πληροφορίας καθορίζει αν οι λέξεις που περιέχονται στα στιγμιότυπά της έχουν διαφορετική βαρύτητα για την περιγραφή του περιεχομένου

των εγγράφων στα οποία περιέχονται³⁹ από τις λέξεις που περιέχονται στα στιγμιότυπα των υπόλοιπων κλάσεων.

5.1.8.4. Στο πεδίο DTD της κλάσης που δημιουργήθηκε αποθηκεύεται ο URL του DTD που περιγράφει τη δομή των στιγμιοτύπων της.

5.1.8.5. Στο πεδίο Δομή της κλάσης εγγράφων που δημιουργήθηκε αποθηκεύεται ο URL του RDF σχήματος που περιγράφει τη δομή της όπως αυτή εξάγεται από τη γλώσσα της πηγής πληροφορίας.

5.1.8.6. Στο πεδίο Κανόνες Αναπαράστασης της κλάσης εγγράφων που δημιουργήθηκε αποθηκεύεται ο URL μέσω του οποίου δίνεται πρόσβαση στους κανόνες αναπαράστασης και διαχείρισης των στιγμιοτύπων της κλάσης.

5.1.8.7. Στο πεδίο Στυλ της κλάσης εγγράφων που δημιουργήθηκε αποθηκεύεται (αν υπάρχει) ο URL του προκαθορισμένου από το διαχειριστή της πηγής πληροφορίας στυλ μορφοποίησης για τα στιγμιότυπα της κλάσης.

5.1.8.8. Στο πεδίο Namespace της κλάσης εγγράφων που δημιουργήθηκε αποθηκεύεται (αν υπάρχει) η namespace του τρέχοντα DTD.

5.2. Αν ο DTD περιγράφει μια κλάση παραδοσιακών δεδομένων, γίνονται τα εξής:

5.2.1. Ο διαχειριστής της πηγής πληροφορίας επιλέγει αν το στοιχείο – ρίζα του DTD θα αποτελέσει οντότητα.

5.2.2. Για κάθε XML στοιχείο που περιγράφεται στο DTD γίνονται τα εξής:

5.2.2.1. Ο διαχειριστής της πηγής πληροφορίας επιλέγει αν το δομικό στοιχείο θα αναπαρίσταται ως χαρακτηριστικό του στοιχείου στο οποίο περιέχεται ή ως ανεξάρτητη οντότητα.

5.2.2.2. Αν το στοιχείο αναπαρίσταται ως χαρακτηριστικό, ο διαχειριστής της πηγής πληροφορίας επιλέγει την κλάση τα στιγμιότυπα της οποίας θα διαθέτουν το χαρακτηριστικό αυτό.

5.2.2.3. Για κάθε δομικό στοιχείο που θα αναπαρίσταται ως κλάση γίνονται τα εξής:

5.2.2.3.1. Ο διαχειριστής της πηγής πληροφορίας επιλέγει την κλάση που θα είναι υπερκλάση της κλάσης που θα αναπαριστά τα στιγμιότυπα του δομικού στοιχείου⁴⁰.

5.2.2.3.2. Δημιουργείται μια κλάση, που αναπαριστά τα στιγμιότυπα του δομικού στοιχείου, και έχει όνομα <όνομα DTD>:<όνομα δομικού στοιχείου>.

5.2.2.3.3. Τα χαρακτηριστικά των στιγμιοτύπων του αντίστοιχου στοιχείου αποτελούν χαρακτηριστικά των στιγμιοτύπων της κλάσης αυτής.

5.2.2.3.4. Χαρακτηριστικά αποτελούν και τα στοιχεία που επιλέχθηκε να αναπαρίστανται ως χαρακτηριστικά των στιγμιοτύπων της κλάσης αυτής.

³⁹ Έτσι, π.χ., σε μια παράγραφο, οι λέξεις που περιέχονται σε ένα στοιχείο που δηλώνει έμφαση (όπως είναι τα στοιχεία , <i>, και της HTML) μπορεί να θεωρηθούν σημαντικότερες για την περιγραφή του εγγράφου που περιέχει την παράγραφο απ' ότι οι υπόλοιπες λέξεις της παραγράφου. Αντίστοιχα, οι λέξεις του τίτλου ενός κεφαλαίου μπορεί να θεωρηθούν σημαντικότερες για την περιγραφή του εγγράφου που περιέχει το κεφάλαιο απ' ότι οι λέξεις που περιγράφονται σε οποιαδήποτε παράγραφο του ίδιου κεφαλαίου.

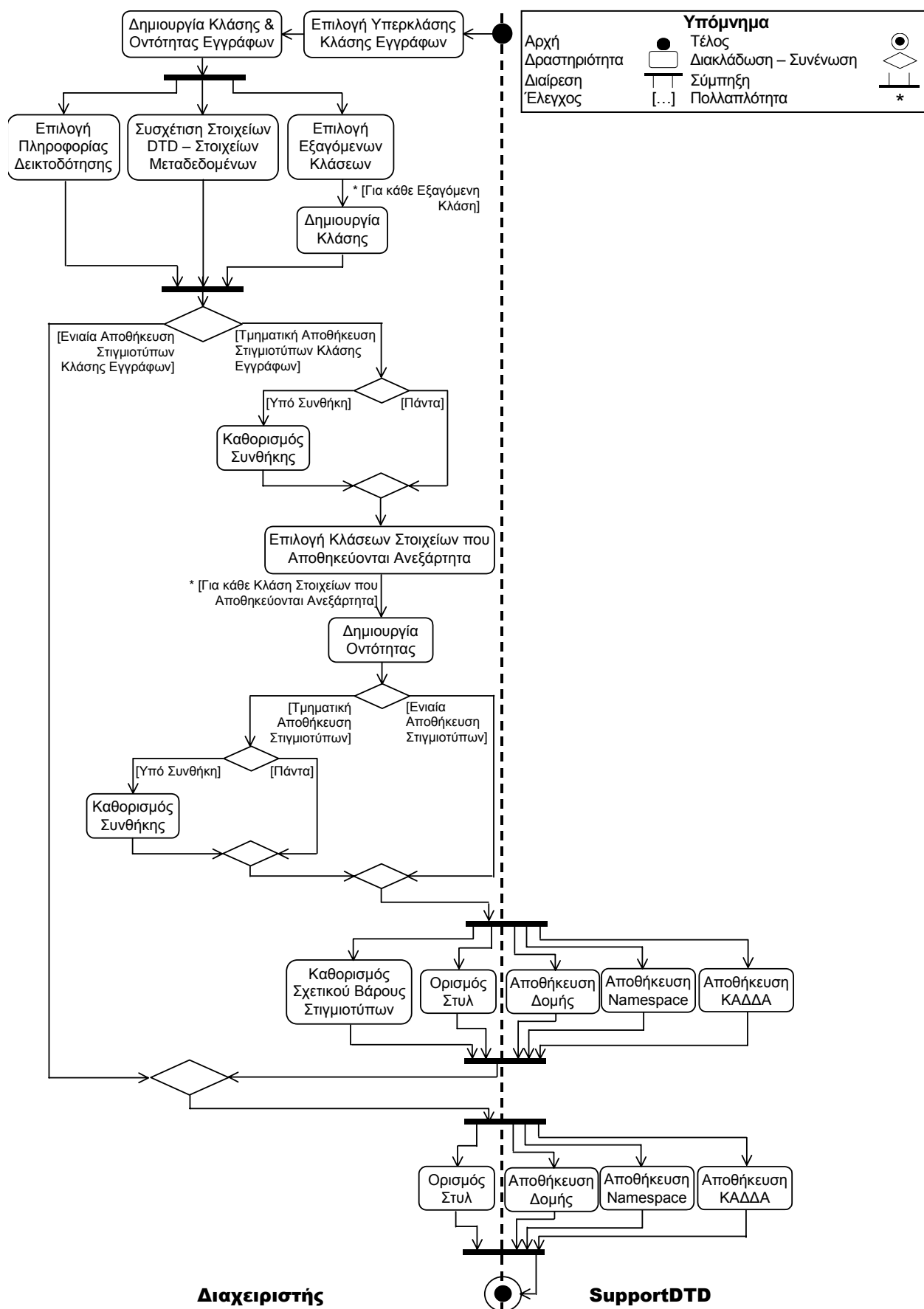
⁴⁰ Η υπερκλάση της κλάσης που θα δημιουργηθεί είναι η κλάση Αντικείμενο ή κάποια υποκλάση της.

- 5.2.2.3.5. Αν το στοιχείο έχει περιεχόμενο σε μορφή κειμένου, αυτό αποθηκεύεται ως χαρακτηριστικό της σχέσης με όνομα «Περιεχόμενο».
- 5.2.2.3.6. Αν το στοιχείο έχει υποστοιχεία που αναπαρίστανται από κλάσεις, για καθένα από αυτά γίνονται τα εξής:
- 5.2.2.3.6.1. Τα στιγμιότυπα της κλάσης που αναπαριστά το υποστοιχείο έχουν ένα ξένο κλειδί που αντιστοιχεί στις ταυτότητες των στιγμιοτύπων της τρέχουσας κλάσης στα οποία περιέχονται.
- 5.2.2.3.6.2. Αν το υποστοιχείο εμφανίζεται το πολύ μια φορά μέσα στο στοιχείο, τα στιγμιότυπα της κλάσης που αναπαριστά το τρέχον στοιχείο έχουν ένα ξένο κλειδί που αντιστοιχεί στις ταυτότητες των στιγμιοτύπων της κλάσης του υποστοιχείου με τα οποία συσχετίζονται.
- 5.2.2.3.7. Ο διαχειριστής της πηγής πληροφορίας καθορίζει τη δομή της κλάσης, όπως αυτή εξάγεται στο μοντέλο πληροφορίας της πηγής πληροφορίας.
- 5.2.2.3.8. Στο πεδίο DTD της κλάσης που δημιουργήθηκε αποθηκεύεται ο URL του DTD που περιγράφει τη δομή των στιγμιοτύπων της.
- 5.2.2.3.9. Στο πεδίο Δομή της κλάσης που δημιουργήθηκε αποθηκεύεται ο URL του RDF σχήματος που περιγράφει τη δομή της κλάσης, όπως αυτή εξάγεται από τη γλώσσα της πηγής πληροφορίας.
- 5.2.2.3.10. Στο πεδίο Κανόνες Αναπαράστασης της κλάσης που δημιουργήθηκε αποθηκεύεται ο URL μέσω του οποίου δίνεται πρόσβαση στους κανόνες αναπαράστασης και διαχείρισης των στιγμιοτύπων της κλάσης.
- 5.2.2.3.11. Στο πεδίο Στυλ της κλάσης που δημιουργήθηκε αποθηκεύεται (αν υπάρχει) ο URL του προκαθορισμένου από το διαχειριστή της πηγής πληροφορίας στυλ μορφοποίησης για τα στιγμιότυπα της κλάσης.
- 5.2.2.3.12. Στο πεδίο Namespace της κλάσης εγγράφων που δημιουργήθηκε αποθηκεύεται (αν υπάρχει) η namespace του τρέχοντα DTD.
- 5.2.2.3.13. Ο διαχειριστής της πηγής πληροφορίας καθορίζει τον τύπο κάθε χαρακτηριστικού.
- 5.2.2.3.14. Ο διαχειριστής της πηγής πληροφορίας καθορίζει το όνομα κάθε χαρακτηριστικού⁴¹.
- 5.2.2.3.15. Ο διαχειριστής της πηγής πληροφορίας καθορίζει ποια από τα χαρακτηριστικά των στιγμιοτύπων της κλάσης θα αποθηκεύονται στην πηγή πληροφορίας για τα δεικτοδοτούμενα στιγμιότυπα της κλάσης.

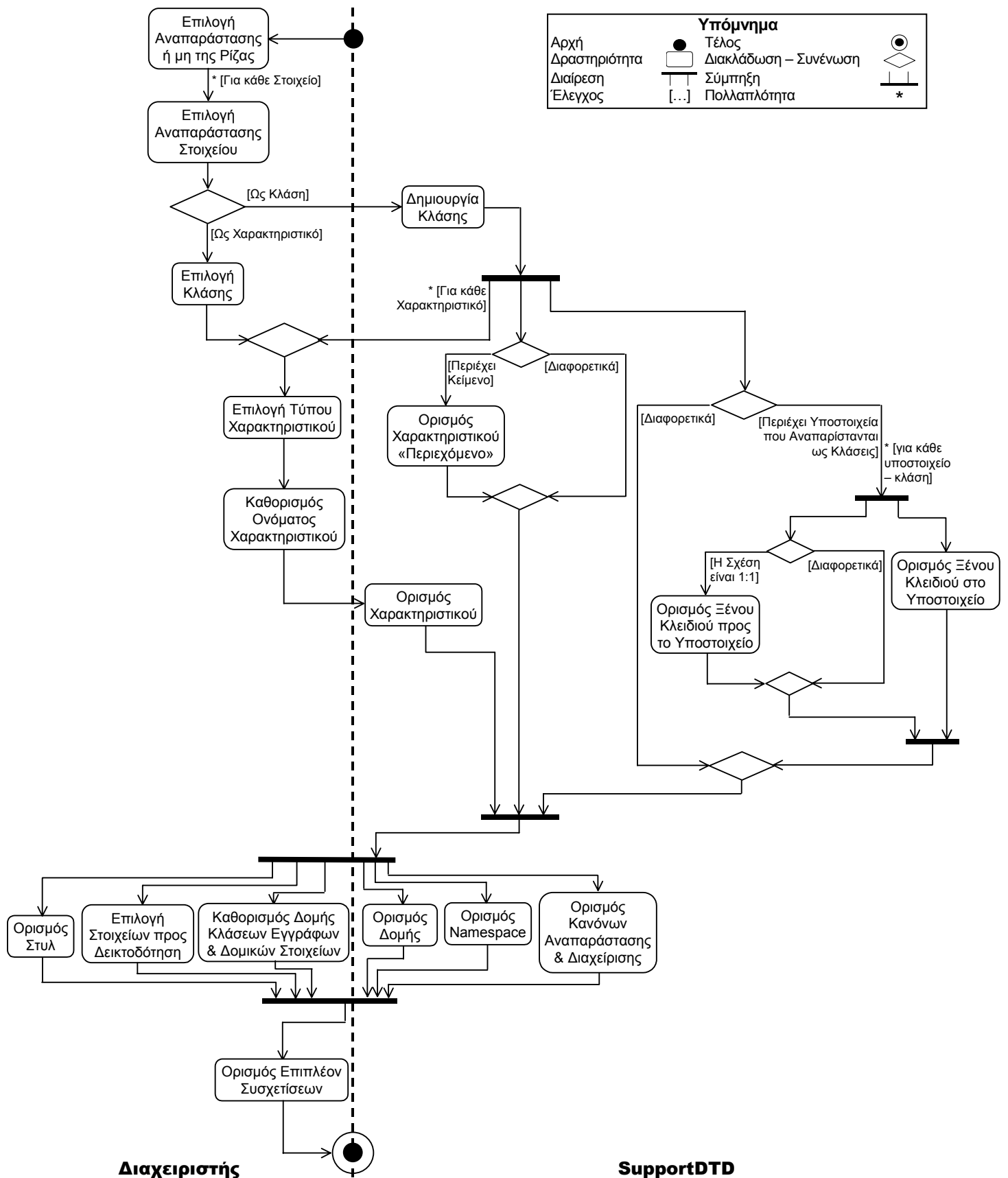
5.2.3. Ο διαχειριστής της πηγής πληροφορίας καθορίζει τις επιπλέον συσχετίσεις που μπορεί να υπάρχουν μεταξύ των οντοτήτων.

5.3. Διαφορετικά τυπώνεται το μήνυμα λάθους «Ο Τύπος της Κλάσης δεν υποστηρίζεται»

⁴¹ Από το εργαλείο SupportDTD προτείνεται κάποιο προκαθορισμένο όνομα, το οποίο ο διαχειριστής της πηγής πληροφορίας αλλάζει, εφ' όσον το επιθυμεί.



Διάγραμμα 7-4: Διάγραμμα Δραστηριοτήτων για τον Ορισμό Κανόνων Αναπαράστασης, Διαχείρισης & Δεικτοδότησης των Στιγμιωτύπων των Κλάσεων Εγγράφων



Διάγραμμα 7-5: Διάγραμμα Δραστηριοτήτων για τον Ορισμό Κανόνων Αναπαράστασης, Διαχείρισης & Δεικτοδότησης των Στιγμιότυπων των Κλάσεων Παραδοσιακών Δεδομένων

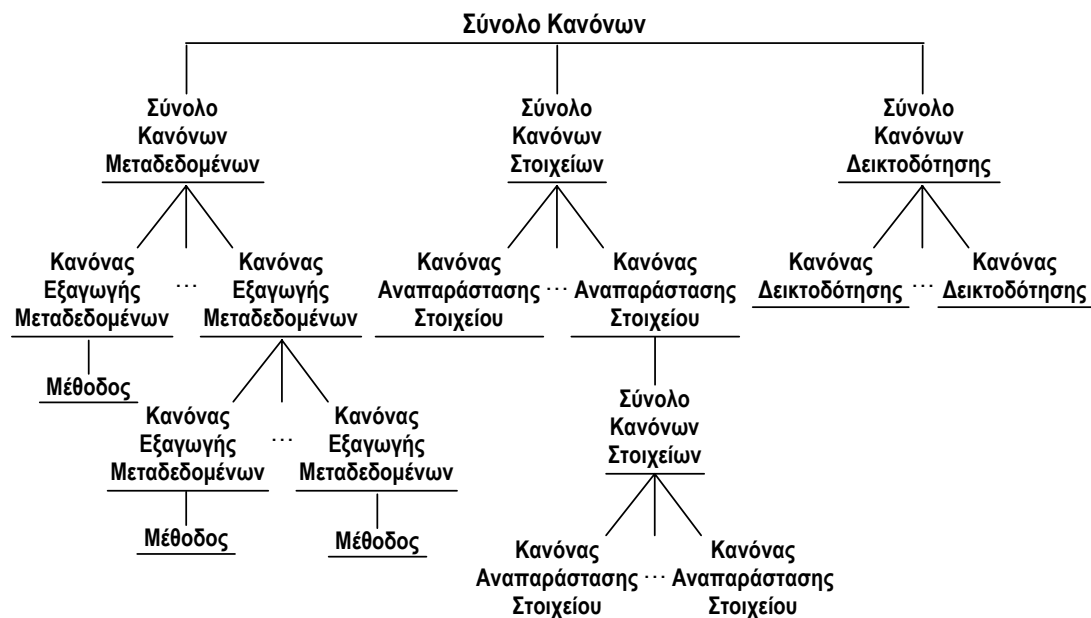
Η συνολική λειτουργία του εργαλείου SupportDTD απεικονίζεται αφαιρετικά στο Διάγραμμα 7-3, που είναι UML διάγραμμα δραστηριοτήτων. Ο ορισμός των κανόνων αναπαράστασης και διαχείρισης των στιγμιοτύπων των κλάσεων εγγράφων απεικονίζεται λεπτομερώς στο Διάγραμμα 7-4, ενώ ο ορισμός των κανόνων αναπαράστασης και διαχείρισης των στιγμιοτύπων των κλάσεων παραδοσιακών δεδομένων φαίνεται λεπτομερώς στο Διάγραμμα 7-5.

Στο σημείο αυτό πρέπει να σημειωθεί ότι μια ελαφρώς διαφορετική χρήση του εργαλείου SupportDTD λαμβάνει χώρα κατά την υποστήριξη εργαλείων όπως οι πράκτορες, που είναι υπεύθυνα για την ενημέρωση των πηγών πληροφορίας για αντικείμενα που γίνονται διαθέσιμα και το περιεχόμενό τους σχετίζεται με αυτό των πηγών πληροφορίας. Λεπτομέρειες για το θέμα αυτό υπάρχουν στο τμήμα 7.6, όπου συζητούνται τα σχετικά με την ενημέρωση των πηγών πληροφορίας θέματα.

7.2.2. Η Γλώσσα ΚΑΔΔΑ (Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Αντικειμένων)

Η γλώσσα ΚΑΔΔΑ (Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Αντικειμένων), που περιγράφεται σ' αυτή την ενότητα, είναι ένα RDF λεξικό για την έκφραση των κανόνων αναπαράστασης, διαχείρισης και δεικτοδότησης αντικειμένων, η δομή των οποίων περιγράφεται από XML DTDs⁴².

Για την αναπαράσταση και τη διαχείριση αντικειμένων των οποίων η δομή περιγράφεται από έναν XML DTD δημιουργείται ένα RDF αρχείο που η σύνταξή του είναι συμβατή με την ΚΑΔΔΑ. Το αρχείο αυτό περιέχει ένα σύνολο κανόνων για την αναπαράσταση και τη διαχείριση της πληροφορίας. Στο Σχήμα 7-1 φαίνεται η δομή ενός *Συνόλου Κανόνων*.



Σχήμα 7-1: Η Δομή ενός ΚΑΔΔΑ Συνόλου Κανόνων

⁴² Όταν η σύνταξη της XML Schema Language σταθεροποιηθεί, η ΚΑΔΔΑ θα τροποποιηθεί ώστε να μπορεί επιπλέον να υποστηρίξει δομημένη πληροφορία η δομή της οποίας περιγράφεται από σχήματα εκφρασμένα στην XML Schema Language.

Όπως φαίνεται στο Σχήμα 7-1, ένα *Σύνολο Κανόνων* απαρτίζεται από ένα *Σύνολο Κανόνων Μεταδεδομένων*, ένα *Σύνολο Κανόνων Στοιχείων* και ένα *Σύνολο Κανόνων Δεικτοδότησης*.

Ένα *Σύνολο Κανόνων Μεταδεδομένων* είναι μια συλλογή *Κανόνων Εξαγωγής Μεταδεδομένων*. Ένας *Κανόνας Εξαγωγής Μεταδεδομένων* επιτρέπει την εξαγωγή της τιμής κάποιου στοιχείου μεταδεδομένων από το περιεχόμενο ενός ή περισσότερων αντικειμένων. Αυτό επιτυγχάνεται με την εφαρμογή μιας *Μεθόδου* στα *Δεδομένα* του Κανόνα. Τα *Δεδομένα* του κανόνα είναι κάποια από τα στοιχεία που απαρτίζουν το αντικείμενο που πρόκειται να περιγραφεί από τα μεταδεδομένα που θα εξαχθούν. Το *Αποτέλεσμα* του Κανόνα είναι το αποτέλεσμα της εφαρμογής της μεθόδου του κανόνα στα δεδομένα του κανόνα. Οι υποστηριζόμενες από την ΚΑΔΔΑ μέθοδοι περιγράφονται στον πίνακα που ακολουθεί.

Όνομα	Παράμετροι	Τελεστής	Περιγραφή
Αντιγραφή		Αντιγραφή	Η τιμή των δεδομένων αντιγράφεται στο αποτέλεσμα
Απαλοιφή Κοινών Όρων	Λεξικό συχνά εμφανιζόμενων όρων	Αντιγραφή-Τομή	Το αποτέλεσμα προκύπτει από την απαλοιφή των συχνά εμφανιζόμενων όρων από τα δεδομένα
Επιλογή Όρων	Λεξικό όρων	Τομή	Το αποτέλεσμα προκύπτει από την τομή των όρων που περιέχονται στο λεξικό με το περιεχόμενο των δεδομένων

Πίνακας 7-1: Οι υποστηριζόμενες από την ΚΑΔΔΑ Μέθοδοι

Ένας *Κανόνας Εξαγωγής Μεταδεδομένων* μπορεί να είναι σύνθετος, οπότε απαρτίζεται από έναν ή περισσότερους κανόνες εξαγωγής μεταδεδομένων. Αυτό συναντάται συχνά κατά την εξαγωγή στοιχείων μεταδεδομένων που διαθέτουν πάνω από ένα πεδία.

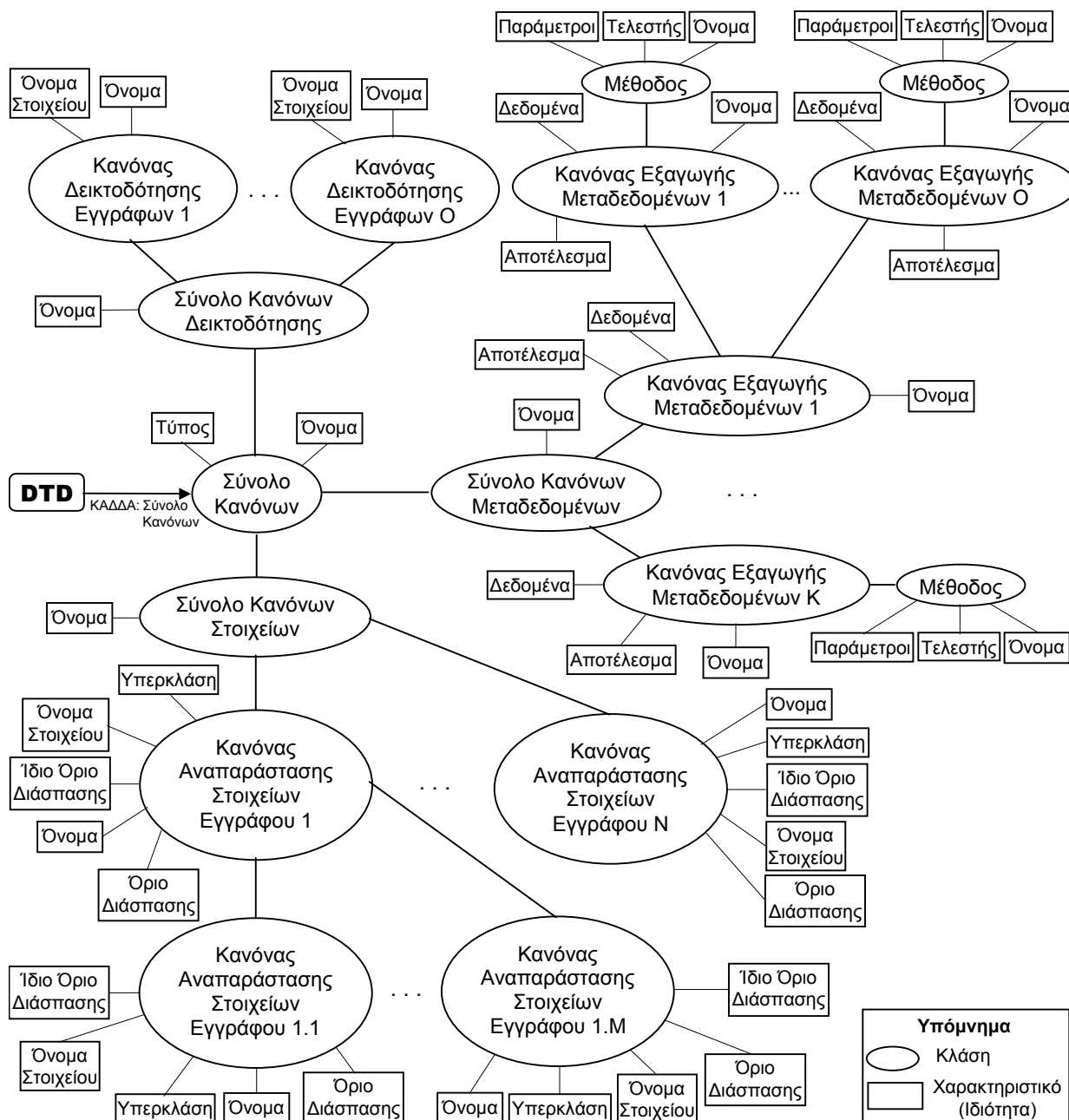
Ένα *Σύνολο Κανόνων Στοιχείων* είναι μια συλλογή *Κανόνων Αναπαράστασης Στοιχείων*. Κάθε *Κανόνας Αναπαράστασης Στοιχείου* καθορίζει την αναπαράσταση, στη βάση δεδομένων της πηγής πληροφορίας, ενός στοιχείου που επιτρέπεται να αποτελεί τμήμα της δομής των αντικειμένων που υπακούουν στον τρέχοντα DTD. Μέσα σε έναν Κανόνα Αναπαράστασης Στοιχείου μπορεί να περιέχεται το *Σύνολο Κανόνων Στοιχείων* για τα στοιχεία που επιτρέπεται να αποτελούν περιεχόμενο του στοιχείου του οποίου η αναπαράσταση καθορίζεται από τον κανόνα⁴³. Στο σημείο αυτό πρέπει να αναφερθεί ότι οι κανόνες αναπαράστασης για κάθε στιγμιότυπο αντικειμένου, η δομή του οποίου καθορίζεται από τον τρέχοντα DTD, καθορίζονται για το στοιχείο που αποτελεί, με βάση το DTD, τη ρίζα του αντικειμένου.

Ένα *Σύνολο Κανόνων Δεικτοδότησης* είναι μια συλλογή *Κανόνων Δεικτοδότησης*, οι οποίοι καθορίζουν ποια χαρακτηριστικά και ποια υποστοιχεία των στιγμιότυπων της

⁴³ Η προσέγγιση αυτή επιλέχθηκε, καθώς δίνει τη δυνατότητα η διαχείριση ενός στοιχείου να είναι διαφορετική, ανάλογα με τον τύπο του στοιχείου στο οποίο περιέχεται κάθε φορά (π.χ. η διαχείριση των παραγράφων που περιέχονται σε κάποια ενότητα μπορεί να διαφέρει από τη διαχείριση των παραγράφων που περιέχονται σε κάποια υποενότητα).

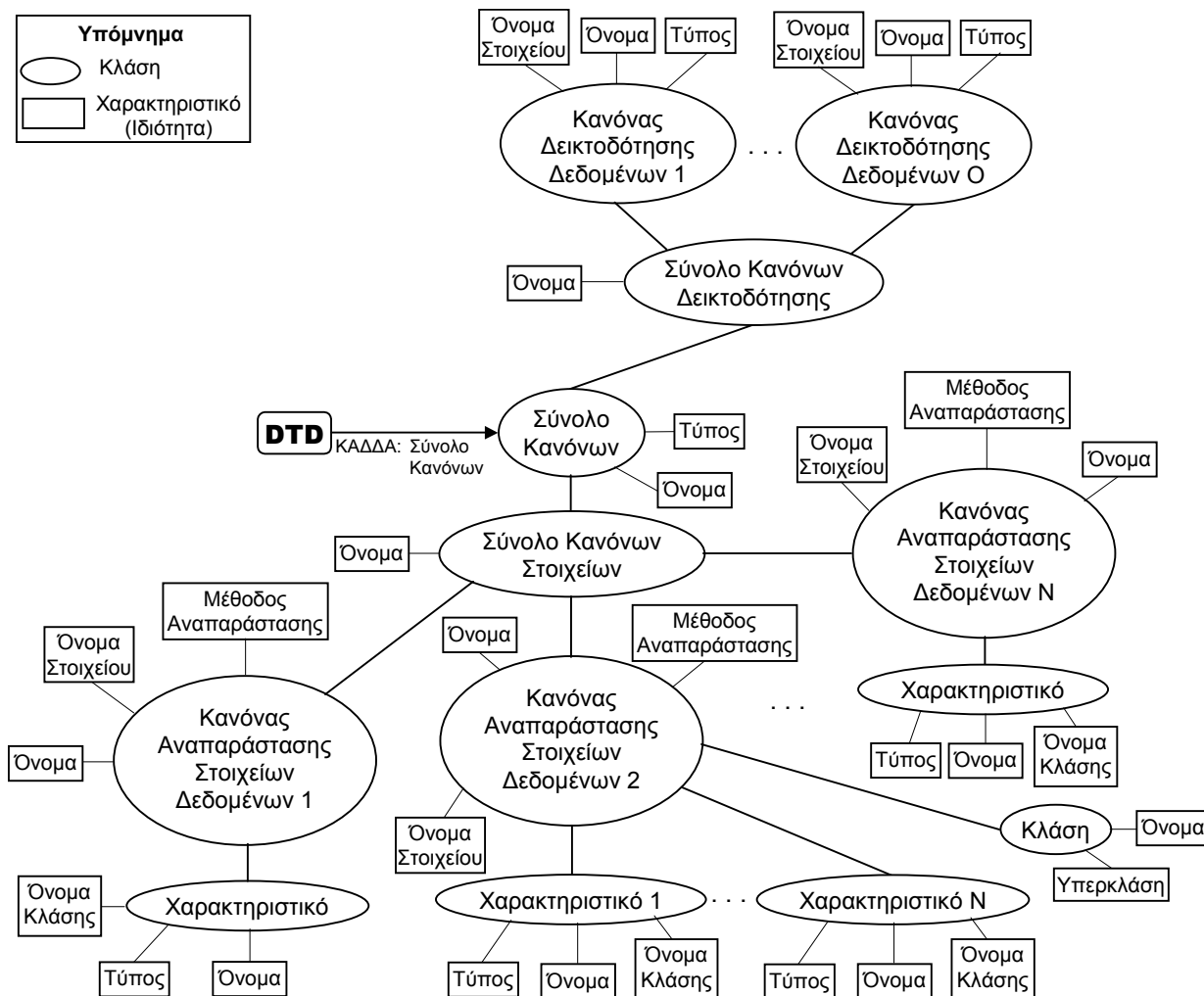
περιγραφόμενης κλάσης θα αποθηκεύονται στη βάση δεδομένων της πηγής πληροφορίας για τα δεικτοδοτούμενα στιγμιότυπά της.

Όπως όλες οι RDF εφαρμογές, η γλώσσα ΚΑΔΔΑ περιγράφεται από ένα RDF μοντέλο. Για να είναι πιο κατανοητό το RDF μοντέλο της γλώσσας ΚΑΔΔΑ, στη συνέχεια παρουσιάζεται χωριστά το RDF μοντέλο των συνόλων κανόνων που αντιστοιχούν σε DTDs που περιγράφουν τη δομή εγγράφων και το RDF μοντέλο των συνόλων κανόνων που αντιστοιχούν σε DTDs που περιγράφουν τη δομή παραδοσιακών δεδομένων. Έτσι, το RDF μοντέλο των συνόλων κανόνων που αντιστοιχούν σε DTDs που περιγράφουν τη δομή εγγράφων φαίνεται στο Σχήμα 7-2.



Σχήμα 7-2: Το RDF Μοντέλο των Συνόλων Κανόνων που αντιστοιχούν σε DTDs που περιγράφουν τη Δομή Εγγράφων

Αντίστοιχα, το RDF μοντέλο των συνόλων κανόνων που αντιστοιχούν σε DTDs που περιγράφουν τη δομή παραδοσιακών δεδομένων φαίνεται στο Σχήμα 7-3.



Σχήμα 7-3: Το RDF Μοντέλο των Συνόλων Κανόνων που αντιστοιχούν σε DTDs που περιγράφουν τη Δομή Παραδοσιακών Δεδομένων

Στη συνέχεια περιγράφονται αναλυτικότερα οι Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Δομημένων Εγγράφων (στην υποενότητα 7.2.2.1), οι Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Παραδοσιακών Δεδομένων (στην υποενότητα 7.2.2.2) και η Σύνταξη της Γλώσσας ΚΑΔΔΑ (στην υποενότητα 7.2.2.3).

7.2.2.1. Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Δομημένων Εγγράφων

Οι Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Δομημένων Εγγράφων που υποστηρίζονται από τη γλώσσα ΚΑΔΔΑ περιγράφονται σ' αυτή την υποενότητα.

Για την περιγραφή μιας κλάσης εγγράφων δημιουργείται ένα σύνολο κανόνων. Ο Πίνακας 7-2 περιέχει τα στοιχεία που απαρτίζουν ένα σύνολο κανόνων για την υποστήριξη κλάσεων δομημένων εγγράφων, μαζί με τα χαρακτηριστικά τους.

Όνομα Στοιχείου	Όνομα	Χαρακτηριστικά Τιμή	Περιγραφή
--------------------	-------	------------------------	-----------

Σύνολο Κανόνων	Τύπος	«Έγγραφα»	Ο τύπος της κλάσης που περιγράφεται από το DTD
	Όνομα	<Όνομα DTD>	Το όνομα του τρέχοντος συνόλου κανόνων
Σύνολο Κανόνων Μεταδεδομένων	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντος συνόλου κανόνων μεταδεδομένων
Κανόνας Εξαγωγής Μεταδεδομένων	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντα κανόνα
	Αποτέλεσμα	<Όνομα Στοιχείου ή Χαρακτηριστικού Στοιχείου Μεταδεδομένων>	Το στοιχείο ή το χαρακτηριστικό στοιχείου μεταδεδομένων το περιεχόμενο του οποίου προκύπτει από τον τρέχοντα κανόνα
	Δεδομένα	Λίστα που περιέχει τα ονόματα των συμβολοσειρών, των στοιχείων και των χαρακτηριστικών που αποτελούν τα δεδομένα του κανόνα	Τα δεδομένα που χρησιμοποιούνται για την εξαγωγή των μεταδεδομένων
Μέθοδος	Όνομα	Συμβολοσειρά	Το όνομα της τρέχουσας μεθόδου
	Τελεστής	«Αντιγραφή-Τομή» / «Αντιγραφή» / «Τομή»	Ο τελεστής που εφαρμόζεται από την τρέχουσα μέθοδο
	Παράμετροι	Συμβολοσειρά που δίνεται από το χρήστη	Οι παράμετροι της τρέχουσας μεθόδου
Σύνολο Κανόνων Στοιχείων	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντος συνόλου κανόνων στοιχείων
Κανόνας Αναπαράστασης Στοιχείων Εγγράφων	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντος κανόνα αναπαράστασης στοιχείων
	Όνομα Στοιχείου	<Όνομα Στοιχείου>	Το όνομα του στοιχείου που η αναπαράστασή του καθορίζεται από τον τρέχοντα κανόνα
	Υπερκλάση	<Όνομα Υπερκλάσης>	Το όνομα της υπερκλάσης της κλάσης όπου θα ανήκουν τα στιγμιότυπα του τρέχοντος στοιχείου
	Όριο Διάσπασης	0, αν θα επιτρέπεται η διάσπαση πάντα -1, αν η διάσπαση δε θα επιτρέπεται ποτέ n (>0), όπου n το κάτω όριο της τιμής του μεγέθους του στοιχείου για να επιτρέπεται η διάσπαση	Το κάτω όριο της τιμής του μεγέθους του στοιχείου που περιέχει τα στιγμιότυπα του τρέχοντος στοιχείου, για να επιτρέπεται η ανεξάρτητη από το περιέχον στοιχείο αποθήκευση των στιγμιοτύπων του τρέχοντος στοιχείου
	Ίδιο Όριο Διάσπασης	0, αν θα επιτρέπεται η διάσπαση πάντα	Το κάτω όριο της τιμής του μεγέθους του τρέχοντος στοιχείου, για να επιτρέπεται η

		-1, αν η διάσπαση δε θα επιτρέπεται ποτέ $n (>0)$, όπου n το κάτω όριο για να επιτρέπεται η διάσπαση	ανεξάρτητη από το στοιχείο στο οποίο περιέχονται αποθήκευση των στιγμιότυπων του τρέχοντος στοιχείου
	Διαφορά Βάρους	$[-1 - 1]$	Η διαφορά του βάρους των όρων που περιέχονται στα στιγμιότυπα του τρέχοντος στοιχείου, σε σχέση με το βάρος του περιεχομένου του στοιχείου που το περιέχει και των στοιχείων που είναι στο ίδιο επίπεδο με τα στιγμιότυπα του τρέχοντος στοιχείου και είναι στιγμιότυπα άλλων στοιχείων
Σύνολο Κανόνων Δεικτοδότησης	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντος συνόλου κανόνων δεικτοδότησης
Κανόνας Δεικτοδότησης	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντα κανόνα δεικτοδότησης
	Όνομα Στοιχείου	<Όνομα Στοιχείου>	Το όνομα του στοιχείου που η παρουσία του στη βάση δεδομένων για τα δεικτοδοτούμενα στιγμιότυπα της κλάσης εγγράφων καθορίζεται από τον τρέχοντα κανόνα

Πίνακας 7-2: Τα Στοιχεία που απαρτίζουν ένα ΚΑΔΔΑ Σύνολο Κανόνων Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Δομημένων Εγγράφων

Όπως φαίνεται στον παραπάνω πίνακα, ένα Σύνολο Κανόνων απαρτίζεται από ένα Σύνολο Κανόνων Μεταδεδομένων και ένα Σύνολο Κανόνων Στοιχείων, ενώ διαθέτει και τα χαρακτηριστικά όνομα και τύπος, που αντιστοιχούν στο όνομα και τον τύπο του τρέχοντος συνόλου κανόνων.

Ένα Σύνολο Κανόνων Μεταδεδομένων απαρτίζεται από Κανόνες Εξαγωγής Μεταδεδομένων, ενώ διαθέτει και το χαρακτηριστικό όνομα που αντιστοιχεί στο όνομα του τρέχοντος συνόλου κανόνων μεταδεδομένων. Κάθε Κανόνας Εξαγωγής Μεταδεδομένων διαθέτει τα χαρακτηριστικά όνομα, δεδομένα και αποτέλεσμα που αντιστοιχούν στο όνομα, τα δεδομένα και το αποτέλεσμα τρέχοντα κανόνα. Επιπλέον, κάθε απλός κανόνας εξαγωγής μεταδεδομένων διαθέτει το υποστοιχείο Μέθοδος, που αντιστοιχεί στη μέθοδο που εφαρμόζεται στα δεδομένα του κανόνα για να προκύψει το αποτέλεσμα, ενώ κάθε σύνθετος κανόνας διαθέτει ένα αριθμό κανόνων εξαγωγής μεταδεδομένων. Κάθε μέθοδος διαθέτει τα χαρακτηριστικά όνομα, τελεστής και παράμετροι που αντιστοιχούν στο όνομα, τον τελεστή και τις παραμέτρους της μεθόδου.

Αντίστοιχα με ένα Σύνολο Κανόνων Μεταδεδομένων, ένα Σύνολο Κανόνων Στοιχείων απαρτίζεται από Κανόνες Αναπαράστασης Στοιχείων Εγγράφου, ενώ διαθέτει και το χαρακτηριστικό όνομα που αντιστοιχεί στο όνομα του τρέχοντος συνόλου κανόνων

στοιχείων. Ένας Κανόνας Αναπαράστασης Στοιχείων Εγγράφου διαθέτει τα εξής χαρακτηριστικά:

- Όνομα, που αντιστοιχεί στο όνομα του τρέχοντα κανόνα.
- Όνομα στοιχείου, που αντιστοιχεί στο όνομα του στοιχείου που η αναπαράστασή του καθορίζεται από τον τρέχοντα κανόνα.
- Υπερκλάση, που αντιστοιχεί στο όνομα της υπερκλάσης της κλάσης όπου θα ανήκουν τα στιγμιότυπα του τρέχοντος στοιχείου.
- Όριο διάσπασης, που αντιστοιχεί στο κάτω όριο της τιμής του μεγέθους του στοιχείου που περιέχει τα στιγμιότυπα του τρέχοντος στοιχείου, για να επιτρέπεται η ανεξάρτητη από το περιέχον στοιχείο αποθήκευση των στιγμιοτύπων του τρέχοντος στοιχείου.
- Ίδιο όριο διάσπασης, που αντιστοιχεί στο κάτω όριο της τιμής του μεγέθους του τρέχοντος στοιχείου, για να επιτρέπεται η ανεξάρτητη από το στοιχείο στο οποίο περιέχονται αποθήκευση των στιγμιοτύπων του τρέχοντος στοιχείου.
- Διαφορά βάρους, που αντιστοιχεί στη διαφορά του βάρους των όρων που περιέχονται στα στιγμιότυπα του τρέχοντος στοιχείου, σε σχέση με το βάρος του περιεχομένου του στοιχείου που το περιέχει και των στοιχείων που είναι στο ίδιο επίπεδο με τα στιγμιότυπα του τρέχοντος στοιχείου και είναι στιγμιότυπα άλλων στοιχείων.

Τέλος, ένα Σύνολο Κανόνων Δεικτοδότησης απαρτίζεται από Κανόνες Δεικτοδότησης Εγγράφων, ενώ διαθέτει και το χαρακτηριστικό όνομα που αντιστοιχεί στο όνομα του τρέχοντος συνόλου κανόνων δεικτοδότησης. Κάθε Κανόνας Δεικτοδότησης Εγγράφων διαθέτει τα χαρακτηριστικά όνομα και όνομα στοιχείου, που αντιστοιχούν στο όνομα του κανόνα και στο όνομα του στοιχείου που η παρουσία του στη βάση δεδομένων για τα δεικτοδοτούμενα στιγμιότυπα της κλάσης εγγράφων καθορίζεται από τον τρέχοντα κανόνα.

Ένα παράδειγμα συνόλου κανόνων, που περιγράφει τους κανόνες αναπαράστασης και διαχείρισης για τα έγγραφα που δομούνται με βάση το DTD 6-2 της υποενότητας 6.3.6.3 φαίνεται στο Παράδειγμα 7-1. Στο Παράδειγμα 7-1 θεωρείται ότι για τα δεικτοδοτούμενα στιγμιότυπα της αντίστοιχης κλάσης εγγράφων, εκτός από τα μεταδεδομένα τους, στη βάση δεδομένων της πηγής πληροφορίας αποθηκεύονται και τα στιγμιότυπα της κλάσης σύνοψη.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema-19990303"
  xmlns:KAΔΔΑ="http://irida.ced.tuc.gr/KAΔΔΑ.rdfs">
  <rdf:Description about="">
    <KAΔΔΑ:σύνολο_κανόνων όνομα="Άρθρο.dtd" τύπος="έγγραφο">
      <KAΔΔΑ:σύνολο_κανόνων_μεταδεδομένων όνομα="κανόνες_μδ_άρθρων">
        <KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων
          όνομα="εξαγωγή_τίτλου" αποτέλεσμα="Τίτλος"
          δεδομένα="Τίτλος">
          <KAΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
            παράμετροι="">
          </KAΔΔΑ:μέθοδος>
        </KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων>
        <KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων
          όνομα="εξαγωγή_περιγραφής" αποτέλεσμα="Περιγραφή">
```

```

δεδομένα="Σύνοψη">
<ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
παράμετροι="">
</ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόνες_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόνες_εξαγωγής_μεταδεδομένων
όνομα="εξαγωγή_διαχειριστή" αποτέλεσμα="Διαχειριστής"
δεδομένα="Συγγραφέας">
<ΚΑΔΔΑ:κανόνες_εξαγωγής_μεταδεδομένων
όνομα="εξαγωγή_ρόλου" αποτέλεσμα="Διαχειριστής.Ρόλος"
δεδομένα="Συγγραφέας">
<ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
παράμετροι="">
</ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόνες_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόνες_εξαγωγής_μεταδεδομένων
όνομα="εξαγωγή_ονόματος" αποτέλεσμα="Διαχειριστής.Όνομα"
δεδομένα="Συγγραφέας.Όνομα">
<ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
παράμετροι="">
</ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόνες_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόνες_εξαγωγής_μεταδεδομένων
όνομα="εξαγωγή_επωνύμου" αποτέλεσμα="Διαχειριστής.Επώνυμο"
δεδομένα="Συγγραφέας.Επώνυμο">
<ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
παράμετροι="">
</ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόνες_εξαγωγής_μεταδεδομένων>
</ΚΑΔΔΑ:κανόνες_εξαγωγής_μεταδεδομένων>
</ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων_μεταδεδομένων>
<ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων όνομα="κανόνες_στ_άρθρων">
<ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_εγγράφου
όνομα="κανόνες_αναπαράστασης_στ_άρθρο"
όνομα_στοιχείου="Άρθρο" όριο_διάσπασης="0"
ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_εγγράφου
όνομα="κανόνες_αναπαράστασης_στ_τίτλος"
όνομα_στοιχείου="Τίτλος" όριο_διάσπασης="-1"
ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0,5">
</ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_εγγράφου
όνομα="κανόνες_αναπαράστασης_στ_συγγραφέας"
όνομα_στοιχείου="Συγγραφέας" όριο_διάσπασης="-1"
ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_εγγράφου
όνομα="κανόνες_αναπαράστασης_στ_σύνοψη"
όνομα_στοιχείου="Σύνοψη" όριο_διάσπασης="-1"
ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0,25">
</ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_εγγράφου>

```

```

<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόνας_αναπαράστασης_στ_ενότητα"
  όνομα_στοιχείου="Ενότητα" όριο_διάστασης="200KB"
  ίδιο_όριο_διάστασης="50KB" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόνας_αναπαρ_στ_υποενότητα"
  όνομα_στοιχείου="Υποενότητα" όριο_διάστασης="-1"
  ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόνας_αναπαρ_στ_παράγραφος"
  όνομα_στοιχείου="Παράγραφος" όριο_διάστασης="-1"
  ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόνας_αναπαράστ_στ_em"
  όνομα_στοιχείου="em" όριο_διάστασης="-1"
  ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0,05">
  </ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόνας_αναπαρ_στ_παράγραφος"
  όνομα_στοιχείου="Παράγραφος" όριο_διάστασης="-1"
  ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόνας_αναπαράστασης_στ_em"
  όνομα_στοιχείου="em" όριο_διάστασης="-1"
  ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0,05">
  </ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων>
<ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης όνομα="σ_κ_δ_άρθρο">
  <ΚΑΔΔΑ:κανόνας_δεικτοδότησης_εγγράφων όνομα="σύνοψη"
    όνομα_στοιχείου="σύνοψη">
  </ΚΑΔΔΑ:κανόνας_δεικτοδότησης_εγγράφων>
</ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης>
</ΚΑΔΔΑ:σύνολο_κανόνων>
</rdf:Description>
</rdf:RDF>

```

Παράδειγμα 7-1: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 6-2

7.2.2.2. Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Παραδοσιακών Δεδομένων

Οι Κανόνες Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Παραδοσιακών Δεδομένων περιγράφονται σ' αυτή την υποενότητα. Όπως και για τις κλάσεις εγγράφων, για κάθε κλάση παραδοσιακών δεδομένων δημιουργείται ένα σύνολο κανόνων. Τα Σύνολα Κανόνων

στην περίπτωση των κανόνων αναπαράστασης και διαχείρισης παραδοσιακών δεδομένων απαρτίζονται από ένα Σύνολο Κανόνων Στοιχείων και ένα Σύνολο Κανόνων Δεικτοδότησης.

Όνομα Στοιχείου	Χαρακτηριστικά		
	Όνομα	Τιμή	Περιγραφή
Σύνολο Κανόνων	Τύπος	«Παραδοσιακά Δεδομένα»	Ο τύπος της κλάσης που περιγράφεται από το DTD
	Όνομα	<Όνομα DTD>	Το όνομα του τρέχοντος συνόλου κανόνων
Σύνολο Κανόνων Στοιχείων	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντος συνόλου κανόνων στοιχείων
Κανόνας Αναπαράστασης Στοιχείων Δεδομένων	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντος κανόνα αναπαράστασης στοιχείων
	Όνομα Στοιχείου	<Όνομα Στοιχείου>	Το όνομα του στοιχείου που η αναπαράστασή του καθορίζεται από τον τρέχοντα κανόνα
	Μέθοδος Αναπαράστασης	«Κλάση» / «Χαρακτηριστικό»	Η μέθοδος που χρησιμοποιείται για την αναπαράσταση του τρέχοντος στοιχείου
Κλάση	Όνομα	Συμβολοσειρά	Το όνομα της κλάσης που αναπαριστά το τρέχον στοιχείο
	Υπερκλάση	<Όνομα Υπερκλάσης>	Το όνομα της υπερκλάσης της κλάσης όπου θα ανήκουν τα στιγμιότυπα του τρέχοντος στοιχείου
Χαρακτηριστικό	Όνομα	Συμβολοσειρά	Το όνομα του χαρακτηριστικού που αναπαριστά το τρέχον στοιχείο
	Όνομα Κλάσης	<Όνομα Κλάσης>	Το όνομα της κλάσης τα στιγμιότυπα της οποίας διαθέτουν το χαρακτηριστικό που αναπαριστά το τρέχον στοιχείο
	Τύπος	<Όνομα Τύπου>	Ο τύπος δεδομένων του χαρακτηριστικού που αναπαριστά το τρέχον στοιχείο
Σύνολο Κανόνων Στοιχείων	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντος συνόλου κανόνων στοιχείων
Σύνολο Κανόνων Δεικτοδότησης	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντος συνόλου κανόνων δεικτοδότησης
Κανόνας Δεικτοδότησης Δεδομένων	Όνομα	Συμβολοσειρά	Το όνομα του τρέχοντα κανόνα δεικτοδότησης
	Όνομα Στοιχείου Πληροφορίας	<Όνομα Στοιχείου Πληροφορίας>	Το όνομα του υποστοιχείου ή του χαρακτηριστικού ⁴⁴ που η παρουσία του στη βάση δεδομένων για τα δεικτοδοτούμενα στιγμιότυπα της κλάσης που περιγράφεται καθορίζεται από τον τρέχοντα κανόνα

⁴⁴ Τα ονόματα των χαρακτηριστικών δίνονται με τη μορφή <όνομα κλάσης>:<όνομα χαρακτηριστικού>, όπου <όνομα χαρακτηριστικού> το όνομα του χαρακτηριστικού και <όνομα κλάσης> το όνομα της κλάσης τα στιγμιότυπα της οποίας διαθέτουν το χαρακτηριστικό

Τύπος	«Κλάση» / «Χαρακτηριστικό»	Η αναπαράσταση του στοιχείου πληροφορίας που η παρουσία του στη βάση δεδομένων για τα δεικτοδοτούμενα στιγμιότυπα της κλάσης που περιγράφεται καθορίζεται από τον τρέχοντα κανόνα από τον τρέχοντα κανόνα
-------	----------------------------	---

Πίνακας 7-3: Τα Στοιχεία που απαρτίζουν ένα ΚΑΔΔΑ Σύνολο Κανόνων Αναπαράστασης, Διαχείρισης & Δεικτοδότησης Παραδοσιακών Δεδομένων

Ο Πίνακας 7-3 περιέχει, μαζί με τα χαρακτηριστικά τους, τα στοιχεία που απαρτίζουν τα ΚΑΔΔΑ σύνολα κανόνων αναπαράστασης, διαχείρισης και δεικτοδότησης παραδοσιακών δεδομένων. Όπως φαίνεται στον παραπάνω πίνακα, ένα Σύνολο Κανόνων απαρτίζεται από ένα Σύνολο Κανόνων Στοιχείων και ένα Σύνολο Κανόνων Δεικτοδότησης, ενώ διαθέτει και τα χαρακτηριστικά όνομα και τύπος, που αντιστοιχούν στο όνομα και τον τύπο του τρέχοντος συνόλου κανόνων.

Ένα Σύνολο Κανόνων Στοιχείων απαρτίζεται από Κανόνες Αναπαράστασης Στοιχείων Δεδομένων, ενώ διαθέτει και το χαρακτηριστικό όνομα που αντιστοιχεί στο όνομα του τρέχοντος συνόλου κανόνων στοιχείων. Ένας Κανόνας Αναπαράστασης Στοιχείων Δεδομένων διαθέτει τα χαρακτηριστικά όνομα, όνομα στοιχείου και μέθοδος αναπαράστασης, που αντιστοιχούν στο όνομα του τρέχοντα κανόνα αναπαράστασης στοιχείων, στο όνομα του στοιχείου που η αναπαράστασή του καθορίζεται από τον τρέχοντα κανόνα και στη μέθοδο αναπαράστασης του στοιχείου. Ανάλογα με τη μέθοδο αναπαράστασης που χρησιμοποιείται κάθε φορά, ο κανόνας διαθέτει είτε το υποστοιχείο Κλάση είτε το υποστοιχείο Χαρακτηριστικό:

- Αν η μέθοδος αναπαράστασης έχει την τιμή «Χαρακτηριστικό», το στοιχείο του οποίου η αναπαράσταση καθορίζεται από τον τρέχοντα κανόνα αναπαρίσταται ως χαρακτηριστικό. Στην περίπτωση αυτή ο κανόνας διαθέτει ένα στιγμιότυπο του στοιχείου Χαρακτηριστικό. Τα στιγμιότυπα του στοιχείου Χαρακτηριστικό διαθέτουν τα χαρακτηριστικά όνομα, όνομα κλάσης και τύπος που αντιστοιχούν στο όνομα του χαρακτηριστικού που αναπαριστά το τρέχον στοιχείο, το όνομα της κλάσης τα στιγμιότυπα της οποίας διαθέτουν το χαρακτηριστικό αυτό και τον τύπο του χαρακτηριστικού.
- Αν η μέθοδος αναπαράστασης έχει την τιμή «Κλάση», το στοιχείο του οποίου η αναπαράσταση καθορίζεται από τον τρέχοντα κανόνα αναπαρίσταται ως κλάση. Στην περίπτωση αυτή ο κανόνας διαθέτει ένα στιγμιότυπο του στοιχείου Κλάση. Τα στιγμιότυπα του στοιχείου Κλάση διαθέτουν τα χαρακτηριστικά όνομα και υπερκλάση, που αντιστοιχούν στο όνομα και την υπερκλάση της κλάσης που αναπαριστά το τρέχον στοιχείο.

Κάποιος κανόνας αναπαράστασης στοιχείου δεδομένων που αναπαρίσταται ως κλάση μπορεί να διαθέτει ένα ή περισσότερα στιγμιότυπα του στοιχείου Χαρακτηριστικό, που αντιστοιχούν στα χαρακτηριστικά του στοιχείου που ο κανόνας αναπαριστά.

Ένα Σύνολο Κανόνων Δεικτοδότησης απαρτίζεται από Κανόνες Δεικτοδότησης Δεδομένων, ενώ διαθέτει και το χαρακτηριστικό όνομα που αντιστοιχεί στο όνομα του τρέχοντος συνόλου κανόνων δεικτοδότησης. Κάθε Κανόνας Δεικτοδότησης Δεδομένων διαθέτει τα

χαρακτηριστικά όνομα, όνομα στοιχείου πληροφορίας και τύπος, που αντιστοιχούν στο όνομα του κανόνα και στο όνομα και τον τύπο του στοιχείου πληροφορίας που η παρουσία του στη βάση δεδομένων για τα δεικτοδοτούμενα στιγμιότυπα της περιγραφόμενης κλάσης καθορίζεται από τον τρέχοντα κανόνα.

Ένα παράδειγμα συνόλου κανόνων, που περιγράφει τους κανόνες αναπαράστασης και διαχείρισης για τα αντικείμενα που δομούνται με βάση το DTD 6-3 της ενότητας 6.4.2 φαίνεται στο Παράδειγμα 7-2. Στο Παράδειγμα 7-2 θεωρείται ότι για τα δεικτοδοτούμενα στιγμιότυπα της κλάσης πρόσωπο, στη βάση δεδομένων της πηγής πληροφορίας αποθηκεύονται τα χαρακτηριστικά «όνομα» και «επώνυμο».

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema-19990303"
  xmlns:ΚΑΔΔΑ="http://irida.ced.tuc.gr/RDF/ΚΑΔΔΑ.rdfs">
  <rdf:Description about="">
    <ΚΑΔΔΑ:σύνολο_κανόνων όνομα="Πρόσωπο.dtd"
      τύπος="παραδοσιακά δεδομένα">
      <ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων όνομα="κανόνες_στ_πρόσωπο">
        <ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων
          όνομα="κανόνες_στ_πρόσωπο" όνομα_στοιχείου="πρόσωπο"
          μέθοδος_αναπαράστασης="κλάση">
          <ΚΑΔΔΑ:Κλάση όνομα="πρόσωπο" υπερκλάση="αντικείμενο">
            </ΚΑΔΔΑ:Κλάση>
          <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="ηλικία"
            κλάση="πρόσωπο" τύπος="αριθμός">
            </ΚΑΔΔΑ:Χαρακτηριστικό>
          <ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων
            όνομα="κανόνες_στ_όνομα" όνομα_στοιχείου="όνομα"
            μέθοδος_αναπαράστασης="χαρακτηριστικό">
            <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="όνομα"
              κλάση="πρόσωπο" τύπος="συμβολοσειρά">
              </ΚΑΔΔΑ:Χαρακτηριστικό>
            </ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων>
          <ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων
            όνομα="κανόνες_στ_επώνυμο" όνομα_στοιχείου="επώνυμο"
            μέθοδος_αναπαράστασης="χαρακτηριστικό">
            <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="επώνυμο"
              κλάση="πρόσωπο" τύπος="συμβολοσειρά">
              </ΚΑΔΔΑ:Χαρακτηριστικό>
            </ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων>
          <ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων
            όνομα="κανόνες_στ_τηλέφωνο" όνομα_στοιχείου="τηλέφωνο"
            μέθοδος_αναπαράστασης="χαρακτηριστικό">
            <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="τηλέφωνο"
              κλάση="πρόσωπο" τύπος="συμβολοσειρά">
              </ΚΑΔΔΑ:Χαρακτηριστικό>
            </ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων>
          <ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων
            όνομα="κανόνες_στ_διεύθυνση" όνομα_στοιχείου="διεύθυνση"
            μέθοδος_αναπαράστασης="χαρακτηριστικό">
            <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="διεύθυνση"
```

```

        κλάση="πρόσωπο" τύπος="συμβολοσειρά">
</ΚΑΔΔΑ:Χαρακτηριστικό>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων>
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων
όνομα="κανόνες_στ_e-mail" όνομα_στοιχείου="e-mail"
μέθοδος_αναπαράστασης="χαρακτηριστικό">
<ΚΑΔΔΑ:Χαρακτηριστικό όνομα="όνομα"
κλάση="e-mail" τύπος="συμβολοσειρά">
</ΚΑΔΔΑ:Χαρακτηριστικό>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων>
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων
όνομα="κανόνες_στ_οργανισμός"
όνομα_στοιχείου="οργανισμός"
μέθοδος_αναπαράστασης="χαρακτηριστικό">
<ΚΑΔΔΑ:Χαρακτηριστικό όνομα="οργανισμός"
κλάση="πρόσωπο" τύπος="συμβολοσειρά">
</ΚΑΔΔΑ:Χαρακτηριστικό>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων>
</ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων>
<ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης όνομα="σ_κ_δ_πρόσωπο">
<ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων όνομα="όνομα"
όνομα_στοιχείου_πληροφορίας="πρόσωπο:όνομα"
τύπος="χαρακτηριστικό">
</ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων>
<ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων όνομα="επώνυμο"
όνομα_στοιχείου_πληροφορίας="πρόσωπο:επώνυμο"
τύπος="χαρακτηριστικό">
</ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων>
</ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης>
</ΚΑΔΔΑ:σύνολο_κανόνων>
</rdf:Description>
</rdf:RDF>

```

Παράδειγμα 7-2: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 6-3

7.2.2.3. Σύνταξη της Γλώσσας ΚΑΔΔΑ

Το θέμα αυτής της υποενότητας είναι η σύνταξη της γλώσσας ΚΑΔΔΑ, βάσει της οποίας συντάσσονται οι κανόνες αναπαράστασης, διαχείρισης και δεικτοδότησης δομημένων αντικειμένων. Όπως έχει ήδη αναφερθεί, η γλώσσα ΚΑΔΔΑ είναι μια εφαρμογή της RDF: Στην πραγματικότητα, η σύνταξη της γλώσσας ΚΑΔΔΑ δίνεται από ένα RDF λεξικό, που υλοποιείται από ένα RDF σχήμα. Η σύνταξη της ΚΑΔΔΑ, με τη μορφή RDF λεξικού, παρατίθεται στο *RDF Σχήμα 7-1* και αντιστοιχεί στο URL "http://irida.ced.tuc.gr/RDF/ΚΑΔΔΑ.rdfs".

```

<?xml version='1.0'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema"
xmlns:ΚΑΔΔΑ="">

```

```

<rdfs:Class ID="σύνολο_κανόνων">
  <rdfs:label>σύνολο_κανόνων</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource="http://www.w3.org/TR/PR-rdf-schema#Resource" />
</rdfs:Class>
<rdf:Property ID="όνομα">
  <rdfs:range
    rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#σύνολο_κανόνων" />
  <rdf:domain rdf:resource="#σύνολο_κανόνων_στοιχείων" />
  <rdf:domain rdf:resource="#σύνολο_κανόνων_μεταδεδομένων" />
  <rdf:domain rdf:resource="#σύνολο_κανόνων_δεικτοδότησης" />
  <rdf:domain rdf:resource="#κανόνας_εξαγωγής_μεταδεδομένων" />
  <rdf:domain rdf:resource="#κανόνας_δεικτοδότησης" />
  <rdf:domain rdf:resource="#μέθοδος" />
  <rdf:domain rdf:resource="#κανόνας_αναπαράστασης_στοιχείου" />
  <rdf:domain rdf:resource="#κλάση" />
  <rdf:domain rdf:resource="#χαρακτηριστικό" />
</rdf:Property>
<rdf:Property ID="τύπος_σκ">
  <rdfs:label>τύπος</rdfs:label>
  <rdfs:range rdf:resource="#τύπος_σκ">
  <rdf:domain rdf:resource="#σύνολο_κανόνων" />
</rdf:Property>
<τύπος_σκ rdf:ID="έγγραφο" />
<τύπος_σκ rdf:ID="παραδοσιακά_δεδομένα" />
<rdfs:Class ID="σύνολο_κανόνων_στοιχείων">
  <rdfs:label>σύνολο_κανόνων_στοιχείων</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource="http://www.w3.org/TR/PR-rdf-schema#Resource" />
  <rdf:domain rdf:resource="#σύνολο_κανόνων" />
</rdfs:Class>
<rdfs:Class ID="σύνολο_κανόνων_μεταδεδομένων">
  <rdfs:label>σύνολο_κανόνων_μεταδεδομένων</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource="http://www.w3.org/TR/PR-rdf-schema#Resource" />
  <rdf:domain rdf:resource="#σύνολο_κανόνων" />
</rdfs:Class>
<rdfs:Class ID="κανόνας_εξαγωγής_μεταδεδομένων">
  <rdfs:label>κανόνας_εξαγωγής_μεταδεδομένων</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource="http://www.w3.org/TR/PR-rdf-schema#Resource" />
  <rdf:domain rdf:resource="#σύνολο_κανόνων_μεταδεδομένων" />
  <rdf:domain rdf:resource="#κανόνας_εξαγωγής_μεταδεδομένων" />
</rdfs:Class>
<rdf:Property ID="δεδομένα">
  <rdfs:range
    rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#κανόνας_εξαγωγής_μεταδεδομένων" />
</rdf:Property>
<rdf:Property ID="αποτέλεσμα">

```

```

<rdfs:range
  rdf:resource="http://www.w3.org/datatypes#String" />
<rdf:domain rdf:resource="#κανόνας_εξαγωγής_μεταδεδομένων" />
</rdf:Property>
<rdfs:Class ID="μέθοδος">
  <rdfs:label>μέθοδος</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource=" http://www.w3.org/TR/PR-rdf-schema#Resource" />
  <rdf:domain rdf:resource="#κανόνας_εξαγωγής_μεταδεδομένων" />
</rdfs:Class>
<rdf:Property ID="τελεστής">
  <rdfs:range rdf:resource="#τελεστής" />
  <rdf:domain rdf:resource="#μέθοδος" />
</rdf:Property>
<τελεστής rdf:ID="αντιγραφή" />
<τελεστής rdf:ID="αντιγραφή-τομή" />
<τελεστής rdf:ID="τομή" />
<rdf:Property ID="παράμετροι">
  <rdfs:range
    rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#μέθοδος" />
</rdf:Property>
<rdfs:Class ID="κανόνας_αναπαράστασης_στοιχείου">
  <rdfs:label>κανόνας_αναπαράστασης_στοιχείου</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource=" http://www.w3.org/TR/PR-rdf-schema#Resource" />
  <rdf:domain rdf:resource="#σύνολο_κανόνων_στοιχείων" />
  <rdf:domain rdf:resource="#κανόνας_αναπαράστασης_στοιχείου" />
</rdfs:Class>
<rdfs:Class ID="κανόνας_αναπαράστασης_στοιχείου_εγγράφου">
  <rdfs:label>κανόνας_αναπαράστασης_στοιχείου_εγγράφου</rdfs:label>
  <rdfs:SubclassOf rdf:resource="#κανόνας_αναπαράστασης_στοιχείου" />
  <rdf:domain rdf:resource="#κανόνας_αναπαράστασης_στοιχείου_εγγράφου" />
</rdfs:Class>
<rdf:Property ID="όνομα_στοιχείου">
  <rdfs:range
    rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#κανόνας_αναπαράστασης_στοιχείου" />
</rdf:Property>
<rdf:Property ID="υπερκλάση">
  <rdfs:range
    rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain
    rdf:resource="#κανόνας_αναπαράστασης_στοιχείου_εγγράφου " />
  <rdf:domain rdf:resource="#κλάση" />
</rdf:Property>
<rdf:Property ID="όριο_διάσπασης">
  <rdfs:range rdf:resource="[-1-1]" />
  <rdf:domain
    rdf:resource="#κανόνας_αναπαράστασης_στοιχείου_εγγράφου " />
</rdf:Property>

```

```

<rdf:Property ID="ίδιο_όριο_διάσπασης">
  <rdfs:range rdf:resource="[-1-1]" />
  <rdf:domain
    rdf:resource="#κανόνας_αναπαράστασης_στοιχείου_εγγράφου " />
</rdf:Property>
<rdfs:Class ID="κανόνας_αναπαράστασης_στοιχείου_δεδομένων">
  <rdfs:label>κανόνας_αναπαράστασης_στοιχείου_δεδομένων</rdfs:label>
  <rdfs:SubclassOf rdf:resource="#κανόνας_αναπαράστασης_στοιχείου" />
  <rdf:domain rdf:resource="#κανόνας_αναπαράστασης_στοιχείου_δεδομένων" />
</rdfs:Class>
<rdf:Property ID="μέθοδος_αναπαράστασης">
  <rdfs:range rdf:resource="#μέθοδος_αναπαράστασης" />
  <rdf:domain
    rdf:resource="#κανόνας_αναπαράστασης_στοιχείου_δεδομένων" />
</rdf:Property>
<μέθοδος_αναπαράστασης rdf:ID="κλάση" />
<μέθοδος_αναπαράστασης rdf:ID="χαρακτηριστικό" />
<rdf:Property ID="τύπος">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes" />
  <rdf:domain
    rdf:resource="#κανόνας_αναπαράστασης_στοιχείου_δεδομένων" />
</rdf:Property>
<rdfs:Class ID="κλάση">
  <rdfs:label>κλάση</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource=" http://www.w3.org/TR/PR-rdf-schema#Resource" />
  <rdf:domain
    rdf:resource="#κανόνας_αναπαράστασης_στοιχείου_δεδομένων" />
</rdfs:Class>
<rdfs:Class ID="χαρακτηριστικό">
  <rdfs:label>χαρακτηριστικό</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource=" http://www.w3.org/TR/PR-rdf-schema#Resource" />
  <rdf:domain
    rdf:resource="#κανόνας_αναπαράστασης_στοιχείου_δεδομένων" />
</rdfs:Class>
<rdf:Property ID="όνομα_κλάσης">
  <rdfs:range
    rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#χαρακτηριστικό" />
</rdf:Property>
<rdfs:Class ID="σύνολο_κανόνων_δεικτοδότησης">
  <rdfs:label>σύνολο_κανόνων_δεικτοδότησης</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource="http://www.w3.org/TR/PR-rdf-schema#Resource" />
  <rdf:domain rdf:resource="#σύνολο_κανόνων" />
</rdfs:Class>
<rdfs:Class ID="κανόνας_δεικτοδότησης">
  <rdfs:label>κανόνας_δεικτοδότησης</rdfs:label>
  <rdfs:SubclassOf
    rdf:resource=" http://www.w3.org/TR/PR-rdf-schema#Resource" />

```

```

<rdf:domain rdf:resource="#σύνολο_κανόνων_δεικτοδότησης" />
</rdfs:Class>
<rdf:Property ID="όνομα_στοιχείου_πληροφορίας">
  <rdfs:range
    rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#κανόνες_δεικτοδότησης" />
</rdf:Property>
<rdfs:Class ID="κανόνες_δεικτοδότησης_εγγράφων">
  <rdfs:label>κανόνες_δεικτοδότησης_εγγράφων</rdfs:label>
  <rdfs:SubclassOf rdf:resource="#σύνολο_κανόνων_δεικτοδότησης" />
</rdfs:Class>
<rdfs:Class ID="κανόνες_δεικτοδότησης_δεδομένων">
  <rdfs:label>κανόνες_δεικτοδότησης_δεδομένων</rdfs:label>
  <rdfs:SubclassOf rdf:resource="#σύνολο_κανόνων_δεικτοδότησης" />
</rdfs:Class>
<rdf:Property ID="τύπος_σκδδ">
  <rdfs:range rdf:resource="#μέθοδος_αναπαράστασης" />
  <rdf:domain rdf:resource="#κανόνες_δεικτοδότησης" />
</rdf:Property>
</rdf:RDF>

```

RDF Σχήμα 7-1: Η Σύνταξη της Γλώσσας ΚΑΔΔΑ

7.2.3. Δομή των Κλάσεων

Η δομή των κλάσεων που υποστηρίζονται από τη βάση δεδομένων μιας πηγής πληροφορίας, όπως αυτή εξάγεται από τη γλώσσα της πηγής πληροφορίας και αναπαρίσταται στο μοντέλο πληροφορίας της πηγής πληροφορίας, είναι το θέμα αυτής της ενότητας. Στο σημείο αυτό πρέπει να σημειωθεί ότι η δομή μιας κλάσης, όπως αυτή εξάγεται από το μοντέλο πληροφορίας της πηγής πληροφορίας, μπορεί να διαφέρει από την αναπαράσταση των στιγμιότυπων της στη βάση δεδομένων της πηγής πληροφορίας: Για παράδειγμα, τα στιγμιότυπα μιας κλάσης δομικών στοιχείων εγγράφων μπορεί να αποθηκεύονται πάντοτε μαζί με το έγγραφο ή το δομικό στοιχείο στο οποίο περιέχονται, η κλάση αυτή όμως μπορεί να εξάγεται στο μοντέλο πληροφορίας της πηγής πληροφορίας πανομοιότυπα με τις κλάσεις δομικών στοιχείων εγγράφων τα στιγμιότυπα των οποίων αποθηκεύονται ανεξάρτητα από το έγγραφο ή το δομικό στοιχείο στο οποίο περιέχονται.

Όπως έχει ήδη αναφερθεί, η δομή μιας κλάσης αποτελεί στην πραγματικότητα ένα RDF σχήμα, το οποίο αποθηκεύεται στο πεδίο Δομή του στιγμιότυπου που αναπαριστά την τρέχουσα κλάση. Το RDF σχήμα που περιγράφει τη δομή μιας κλάσης παράγεται από το εργαλείο SupportDTD, μετά την αλληλεπίδραση του εργαλείου με το διαχειριστή της πηγής πληροφορίας και με βάση το DTD στον οποίο υπακούει η σύνταξη των στιγμιότυπων της και τους κανόνες αναπαράστασης και διαχείρισης που αντιστοιχούν στο DTD αυτό.

Έτσι, η παραγωγή του RDF σχήματος που περιγράφει τη δομή μιας κλάσης εγγράφων και των κλάσεων δομικών στοιχείων εγγράφων που τα στιγμιότυπά τους απαρτίζουν τα στιγμιότυπα της κλάσης εγγράφων γίνεται ως εξής:

- Το στοιχείο που αντιστοιχεί στην κλάση εγγράφων αναπαρίσταται από την ομώνυμη RDF κλάση, που είναι υποκλάση της κλάσης Έγγραφο (ή κάποιας υποκλάσης της).
- Τα στοιχεία που αντιστοιχούν στις κλάσεις δομικών στοιχείων εγγράφων που από τα στιγμιότυπά τους απαρτίζονται τα στιγμιότυπα της κλάσης εγγράφων, αν επιτρέπεται να εξαχθούν στο μοντέλο πληροφορίας, αναπαρίστανται από τις ομώνυμες RDF κλάσεις, που είναι υποκλάσεις της κλάσης Δομικό Στοιχείο Εγγράφου (ή κάποιας υποκλάσης της). Οι κλάσεις αυτές έχουν πεδίο (domain) την περιγραφόμενη στο DTD κλάση εγγράφων και δημιουργούνται ανεξάρτητα από την πολιτική αναπαράστασης και διαχείρισης των στιγμιότυπων τους.
- Κάθε χαρακτηριστικό στοιχείου αναπαρίσταται από την ομώνυμη RDF ιδιότητα (RDF property). Οι ιδιότητες αυτές έχουν πεδίο (domain) την κλάση (ή τις κλάσεις) που τα στιγμιότυπά τους διαθέτουν τα ομώνυμα χαρακτηριστικά.
- Κάθε στοιχείο μεταδεδομένων για το οποίο έχει οριστεί κάποιος κανόνας εξαγωγής αναπαρίσταται από την ομώνυμη RDF ιδιότητα (RDF property). Οι ιδιότητες αυτές έχουν πεδίο (domain) την κλάση του εγγράφου.

Μ' αυτό τον τρόπο, το RDF σχήμα που αντιστοιχεί στο DTD 6-2 της υποενότητας 6.3.6.3, αν υποτεθεί ότι έχει επιλεγεί να μην εξάγονται στο μοντέλο πληροφορίας οι κλάσεις που αντιστοιχούν στα στοιχεία Τίτλος και em φαίνεται στο RDF Σχήμα 7-2 και αντιστοιχεί στο URL "http://irida.ced.tuc.gr/RDF/Άρθρο.rdfs".

```
<?xml version='1.0'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema"
  xmlns:rdfs="http://irida.ced.tuc.gr/RDF/ΠηγήΠληροφορίαςΕγγράφων.rdfs"
  xmlns:Άρθρο="">
  <rdfs:Class ID="Άρθρο">
    <rdfs:label>Άρθρο</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
      ΠηγήΠληροφορίαςΕγγράφων.rdfs#έγγραφο"/>
  </rdfs:Class>
  <rdf:Property ID="τίτλος">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#Άρθρο" />
  </rdf:Property>
  <rdf:Property ID="περιγραφή">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#Άρθρο" />
  </rdf:Property>
  <rdfs:Class ID="συγγραφέας">
    <rdfs:label>συγγραφέας</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
      ΠηγήΠληροφορίαςΕγγράφων.rdfs#αντικείμενο"/>
    <rdfs:domain rdf:resource="#Άρθρο" />
  </rdfs:Class>
  <rdf:Property ID="όνομα">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#συγγραφέας" />
  </rdf:Property>
```

```

</rdf:Property>
<rdf:Property ID="επώνυμο">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#συγγραφέας" />
</rdf:Property>
<rdfs:Class ID="σύννοψη">
  <rdfs:label>σύννοψη</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
  ΠηγήΠληροφορίαςΕγγράφων.rdfs#τμήμα_εγγράφου"/>
  <rdf:domain rdf:resource="#άρθρο" />
</rdfs:Class>
<rdfs:Class ID="ενότητα">
  <rdfs:label>ενότητα</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
  ΠηγήΠληροφορίαςΕγγράφων.rdfs#τμήμα_εγγράφου"/>
  <rdf:domain rdf:resource="#άρθρο" />
</rdfs:Class>
<rdfs:Class ID="υποενότητα">
  <rdfs:label>υποενότητα</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
  ΠηγήΠληροφορίαςΕγγράφων.rdfs#τμήμα_εγγράφου"/>
  <rdf:domain rdf:resource="#ενότητα" />
</rdfs:Class>
<rdfs:Class ID="παράγραφος">
  <rdfs:label>παράγραφος</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
  ΠηγήΠληροφορίαςΕγγράφων.rdfs#τμήμα_εγγράφου"/>
  <rdf:domain rdf:resource="#ενότητα" />
  <rdf:domain rdf:resource="#υποενότητα" />
</rdfs:Class>
</rdf:RDF>

```

RDF Σχήμα 7-2: Το RDF Σχήμα που αντιστοιχεί στο DTD 6-2

Αντίστοιχα, η παραγωγή του RDF σχήματος που περιγράφει τη δομή κάποιας κλάσης παραδοσιακών δεδομένων γίνεται ως εξής:

- Για καθένα από τα στοιχεία του DTD που αναπαρίστανται ως κλάσεις, δημιουργείται η ομώνυμη RDF κλάση, που είναι υποκλάση της κλάσης Αντικείμενο (ή κάποιας υποκλάσης της).
- Οι κλάσεις που τα στιγμιότυπά τους αποτελούν τμήματα των στιγμιότυπων κάποιας άλλης κλάσης έχουν πεδίο (domain) την κλάση στα στιγμιότυπα της οποίας περιέχονται τα στιγμιότυπα των κλάσεων αυτών.
- Κάθε χαρακτηριστικό στοιχείου αναπαρίσταται από την ομώνυμη RDF ιδιότητα (RDF property). Οι ιδιότητες αυτές έχουν πεδίο (domain) την κλάση (ή τις κλάσεις) που τα στιγμιότυπά τους διαθέτουν τα ομώνυμα χαρακτηριστικά.

Με βάση τα παραπάνω, το RDF σχήμα που περιγράφει τη δομή της κλάσης Πρόσωπο που περιγράφεται στο DTD 6-3 φαίνεται στο RDF Σχήμα 7-3 και αντιστοιχεί στο URL "http://irida.ced.tuc.gr/RDF/Πρόσωπο.rdfs".

```

<?xml version='1.0'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema"
  xmlns:rdfs="http://irida.ced.tuc.gr/RDF/ΠηγήΠληροφορίαςΔεδομένων.rdfs"
  xmlns:Πρόσωπο=""/>
<rdfs:Class ID="Πρόσωπο">
  <rdfs:label>Πρόσωπο</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
    ΠηγήΠληροφορίαςΔεδομένων.rdfs#αντικείμενο"/>
</rdfs:Class>
<rdf:Property ID="ηλικία">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#Number" />
  <rdf:domain rdf:resource="#Πρόσωπο" />
</rdf:Property>
<rdf:Property ID="όνομα">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#Πρόσωπο" />
</rdf:Property>
<rdf:Property ID="επώνυμο">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#Πρόσωπο" />
</rdf:Property>
<rdf:Property ID="τηλέφωνο">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#Πρόσωπο" />
</rdf:Property>
<rdf:Property ID="διεύθυνση">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#Πρόσωπο" />
</rdf:Property>
<rdf:Property ID="e-mail">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#Πρόσωπο" />
</rdf:Property>
<rdf:Property ID="οργανισμός">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#Πρόσωπο" />
</rdf:Property>
</rdf:RDF>

```

RDF Σχήμα 7-3: RDF σχήμα που περιγράφει τη Δομή της Κλάσης Πρόσωπο

7.3. Εισαγωγή Αντικειμένων

Στο τμήμα αυτό θα περιγραφεί η υποστήριξη της εισαγωγής αντικειμένων στις σχεσιακές βάσεις δεδομένων που υποστηρίζουν τις πηγές πληροφορίας. Η εισαγωγή αντικειμένων στις βάσεις δεδομένων τόσο των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα όσο και των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα μπορεί να γίνει

σε διαφορετικές φάσεις από προγράμματα που έχουν διαφορετικό επίπεδο διεπαφής. Αυτό συμβαίνει καθώς η εισαγωγή αντικειμένων σε μια πηγή πληροφορίας μπορεί να γίνει:

- Από προνομιούχους απομακρυσμένους χρήστες που έχουν τη δυνατότητα να συνεισφέρουν υλικό στις πηγές πληροφορίας, με τη χρήση ενός επιπέδου διεπαφής που είναι διαθέσιμο στον Παγκόσμιο Ιστό.
- Από τους διαχειριστές των πηγών πληροφορίας και τους προνομιούχους τοπικούς χρήστες που έχουν τη δυνατότητα εισαγωγής αντικειμένων. Αυτή η κατηγορία χρηστών χρησιμοποιεί ένα παραθυρικό επίπεδο διεπαφής στους υπολογιστές που βρίσκονται στο ίδιο τοπικό δίκτυο με τον υπολογιστή όπου έχει στηθεί η πηγή πληροφορίας.
- Από προγράμματα, όπως οι πράκτορες, που ενημερώνουν τις πηγές πληροφορίας για αντικείμενα που γίνονται διαθέσιμα και σχετίζονται με το περιεχόμενο των συγκεκριμένων πηγών πληροφορίας.

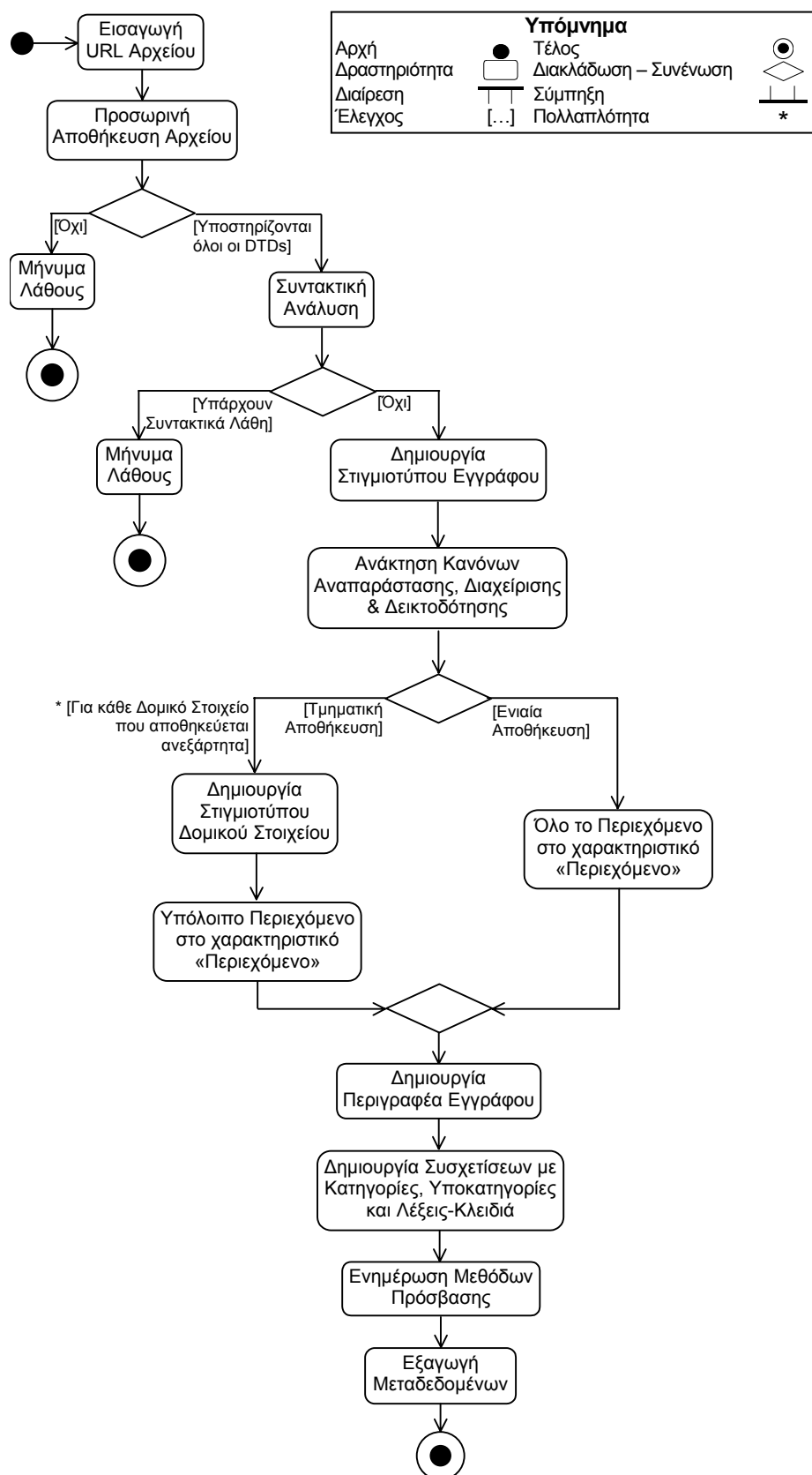
Για να γίνει δυνατή η υποστήριξη όλων των παραπάνω εφαρμογών, έχουν υλοποιηθεί μηχανισμοί που επιτρέπουν την εισαγωγή στις πηγές πληροφορίας αντικειμένων που είναι σε μορφή XML. Οι μηχανισμοί αυτοί μπορούν να συνεργαστούν τόσο με προγράμματα όσο και με επίπεδα διεπαφής. Η λειτουργία των μηχανισμών που υποστηρίζουν την εισαγωγή στις πηγές πληροφορίας αντικειμένων που είναι σε μορφή XML περιγράφεται στην ενότητα 7.3.1. Τέλος, στην ενότητα 7.3.2 γίνεται αναφορά στις εφαρμογές που υποστηρίζουν την εισαγωγή αντικειμένων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

7.3.1. Λειτουργία των Μηχανισμών Εισαγωγής Αντικειμένων

Το θέμα αυτής της ενότητας είναι η λειτουργία των μηχανισμών που υποστηρίζουν την εισαγωγή αντικειμένων που είναι σε μορφή XML στις σχεσιακές βάσεις δεδομένων που υποστηρίζουν τόσο τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα όσο και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Η λειτουργία του μηχανισμού που υποστηρίζει την εισαγωγή δομημένων εγγράφων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα απεικονίζεται στο Διάγραμμα 7-6, που είναι UML διάγραμμα δραστηριοτήτων.

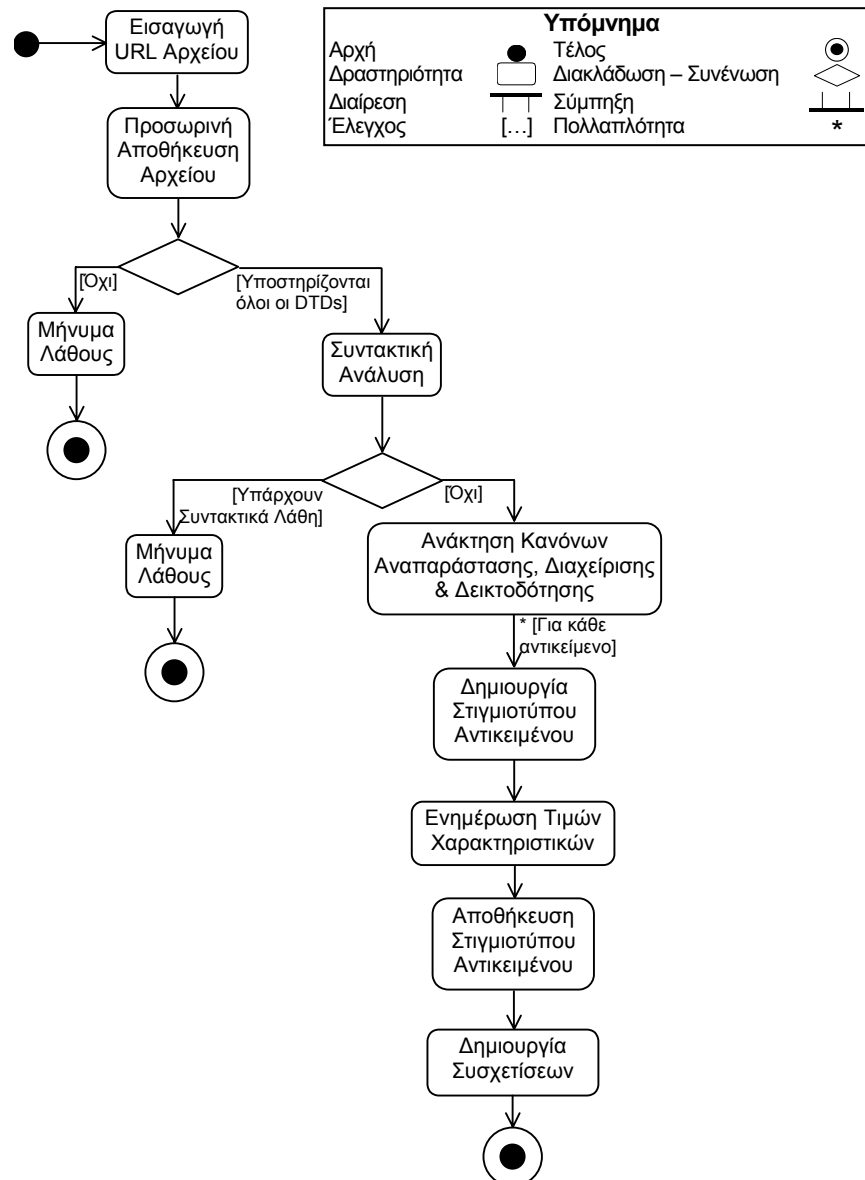
Όπως φαίνεται στο Διάγραμμα 7-6, ο μηχανισμός που υποστηρίζει την εισαγωγή δομημένων εγγράφων στις βάσεις δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα λειτουργεί με βάση τον εξής αλγόριθμο:



Διάγραμμα 7-6: Διάγραμμα Δραστηριοτήτων του Μηχανισμού που υποστηρίζει την Εισαγωγή Δομημένων Εγγράφων στις Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα

1. Δέχεται ως είσοδο το URL του αρχείου όπου περιέχεται σε XML μορφή το έγγραφο που πρόκειται να εισαχθεί.

2. Το αρχείο αποθηκεύεται προσωρινά στο δίσκο του υπολογιστή όπου έχει στηθεί η πηγή πληροφορίας.
3. Ελέγχεται αν οι DTDs που περιγράφουν τη δομή του εγγράφου υποστηρίζονται από την πηγή πληροφορίας.
 - 3.1. Αν κάποιος DTD δεν υποστηρίζεται από την πηγή πληροφορίας, επιστρέφεται ο κωδικός του μηνύματος λάθους "Δεν υποστηρίζεται ο DTD <DTD_URL>" και η διαδικασία σταματά.
 - 3.2. Αν όλοι οι DTDs υποστηρίζονται από την πηγή πληροφορίας γίνονται τα εξής:
 - 3.2.1. Γίνεται συντακτικός έλεγχος των περιεχομένων του αρχείου.
 - 3.2.1.1. Αν υπάρχουν συντακτικά λάθη, επιστρέφεται ο κωδικός του μηνύματος λάθους που περιγράφει το συντακτικό λάθος και η διαδικασία σταματά.
 - 3.2.1.2. Αν δεν υπάρχουν συντακτικά λάθη γίνονται τα εξής:
 - 3.2.1.2.1. Δημιουργείται ένα στιγμιότυπο της κλάσης εγγράφων στην οποία θα ενταχθεί το έγγραφο που περιέχεται στο αρχείο. Το στιγμιότυπο αυτό αναπαριστά το τρέχον έγγραφο.
 - 3.2.1.2.2. Ανακτώνται οι κανόνες αναπαράστασης και διαχείρισης της κλάσης εγγράφων στην οποία το τρέχον στιγμιότυπο θα ενταχθεί.
 - 3.2.1.2.2.1. Αν το έγγραφο πρέπει να αποθηκευτεί τμηματικά γίνονται τα εξής:
 - 3.2.1.2.2.1.1. Δημιουργείται, για κάθε δομικό στοιχείο του εγγράφου που αποθηκεύεται ανεξάρτητα, ένα στιγμιότυπο της αντίστοιχης κλάσης δομικών στοιχείων, το οποίο και το αναπαριστά.
 - 3.2.1.2.2.1.2. Το υπόλοιπο του περιεχόμενου του περιέχοντος (δομικού στοιχείου ή εγγράφου) αποθηκεύεται στο πεδίο «Περιεχόμενο» του περιέχοντος.
 - 3.2.1.2.2.2. Διαφορετικά, όλο το περιεχόμενο του εγγράφου αποθηκεύεται στο πεδίο «Περιεχόμενο».
 - 3.2.1.2.3. Δημιουργείται και ενημερώνεται ο περιγραφέας του εγγράφου.
 - 3.2.1.2.4. Με βάση το περιεχόμενο του εγγράφου και των δομικών στοιχείων από τα οποία απαρτίζεται, δημιουργούνται συσχετίσεις του εγγράφου και των δομικών στοιχείων με κατηγορίες, υποκατηγορίες και λέξεις-κλειδιά.
 - 3.2.1.2.5. Ενημερώνονται οι δομές που υποστηρίζουν τις μεθόδους πρόσβασης.
 - 3.2.1.2.6. Ενημερώνονται οι τιμές των χαρακτηριστικών του εγγράφου που αναπαριστούν χαρακτηριστικά της Dublin Core μετά την κατάλληλη επεξεργασία των αντίστοιχων στοιχείων του εγγράφου (αν αυτά υπάρχουν και είναι διαθέσιμα στο συγκεκριμένο έγγραφο).



Διάγραμμα 7-7: Διάγραμμα Δραστηριοτήτων του Μηχανισμού που υποστηρίζει την Εισαγωγή Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

Αντίστοιχα με το μηχανισμό που υποστηρίζει την εισαγωγή δομημένων εγγράφων στις βάσεις δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, ο μηχανισμός που υποστηρίζει την εισαγωγή αντικειμένων στις βάσεις δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα λειτουργεί ως εξής:

1. Δέχεται ως είσοδο το URL του αρχείου όπου περιέχονται σε XML μορφή τα αντικείμενα που πρόκειται να εισαχθούν.
2. Το αρχείο αποθηκεύεται προσωρινά στο δίσκο του υπολογιστή όπου έχει στηθεί η πηγή πληροφορίας.
3. Ελέγχεται αν οι DTDs που περιγράφουν τα αντικείμενα υποστηρίζονται από την πηγή πληροφορίας.
 - 3.1. Αν κάποιος DTD δεν υποστηρίζεται από την πηγή πληροφορίας, επιστρέφεται ο κωδικός του μηνύματος λάθους "Δεν υποστηρίζεται ο DTD <DTD_URL>" και η διαδικασία σταματά.
 - 3.2. Αν όλοι οι DTDs υποστηρίζονται από την πηγή πληροφορίας γίνονται τα εξής:

3.2.1. Γίνεται συντακτικός έλεγχος των περιεχομένων του αρχείου.

3.2.1.1. Αν υπάρχουν συντακτικά λάθη, επιστρέφεται ο κωδικός του μηνύματος λάθους που περιγράφει το συντακτικό λάθος και η διαδικασία σταματά.

3.2.1.2. Αν δεν υπάρχουν συντακτικά λάθη γίνονται τα εξής:

3.2.1.2.1. Ανακτώνται οι κανόνες αναπαράστασης και διαχείρισης της κλάσης στην οποία εντάσσεται το τρέχον αντικείμενο.

3.2.1.2.2. Για κάθε αντικείμενο (το κυρίως αντικείμενο ή κάποιο αντικείμενο – υποστοιχείο του) γίνονται τα εξής:

3.2.1.2.2.1. Δημιουργείται ένα στιγμιότυπο της κλάσης στην οποία εντάσσεται το τρέχον αντικείμενο.

3.2.1.2.2.2. Ενημερώνονται οι τιμές των χαρακτηριστικών του στιγμιότυπου.

3.2.1.2.2.3. Εισάγεται στη βάση δεδομένων της πηγής πληροφορίας ένα στιγμιότυπο της αντίστοιχης κλάσης.

3.2.1.2.2.4. Δημιουργούνται οι συσχετίσεις του αντικειμένου με τα υπόλοιπα περιεχόμενα της βάσης δεδομένων.

Η λειτουργία του μηχανισμού που υποστηρίζει την εισαγωγή αντικειμένων στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα απεικονίζεται στο Διάγραμμα 7-7, που είναι UML διάγραμμα δραστηριοτήτων.

7.3.2. Εφαρμογές για την Εισαγωγή Αντικειμένων

Σ' αυτή την ενότητα παρουσιάζονται οι εφαρμογές που χρησιμοποιούνται για την εισαγωγή αντικειμένων στις πηγές πληροφορίας. Οι εφαρμογές αυτές χρησιμοποιούν τους μηχανισμούς που υποστηρίζουν την εισαγωγή αντικειμένων που περιγράφηκαν στην ενότητα 7.3.1.

Όπως έχει ήδη αναφερθεί, τα εργαλεία που χρησιμοποιούνται για την εισαγωγή αντικειμένων στις πηγές πληροφορίας εντάσσονται στις εξής τρεις κατηγορίες:

- *Εργαλεία που επιτρέπουν την εισαγωγή αντικειμένων από προνομιούχους απομακρυσμένους χρήστες μέσω του Παγκοσμίου Ιστού.* Το επίπεδο διεπαφής των εργαλείων αυτών απαρτίζεται από φόρμες που παρουσιάζονται από τα προγράμματα πλοήγησης του Παγκοσμίου Ιστού. Μέσω των φορμών αυτών επιτρέπεται η αποστολή, σε μορφή πραγματικών ή ιδεατών XML αρχείων⁴⁵, της πληροφορίας που εισάγεται στη βάση δεδομένων της πηγής πληροφορίας. Η πληροφορία αυτή στέλνεται στο κατάλληλο πρόγραμμα⁴⁶, το οποίο αναλαμβάνει την κλήση του μηχανισμού που είναι υπεύθυνος για την εισαγωγή αντικειμένων στη βάση δεδομένων της πηγής πληροφορίας.

⁴⁵ Αν στη φόρμα δίνεται ο URL του αρχείου που περιέχει την πληροφορία που πρόκειται να εισαχθεί, τότε το αρχείο είναι πραγματικό. Αν όμως στη φόρμα δίνονται τα δεδομένα που απαρτίζουν τα προς εισαγωγή αντικείμενα (είτε σε XML μορφή είτε ως μη μορφοποιημένα δεδομένα), το αρχείο είναι ιδεατό.

⁴⁶ Συνήθως κάποιο CGI script ή κάποια ASP σελίδα.

- *Εργαλεία που επιτρέπουν την εισαγωγή αντικειμένων από τοπικούς χρήστες και το διαχειριστή της πηγής πληροφορίας μέσω ενός παραθυρικού συστήματος διεπαφής. Το παραθυρικό επίπεδο διεπαφής των εργαλείων αυτών επικοινωνεί άμεσα με το μηχανισμό που είναι υπεύθυνος για την εισαγωγή αντικειμένων στη βάση δεδομένων της πηγής πληροφορίας. Η πληροφορία και στην περίπτωση αυτή έχει τη μορφή πραγματικών ή ιδεατών XML αρχείων.*
- *Προγράμματα που είναι υπεύθυνα για την ενημέρωση των πηγών πληροφορίας για αντικείμενα που γίνονται διαθέσιμα και το περιεχόμενό τους σχετίζεται με αυτό των πηγών πληροφορίας. Παράδειγμα προγραμμάτων αυτής της κατηγορίας είναι οι πράκτορες. Τα προγράμματα αυτά, όταν εντοπίσουν κάποιο αντικείμενο που σχετίζεται με το περιεχόμενο των πηγών πληροφορίας για την ενημέρωση των οποίων είναι υπεύθυνα, αποφασίζουν αν θα εισαχθεί εξ' ολοκλήρου ή τμηματικά το αντικείμενο αυτό στην πηγή πληροφορίας ή αν απλώς θα δεικτοδοτηθεί. Στην περίπτωση της (ολικής ή μερικής) εισαγωγής του αντικειμένου, καλείται ο μηχανισμός που είναι υπεύθυνος για την εισαγωγή αντικειμένων στη βάση δεδομένων της πηγής πληροφορίας. Περισσότερες πληροφορίες για τα προγράμματα αυτής της κατηγορίας υπάρχουν στο τμήμα 7.6.*

7.4. Αναζήτηση & Ανάκτηση Αντικειμένων

Το θέμα αυτού του τμήματος είναι η υποστήριξη της αναζήτησης και της ανάκτησης αντικειμένων και τμημάτων αντικειμένων τόσο από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα όσο και από τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Στο σημείο αυτό πρέπει να τονιστεί ότι η παρούσα διατριβή δεν ασχολήθηκε με την ανάπτυξη μεθόδων πρόσβασης ούτε με το σχεδιασμό και την υλοποίηση μεθοδολογιών υποστήριξης ερωτήσεων ομοιότητας. Τα παραπάνω θεωρούνται δεδομένα από την παρούσα διατριβή, η οποία επικεντρώνει, σε ότι αφορά την αναζήτηση και την ανάκτηση αντικειμένων, στην υποστήριξη ερωτήσεων μέσω των γλωσσών των πηγών πληροφορίας.

Ο μηχανισμός μέσω του οποίου πραγματοποιούνται οι αναζητήσεις και ικανοποιούνται οι αιτήσεις σε μια πηγή πληροφορίας είναι η γλώσσα της πηγής πληροφορίας. Πιο συγκεκριμένα, ο επεξεργαστής της γλώσσας της πηγής πληροφορίας αναλαμβάνει κάθε φορά να κάνει τις κατάλληλες ενέργειες ώστε να εκτελεστεί η αίτηση ή η ερώτηση.

Στη συνέχεια, στην ενότητα 7.4.1 γίνεται εκτενέστερη αναφορά σε *Αναζήτηση & Ανάκτηση Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα*, ενώ στην ενότητα 7.4.2 περιγράφεται η *Αναζήτηση & Ανάκτηση Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα*.

7.4.1. Αναζήτηση & Ανάκτηση Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Το θέμα αυτής της ενότητας είναι η αναζήτηση και η ανάκτηση αντικειμένων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα. Όπως έχει ήδη αναφερθεί, ο μηχανισμός

μέσω του οποίου πραγματοποιούνται οι αναζητήσεις και ικανοποιούνται οι αιτήσεις σε μια πηγή πληροφορίας είναι η γλώσσα της πηγής πληροφορίας, ενώ ο επεξεργαστής της γλώσσας της πηγής πληροφορίας αναλαμβάνει κάθε φορά να κάνει τις κατάλληλες ενέργειες ώστε να εκτελεστεί η αίτηση ή η ερώτηση.

Έτσι, στην περίπτωση των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, ο επεξεργαστής της γλώσσας της πηγής πληροφορίας αντιμετωπίζει τις εξής απλές περιπτώσεις:

- Αν πρόκειται για ερώτηση ή αίτηση που περιλαμβάνει μόνο κριτήρια που περιέχουν συνθήκες δίτιμης λογικής για τα χαρακτηριστικά και/ή τη δομή εγγράφων ή τμημάτων εγγράφων, εκτελούνται οι XQL μέθοδοι και το αποτέλεσμα που προκύπτει μετατρέπεται σε SQL μορφή μέσω του επιπέδου διεπαφής με XML έγγραφα. Η ερώτηση αυτή εκτελείται στη βάση δεδομένων της πηγής πληροφορίας.
- Αν πρόκειται για ερώτηση που περιλαμβάνει τον εντοπισμό εγγράφων και/ή τμημάτων εγγράφων βάσει του περιεχομένου τους, με κριτήρια ομοιότητας και/ή δίτιμης λογικής, χρησιμοποιούνται οι μέθοδοι πρόσβασης που αναπτύχθηκαν στα πλαίσια της διατριβής [Πουτ2000]. Αφού τα αντικείμενα εντοπιστούν, μέσω του επιπέδου διεπαφής αποστέλλονται αιτήσεις ανάκτησης των χαρακτηριστικών και/ή του περιεχομένου τους.

Ανάλογα με την ερώτηση που κάθε φορά εκτελείται, ο επεξεργαστής της γλώσσας της πηγής πληροφορίας μπορεί να συνδυάσει τις παραπάνω ενέργειες ως εξής: Αν απαιτείται η χρήση των μεθόδων πρόσβασης, αυτή λαμβάνει χώρα πρώτη έτσι ώστε να επιστρέψει τα αντικείμενα στα οποία θα εφαρμοστεί η ερώτηση που τα κριτήριά της σχετίζονται με τα χαρακτηριστικά και/ή τη δομή των αντικειμένων αυτών.

Σε κάθε περίπτωση, το αποτέλεσμα της ερώτησης ή της αίτησης επιστρέφεται αφού μετατραπεί σε XML μορφή σύμφωνα με το DTD 6-1, που περιγράφει τη δομή της πληροφορίας που επιστρέφεται, ως αποτέλεσμα ερωτήσεων ή αιτήσεων, από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα.

7.4.2. Αναζήτηση & Ανάκτηση Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

Το θέμα αυτής της ενότητας είναι η αναζήτηση και η ανάκτηση αντικειμένων στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Όπως και για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, ο μηχανισμός μέσω του οποίου πραγματοποιούνται οι αναζητήσεις και ικανοποιούνται οι αιτήσεις σε μια πηγή πληροφορίας που περιέχει παραδοσιακά δεδομένα είναι η γλώσσα της πηγής πληροφορίας, ενώ ο επεξεργαστής της γλώσσας της πηγής πληροφορίας αναλαμβάνει κάθε φορά να κάνει τις κατάλληλες ενέργειες ώστε να εκτελεστεί η αίτηση ή η ερώτηση.

Έτσι, η πρώτη ενέργεια που εκτελείται είναι, στην περίπτωση που κάποια ερώτηση ή αίτηση αναφέρεται σε κάποια κλάση και τις υποκλάσεις της, η αντικατάσταση του ονόματος της κλάσης με μια λίστα ονομάτων: Η λίστα αυτή περιέχει τα ονόματα της κλάσης και των υποκλάσεών της. Στη συνέχεια, η ερώτηση ή η αίτηση υποβάλλεται στη *Μηχανή Αναζήτησης (Retrieval Engine)* που αναπτύχθηκε στα πλαίσια της διατριβής [Μουμ97] και της διπλωματικής εργασίας [Ανεσ97]. Η Μηχανή Αναζήτησης έχει τη δυνατότητα

υποστήριξης ερωτήσεων που περιέχουν όρους ομοιότητας και/ή δίτιμης λογικής για παραδοσιακά δεδομένα

Τέλος, το αποτέλεσμα της ερώτησης ή της αίτησης που επιστρέφεται από τη Μηχανή Αναζήτησης προωθείται στο περιτύλιγμα της πηγής πληροφορίας αφού μετατραπεί σε XML μορφή σύμφωνη με τον DTD 6-1, που περιγράφει τη δομή της πληροφορίας που επιστρέφεται, ως αποτέλεσμα ερωτήσεων ή αιτήσεων, από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα.

7.5. Δεικτοδότηση Απομακρυσμένων Αντικειμένων

Στο τμήμα αυτό περιγράφονται οι μηχανισμοί που υποστηρίζουν τη δεικτοδότηση απομακρυσμένων αντικειμένων από τις πηγές πληροφορίας. Οι μηχανισμοί αυτοί είναι υπεύθυνοι για την ενημέρωση των δομών δεικτοδότησης για τα δεικτοδοτούμενα αντικείμενα και για την αντιγραφή, εφ' όσον δεν υπάρχουν προβλήματα σχετικά με τα πνευματικά δικαιώματα του περιεχομένου των αντικειμένων, στις πηγές πληροφορίας των χαρακτηριστικών και των τμημάτων των αντικειμένων που καθορίζεται, με βάση τους κανόνες δεικτοδότησης για την κλάση στην οποία ανήκει κάθε αντικείμενο, ότι θα αποθηκεύονται στην πηγή πληροφορίας για τα δεικτοδοτούμενα στιγμιότυπα της κλάσης.

Οι μηχανισμοί που υποστηρίζουν τη δεικτοδότηση απομακρυσμένων αντικειμένων στις πηγές πληροφορίας χρησιμοποιούνται κυρίως από τους πράκτορες, για τα αντικείμενα που το περιεχόμενό τους σχετίζεται με το περιεχόμενο πηγών πληροφορίας για την ενημέρωση των οποίων υπεύθυνοι είναι οι πράκτορες, αλλά όχι αρκετά ώστε τα αντικείμενα να εισαχθούν εξ' ολοκλήρου σε αυτές.

Καθώς η λειτουργία του μηχανισμού δεικτοδότησης για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα διαφέρει σε σημαντικό βαθμό από τη λειτουργία του μηχανισμού δεικτοδότησης για τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, οι μηχανισμοί αυτοί εξετάζονται χωριστά, στις ενότητες 7.5.1 και 7.5.2 αντίστοιχα.

7.5.1. Δεικτοδότηση Απομακρυσμένων Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα

Σ' αυτή την ενότητα περιγράφεται ο μηχανισμός δεικτοδότησης απομακρυσμένων αντικειμένων (εγγράφων και τμημάτων εγγράφων) στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα.

Ο μηχανισμός δεικτοδότησης απομακρυσμένων αντικειμένων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα είναι υπεύθυνος για την εκτέλεση των εξής λειτουργιών για κάθε δεικτοδοτούμενο έγγραφο:

- Τη δημιουργία ενός στιγμιότυπου της κλάσης εγγράφων στην οποία εντάσσεται το τρέχον έγγραφο, το οποίο αντιστοιχεί στο έγγραφο αυτό. Τα χαρακτηριστικά του στιγμιότυπου που αντιστοιχούν στα στοιχεία μεταπληροφορίας που χρησιμοποιούνται για την περιγραφή του εξάγονται και αποθηκεύονται, όπως και για τα έγγραφα που αποθηκεύονται στη βάση δεδομένων της πηγής πληροφορίας, με βάση τους κανόνες αναπαράστασης και διαχείρισης για τα στιγμιότυπα της κλάσης εγγράφων στην οποία

εντάσσεται το τρέχον έγγραφο. Αντίστοιχα με τα χαρακτηριστικά του στιγμιοτύπου που αντιστοιχούν στα στοιχεία μεταπληροφορίας που χρησιμοποιούνται για την περιγραφή του, τα υπόλοιπα χαρακτηριστικά του στιγμιοτύπου αποθηκεύονται στη βάση δεδομένων.

Σε αντίθεση με τα χαρακτηριστικά του στιγμιοτύπου, το περιεχόμενό του δεν αποθηκεύεται στη βάση δεδομένων της πηγής πληροφορίας: Αντί του περιεχομένου του, αποθηκεύεται η διεύθυνση στο Διαδίκτυο, όπου βρίσκεται το προς δεικτοδότηση έγγραφο. Αντίστοιχα, ο περιγραφέας του εγγράφου, αντί των ταυτοτήτων των δομικών στοιχείων από τα οποία αυτό απαρτίζεται, περιέχει τις διευθύνσεις στο Διαδίκτυο που αντιστοιχούν σε αυτά.

Εξαίρεση σε αυτό το γενικό κανόνα αποτελούν τα δομικά στοιχεία του εγγράφου τα οποία εντάσσονται σε κλάσεις για τα στιγμιότυπα των οποίων έχει οριστεί, από τους κανόνες δεικτοδότησης, ότι θα εισάγονται στη βάση δεδομένων της πηγής πληροφορίας. Επιπλέον, μπορούν να εισαχθούν στη βάση δεδομένων της πηγής πληροφορίας δομικά στοιχεία του εγγράφου για τα οποία εκφράζεται η αίτηση αυτή κατά την έκφραση της αίτησης δεικτοδότησης του εγγράφου. Η τελευταία κατάσταση συναντάται συχνά κατά την αίτηση δεικτοδότησης εγγράφων από πράκτορες: Ο πράκτορας μπορεί να ζητήσει την εισαγωγή στη βάση δεδομένων της πηγής πληροφορίας συγκεκριμένων δομικών στοιχείων αν διαπιστώσει μεγάλη συσχέτιση του περιεχομένου των συγκεκριμένων δομικών στοιχείων του εγγράφου με το περιεχόμενο της πηγής πληροφορίας, παρ' όλο που ο βαθμός συσχέτισης του περιεχομένου ολόκληρου του εγγράφου με το περιεχόμενο της πηγής πληροφορίας είναι μικρότερος και δε δικαιολογεί την εισαγωγή ολόκληρου του εγγράφου.

- Την ενημέρωση των δομών δεικτοδότησης της πηγής πληροφορίας για το περιεχόμενο του εγγράφου και των δομικών στοιχείων από τα οποία αυτό απαρτίζεται.

Η λειτουργία του μηχανισμού δεικτοδότησης απομακρυσμένων αντικειμένων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα απεικονίζεται στο Διάγραμμα 7-8, που είναι UML διάγραμμα δραστηριοτήτων.

Όπως φαίνεται στο Διάγραμμα 7-8, η λειτουργία του μηχανισμού δεικτοδότησης απομακρυσμένων αντικειμένων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα περιγράφεται από τον εξής αλγόριθμο:

1. Δέχεται ως είσοδο το URL του αρχείου όπου περιέχεται σε XML μορφή το έγγραφο που πρόκειται να δεικτοδοτηθεί.
2. Το αρχείο αποθηκεύεται προσωρινά στο δίσκο του υπολογιστή όπου έχει στηθεί η πηγή πληροφορίας.
3. Ελέγχεται αν οι DTDs που περιγράφουν τη δομή του εγγράφου υποστηρίζονται από την πηγή πληροφορίας.
 - 3.1. Αν κάποιος DTD δεν υποστηρίζεται από την πηγή πληροφορίας, επιστρέφεται ο κωδικός του μηνύματος λάθους "Δεν υποστηρίζεται ο DTD <DTD_URL>"⁴⁷ και η διαδικασία σταματά.
 - 3.2. Αν όλοι οι DTDs υποστηρίζονται από την πηγή πληροφορίας γίνονται τα εξής:

⁴⁷ Όπου DTD_URL η διεύθυνση του DTD.

3.2.1. Γίνεται συντακτικός έλεγχος των περιεχομένων του αρχείου.

3.2.1.1. Αν υπάρχουν συντακτικά λάθη, επιστρέφεται ο κωδικός του μηνύματος λάθους που περιγράφει το συντακτικό λάθος και η διαδικασία σταματά.

3.2.1.2. Αν δεν υπάρχουν συντακτικά λάθη γίνονται τα εξής:

3.2.1.2.1. Δημιουργείται ένα στιγμιότυπο της κλάσης εγγράφων στην οποία θα ενταχθεί το έγγραφο που περιέχεται στο αρχείο. Το στιγμιότυπο αυτό αναπαριστά το τρέχον έγγραφο.

3.2.1.2.2. Ανακτώνται οι κανόνες αναπαράστασης, διαχείρισης και δεικτοδότησης της κλάσης εγγράφων στην οποία το τρέχον στιγμιότυπο θα ενταχθεί.

3.2.1.2.3. Αν υπάρχουν δομικά στοιχεία του εγγράφου που θα αποθηκευθούν στη βάση δεδομένων της πηγής πληροφορίας (είτε λόγω των κανόνων δεικτοδότησης είτε μετά από αίτηση που συνοδεύει την αίτηση δεικτοδότησης του εγγράφου) γίνονται τα εξής:

3.2.1.2.3.1. Αν το έγγραφο πρέπει να αποθηκευτεί τμηματικά γίνονται τα εξής:

3.2.1.2.3.1.1. Δημιουργείται, για κάθε αντιγραφόμενο δομικό στοιχείο του εγγράφου που αποθηκεύεται ανεξάρτητα, ένα στιγμιότυπο της αντίστοιχης κλάσης δομικών στοιχείων, το οποίο και το αναπαριστά.

3.2.1.2.3.1.2. Το υπόλοιπο του περιεχόμενου του περιέχοντος (δομικού στοιχείου ή εγγράφου) που αντιγράφεται στην πηγή πληροφορίας αποθηκεύεται στο πεδίο «Περιεχόμενο» του περιέχοντος.

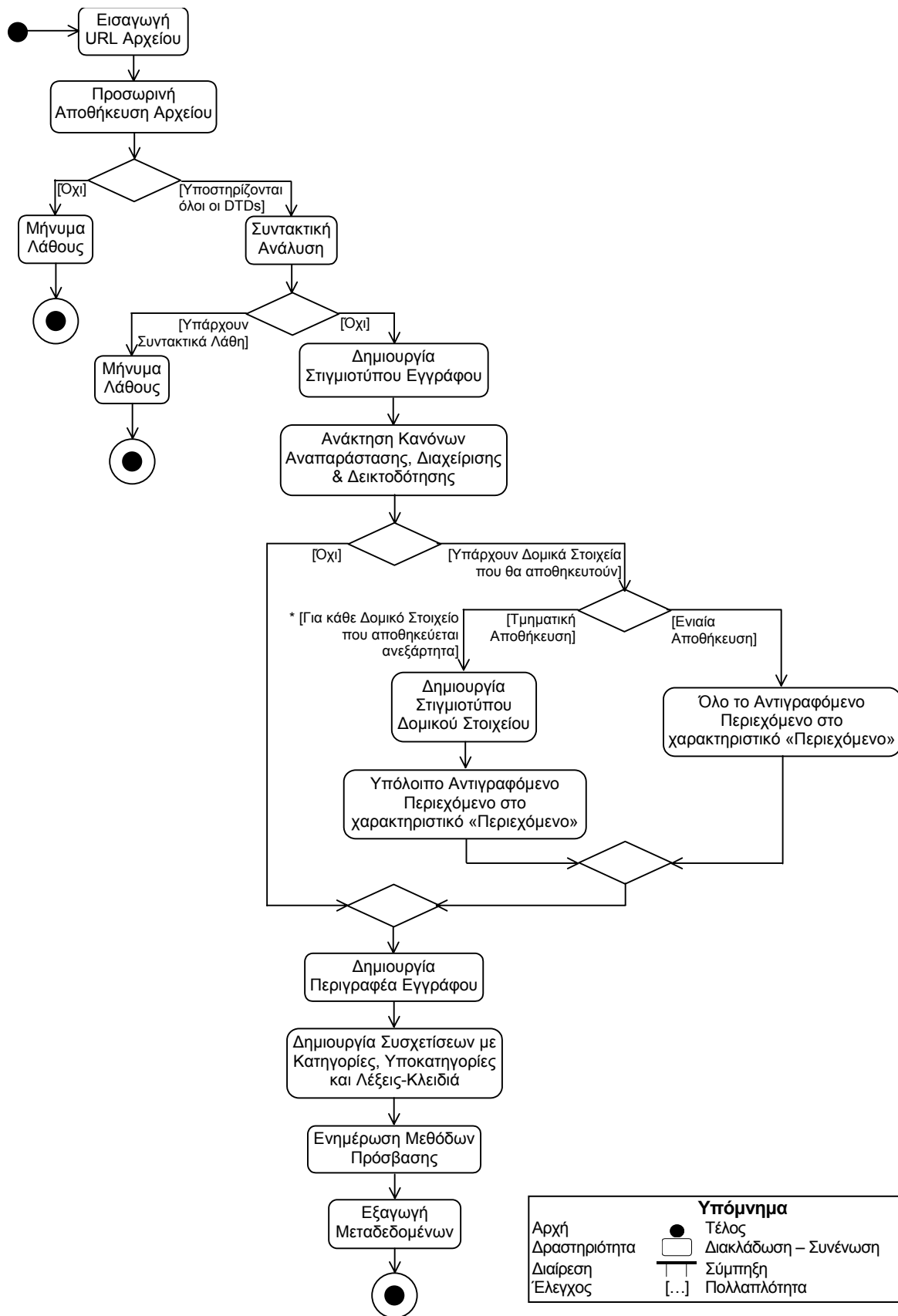
3.2.1.2.3.2. Διαφορετικά, όλο το αντιγραφόμενο περιεχόμενο του εγγράφου αποθηκεύεται στο πεδίο «Περιεχόμενο».

3.2.1.2.4. Δημιουργείται και ενημερώνεται ο περιγραφέας του εγγράφου.

3.2.1.2.5. Με βάση το περιεχόμενο του εγγράφου και των δομικών στοιχείων από τα οποία απαρτίζεται, δημιουργούνται συσχετίσεις του εγγράφου και των δομικών στοιχείων με κατηγορίες, υποκατηγορίες και λέξεις-κλειδιά.

3.2.1.2.6. Ενημερώνονται οι δομές που υποστηρίζουν τις μεθόδους πρόσβασης.

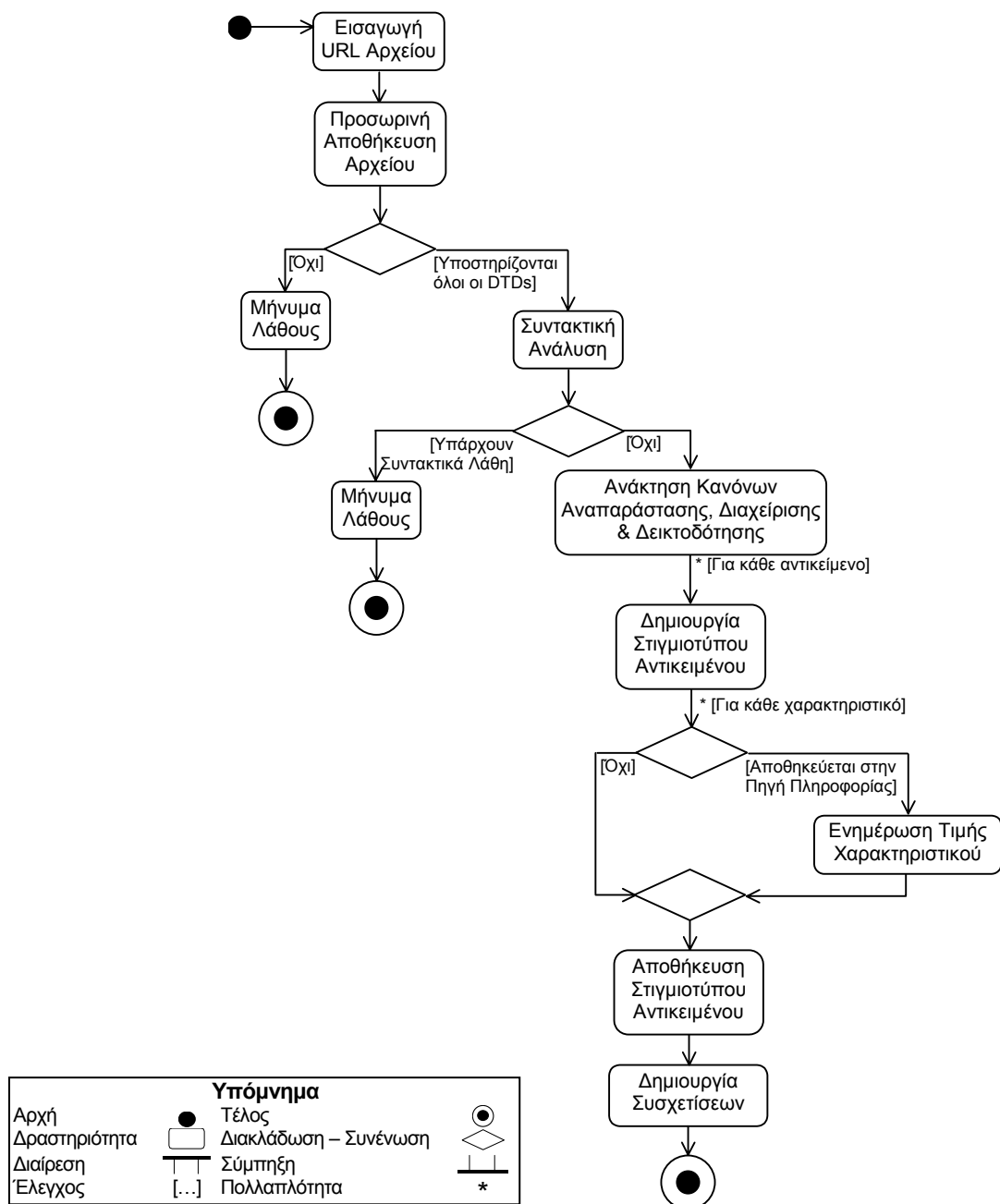
3.2.1.2.7. Ενημερώνονται οι τιμές των χαρακτηριστικών του εγγράφου που αναπαριστούν χαρακτηριστικά της Dublin Core μετά την κατάλληλη επεξεργασία των αντίστοιχων στοιχείων του εγγράφου (αν αυτά υπάρχουν και είναι διαθέσιμα στο συγκεκριμένο έγγραφο).



Διάγραμμα 7-8: UML Διάγραμμα Δραστηριοτήτων που περιγράφει τη λειτουργία του Μηχανισμού Δεικτοδότησης Απομακρυσμένων Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Δομημένα Έγγραφα

7.5.2. Δεικτοδότηση Απομακρυσμένων Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

Το θέμα αυτής της ενότητας είναι η περιγραφή του μηχανισμού δεικτοδότησης απομακρυσμένων αντικειμένων στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα.



Διάγραμμα 7-9: UML Διάγραμμα Δραστηριοτήτων που περιγράφει τη λειτουργία του Μηχανισμού Δεικτοδότησης Απομακρυσμένων Αντικειμένων στις Πηγές Πληροφορίας που περιέχουν Παραδοσιακά Δεδομένα

Ο μηχανισμός δεικτοδότησης απομακρυσμένων αντικειμένων στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα είναι υπεύθυνος για την εκτέλεση των εξής λειτουργιών για κάθε δεικτοδοτούμενο αντικείμενο:

- Τη δημιουργία ενός στιγμιοτύπου της κλάσης στην οποία εντάσσεται το αντικείμενο, το οποίο αντιστοιχεί στο δεικτοδοτούμενο αντικείμενο. Από τα χαρακτηριστικά του αντικειμένου αυτού τιμές έχουν τα χαρακτηριστικά εκείνα για τα οποία καθορίζεται από τους κανόνες δεικτοδότησης ότι η τιμή τους θα αντιγράφεται στη βάση δεδομένων της πηγής πληροφορίας, ενώ για τα υπόλοιπα αντί τιμής αποθηκεύεται η διεύθυνση στο Διαδίκτυο που αντιστοιχεί στην τιμή του χαρακτηριστικού. Με τον τρόπο αυτό δημιουργείται στη βάση δεδομένων της πηγής πληροφορίας μια *Αποψη (View)* των απομακρυσμένων αντικειμένων της κλάσης στην οποία το τρέχον αντικείμενο εντάσσεται.
- Τη δημιουργία, σύμφωνα πάντα με τους κανόνες δεικτοδότησης, στιγμιοτύπων που αντιστοιχούν σε τιμές του τρέχοντος αντικειμένου: Τα στιγμιότυπα αυτά στην XML μορφή του αντικειμένου αντιστοιχούν σε «υποστοιχεία» του, τα οποία με βάση τους κανόνες αναπαράστασης και διαχείρισης της κλάσης στην οποία αυτό εντάσσεται αναπαρίστανται από στιγμιότυπα κλάσεων.
- Την ενημέρωση των δομών δεικτοδότησης της πηγής πληροφορίας για το περιεχόμενο του αντικειμένου.

Η λειτουργία του μηχανισμού δεικτοδότησης απομακρυσμένων αντικειμένων στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα απεικονίζεται στο Διάγραμμα 7-9, που είναι UML διάγραμμα δραστηριότητας.

Όπως φαίνεται στο Διάγραμμα 7-9, η λειτουργία του μηχανισμού δεικτοδότησης απομακρυσμένων αντικειμένων στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα περιγράφεται από τον εξής αλγόριθμο:

1. Δέχεται ως είσοδο το URL του αρχείου όπου περιέχονται σε XML μορφή τα αντικείμενα που πρόκειται να εισαχθούν.
2. Το αρχείο αποθηκεύεται προσωρινά στο δίσκο του υπολογιστή όπου έχει στηθεί η πηγή πληροφορίας.
3. Ελέγχεται αν οι DTDs που περιγράφουν τα αντικείμενα υποστηρίζονται από την πηγή πληροφορίας.
 - 3.1. Αν κάποιος DTD δεν υποστηρίζεται από την πηγή πληροφορίας, επιστρέφεται ο κωδικός του μηνύματος λάθους "Δεν υποστηρίζεται ο DTD <DTD_URL>"⁴⁸ και η διαδικασία σταματά.
 - 3.2. Αν όλοι οι DTDs υποστηρίζονται από την πηγή πληροφορίας γίνονται τα εξής:
 - 3.2.1. Γίνεται συντακτικός έλεγχος των περιεχομένων του αρχείου.
 - 3.2.1.1. Αν υπάρχουν συντακτικά λάθη, επιστρέφεται ο κωδικός του μηνύματος λάθους που περιγράφει το συντακτικό λάθος και η διαδικασία σταματά.
 - 3.2.1.2. Αν δεν υπάρχουν συντακτικά λάθη γίνονται τα εξής:
 - 3.2.1.2.1. Ανακτώνται οι κανόνες αναπαράστασης και διαχείρισης της κλάσης στην οποία εντάσσεται το τρέχον αντικείμενο.
 - 3.2.1.2.2. Για κάθε αντικείμενο (το κυρίως αντικείμενο ή κάποιο αντικείμενο – υποστοιχείο του) γίνονται τα εξής:

⁴⁸ Όπου DTD_URL η διεύθυνση του DTD.

3.2.1.2.2.1. Δημιουργείται ένα στιγμιότυπο της κλάσης στην οποία εντάσσεται το τρέχον αντικείμενο.

3.2.1.2.2.2. Για κάθε χαρακτηριστικό του αντικειμένου, ελέγχεται αν, βάσει των κανόνων δεικτοδότησης, αυτό πρέπει να αποθηκευθεί στη βάση δεδομένων.

3.2.1.2.2.2.1. Αν ναι, ενημερώνεται η τιμή του συγκεκριμένου χαρακτηριστικού του στιγμιότυπου.

3.2.1.2.2.3. Εισάγεται στη βάση δεδομένων της πηγής πληροφορίας ένα στιγμιότυπο της αντίστοιχης κλάσης.

3.2.1.2.2.4. Δημιουργούνται οι συσχετίσεις του αντικειμένου με τα υπόλοιπα περιεχόμενα της βάσης δεδομένων.

7.6. Ενημέρωση των Πηγών Πληροφορίας

Το τμήμα αυτό ασχολείται με την ενημέρωση των πηγών πληροφορίας για αντικείμενα που γίνονται διαθέσιμα και σχετίζονται με το περιεχόμενο των πηγών πληροφορίας.

Για την ενημέρωση κάθε πηγής πληροφορίας υπεύθυνοι είναι ένας ή περισσότεροι *Πράκτορες (Agents)*. Οι πράκτορες αυτοί είναι προγράμματα που διασχίζουν το Διαδίκτυο και εντοπίζουν αντικείμενα των οποίων το περιεχόμενο σχετίζεται με το περιεχόμενο των πηγών πληροφορίας για την ενημέρωση των οποίων είναι υπεύθυνοι οι πράκτορες. Όταν κάποιος πράκτορας εντοπίσει κάποιο αντικείμενο που το περιεχόμενό του σχετίζεται με το περιεχόμενο κάποιων από τις πηγές πληροφορίας για τις οποίες ο πράκτορας είναι υπεύθυνος, στέλνει τα αντίστοιχα αιτήματα στις πηγές πληροφορίας, είτε για τη δεικτοδότηση του αντικειμένου είτε για την εισαγωγή του (ολική ή μερική).

Η διαδικασία αυτή υλοποιείται, στα πλαίσια της παρούσας διατριβής, για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα, από το εργαλείο ISAgent, το οποίο επισκέπτεται διευθύνσεις στο Διαδίκτυο και ενημερώνει τις πηγές πληροφορίας για την ύπαρξη αντικειμένων που το περιεχόμενό τους σχετίζεται με αυτό των πηγών πληροφορίας. Πριν περιγραφεί η λειτουργία του εργαλείου ISAgent, πρέπει να σημειωθεί ότι, μετά την επίσκεψη μιας διεύθυνσης στο Διαδίκτυο από κάποιο στιγμιότυπο του εργαλείου, αποθηκεύονται οι βαθμοί συσχέτισης του περιεχομένου της με το περιεχόμενο καθεμιάς από τις πηγές πληροφορίας, για την ενημέρωση των οποίων υπεύθυνο είναι το τρέχον στιγμιότυπο. Αυτό γίνεται για να αναδιοργανωθεί, μετά από κάθε κύκλο επισκέψεων και με βάση τους βαθμούς συσχέτισης των περιεχομένων των πηγών πληροφορίας με το περιεχόμενο των διευθύνσεων, η λίστα των διευθύνσεων που επισκέπτεται το τρέχον στιγμιότυπο και να επανακαθοριστούν οι προτεραιότητες επίσκεψης των διευθύνσεων από τις οποίες η λίστα απαρτίζεται.

Στο σημείο αυτό πρέπει να σημειωθεί ότι η παρούσα διατριβή δεν ασχολείται ούτε με την κατάρτιση και την ενημέρωση της λίστας των διευθύνσεων στο Διαδίκτυο, τις οποίες επισκέπτονται οι πράκτορες, ούτε με τον καθορισμό των κανόνων προτεραιότητας, με βάση

τους οποίους θα γίνονται από κάθε στιγμιότυπο του εργαλείου ISAgent οι επισκέψεις των διευθύνσεων στο Διαδίκτυο ούτε με θέματα επικοινωνίας και συνεργασίας μεταξύ πρακτόρων και με πολυπρακτορικά περιβάλλοντα, καθώς τα θέματα αυτά αφήνονται ως μελλοντικές επεκτάσεις. Έτσι, στη συνέχεια αυτού του τμήματος η συζήτηση επικεντρώνει στην πολιτική που ακολουθείται από κάθε πηγή πληροφορίας για τη *Διαχείριση Απομακρυσμένων Αντικειμένων* (ενότητα 7.6.1) και τη *Λειτουργία του Εργαλείου ISAgent* (ενότητα 7.6.2).

7.6.1. Διαχείριση Απομακρυσμένων Αντικειμένων

Το θέμα αυτής της ενότητας είναι η περιγραφή της πολιτικής που ακολουθείται από τις πηγές πληροφορίας για τη διαχείριση των αντικειμένων που εντοπίζονται από τους πράκτορες που είναι υπεύθυνοι για την ενημέρωσή τους.

Οι παράγοντες που καθορίζουν την πολιτική που ακολουθείται για τα αντικείμενα που εντοπίζονται από κάποιο συγκεκριμένο πράκτορα σε κάθε πηγή πληροφορίας είναι οι εξής:

- Απαγορεύσεις αντιγραφής του περιεχομένου των αντικειμένων από αυτούς που έχουν τα πνευματικά δικαιώματα για το περιεχόμενο των αντικειμένων.
- Ο βαθμός συσχέτισης του περιεχομένου του αντικειμένου με το περιεχόμενο της πηγής πληροφορίας.
- Αν πρόκειται για σύνθετο αντικείμενο (π.χ. έγγραφο), οι βαθμοί συσχέτισης του περιεχομένου των δομικών στοιχείων του αντικειμένου με το περιεχόμενο της πηγής πληροφορίας.
- Η πολιτική που ακολουθείται από την πηγή πληροφορίας στο θέμα των απομακρυσμένων αντικειμένων. Η πολιτική αυτή καθορίζει ποια απομακρυσμένα αντικείμενα θα δεικτοδοτηθούν στην πηγή πληροφορίας, ποια θα αποθηκευτούν σε αυτήν ολικά και ποια μερικά. Οι παράμετροι που καθορίζουν την πολιτική κάθε πηγής πληροφορίας είναι δύο:
 - Το *κάτω όριο δεικτοδότησης*, που είναι η ελάχιστη τιμή που πρέπει να έχει ο βαθμός συσχέτισης του περιεχομένου κάποιου αντικειμένου με το περιεχόμενο της πηγής πληροφορίας για να δεικτοδοτηθεί το αντικείμενο από την πηγή πληροφορίας.
 - Το *κάτω όριο αποθήκευσης*, που είναι η ελάχιστη τιμή που πρέπει να έχει ο βαθμός συσχέτισης του περιεχομένου κάποιου αντικειμένου με το περιεχόμενο της πηγής πληροφορίας για να αποθηκευτεί το αντικείμενο στην πηγή πληροφορίας.

Με βάση τα παραπάνω, ο αλγόριθμος που εκτελείται για να αποφασιστεί ο τρόπος διαχείρισης κάποιου αντικειμένου είναι ο εξής:

1. Αν πρόκειται για απλό αντικείμενο κάνε τα εξής:

1.1. Αν ο βαθμός συσχέτισης του περιεχομένου του αντικειμένου με το περιεχόμενο της πηγής πληροφορίας είναι μικρότερος από το κάτω όριο δεικτοδότησης, αγνόησε το αντικείμενο.

1.2. Διαφορετικά κάνε τα εξής:

1.2.1. Αν απαγορεύεται η αντιγραφή του αντικειμένου ή ο βαθμός συσχέτισης του περιεχομένου του αντικειμένου με το περιεχόμενο της πηγής πληροφορίας

είναι μικρότερος από το κάτω όριο αποθήκευσης, το αντικείμενο απλώς θα δεικτοδοτηθεί στην πηγή πληροφορίας.

1.2.2. Διαφορετικά, το αντικείμενο θα εισαχθεί στην πηγή πληροφορίας.

2. Διαφορετικά, αν πρόκειται για σύνθετο αντικείμενο κάνε τα εξής:

2.1. Αν τόσο ο βαθμός συσχέτισης του περιεχομένου του αντικειμένου όσο και οι βαθμοί συσχέτισης του περιεχομένου των δομικών στοιχείων του με το περιεχόμενο της πηγής πληροφορίας είναι μικρότεροι από το κάτω όριο δεικτοδότησης, αγνόησε το αντικείμενο.

2.2. Διαφορετικά κάνε τα εξής:

2.2.1. Αν ο βαθμός συσχέτισης του περιεχομένου του αντικειμένου με το περιεχόμενο της πηγής πληροφορίας είναι μικρότερος από το κάτω όριο δεικτοδότησης κάνε τα εξής:

2.2.1.1. Αν ο βαθμός συσχέτισης του περιεχομένου όλων των δομικών στοιχείων του αντικειμένου με το περιεχόμενο της πηγής πληροφορίας είναι μικρότερος από το κάτω όριο αποθήκευσης, δεικτοδότησε μόνο τα στοιχεία που ο βαθμός συσχέτισης του περιεχομένου τους είναι μεγαλύτερος από το κάτω όριο δεικτοδότησης.

2.2.1.2. Διαφορετικά κάνε τα εξής:

2.2.1.2.1. Ολόκληρο το αντικείμενο θα δεικτοδοτηθεί.

2.2.1.2.2. Αν δεν απαγορεύεται η αντιγραφή του αντικειμένου, θα εισαχθούν στην πηγή πληροφορίας τα τμήματα του αντικειμένου που ο βαθμός συσχέτισης του περιεχομένου τους με το περιεχόμενο της πηγής πληροφορίας είναι μεγαλύτερος από ή ίσος με το κάτω όριο αποθήκευσης.

2.2.2. Διαφορετικά κάνε τα εξής:

2.2.2.1. Αν ο βαθμός συσχέτισης του περιεχομένου του αντικειμένου με το περιεχόμενο της πηγής πληροφορίας είναι μικρότερος από το κάτω όριο αποθήκευσης κάνε τα εξής:

2.2.2.1.1. Ολόκληρο το αντικείμενο θα δεικτοδοτηθεί.

2.2.2.1.2. Αν δεν απαγορεύεται η αντιγραφή του αντικειμένου, θα εισαχθούν στην πηγή πληροφορίας τα τμήματα του αντικειμένου που ο βαθμός συσχέτισης του περιεχομένου τους με το περιεχόμενο της πηγής πληροφορίας είναι μεγαλύτερος από ή ίσος με το κάτω όριο αποθήκευσης.

2.2.2.2. Διαφορετικά το αντικείμενο θα εισαχθεί στην πηγή πληροφορίας.

7.6.2. Λειτουργία του Εργαλείου ISAgent

Το θέμα αυτής της ενότητας είναι η περιγραφή της λειτουργίας του εργαλείου ISAgent, τα στιγμιότυπα του οποίου είναι υπεύθυνα για την ενημέρωση των πηγών πληροφορίας. Η λειτουργία του εργαλείου ISAgent απεικονίζεται στο Διάγραμμα 7-10, που είναι ένα UML διάγραμμα δραστηριοτήτων.

Όπως φαίνεται στο Διάγραμμα 7-10, η λειτουργία του εργαλείου ISAgent συνίσταται από την παράλληλη εκτέλεση δυο δραστηριοτήτων:

- Της δραστηριότητας εντοπισμού, μέσω περιοδικών διασχίσεων του Διαδικτύου, νέων αντικειμένων και ενημέρωσης για την ύπαρξή τους των πηγών πληροφορίας που το περιεχόμενό τους σχετίζεται με αυτό των αντικειμένων.

- Της δραστηριότητας διαχείρισης των εκκρεμών, προς τις πηγές πληροφορίας, αιτήσεων.

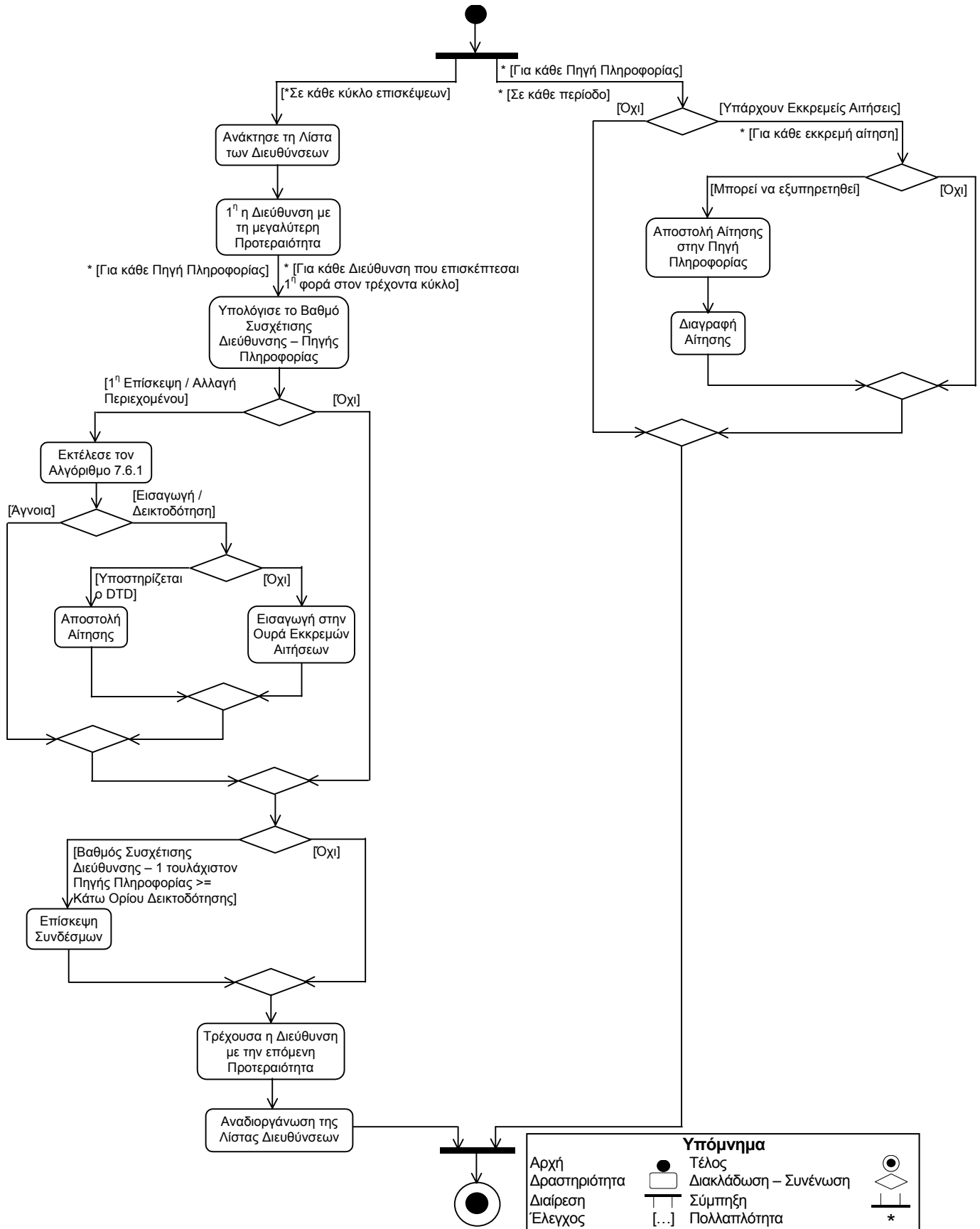
Με βάση το Διάγραμμα 7-10, η λειτουργία της δραστηριότητας εντοπισμού νέων αντικειμένων και ενημέρωσης των πηγών πληροφορίας σε κάθε κύκλο επισκέψεων περιγράφεται από τον εξής αλγόριθμο:

1. Ανάκτησε τη λίστα διευθύνσεων που θα επισκεφθείς.
2. Τρέχουσα διεύθυνση είναι αυτή με τη μεγαλύτερη προτεραιότητα επίσκεψης.
3. Για κάθε διεύθυνση που επισκέπτεσαι πρώτη φορά σ' αυτό τον κύκλο επισκέψεων κάνε τα εξής:
 - 3.1. Για κάθε πηγή πληροφορίας κάνε τα εξής:
 - 3.1.1. Υπολόγισε το βαθμό συσχέτισης του περιεχομένου της διεύθυνσης με το περιεχόμενο της πηγής πληροφορίας.
 - 3.1.2. Αν δεν έχεις επισκεφθεί τη διεύθυνση ξανά ή το περιεχόμενό της έχει αλλάξει από την προηγούμενη επίσκεψή σου κάνε τα εξής:
 - 3.1.2.1. Με βάση τον αλγόριθμο της ενότητας 7.6.1 αποφάσισε αν θα αγνοήσεις το περιεχόμενο της τρέχουσας διεύθυνσης για την τρέχουσα πηγή πληροφορίας.
 - 3.1.2.2. Αν όχι, κάνε τα εξής:
 - 3.1.2.2.1. Αν ο DTD που περιγράφει τη δομή του αντικειμένου που περιέχεται στην τρέχουσα διεύθυνση υποστηρίζεται από την τρέχουσα πηγή πληροφορίας, στείλε στην πηγή πληροφορίας τις κατόλληλες αιτήσεις εισαγωγής και/ή δεικτοδότησης.
 - 3.1.2.2.2. Διαφορετικά κάνε τα εξής:
 - 3.1.2.2.2.1. Τοποθέτησε την τρέχουσα αίτηση στην ουρά των εκκρεμών αιτήσεων για την τρέχουσα πηγή πληροφορίας.
 - 3.1.2.2.2.2. Στείλε ένα αίτημα υποστήριξης του DTD στο διαχειριστή της πηγής πληροφορίας
 - 3.1.3. Αν ο βαθμός συσχέτισης του περιεχομένου της διεύθυνσης με το περιεχόμενο μιας τουλάχιστον πηγής πληροφορίας είναι μεγαλύτερος ή ίσος του κάτω ορίου δεικτοδότησης της πηγής πληροφορίας, επισκέψου τις διευθύνσεις προς τις οποίες υπάρχουν σύνδεσμοι από την τρέχουσα διεύθυνση.
 - 3.2. Τρέχουσα διεύθυνση είναι αυτή με την αμέσως μεγαλύτερη προτεραιότητα επίσκεψης.
4. Αναδιοργάνωσε τη λίστα των διευθύνσεων.

Αντίστοιχα, η λειτουργία της δραστηριότητας διαχείρισης των εκκρεμών αιτήσεων προς τις πηγές πληροφορίας περιγράφεται από τον εξής αλγόριθμο, που εκτελείται περιοδικά:

1. Για κάθε πηγή πληροφορίας κάνε τα εξής:
 - 1.1. Έλεγχε αν υπάρχουν εκκρεμείς αιτήσεις γι' αυτή την πηγή πληροφορίας.
 - 1.2. Για κάθε εκκρεμή αίτηση κάνε τα εξής:
 - 1.2.1. Έλεγχε αν η αίτηση μπορεί να εξυπηρετηθεί.
 - 1.2.2. Αν ναι, κάνε τα εξής:
 - 1.2.2.1. Στείλε την αίτηση στην πηγή πληροφορίας.
 - 1.2.2.2. Διέγραψε την αίτηση από την ουρά των εκκρεμών αιτήσεων για την τρέχουσα πηγή πληροφορίας.
 - 1.2.3. Προχώρησε στην επόμενη αίτηση.

1.3. Προχώρησε στην επόμενη πηγή πληροφορίας.



Διάγραμμα 7-10: UML Διάγραμμα Δραστηριοτήτων που περιγράφει τη λειτουργία του Εργαλείου ISAgent

7.7. Ανακεφαλαίωση

Στο κεφάλαιο αυτό έγινε αναφορά στη διαχείριση τόσο των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα όσο και των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα, καθώς και στα εργαλεία που απαιτούνται για τη δημιουργία και διαχείριση αυτών των πηγών πληροφορίας.

Ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας ΚΑΔΔΑ, που περιγράφει τους κανόνες διαχείρισης, αναπαράστασης και δεικτοδότησης που ισχύουν για τα στιγμιότυπα κλάσεων που περιγράφονται από XML DTDs και επιτρέπει μ' αυτό τον τρόπο την εξέλιξη των σχημάτων και των μοντέλων πληροφορίας των βάσεων δεδομένων των πηγών πληροφορίας μέσω της υποστήριξης νέων κλάσεων. Τόσο η γλώσσα ΚΑΔΔΑ όσο και ο σχεδιασμός των εργαλείων και των μηχανισμών που βασίζονται στη χρήση της (εργαλεία SupportDTD και ISAgent και μηχανισμοί εισαγωγής και δεικτοδότησης) αποτελούν συνεισφορά της παρούσας διατριβής.

Αρχικά, στο τμήμα 7.1 περιγράφηκε η *Δημιουργία Πηγών Πληροφορίας* ενώ στο τμήμα 7.2 περιγράφηκε η δυνατότητα που παρέχεται για την *Υποστήριξη Νέων Κλάσεων από τις Πηγές Πληροφορίας*. Στη συνέχεια, στο τμήμα 7.3 περιγράφηκε η *Εισαγωγή Αντικειμένων* στις βάσεις δεδομένων που υποστηρίζουν τις πηγές πληροφορίας. Θέματα που σχετίζονται με την *Αναζήτηση & Ανάκτηση Αντικειμένων* στις πηγές πληροφορίας συζητήθηκαν στο τμήμα 7.4, ενώ συζήτηση για τη *Δεικτοδότηση Απομακρυσμένων Αντικειμένων* στις πηγές πληροφορίας έγινε στο τμήμα 7.5. Τέλος, στο τμήμα 7.6 περιγράφηκε η *Ενημέρωση των Πηγών Πληροφορίας* για αντικείμενα που γίνονται διαθέσιμα και σχετίζονται με το περιεχόμενο των συγκεκριμένων πηγών πληροφορίας.

Στο 8^ο κεφάλαιο, που ακολουθεί, περιγράφεται μια εφαρμογή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», με στόχο να παρουσιαστεί η λειτουργία του συστήματος και να γίνουν εμφανείς η χρησιμότητα και η εφαρμοσιμότητά του.

8^ο Κεφάλαιο

Εφαρμογή του Συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»

Σ' αυτό το κεφάλαιο δίνεται ένα παράδειγμα εφαρμογής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» σε ακαδημαϊκό περιβάλλον, με στόχο να παρουσιαστεί η λειτουργία του συστήματος και να γίνουν εμφανείς η χρησιμότητα και η εφαρμοσιμότητα του συστήματος. Έτσι, στο τμήμα 8.1 περιγράφεται το *Περιβάλλον της Εφαρμογής*, στο τμήμα 8.2 περιγράφεται η *Αρχιτεκτονική της Εφαρμογής* και στο τμήμα 8.3 γίνεται αναφορά στη *Λειτουργία της Εφαρμογής*. Το κεφάλαιο κλείνει με μια σύντομη *Ανακεφαλαίωση* στο τμήμα 8.4.

8.1. Περιβάλλον της Εφαρμογής

Το θέμα αυτού του τμήματος είναι η περιγραφή του περιβάλλοντος στο οποίο λαμβάνει χώρα η εφαρμογή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» που δίνεται ως παράδειγμα σ' αυτό το κεφάλαιο.

Το περιβάλλον στο οποίο λαμβάνει χώρα η εφαρμογή είναι ακαδημαϊκό: Η εφαρμογή έρχεται να καλύψει τις ανάγκες καθηγητών, ερευνητών και φοιτητών (προπτυχιακών και μεταπτυχιακών) των ελληνικών ΑΕΙ (Πανεπιστημίων και Πολυτεχνείων), σε ότι αφορά τον εντοπισμό υλικού σχετικού με τα ερευνητικά και επιστημονικά θέματα με τα οποία αυτοί ασχολούνται.

Υλικό σχετικό με τα θέματα αυτά διαθέτουν τα ΑΕΙ, τα οποία έχουν συμφωνήσει να ανταλλάσσουν υλικό και τεχνογνωσία. Το υλικό αυτό απαρτίζεται από διπλωματικές εργασίες, μεταπτυχιακές και διδακτορικές διατριβές, επιστημονικά άρθρα, τεχνικές αναφορές, παρουσιάσεις και παραπομπές σε σχετική πληροφορία στο Διαδίκτυο. Επιπλέον, ο εκδοτικός οίκος «Πανεπιστημιακές Εκδόσεις Κρήτης», που ειδικεύεται στις ακαδημαϊκές και τεχνικές εκδόσεις έχει συμφωνήσει να παρέχει στα μέλη των ακαδημαϊκών κοινοτήτων των ελληνικών ΑΕΙ δωρεάν πρόσβαση στη βάση δεδομένων που έχει σκοπό να δημιουργήσει. Η βάση δεδομένων του εκδοτικού οίκου «Πανεπιστημιακές Εκδόσεις Κρήτης» θα περιέχει πληροφορία για την περιγραφή των τίτλων (βιβλίων, περιοδικών κλπ.) που θα γίνονται διαθέσιμοι απ' αυτόν καθώς και τις συνόψεις των τίτλων αυτών⁴⁹.

Ο στόχος της εφαρμογής που θα αναπτυχθεί είναι η παροχή στα μέλη των ακαδημαϊκών κοινοτήτων των ελληνικών ΑΕΙ των εξής δυνατοτήτων:

- Της αναζήτησης με ενιαίο και διαφανή τρόπο, μέσω ερωτήσεων που θα περιέχουν όρους δίτιμης λογικής και/ή όρους ομοιότητας για το περιεχόμενο και/ή τη δομή των αντικειμένων, υλικού που σχετίζεται με τα ενδιαφέροντά τους. Η αναζήτηση του υλικού θα λαμβάνει χώρα στις πηγές πληροφορίας όπου θα περιέχεται το υλικό που κατέχουν τα ελληνικά ΑΕΙ και στη βάση δεδομένων του εκδοτικού οίκου «Πανεπιστημιακές Εκδόσεις Κρήτης».
- Της ανάκτησης του υλικού, που εντοπίζεται στις πηγές πληροφορίας των ΑΕΙ, και της σχετικής με τους τίτλους πληροφορίας, που εντοπίζεται στην πηγή πληροφορίας των «Πανεπιστημιακών Εκδόσεων Κρήτης».
- Της αναζήτησης παραπομπών σε αντικείμενα στο Διαδίκτυο των οποίων το περιεχόμενο σχετίζεται με τα ενδιαφέροντά τους.
- Της ενημέρωσης των ενδιαφερόμενων μελών των ακαδημαϊκών κοινοτήτων, όταν κάποιο αντικείμενο που το περιεχόμενό του σχετίζεται με τα ενδιαφέροντά τους γίνει διαθέσιμο.

Από τα παραπάνω φαίνεται ότι η κάλυψη των απαιτήσεων που τίθενται συμπίπτει με την κάλυψη των απαιτήσεων που περιγράφηκαν για τους χρήστες του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» στο 4^ο κεφάλαιο. Έτσι, για την κάλυψη των αναγκών που προαναφέρθηκαν προτείνεται η δημιουργία μιας εφαρμογής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», η αρχιτεκτονική της οποίας περιγράφεται στο τμήμα 8.2.

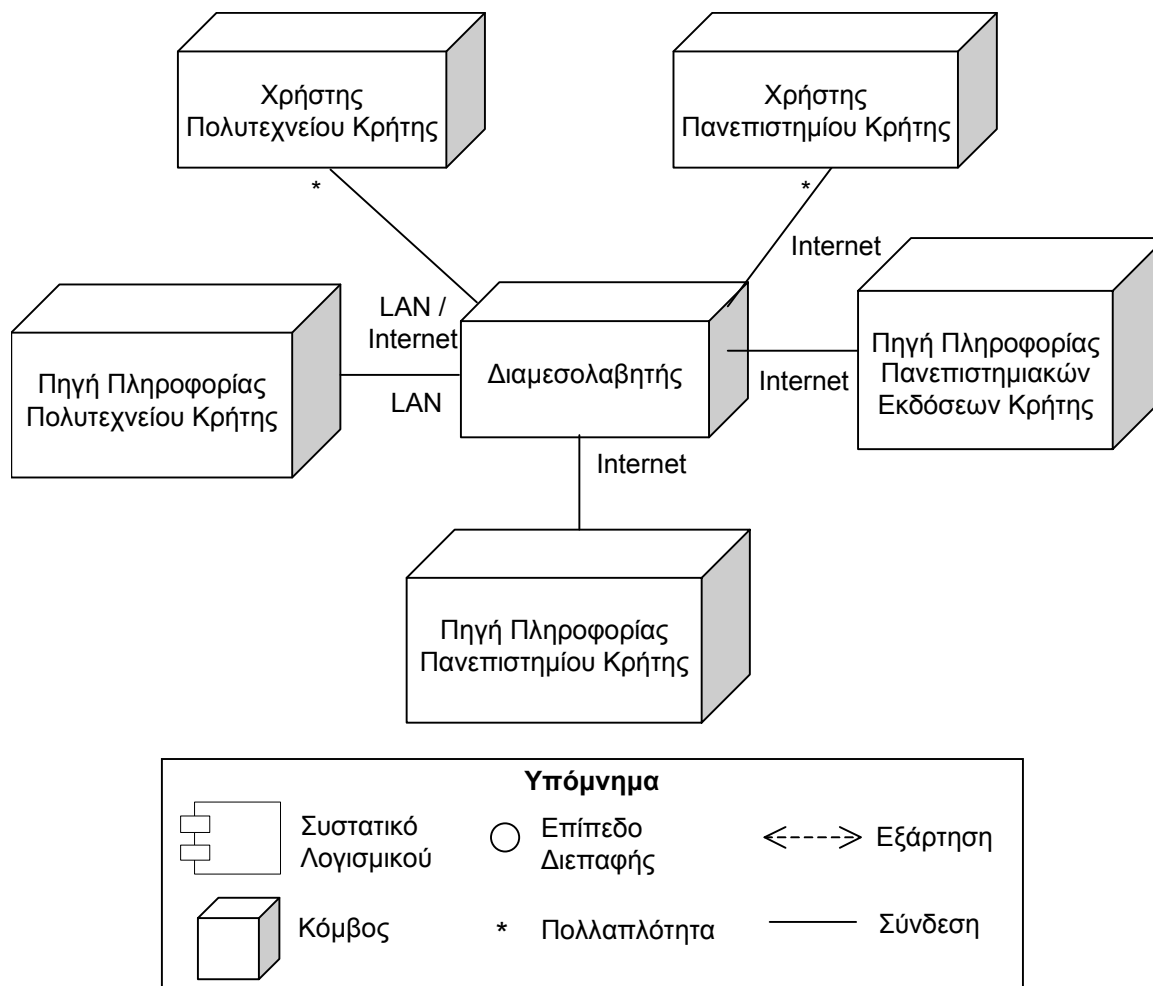
8.2. Αρχιτεκτονική της Εφαρμογής

Σ' αυτό το τμήμα περιγράφεται η αρχιτεκτονική της εφαρμογής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», που έρχεται να καλύψει τις ανάγκες των ακαδημαϊκών κοινοτήτων των ελληνικών ΑΕΙ για τον εντοπισμό πληροφορίας σχετικής με τα ερευνητικά και επιστημονικά θέματα με τα οποία τα μέλη τους ασχολούνται. Η περιγραφή της

⁴⁹ Η εφαρμογή που περιγράφεται στο κεφάλαιο αυτό θα μπορούσε να παρέχει παγκόσμια κάλυψη. Καθώς όμως αυτό θα έκανε την περιγραφή της πολύ πιο περίπλοκη χωρίς να προσθέτει κάτι από πλευράς ουσίας, επιλέχθηκε η εφαρμογή να περιγραφεί με πανελλήνια αντί για παγκόσμια εμβέλεια.

εφαρμογής δίνεται για να παρουσιαστεί η λειτουργία του συστήματος και να γίνουν εμφανείς η χρησιμότητα και η εφαρμοσιμότητα του.

Έστω τώρα, χωρίς βλάβη της γενικότητας, ότι αρχικά δημιουργείται μια πιλοτική εφαρμογή, στην οποία συμμετέχουν το Πολυτεχνείο Κρήτης και το Πανεπιστήμιο Κρήτης και οι «Πανεπιστημιακές Εκδόσεις Κρήτης». Στην περίπτωση αυτή, η προτεινόμενη για την εφαρμογή αρχιτεκτονική απεικονίζεται στο Διάγραμμα 8-1, που είναι UML διάγραμμα παράταξης.



Διάγραμμα 8-1: UML Διάγραμμα Παράταξης για την Εφαρμογή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση»

Όπως φαίνεται στο Διάγραμμα 8-1, προτείνεται η δημιουργία ενός στιγμιοτύπου του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», το οποίο απαρτίζεται από ένα διαμεσολαβητή και τρεις πηγές πληροφορίας.

Ο διαμεσολαβητής της εφαρμογής επιτρέπει στους χρήστες να προσπελάσουν με ενιαίο και διαφανή τρόπο, μέσω των υπηρεσιών που παρέχει, την πληροφορία που γίνεται διαθέσιμη από τις πηγές πληροφορίας. Παρ' όλο που ο διαμεσολαβητής θα μπορούσε να «στεγαστεί» τόσο στο Πολυτεχνείο Κρήτης όσο και στο Πανεπιστήμιο Κρήτης, για το υπόλοιπο του τρέχοντος κεφαλαίου θα θεωρηθεί ότι στα πλαίσια της περιγραφόμενης εφαρμογής ο διαμεσολαβητής θα στεγάζεται στο Πολυτεχνείο Κρήτης.

Οι πηγές πληροφορίας που περιλαμβάνει η προτεινόμενη αρχιτεκτονική της εφαρμογής είναι οι εξής:

- Η πηγή πληροφορίας του Πολυτεχνείου Κρήτης, που είναι μια πηγή πληροφορίας που περιέχει δομημένα έγγραφα και στεγάζεται στο Πολυτεχνείο Κρήτης. Το περιεχόμενό της απαρτίζεται από διπλωματικές εργασίες, μεταπτυχιακές και διδακτορικές διατριβές, επιστημονικά άρθρα, τεχνικές αναφορές, παρουσιάσεις και παραπομπές στο Διαδίκτυο.
- Η πηγή πληροφορίας του Πανεπιστημίου Κρήτης, που είναι μια πηγή πληροφορίας που περιέχει δομημένα έγγραφα και στεγάζεται στο Πανεπιστήμιο Κρήτης. Το περιεχόμενό της είναι αντίστοιχο με αυτό της πηγής πληροφορίας του Πολυτεχνείου Κρήτης: Διπλωματικές εργασίες, μεταπτυχιακές και διδακτορικές διατριβές, επιστημονικά άρθρα, τεχνικές αναφορές, παρουσιάσεις και παραπομπές στο Διαδίκτυο.
- Η πηγή πληροφορίας του εκδοτικού οίκου «Πανεπιστημιακές Εκδόσεις Κρήτης», που είναι μια πηγή πληροφορίας που περιέχει παραδοσιακά δεδομένα και στεγάζεται στον εκδοτικό οίκο «Πανεπιστημιακές Εκδόσεις Κρήτης». Το περιεχόμενό της απαρτίζεται από πληροφορία για την περιγραφή των τίτλων (βιβλίων, περιοδικών κλπ.) που θα γίνονται διαθέσιμοι απ' αυτήν καθώς και τις συνόψεις των τίτλων αυτών.

Καθώς οι πηγές πληροφορίας του Πολυτεχνείου Κρήτης και του Πανεπιστημίου Κρήτης είναι πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, ενώ αναμένεται σε κάποιο ποσοστό να περιέχουν υλικό που εντάσσεται στις ίδιες κατηγορίες περιεχομένου⁵⁰, ο ίδιος πράκτορας είναι υπεύθυνος για την ενημέρωσή τους για έγγραφα που γίνονται διαθέσιμα και το περιεχόμενό τους σχετίζεται με αυτό των πηγών πληροφορίας. Αντίθετα, ο εκδοτικός οίκος «Πανεπιστημιακές Εκδόσεις Κρήτης» ενημερώνεται μέσω ενός πράκτορα που είναι υπεύθυνος μόνο για την ενημέρωσή του.

Οι χρήστες της εφαρμογής προσπελαίνουν τις υπηρεσίες που παρέχονται απ' αυτήν μέσω ενός προγράμματος πλοήγησης του Παγκοσμίου Ιστού, το οποίο χρησιμοποιούν για να αλληλεπιδράσουν με το διαμεσολαβητή.

8.3. Λειτουργία της Εφαρμογής

Σ' αυτό το τμήμα περιγράφεται η λειτουργία της εφαρμογής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» που έρχεται να καλύψει τις ανάγκες των ακαδημαϊκών κοινοτήτων των ελληνικών ΑΕΙ για τον εντοπισμό πληροφορίας σχετικής με τα ερευνητικά και επιστημονικά θέματα με τα οποία τα μέλη τους ασχολούνται.

Για να γίνει κατανοητή η λειτουργία της εφαρμογής (και κατ' επέκταση και του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», στη συνέχεια περιγράφεται όχι μόνο η αλληλεπίδραση της εφαρμογής με τους χρήστες της, αλλά η λειτουργία της εφαρμογής σε όλες τις φάσεις: Στην ενότητα 8.3.1 περιγράφονται η *Αρχικοποίηση & Εξέλιξη των Βάσεων Δεδομένων των Πηγών Πληροφορίας*, στην ενότητα 8.3.2 γίνεται αναφορά στην

⁵⁰ Για παράδειγμα, είναι πολύ πιθανό υλικό που προέρχεται από τα τμήματα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης και Ηλεκτρονικής & Μηχανικών Η.Υ. του Πολυτεχνείου Κρήτης να εντάσσεται στις ίδιες κατηγορίες περιεχομένου (π.χ. Βάσεις Δεδομένων, Δίκτυα Υπολογιστών, Τεχνητή Νοημοσύνη κλπ.).

Ενημέρωση των Πηγών Πληροφορίας και στην ενότητα 8.3.3 περιγράφεται η Αλληλεπίδραση με τους Χρήστες.

8.3.1. Αρχικοποίηση & Εξέλιξη των Βάσεων Δεδομένων των Πηγών Πληροφορίας

Το θέμα αυτής της ενότητας είναι η περιγραφή των βάσεων δεδομένων των πηγών πληροφορίας που παίρνουν μέρος στην εφαρμογή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» μετά την αρχικοποίησή τους και κατά την εξέλιξη του σχήματός τους.

Αμέσως μετά την αρχικοποίησή τους, οι πηγές πληροφορίας του Πολυτεχνείου Κρήτης και του Πανεπιστημίου Κρήτης εξάγουν μέσω του διαμεσολαβητή αντίγραφα του μοντέλου πληροφορίας που φαίνεται στο RDF Σχήμα 6-1. Αντίστοιχα, μετά την αρχικοποίησή της η πηγή πληροφορίας του εκδοτικού οίκου «Πανεπιστημιακές Εκδόσεις Κρήτης» εξάγει μέσω του διαμεσολαβητή ένα αντίγραφο του μοντέλου πληροφορίας που φαίνεται στο RDF Σχήμα 6-2.

Έστω τώρα ότι ο πρώτος DTD που υποστηρίζεται από την πηγή πληροφορίας του Πολυτεχνείου Κρήτης είναι ο DTD 6-2, που περιγράφει τη δομή Άρθρων. Μετά από αλληλεπίδραση του διαχειριστή της με το εργαλείο SupportDTD επιλέγεται οι κανόνες διαχείρισης, αναπαράστασης και δεικτοδότησης των στιγμιοτύπων της κλάσης Άρθρο, που αναπαριστά τα έγγραφα που υπακούουν στο DTD 6-2 να είναι αυτοί που φαίνονται στο Παράδειγμα 7-1 και η δομή της να εξάγεται μέσω της γλώσσας της πηγής πληροφορίας με τη μορφή που φαίνεται στο RDF Σχήμα 7-2. Στη συνέχεια, έστω ότι υποστηρίζεται από την πηγή πληροφορίας ο DTD 8-1, που περιγράφει τη δομή Τεχνικών Αναφορών.

```
<!ELEMENT ΑΝΑΦΟΡΑ (ΤΙΤΛΟΣ, ΣΥΓΓΡΑΦΕΑΣ, ΣΥΝΟΨΗ, ΕΝΟΤΗΤΑ*)>
<!ELEMENT ΣΥΓΓΡΑΦΕΑΣ EMPTY>
<!ATTLIST ΣΥΓΓΡΑΦΕΑΣ      ΟΝΟΜΑ CDATA #IMPLIED
                        ΕΠΩΝΥΜΟ CDATA #IMPLIED>
<!ELEMENT ΕΝΟΤΗΤΑ (ΠΑΡΑΓΡΑΦΟΣ*, ΥΠΟΕΝΟΤΗΤΑ*)>
<!ELEMENT ΥΠΟΕΝΟΤΗΤΑ (ΠΑΡΑΓΡΑΦΟΣ*)>
<!ELEMENT ΠΑΡΑΓΡΑΦΟΣ (#PCDATA|EM)*>
<!ELEMENT EM (#PCDATA)>
<!ELEMENT ΣΥΝΟΨΗ (#PCDATA)>
<!ELEMENT ΤΙΤΛΟΣ (#PCDATA)>
```

DTD 8-1: Ο DTD Αναφορά.dtd που περιγράφει τη δομή και τη σύνταξη Αναφορών

Ο διαχειριστής της πηγής πληροφορίας στην περίπτωση αυτή έστω ότι αποφασίζει ότι οι κανόνες βάσει των οποίων θα αποθηκεύονται τα έγγραφα που υπακούουν στο DTD 8-1, είναι οι εξής:

- Η κλάση Αναφορά αποτελεί υποκλάση της κλάσης Έγγραφο.
- Επιτρέπεται η τμηματική αποθήκευση των εγγράφων που υπακούουν στο DTD Αναφορά.dtd, τα οποία είναι και στιγμιότυπα της κλάσης Αναφορά, ανεξαρτήτως του μεγέθους τους.

- Από τα στιγμιότυπα της κλάσης Ενότητα επιτρέπεται να αποθηκεύονται τμηματικά μόνο αυτά που το μέγεθός τους υπερβαίνει τα 60 KB.
- Τα στιγμιότυπα των κλάσεων Τίτλος, Συγγραφέας Παράγραφος, Υποενότητα και Σύνοψη αποθηκεύονται πάντα μαζί με το αντικείμενο στο οποίο περιέχονται.
- Τα στιγμιότυπα της κλάσης Τίτλος, που αποτελούν τους τίτλους των αντίστοιχων άρθρων, αντιστοιχούν και στο ομώνυμο στοιχείο μεταπληροφορίας και αποθηκεύονται ως χαρακτηριστικά.
- Τα στιγμιότυπα της κλάσης Σύνοψη, που αποτελούν τις συνόψεις των αντίστοιχων άρθρων, αντιστοιχούν και στο στοιχείο μεταπληροφορίας «Περιγραφή» και αποθηκεύονται ως χαρακτηριστικά.
- Τα στιγμιότυπα της κλάσης Συγγραφέας, που αποτελούν τους συγγραφείς των αντίστοιχων άρθρων, αντιστοιχούν και στο στοιχείο μεταπληροφορίας «Διαχειριστής» με Ρόλο «Συγγραφέας».
- Η βαρύτητα των όρων που περιέχονται στο δομικό στοιχείο Τίτλος κάθε εγγράφου είναι 50% μεγαλύτερη από αυτή των όρων που περιέχονται στα υπόλοιπα δομικά στοιχεία που απαρτίζουν το έγγραφο.
- Για τα δεικτοδοτούμενα στιγμιότυπα της αντίστοιχης κλάσης εγγράφων, εκτός από τα μεταδεδομένα τους, στη βάση δεδομένων της πηγής πληροφορίας αποθηκεύονται και τα στιγμιότυπα της κλάσης σύνοψη.
- Δεν εξάγονται στο μοντέλο πληροφορίας οι κλάσεις που αντιστοιχούν στα στοιχεία Τίτλος και em.

Στην περίπτωση αυτή, οι κανόνες διαχείρισης, αναπαράστασης και δεικτοδότησης των στιγμιότυπων της κλάσης Αναφορά, που αναπαριστά τα έγγραφα που υπακούουν στο DTD 8-1 είναι αυτοί που φαίνονται στο Παράδειγμα 8-1 και η δομή της εξάγεται μέσω της γλώσσας της πηγής πληροφορίας με τη μορφή που φαίνεται στο RDF Σχήμα 8-1.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema-1990303"
  xmlns:KAΔΔΑ="http://irida.ced.tuc.gr/KAΔΔΑ.rdfs">
  <rdf:Description about="">
    <KAΔΔΑ:σύνολο_κανόνων όνομα="Αναφορά.dtd" τύπος="έγγραφο">
      <KAΔΔΑ:σύνολο_κανόνων_μεταδεδομένων όνομα="κανόνες_μδ_αναφορά">
        <KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων
          όνομα="εξαγωγή_τίτλου" αποτέλεσμα="Τίτλος"
          δεδομένα="Τίτλος">
          <KAΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
            παράμετροι="">
          </KAΔΔΑ:μέθοδος>
        </KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων>
        <KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων
          όνομα="εξαγωγή_περιγραφής" αποτέλεσμα="Περιγραφή"
          δεδομένα="Σύνοψη">
          <KAΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
            παράμετροι="">
```

```

</ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_διαχειριστή" αποτέλεσμα="Διαχειριστής"
  δεδομένα="Συγγραφέας">
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_ρόλου" αποτέλεσμα="Διαχειριστής.Ρόλος"
  δεδομένα="Συγγραφέας">
<ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
  παράμετροι="">
</ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_ονόματος" αποτέλεσμα="Διαχειριστής.Όνομα"
  δεδομένα="Συγγραφέας.Όνομα">
<ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
  παράμετροι="">
</ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_επωνύμου" αποτέλεσμα="Διαχειριστής.Επώνυμο"
  δεδομένα="Συγγραφέας.Επώνυμο">
<ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
  παράμετροι="">
</ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
</ΚΑΔΔΑ:σύνολο_κανόνων_μεταδεδομένων>
<ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων όνομα="κανόνες_στ_αναφορά">
<ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόννας_αναπαράστασης_στ_αναφορά "
  όνομα_στοιχείου="Αναφορά" όριο_διάσπασης="0"
  ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόννας_αναπαράστασης_στ_τίτλος"
  όνομα_στοιχείου="Τίτλος" όριο_διάσπασης="-1"
  ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0,5">
</ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόννας_αναπαράστασης_στ_συγγραφέας"
  όνομα_στοιχείου="Συγγραφέας" όριο_διάσπασης="-1"
  ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόννας_αναπαράστασης_στ_σύνοψη"
  όνομα_στοιχείου="Σύνοψη" όριο_διάσπασης="-1"
  ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
  όνομα="κανόννας_αναπαράστασης_στ_ενότητα"
  όνομα_στοιχείου="Ενότητα" όριο_διάσπασης="0"

```

```

        ίδιο_όριο_διάστασης="60KB" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόνας_αναπαρ_στ_υποενότητα"
    όνομα_στοιχείου="Υποενότητα" όριο_διάστασης="-1"
    ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόνας_αναπαρ_στ_παράγραφος"
    όνομα_στοιχείου="Παράγραφος" όριο_διάστασης="-1"
    ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόνας_αναπαράστ_στ_em"
    όνομα_στοιχείου="em" όριο_διάστασης="-1"
    ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόνας_αναπαρ_στ_παράγραφος"
    όνομα_στοιχείου="Παράγραφος" όριο_διάστασης="-1"
    ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόνας_αναπαράστασης_στ_em"
    όνομα_στοιχείου="em" όριο_διάστασης="-1"
    ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων>
<ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης όνομα="σ_κ_δ_αναφορά">
    <ΚΑΔΔΑ:κανόνας_δεικτοδότησης_εγγράφων όνομα="σύνοψη"
        όνομα_στοιχείου="σύνοψη">
    </ΚΑΔΔΑ:κανόνας_δεικτοδότησης_εγγράφων>
</ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης>
</ΚΑΔΔΑ:σύνολο_κανόνων>
</rdf:Description>
</rdf:RDF>

```

Παράδειγμα 8-1: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 8-1

```

<?xml version='1.0'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema"
    xmlns:rdfs="http://irida.ced.tuc.gr/RDF/ΠηγήΠληροφορίαςΕγγράφων.rdfs"
    xmlns:Αναφορά="">
<rdfs:Class ID="αναφορά">
    <rdfs:label>αναφορά</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
        ΠηγήΠληροφορίαςΕγγράφων.rdfs#έγγραφο"/>
</rdfs:Class>

```

```

<rdf:Property ID="περιγραφή">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#αναφορά" />
</rdf:Property>
<rdfs:Class ID="συγγραφέας">
  <rdfs:label>συγγραφέας</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
  ΠηγήΠληροφορίαςΕγγράφων.rdfs#αντικείμενο"/>
  <rdf:domain rdf:resource="#αναφορά" />
</rdfs:Class>
<rdf:Property ID="όνομα">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#συγγραφέας" />
</rdf:Property>
<rdf:Property ID="επώνυμο">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#συγγραφέας" />
</rdf:Property>
<rdf:Property ID="τίτλος">
  <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
  <rdf:domain rdf:resource="#αναφορά" />
</rdf:Property>
<rdfs:Class ID="σύνοψη">
  <rdfs:label>σύνοψη</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
  ΠηγήΠληροφορίαςΕγγράφων.rdfs#τιμήμα_εγγράφου"/>
  <rdf:domain rdf:resource="#αναφορά" />
</rdfs:Class>
<rdfs:Class ID="ενότητα">
  <rdfs:label>ενότητα</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
  ΠηγήΠληροφορίαςΕγγράφων.rdfs#τιμήμα_εγγράφου"/>
  <rdf:domain rdf:resource="#αναφορά" />
</rdfs:Class>
<rdfs:Class ID="υποενότητα">
  <rdfs:label>υποενότητα</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
  ΠηγήΠληροφορίαςΕγγράφων.rdfs#τιμήμα_εγγράφου"/>
  <rdf:domain rdf:resource="#ενότητα" />
</rdfs:Class>
<rdfs:Class ID="παράγραφος">
  <rdfs:label>παράγραφος</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
  ΠηγήΠληροφορίαςΕγγράφων.rdfs#τιμήμα_εγγράφου"/>
  <rdf:domain rdf:resource="#ενότητα" />
  <rdf:domain rdf:resource="#υποενότητα" />
</rdfs:Class>
</rdf:RDF>

```

RDF Σχήμα 8-1: Το RDF Σχήμα που αντιστοιχεί στο DTD 8-1

Μετά τις παραπάνω ενέργειες, στην πηγή πληροφορίας του Πολυτεχνείου Κρήτης μπορούν να εισαχθούν και να δεικτοδοτηθούν στιγμιότυπα των κλάσεων εγγράφων Άρθρο και Αναφορά, ενώ το μοντέλο πληροφορίας της πηγής πληροφορίας, όπως αυτό εξάγεται μέσω του διαμεσολαβητή στους χρήστες απαρτίζεται από το RDF σχήμα που προκύπτει αν συνενωθεί το RDF Σχήμα 6-1 με το RDF Σχήμα 7-2 και το RDF Σχήμα 8-1.

Σε ότι αφορά την πηγή πληροφορίας του Πανεπιστημίου Κρήτης, έστω ότι πρώτος υποστηρίζεται ο DTD 6-2, που περιγράφει τη δομή Άρθρων. Μετά από αλληλεπίδραση του διαχειριστή της με το εργαλείο SupportDTD επιλέγεται οι κανόνες διαχείρισης, αναπαράστασης και δεικτοδότησης των στιγμιοτύπων της κλάσης Άρθρο να είναι ελαφρώς διαφορετικοί απ' αυτούς που ισχύουν στο Πολυτεχνείο Κρήτης: Η βαρύτητα των όρων που περιέχονται στο δομικό στοιχείο Τίτλος κάθε εγγράφου είναι 25% μεγαλύτερη από αυτή των όρων που περιέχονται στα υπόλοιπα δομικά στοιχεία που απαρτίζουν το έγγραφο, ενώ η βαρύτητα των όρων που περιέχονται στα δομικά στοιχεία Σύνοψη και em είναι ίδια με αυτή των όρων που περιέχονται στα υπόλοιπα δομικά στοιχεία που απαρτίζουν το έγγραφο. Στην περίπτωση αυτή, η δομή της κλάσης Άρθρο συνεχίζει να εξάγεται μέσω της γλώσσας της πηγής πληροφορίας με τη μορφή που φαίνεται στο RDF Σχήμα 7-2, αλλά οι κανόνες διαχείρισης, αναπαράστασης και δεικτοδότησης των στιγμιοτύπων της κλάσης είναι αυτοί που φαίνονται στο Παράδειγμα 8-2.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema-19990303"
  xmlns:KAΔΔΑ="http://irida.ced.tuc.gr/KAΔΔΑ.rdfs">
  <rdf:Description about="">
    <KAΔΔΑ:σύνολο_κανόνων όνομα="Άρθρο.dtd" τύπος="έγγραφο">
      <KAΔΔΑ:σύνολο_κανόνων_μεταδεδομένων όνομα="κανόνες_μδ_άρθρων">
        <KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων
          όνομα="εξαγωγή_τίτλου" αποτέλεσμα="Τίτλος"
          δεδομένα="Τίτλος">
          <KAΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
            παράμετροι="">
          </KAΔΔΑ:μέθοδος>
        </KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων>
        <KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων
          όνομα="εξαγωγή_περιγραφής" αποτέλεσμα="Περιγραφή"
          δεδομένα="Σύνοψη">
          <KAΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
            παράμετροι="">
          </KAΔΔΑ:μέθοδος>
        </KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων>
        <KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων
          όνομα="εξαγωγή_διαχειριστή" αποτέλεσμα="Διαχειριστής"
          δεδομένα="Συγγραφέας">
          <KAΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων
            όνομα="εξαγωγή_ρόλου" αποτέλεσμα="Διαχειριστής.Ρόλος"
            δεδομένα="Συγγραφέας">
            <KAΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
              παράμετροι="">
            </KAΔΔΑ:μέθοδος>
```

```

</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_ονόματος" αποτέλεσμα="Διαχειριστής.Όνομα"
  δεδομένα="Συγγραφέας.Όνομα">
  <ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
    παράμετροι="">
  </ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_επωνύμου" αποτέλεσμα="Διαχειριστής.Επώνυμο"
  δεδομένα="Συγγραφέας.Επώνυμο">
  <ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
    παράμετροι="">
  </ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
</ΚΑΔΔΑ:σύνολο_κανόνων_μεταδεδομένων>
<ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων όνομα="κανόνες_στ_άρθρων">
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαράστασης_στ_άρθρο"
    όνομα_στοιχείου="Άρθρο" όριο_διάσπασης="0"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαράστασης_στ_τίτλος"
    όνομα_στοιχείου="Τίτλος" όριο_διάσπασης="-1"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0,25">
  </ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου>
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαράστασης_στ_συγγραφέας"
    όνομα_στοιχείου="Συγγραφέας" όριο_διάσπασης="-1"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
  </ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου>
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαράστασης_στ_σύνοψη"
    όνομα_στοιχείου="Σύνοψη" όριο_διάσπασης="-1"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
  </ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου>
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαράστασης_στ_ενότητα"
    όνομα_στοιχείου="Ενότητα" όριο_διάσπασης="200KB"
    ίδιο_όριο_διάσπασης="50KB" διαφορά_βάρους="0">
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαρ_στ_υποενότητα"
    όνομα_στοιχείου="Υποενότητα" όριο_διάσπασης="-1"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαρ_στ_παράγραφος"
    όνομα_στοιχείου="Παράγραφος" όριο_διάσπασης="-1"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαράστ_στ_em"

```

```

        όνομα_στοιχείου="em" όριο_διάσπασης="-1"
        ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόνας_αναπαρ_στ_παράγραφος"
    όνομα_στοιχείου="Παράγραφος" όριο_διάσπασης="-1"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόνας_αναπαράστασης_στ_em"
    όνομα_στοιχείου="em" όριο_διάσπασης="-1"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων>
<ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης όνομα="σ_κ_δ_άρθρο">
    <ΚΑΔΔΑ:κανόνας_δεικτοδότησης_εγγράφων όνομα="σύνοψη"
        όνομα_στοιχείου="σύνοψη">
    </ΚΑΔΔΑ:κανόνας_δεικτοδότησης_εγγράφων>
</ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης>
</ΚΑΔΔΑ:σύνολο_κανόνων>
</rdf:Description>
</rdf:RDF>

```

Παράδειγμα 8-2: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 6-2

Έστω τώρα ότι ο επόμενος DTD που υποστηρίζεται από την πηγή πληροφορίας του Πανεπιστημίου Κρήτης είναι ο DTD 8-2, ο οποίος περιγράφει τη δομή και τη σύνταξη Μεταπτυχιακών και Διδακτορικών Διατριβών.

```

<!ELEMENT ΔΙΑΤΡΙΒΗ (ΤΙΤΛΟΣ,ΣΥΓΓΡΑΦΕΑΣ,ΣΥΝΟΨΗ,ΚΕΦΑΛΑΙΟ*)>
<!ELEMENT ΣΥΓΓΡΑΦΕΑΣ EMPTY>
<!ATTLIST ΣΥΓΓΡΑΦΕΑΣ    ΟΝΟΜΑ CDATA #IMPLIED
                        ΕΠΩΝΥΜΟ CDATA #IMPLIED>
<!ELEMENT ΚΕΦΑΛΑΙΟ (ΠΑΡΑΓΡΑΦΟΣ*, ΕΝΟΤΗΤΑ*)>
<!ELEMENT ΕΝΟΤΗΤΑ (ΠΑΡΑΓΡΑΦΟΣ*, ΥΠΟΕΝΟΤΗΤΑ*)>
<!ELEMENT ΥΠΟΕΝΟΤΗΤΑ (ΠΑΡΑΓΡΑΦΟΣ*)>
<!ELEMENT ΠΑΡΑΓΡΑΦΟΣ (#PCDATA|EM)*>
<!ELEMENT EM (#PCDATA)>
<!ELEMENT ΣΥΝΟΨΗ (#PCDATA)>
<!ELEMENT ΤΙΤΛΟΣ (#PCDATA)>

```

DTD 8-2: Ο DTD Διατριβή.dtd που περιγράφει τη δομή και τη σύνταξη Διατριβών

Ο διαχειριστής της πηγής πληροφορίας στην περίπτωση αυτή έστω ότι αποφασίζει ότι οι κανόνες βάσει των οποίων θα αποθηκεύονται τα έγγραφα που υπακούουν στο DTD 8-2, είναι οι εξής:

- Η κλάση Διατριβή αποτελεί υποκλάση της κλάσης Έγγραφο.
- Επιτρέπεται η τμηματική αποθήκευση των εγγράφων που υπακούουν στο DTD Διατριβή.dtd, τα οποία είναι και στιγμιότυπα της κλάσης Διατριβή, ανεξαρτήτως του μεγέθους τους.
- Όλα τα στιγμιότυπα της κλάσης Κεφάλαιο θα αποθηκεύονται τμηματικά.
- Από τα στιγμιότυπα της κλάσης Ενότητα επιτρέπεται να αποθηκεύονται τμηματικά μόνο αυτά που το μέγεθός τους υπερβαίνει τα 100 KB.
- Τα στιγμιότυπα των κλάσεων Τίτλος, Παράγραφος, Υποενότητα και Σύνοψη αποθηκεύονται πάντα μαζί με το αντικείμενο στο οποίο περιέχονται.
- Τα στιγμιότυπα της κλάσης Τίτλος, που αποτελούν τους τίτλους των αντίστοιχων άρθρων, αντιστοιχούν και στο ομώνυμο στοιχείο μεταπληροφορίας και αποθηκεύονται ως χαρακτηριστικά.
- Τα στιγμιότυπα της κλάσης Σύνοψη, που αποτελούν τις συνόψεις των αντίστοιχων άρθρων, αντιστοιχούν και στο στοιχείο μεταπληροφορίας «Περιγραφή» και αποθηκεύονται ως χαρακτηριστικά.
- Τα στιγμιότυπα της κλάσης Συγγραφέας, που αποτελούν τους συγγραφείς των αντίστοιχων άρθρων, αντιστοιχούν και στο στοιχείο μεταπληροφορίας «Διαχειριστής» με Ρόλο «Συγγραφέας».
- Η βαρύτητα των όρων που περιέχονται στο δομικό στοιχείο Τίτλος κάθε εγγράφου είναι 50% μεγαλύτερη από αυτή των όρων που περιέχονται στα υπόλοιπα δομικά στοιχεία που απαρτίζουν το έγγραφο.
- Για τα δεικτοδοτούμενα στιγμιότυπα της αντίστοιχης κλάσης εγγράφων, εκτός από τα μεταδεδομένα τους, στη βάση δεδομένων της πηγής πληροφορίας αποθηκεύονται και τα στιγμιότυπα της κλάσης σύνοψη.
- Δεν εξάγονται στο μοντέλο πληροφορίας οι κλάσεις που αντιστοιχούν στα στοιχεία Τίτλος και em

Στην περίπτωση αυτή, οι κανόνες διαχείρισης, αναπαράστασης και δεικτοδότησης των στιγμιότυπων της κλάσης Διατριβή, που αναπαριστά τα έγγραφα που υπακούουν στο DTD 8-2 είναι αυτοί που φαίνονται στο Παράδειγμα 8-3 και η δομή της εξάγεται μέσω της γλώσσας της πηγής πληροφορίας με τη μορφή που φαίνεται στο RDF Σχήμα 8-2.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema-19990303"
  xmlns:ΚΑΔΔΑ="http://irida.ced.tuc.gr/ΚΑΔΔΑ.rdfs">
  <rdf:Description about="">
    <ΚΑΔΔΑ:σύνολο_κανόνων όνομα="Διατριβή.dtd" τύπος="έγγραφο">
      <ΚΑΔΔΑ:σύνολο_κανόνων_μεταδεδομένων όνομα="κανόνες_μδ_διατριβή">
        <ΚΑΔΔΑ:κανόνας_εξαγωγής_μεταδεδομένων
          όνομα="εξαγωγή_τίτλου" αποτέλεσμα="Τίτλος"
          δεδομένα="Τίτλος">
          <ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
            παράμετροι="">
```

```

</ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_περιγραφής" αποτέλεσμα="Περιγραφή"
  δεδομένα="Σύνοψη">
  <ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
    παράμετροι="">
  </ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_διαχειριστή" αποτέλεσμα="Διαχειριστής"
  δεδομένα="Συγγραφέας">
  <ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
    όνομα="εξαγωγή_ρόλου" αποτέλεσμα="Διαχειριστής.Ρόλος"
    δεδομένα="">
  <ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
    παράμετροι="">
  </ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_ονόματος" αποτέλεσμα="Διαχειριστής.Όνομα"
  δεδομένα="Συγγραφέας.Όνομα">
  <ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
    παράμετροι="">
  </ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
<ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων
  όνομα="εξαγωγή_επωνύμου" αποτέλεσμα="Διαχειριστής.Επώνυμο"
  δεδομένα="Συγγραφέας.Επώνυμο">
  <ΚΑΔΔΑ:μέθοδος όνομα="αντιγραφή" τελεστής="αντιγραφή"
    παράμετροι="">
  </ΚΑΔΔΑ:μέθοδος>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
</ΚΑΔΔΑ:κανόννας_εξαγωγής_μεταδεδομένων>
</ΚΑΔΔΑ:σύνολο_κανόνων_μεταδεδομένων>
<ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων όνομα="κανόνες_στ_διατριβή">
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαράστασης_στ_διατριβή "
    όνομα_στοιχείου="Διατριβή" όριο_διάσπασης="0"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαράστασης_στ_τίτλος"
    όνομα_στοιχείου="Τίτλος" όριο_διάσπασης="-1"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0,5">
  </ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου>
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου
    όνομα="κανόννας_αναπαράστασης_στ_συγγραφέας"
    όνομα_στοιχείου="Συγγραφέας" όριο_διάσπασης="-1"
    ίδιο_όριο_διάσπασης="-1" διαφορά_βάρους="0">
  </ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου>
  <ΚΑΔΔΑ:κανόννας_αναπαράστασης_στοιχείου_εγγράφου

```

```

        όνομα="κανόνας_αναπαράστασης_στ_σύννοψη"
        όνομα_στοιχείου="Σύννοψη" όριο_διάστασης="-1"
        ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
        όνομα="κανόνας_αναπαράστασης_στ_κεφάλαιο"
        όνομα_στοιχείου="Κεφάλαιο" όριο_διάστασης="0"
        ίδιο_όριο_διάστασης="0" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
        όνομα="κανόνας_αναπαράστασης_στ_ενότητα"
        όνομα_στοιχείου="Ενότητα" όριο_διάστασης="0"
        ίδιο_όριο_διάστασης="100KB" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
        όνομα="κανόνας_αναπαρ_στ_υποενότητα"
        όνομα_στοιχείου="Υποενότητα" όριο_διάστασης="-1"
        ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
        όνομα="κανόνας_αναπαρ_στ_παράγραφος"
        όνομα_στοιχείου="Παράγραφος" όριο_διάστασης="-1"
        ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
        όνομα="κανόνας_αναπαράστ_στ_em"
        όνομα_στοιχείου="em" όριο_διάστασης="-1"
        ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
        όνομα="κανόνας_αναπαρ_στ_παράγραφος"
        όνομα_στοιχείου="Παράγραφος" όριο_διάστασης="-1"
        ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου
        όνομα="κανόνας_αναπαράστασης_στ_em"
        όνομα_στοιχείου="em" όριο_διάστασης="-1"
        ίδιο_όριο_διάστασης="-1" διαφορά_βάρους="0">
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_εγγράφου>
</ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων>
<ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης όνομα="σ_κ_δ_διατριβή">
<ΚΑΔΔΑ:κανόνας_δεικτοδότησης_εγγράφων όνομα="σύννοψη"
        όνομα_στοιχείου="σύννοψη">
</ΚΑΔΔΑ:κανόνας_δεικτοδότησης_εγγράφων>
</ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης>
</ΚΑΔΔΑ:σύνολο_κανόνων>
</rdf:Description>
</rdf:RDF>

```

Παράδειγμα 8-3: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 8-2

```

<?xml version='1.0'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema"
  xmlns:rdfs="http://irida.ced.tuc.gr/RDF/ΠηγήΠληροφορίαςΕγγράφων.rdfs"
  xmlns:Διατριβή="">
  <rdfs:Class ID="Διατριβή">
    <rdfs:label>Διατριβή</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
      ΠηγήΠληροφορίαςΕγγράφων.rdfs#έγγραφο"/>
  </rdfs:Class>
  <rdf:Property ID="περιγραφή">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#αναφορά" />
  </rdf:Property>
  <rdfs:Class ID="συγγραφέας">
    <rdfs:label>συγγραφέας</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
      ΠηγήΠληροφορίαςΕγγράφων.rdfs#αντικείμενο"/>
    <rdfs:domain rdf:resource="#αναφορά" />
  </rdfs:Class>
  <rdf:Property ID="όνομα">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#συγγραφέας" />
  </rdf:Property>
  <rdf:Property ID="επώνυμο">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#συγγραφέας" />
  </rdf:Property>
  <rdf:Property ID="τίτλος">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdfs:domain rdf:resource="#Διατριβή" />
  </rdf:Property>
  <rdfs:Class ID="σύνοψη">
    <rdfs:label>σύνοψη</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
      ΠηγήΠληροφορίαςΕγγράφων.rdfs#τμήμα_εγγράφου"/>
    <rdfs:domain rdf:resource="#Διατριβή" />
  </rdfs:Class>
  <rdfs:Class ID="κεφάλαιο">
    <rdfs:label>κεφάλαιο</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
      ΠηγήΠληροφορίαςΕγγράφων.rdfs#τμήμα_εγγράφου"/>
    <rdfs:domain rdf:resource="#Διατριβή" />
  </rdfs:Class>
  <rdfs:Class ID="ενότητα">
    <rdfs:label>ενότητα</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
      ΠηγήΠληροφορίαςΕγγράφων.rdfs#τμήμα_εγγράφου"/>
    <rdfs:domain rdf:resource="#κεφάλαιο" />
  </rdfs:Class>
  <rdfs:Class ID="υποενότητα">
    <rdfs:label>υποενότητα</rdfs:label>

```

```

<rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
ΠηγήΠληροφορίαςΕγγράφων.rdfs#τιμήμα_εγγράφου"/>
<rdf:domain rdf:resource="#ενότητα" />
</rdfs:Class>
<rdfs:Class ID="παράγραφος">
  <rdfs:label>παράγραφος</rdfs:label>
  <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
ΠηγήΠληροφορίαςΕγγράφων.rdfs#τιμήμα_εγγράφου"/>
  <rdf:domain rdf:resource="#κεφάλαιο" />
  <rdf:domain rdf:resource="#ενότητα" />
  <rdf:domain rdf:resource="#υποενότητα" />
</rdfs:Class>
</rdf:RDF>

```

RDF Σχήμα 8-2: Το RDF Σχήμα που αντιστοιχεί στο DTD 8-2

Μετά τις παραπάνω ενέργειες, στην πηγή πληροφορίας του Πανεπιστημίου Κρήτης μπορούν να εισαχθούν και να δεικτοδοτηθούν στιγμιότυπα των κλάσεων εγγράφων Άρθρο και Διατριβή, ενώ το μοντέλο πληροφορίας της πηγής πληροφορίας, όπως αυτό εξάγεται μέσω του διαμεσολαβητή στους χρήστες απαρτίζεται από το RDF σχήμα που προκύπτει αν συνενωθεί το RDF Σχήμα 6-1 με το RDF Σχήμα 7-2 και το RDF Σχήμα 8-2.

Έστω τέλος ότι ο πρώτος DTD που υποστηρίζεται από την πηγή πληροφορίας του εκδοτικού οίκου «Πανεπιστημιακές Εκδόσεις Κρήτης» είναι ο DTD 8-3, που αναπαριστά τις περιγραφές βιβλίων που διατίθενται από τον εκδοτικό οίκο.

```

<!ELEMENT ΠΕΡΙΓΡΑΦΗ_ΒΙΒΛΙΟΥ (ΤΙΤΛΟΣ,ΣΥΓΓΡΑΦΕΑΣ,ΣΥΝΟΨΗ)>
<!ATTLIST ΠΕΡΙΓΡΑΦΗ_ΒΙΒΛΙΟΥ      Τιμή CDATA   #REQUIRED
                                   ISBN CDATA   #REQUIRED >
<!ELEMENT ΣΥΝΟΨΗ                (#PCDATA)>
<!ELEMENT ΣΥΓΓΡΑΦΕΑΣ            EMPTY>
<!ATTLIST ΣΥΓΓΡΑΦΕΑΣ            ΟΝΟΜΑ CDATA   #IMPLIED
                                   ΕΠΩΝΥΜΟ CDATA #IMPLIED>
<!ELEMENT ΤΙΤΛΟΣ                 (#PCDATA)>

```

DTD 8-3: Ο DTD ΠεριγραφήΒιβλίου.dtd που περιγράφει τη δομή και τη σύνταξη Περιγραφών Βιβλίων

Έστω τώρα ότι ο διαχειριστής της πηγής πληροφορίας κάνει τις ακόλουθες επιλογές σχετικά με την αναπαράσταση και τη διαχείριση των αντικειμένων που η δομή τους περιγράφεται από το DTD 8-3 είναι οι εξής:

- Τα αντικείμενα που η δομή τους περιγράφεται από το DTD 8-3 αναπαρίστανται ως στιγμιότυπα της κλάσης «Περιγραφή_Βιβλίου», που είναι υποκλάση της κλάσης «Αντικείμενο».
- Τα στιγμιότυπα του στοιχείου συγγραφέας αναπαρίστανται ως στιγμιότυπα της κλάσης «Συγγραφέας», που είναι υποκλάση της κλάσης «Αντικείμενο».
- Τα στοιχεία τίτλος και σύνοψη αναπαρίστανται ως χαρακτηριστικά των στιγμιότυπων της κλάσης «Περιγραφή_Βιβλίου».

- Για τα δεικτοδοτούμενα στιγμιότυπα της κλάσης «Περιγραφή_Βιβλίου» αποθηκεύονται στη βάση δεδομένων της πηγής πληροφορίας όλα τα χαρακτηριστικά τους εκτός από τη σύνοψη.

Στην περίπτωση αυτή, οι κανόνες διαχείρισης, αναπαράστασης και δεικτοδότησης των στιγμιότυπων της κλάσης Περιγραφή_Βιβλίου, που αναπαριστά τα έγγραφα που υπακούουν στο DTD 8-3 είναι αυτοί που φαίνονται στο Παράδειγμα 8-4 και η δομή της εξάγεται μέσω της γλώσσας της πηγής πληροφορίας με τη μορφή που φαίνεται στο RDF Σχήμα 8-3.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema-19990303"
  xmlns:ΚΑΔΔΑ="http://irida.ced.tuc.gr/RDF/ΚΑΔΔΑ.rdfs">
  <rdf:Description about="">
    <ΚΑΔΔΑ:σύνολο_κανόνων όνομα="ΠεριγραφήΒιβλίου.dtd"
      τύπος="παραδοσιακά δεδομένα">
      <ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων όνομα="κ_στ_περιγραφή_βιβλίου">
        <ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_δεδομένων
          όνομα="κανόνες_στ_περιγραφή_βιβλίου"
          όνομα_στοιχείου="περιγραφή_βιβλίου"
          μέθοδος_αναπαράστασης="κλάση">
          <ΚΑΔΔΑ:Κλάση όνομα="περιγραφή_βιβλίου"
            υπερκλάση="αντικείμενο">
          </ΚΑΔΔΑ:Κλάση>
          <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="τιμή"
            κλάση="περιγραφή_βιβλίου" τύπος="αριθμός">
          </ΚΑΔΔΑ:Χαρακτηριστικό>
          <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="ISBN"
            κλάση="περιγραφή_βιβλίου" τύπος="συμβολοσειρά">
          </ΚΑΔΔΑ:Χαρακτηριστικό>
          <ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_δεδομένων
            όνομα="κανόνες_στ_τίτλος" όνομα_στοιχείου="τίτλος"
            μέθοδος_αναπαράστασης="χαρακτηριστικό">
          <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="τίτλος"
            κλάση="περιγραφή_βιβλίου" τύπος="συμβολοσειρά">
          </ΚΑΔΔΑ:Χαρακτηριστικό>
          </ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_δεδομένων>
          <ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_δεδομένων
            όνομα="κανόνες_στ_συγγραφέας"
            όνομα_στοιχείου="συγγραφέας"
            μέθοδος_αναπαράστασης="κλάση">
          <ΚΑΔΔΑ:Κλάση όνομα="συγγραφέας"
            υπερκλάση="αντικείμενο">
          </ΚΑΔΔΑ:Κλάση>
          <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="όνομα"
            κλάση="συγγραφέας" τύπος="συμβολοσειρά">
          </ΚΑΔΔΑ:Χαρακτηριστικό>
          <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="επώνυμο"
            κλάση="συγγραφέας" τύπος="συμβολοσειρά">
          </ΚΑΔΔΑ:Χαρακτηριστικό>
          </ΚΑΔΔΑ:κανόνες_αναπαράστασης_στοιχείου_δεδομένων>
```

```

<ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων
  όνομα="κανόνες_στ_σύνοψη" όνομα_στοιχείου="σύνοψη"
  μέθοδος_αναπαράστασης="χαρακτηριστικό">
  <ΚΑΔΔΑ:Χαρακτηριστικό όνομα="σύνοψη"
    κλάση="περιγραφή_βιβλίου" τύπος="συμβολοσειρά">
    </ΚΑΔΔΑ:Χαρακτηριστικό>
  </ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων>
</ΚΑΔΔΑ:κανόνας_αναπαράστασης_στοιχείου_δεδομένων>
</ΚΑΔΔΑ:σύνολο_κανόνων_στοιχείων>
<ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης όνομα="σ_κ_μδ_περ_βιβλίου">
  <ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων όνομα="τίτλος"
    όνομα_στοιχείου_πληροφορίας="τίτλος"
    τύπος="χαρακτηριστικό">
  </ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων>
  <ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων όνομα="συγγραφέας"
    όνομα_στοιχείου_πληροφορίας="συγγραφέας"
    τύπος="χαρακτηριστικό">
  </ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων>
  <ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων όνομα="ISBN"
    όνομα_στοιχείου_πληροφορίας="ISBN"
    τύπος="χαρακτηριστικό">
  </ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων>
  <ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων όνομα="τιμή"
    όνομα_στοιχείου_πληροφορίας="τιμή"
    τύπος="χαρακτηριστικό">
  </ΚΑΔΔΑ:κανόνας_δεικτοδότησης_δεδομένων>
</ΚΑΔΔΑ:σύνολο_κανόνων_δεικτοδότησης>
</ΚΑΔΔΑ:σύνολο_κανόνων>
</rdf:Description>
</rdf:RDF>

```

Παράδειγμα 8-4: Σύνολο ΚΑΔΔΑ Κανόνων για το DTD 8-3

```

<?xml version='1.0'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/PR-rdf-schema"
  xmlns:rdfs="http://irida.ced.tuc.gr/RDF/ΠηγήΠληροφορίαςΔεδομένων.rdfs"
  xmlns:Περιγραφή_Βιβλίου="">
  <rdfs:Class ID="Περιγραφή_βιβλίου">
    <rdfs:label>Περιγραφή_βιβλίου</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
      ΠηγήΠληροφορίαςΔεδομένων.rdfs#αντικείμενο"/>
  </rdfs:Class>
  <rdf:Property ID="τιμή">
    <rdfs:range rdf:resource="http://www.w3.org/datatypes#Number" />
    <rdf:domain rdf:resource="#Περιγραφή_βιβλίου" />
  </rdf:Property>
  <rdfs:Class ID="συγγραφέας">
    <rdfs:label>συγγραφέας</rdfs:label>
    <rdfs:SubclassOf rdf:resource="http://irida.ced.tuc.gr/RDF/
      ΠηγήΠληροφορίαςΔεδομένων.rdfs#αντικείμενο"/>
  </rdfs:Class>

```

```

    <rdf:domain rdf:resource="#αναφορά" />
</rdf:Class>
<rdf:Property ID="όνομα">
    <rdf:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#συγγραφέας" />
</rdf:Property>
<rdf:Property ID="επώνυμο">
    <rdf:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#συγγραφέας" />
</rdf:Property>
<rdf:Property ID="τίτλος">
    <rdf:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#διατριβή" />
</rdf:Property>
<rdf:Property ID="ISBN">
    <rdf:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#περιγραφή_βιβλίου" />
</rdf:Property>
<rdf:Property ID="σύνοψη">
    <rdf:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#περιγραφή_βιβλίου" />
</rdf:Property>
<rdf:Property ID="τίτλος">
    <rdf:range rdf:resource="http://www.w3.org/datatypes#String" />
    <rdf:domain rdf:resource="#περιγραφή_βιβλίου" />
</rdf:Property>
</rdf:RDF>

```

RDF Σχήμα 8-3: RDF σχήμα που περιγράφει τη Δομή της Κλάσης Περιγραφή_Βιβλίου

Μετά τις παραπάνω ενέργειες, στην πηγή πληροφορίας μπορούν να εισαχθούν και να δεικτοδοτηθούν στιγμιότυπα της κλάσης Περιγραφή_Βιβλίου, ενώ το μοντέλο πληροφορίας της πηγής πληροφορίας, όπως αυτό εξάγεται μέσω του διαμεσολαβητή στους χρήστες της απαρτίζεται από το RDF σχήμα που προκύπτει αν συνενωθεί το RDF Σχήμα 6-2 με το RDF Σχήμα 8-3.

8.3.2. Ενημέρωση των Πηγών Πληροφορίας

Το θέμα αυτής της ενότητας είναι η περιγραφή των λειτουργιών ενημέρωσης των πηγών πληροφορίας που παίρνουν μέρος στην εφαρμογή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» σε ακαδημαϊκά περιβάλλοντα. Οι λειτουργίες αυτές είναι η (μέσω του διαχειριστή, προνομιούχων χρηστών ή πρακτόρων) εισαγωγή αντικειμένων στις πηγές πληροφορίας και η δεικτοδότηση απομακρυσμένων αντικειμένων σ' αυτές. Έτσι, στη συνέχεια περιγράφονται τα αποτελέσματα των εξής λειτουργιών:

- Της εισαγωγής ενός εγγράφου σε μια πηγή πληροφορίας που περιέχει δομημένα έγγραφα. Πιο συγκεκριμένα, στην υποενότητα 8.3.2.1 περιγράφεται η εισαγωγή ενός Άρθρου στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης.

- Της δεικτοδότησης ενός εγγράφου σε μια πηγή πληροφορίας που περιέχει δομημένα έγγραφα. Πιο συγκεκριμένα, στην υποενότητα 8.3.2.2 περιγράφεται η δεικτοδότηση ενός Άρθρου στην Πηγή Πληροφορίας του Πανεπιστημίου Κρήτης.
- Της εισαγωγής ενός αντικειμένου σε μια πηγή πληροφορίας που περιέχει παραδοσιακά δεδομένα. Πιο συγκεκριμένα, στην υποενότητα 8.3.2.3 περιγράφεται η εισαγωγή μιας Περιγραφής Βιβλίου στην Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης.
- Της δεικτοδότησης ενός αντικειμένου σε μια πηγή πληροφορίας που περιέχει παραδοσιακά δεδομένα. Πιο συγκεκριμένα, στην υποενότητα 8.3.2.4 περιγράφεται η δεικτοδότηση μιας Περιγραφής Βιβλίου στην Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης.

Πριν περιγραφούν τα αποτελέσματα της εκτέλεσης των λειτουργιών αυτών, πρέπει να σημειωθεί ότι αμέσως μετά είτε την εισαγωγή είτε τη δεικτοδότηση κάποιου αντικειμένου σε κάποια πηγή πληροφορίας ενημερώνονται, με τον τρόπο που περιγράφεται στο Διάγραμμα 5-9, όλοι οι πιθανώς ενδιαφερόμενοι για το περιεχόμενό του χρήστες.

8.3.2.1. Εισαγωγή Εγγράφου σε Πηγή Πληροφορίας που περιέχει Δομημένα Έγγραφα

Σ' αυτή την υποενότητα περιγράφεται το αποτέλεσμα της εισαγωγής ενός εγγράφου (και συγκεκριμένα ενός Άρθρου) στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης, υποθέτοντας ότι η πηγή πληροφορίας έχει τη μορφή που προέκυψε μετά την εκτέλεση των λειτουργιών που περιγράφηκαν στην ενότητα 8.3.1. Αυτό σημαίνει ότι στην πηγή πληροφορίας μπορούν να εισαχθούν XML έγγραφα που δομούνται με βάση το DTD 6-2. όπως είναι το έγγραφο που φαίνεται στο Παράδειγμα 8-5.

```
<Άρθρο>
  <Τίτλος>Τεχνολογία Αιχμής στην Επικοινωνία Ανθρώπων-Υπολογιστών</Τίτλος>
  <Συγγραφέας>
    <Όνομα>X</Όνομα>
    <Επώνυμο>X</Επώνυμο>
  </Συγγραφέας>
  <Σύνοψη>Σύνοψη του Άρθρου</Σύνοψη>
  <Ενότητα>
    <Παράγραφος>1η Ενότητα</Παράγραφος>
  </Ενότητα>
  <Ενότητα>
    <Υποενότητα>
      <Παράγραφος>2η Ενότητα</Παράγραφος>
      <Παράγραφος>κείμενο...</Παράγραφος>
    </Υποενότητα>
  </Ενότητα>
</Άρθρο>
```

Παράδειγμα 8-5: XML Έγγραφο που δομείται με βάση το DTD 6-2

Αν υποτεθεί ότι το μέγεθος της 2^{ης} ενότητας υπερβαίνει τα 50KB και το μέγεθος ολόκληρου του άρθρου υπερβαίνει τα 200KB, η εισαγωγή του άρθρου που φαίνεται στο Παράδειγμα 8-5 έχει τις εξής συνέπειες:

- Τη δημιουργία ενός στιγμιότυπου της κλάσης διαχειριστής, ως εξής:
 - Το πεδίο ρόλος του διαχειριστή έχει την τιμή «Συγγραφέας».
 - Το πεδίο όνομα του διαχειριστή έχει την τιμή «Χ».
 - Το πεδίο επώνυμο του διαχειριστή έχει την τιμή «Χ».
- Τη δημιουργία ενός στιγμιότυπου της κλάσης άρθρο, ως εξής:
 - Το πεδίο τίτλος του άρθρου έχει την τιμή «Τεχνολογία Αιχμής στην Επικοινωνία Ανθρώπων-Υπολογιστών».
 - Στο πεδίο περιεχόμενο του άρθρου περιέχεται όλο το περιεχόμενο του άρθρου εκτός από αυτό της 2^{ης} ενότητας, προς την οποία υπάρχει σύνδεσμος.
 - Στο πεδίο περιγραφείας του άρθρου περιέχονται αναφορές στη σύνοψη, την 1^η ενότητα και την παράγραφο που περιέχεται σ' αυτή (που αποθηκεύονται στο πεδίο περιεχόμενο του άρθρου) και στη 2^η ενότητα, την υποενότητα και τις παραγράφους που περιέχονται σ' αυτήν (μέσω αναφοράς στο στιγμιότυπο της κλάσης ενότητα όπου αποθηκεύεται η 2^η ενότητα).
- Τη συσχέτιση του στιγμιότυπου της κλάσης άρθρο με το στιγμιότυπο της κλάσης διαχειριστής.
- Τη δημιουργία ενός στιγμιότυπου της κλάσης ενότητα, ως εξής:
 - Στο πεδίο περιεχόμενο της ενότητας περιέχεται όλο το περιεχόμενο της 2^{ης} ενότητας.
 - Στο πεδίο περιγραφείας της ενότητας περιέχονται αναφορές την υποενότητα και τις παραγράφους που περιέχονται σ' αυτήν (που αποθηκεύονται στο πεδίο περιεχόμενο της ενότητας).

8.3.2.2. Δεικτοδότηση Εγγράφου σε Πηγή Πληροφορίας που περιέχει Δομημένα Έγγραφα

Το θέμα αυτής της υποενότητας είναι η περιγραφή του αποτελέσματος της δεικτοδότησης ενός εγγράφου (και συγκεκριμένα ενός Άρθρου) στην Πηγή Πληροφορίας του Πανεπιστημίου Κρήτης, υποθέτοντας ότι η πηγή πληροφορίας έχει τη μορφή που προέκυψε μετά την εκτέλεση των λειτουργιών που περιγράφηκαν στην ενότητα 8.3.1. Αυτό σημαίνει ότι στην πηγή πληροφορίας μπορούν να δεικτοδοτηθούν XML έγγραφα που δομούνται με βάση το DTD 6-2. όπως είναι το έγγραφο που φαίνεται στο Παράδειγμα 8-5.

Έστω τώρα ότι:

- Το μέγεθος της 2^{ης} ενότητας υπερβαίνει τα 100KB.
- Το κάτω όριο δεικτοδότησης της πηγής πληροφορίας είναι 50%.
- Το κάτω όριο αποθήκευσης της πηγής πληροφορίας είναι 75%.

- Ο βαθμός συσχέτισης του περιεχομένου της πηγής πληροφορίας με το περιεχόμενο του άρθρου είναι 55%.
- Ο βαθμός συσχέτισης του περιεχομένου της πηγής πληροφορίας με το περιεχόμενο της 2^{ης} ενότητας του άρθρου είναι 80%.

Στην περίπτωση αυτή η δεικτοδότηση του άρθρου που φαίνεται στο Παράδειγμα 8-5 έχει τις εξής συνέπειες:

- Τη δημιουργία ενός στιγμιοτύπου της κλάσης διαχειριστής, ως εξής:
 - Το πεδίο ρόλος του διαχειριστή έχει την τιμή «Συγγραφέας».
 - Το πεδίο όνομα του διαχειριστή έχει την τιμή «Χ».
 - Το πεδίο επώνυμο του διαχειριστή έχει την τιμή «Χ».
- Τη δημιουργία ενός στιγμιοτύπου της κλάσης άρθρο, ως εξής:
 - Το πεδίο τίτλος του άρθρου έχει την τιμή «Τεχνολογία Αιχμής στην Επικοινωνία Ανθρώπων-Υπολογιστών».
 - Στο πεδίο περιεχόμενο του άρθρου περιέχεται η σύνοψη του άρθρου.
 - Στο πεδίο περιγραφές του άρθρου περιέχονται αναφορές στη σύνοψη (που αποθηκεύεται στο πεδίο περιεχόμενο του άρθρου), την 1^η ενότητα και την παράγραφο που περιέχεται σ' αυτή (μέσω αναφορών στις διευθύνσεις τους στο Διαδίκτυο) και στη 2^η ενότητα, την υποενότητα και τις παραγράφους που περιέχονται σ' αυτήν (μέσω αναφοράς στο στιγμιότυπο της κλάσης ενότητα όπου αποθηκεύεται η 2^η ενότητα).
- Τη συσχέτιση του στιγμιοτύπου της κλάσης άρθρο με το στιγμιότυπο της κλάσης διαχειριστής.
- Τη δημιουργία ενός στιγμιοτύπου της κλάσης ενότητα, ως εξής:
 - Στο πεδίο περιεχόμενο της ενότητας περιέχεται όλο το περιεχόμενο της 2^{ης} ενότητας.
 - Στο πεδίο περιγραφές της ενότητας περιέχονται αναφορές την υποενότητα και τις παραγράφους που περιέχονται σ' αυτήν (που αποθηκεύονται στο πεδίο περιεχόμενο της ενότητας).

8.3.2.3. Εισαγωγή Αντικειμένου σε Πηγή Πληροφορίας που περιέχει Παραδοσιακά Δεδομένα

Σ' αυτή την υποενότητα περιγράφεται το αποτέλεσμα της εισαγωγής ενός αντικειμένου (και συγκεκριμένα μιας Περιγραφής Βιβλίου) στην Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης, υποθέτοντας ότι η πηγή πληροφορίας έχει τη μορφή που προέκυψε μετά την εκτέλεση των λειτουργιών που περιγράφηκαν στην ενότητα 8.3.1. Αυτό σημαίνει ότι στην πηγή πληροφορίας μπορούν να εισαχθούν αντικείμενα που περιγράφονται από XML έγγραφα που δομούνται με βάση το DTD 8-3. όπως είναι το αντικείμενο που φαίνεται στο Παράδειγμα 8-6.

<ΠΕΡΙΓΡΑΦΗ_ΒΙΒΛΙΟΥ ISBN=3-222-111-333 Τιμή=3500>

```

<ΣΥΓΓΡΑΦΕΑΣ>
  <ΟΝΟΜΑ>Υ</ΟΝΟΜΑ>
  <ΕΠΩΝΥΜΟ>Χ</ΕΠΩΝΥΜΟ>
</ΣΥΓΓΡΑΦΕΑΣ>
<ΤΙΤΛΟΣ>Τεχνικές Επικοινωνίας Ανθρώπων-Υπολογιστών</ΤΙΤΛΟΣ>
<ΣΥΝΟΨΗ>Τεχνικές Επικοινωνίας Ανθρώπων-Υπολογιστών...</ΣΥΝΟΨΗ>
</ΠΕΡΙΓΡΑΦΗ_ΒΙΒΛΙΟΥ>

```

Παράδειγμα 8-6: Αντικείμενο που δομείται με βάση το DTD 8-3

Η εισαγωγή του άρθρου που φαίνεται στο Παράδειγμα 8-6 έχει ως συνέπεια τη δημιουργία :

- Τη δημιουργία ενός στιγμιότυπου της κλάσης Συγγραφέας για το οποίο το χαρακτηριστικό Όνομα παίρνει την τιμή Υ και το χαρακτηριστικό Επώνυμο παίρνει την τιμή Χ.
- Τη δημιουργία ενός στιγμιότυπου της κλάσης Περιγραφή Βιβλίου, για το οποίο το χαρακτηριστικό ISBN παίρνει την τιμή 3-222-111-333, το χαρακτηριστικό Τιμή παίρνει την τιμή 3500, το χαρακτηριστικό Τίτλος παίρνει την τιμή Επικοινωνία Ανθρώπων-Υπολογιστών και το χαρακτηριστικό Σύνοψη παίρνει την τιμή «Τεχνικές Επικοινωνίας Ανθρώπων-Υπολογιστών...».

8.3.2.4. Δεικτοδότηση Αντικειμένου σε Πηγή Πληροφορίας που περιέχει Παραδοσιακά Δεδομένα

Το θέμα αυτής της υποενότητας είναι η περιγραφή του αποτελέσματος της δεικτοδότησης μιας Περιγραφής Βιβλίου στην Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης, υποθέτοντας ότι η πηγή πληροφορίας έχει τη μορφή που προέκυψε μετά την εκτέλεση των λειτουργιών που περιγράφηκαν στην ενότητα 8.3.1. Αυτό σημαίνει ότι στην πηγή πληροφορίας μπορούν να δεικτοδοτηθούν αντικείμενα που περιγράφονται από XML έγγραφα που δομούνται με βάση το DTD 8-3. όπως είναι το αντικείμενο που φαίνεται στο Παράδειγμα 8-7.

```

<ΠΕΡΙΓΡΑΦΗ_ΒΙΒΛΙΟΥ ISBN=3-222-111-334 Τιμή=3500>
<ΣΥΓΓΡΑΦΕΑΣ>
  <ΟΝΟΜΑ>Υ</ΟΝΟΜΑ>
  <ΕΠΩΝΥΜΟ>Χ</ΕΠΩΝΥΜΟ>
</ΣΥΓΓΡΑΦΕΑΣ>
<ΤΙΤΛΟΣ>Επικοινωνία Ανθρώπων-Υπολογιστών</ΤΙΤΛΟΣ>
<ΣΥΝΟΨΗ>Επικοινωνία Ανθρώπων-Υπολογιστών...</ΣΥΝΟΨΗ>
</ΠΕΡΙΓΡΑΦΗ_ΒΙΒΛΙΟΥ>

```

Παράδειγμα 8-7: Αντικείμενο που δομείται με βάση το DTD 8-3

Η δεικτοδότηση της περιγραφής βιβλίου που φαίνεται στο Παράδειγμα 8-7 έχει ως συνέπεια:

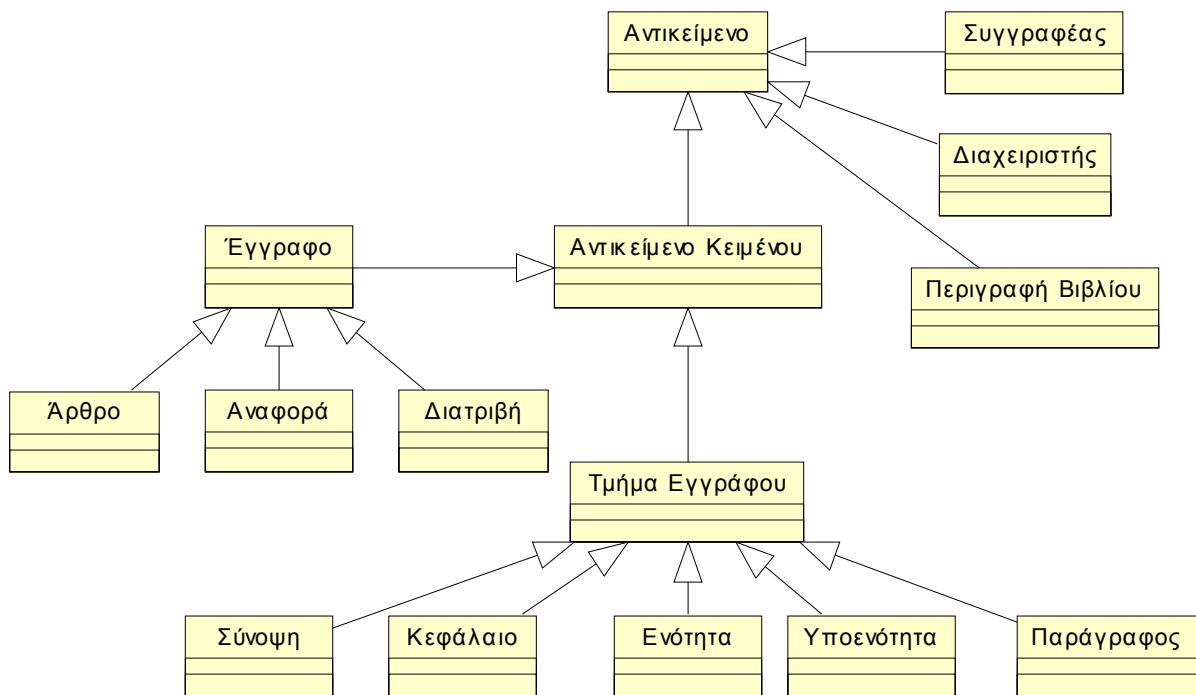
- Τη δημιουργία ενός στιγμιοτύπου της κλάσης Συγγραφέας για το οποίο το χαρακτηριστικό Όνομα παίρνει την τιμή Υ και το χαρακτηριστικό Επώνυμο παίρνει την τιμή Χ.
- Τη δημιουργία ενός στιγμιοτύπου της κλάσης Περιγραφή Βιβλίου, για το οποίο το χαρακτηριστικό ISBN παίρνει την τιμή 3-222-111-334, το χαρακτηριστικό Τιμή παίρνει την τιμή 3500, το χαρακτηριστικό Τίτλος παίρνει την τιμή Επικοινωνία Ανθρώπων-Υπολογιστών και το χαρακτηριστικό Σύνοψη παίρνει την τιμή της διεύθυνσης που έχει η σύνοψη του XML εγγράφου στο Διαδίκτυο.

8.3.3. Αλληλεπίδραση με τους Χρήστες

Σ' αυτή την ενότητα περιγράφεται η αλληλεπίδραση των χρηστών με την εφαρμογή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» σε ακαδημαϊκά περιβάλλοντα. Όπως έχει ήδη αναφερθεί, ο διαμεσολαβητής παίζει το ρόλο του επιπέδου διεπαφής μεταξύ των χρηστών της εφαρμογής και των πηγών πληροφορίας: Ο διαμεσολαβητής παρουσιάζει στους χρήστες ένα μοντέλο πληροφορίας που προέρχεται από τη σύνθεση των μοντέλων πληροφορίας ενώ οι ερωτήσεις και οι αιτήσεις που υποβάλλουν οι χρήστες, πριν μεταβιαστούν στις πηγές πληροφορίας, μετατρέπονται σε μορφή συμβατή με τη γλώσσα του διαμεσολαβητή. Επικεντρώνοντας στην εφαρμογή, ο διαμεσολαβητής επιτρέπει στους χρήστες να αναζητήσουν πληροφορία που σχετίζεται με τα επιστημονικά και/ή ερευνητικά ενδιαφέροντά τους είτε στις πηγές πληροφορίας του Πολυτεχνείου Κρήτης και του Πανεπιστημίου Κρήτης, είτε σε μια μόνο απ' αυτές και/ή την πηγή πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης είτε μόνο στην πηγή πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης είτε και στις τρεις πηγές πληροφορίας παράλληλα.

Ας υποθεθεί ότι κατά τη διάρκεια των αλληλεπιδράσεων των χρηστών με την εφαρμογή οι οποίες περιγράφονται σ' αυτή την ενότητα τα μοντέλα των πηγών πληροφορίας έχουν τη μορφή που προέκυψε μετά την εκτέλεση των λειτουργιών που περιγράφηκαν στην ενότητα 8.3.1. Στην περίπτωση αυτή, στους χρήστες που επιθυμούν να αναζητήσουν πληροφορία παράλληλα στις πηγές πληροφορίας του Πολυτεχνείου Κρήτης, του Πανεπιστημίου Κρήτης και των Πανεπιστημιακών Εκδόσεων Κρήτης δίνεται η δυνατότητα να εκφράσουν τις αιτήσεις και τις ερωτήσεις τους με βάση το μοντέλο πληροφορίας που φαίνεται στο Διάγραμμα 8-2, το οποίο είναι ένα αφαιρετικό διάγραμμα κλάσεων όπου δε φαίνονται τα χαρακτηριστικά των κλάσεων και οι μεταξύ τους συσχετίσεις (με εξαίρεση τις σχέσεις ISA). Αν κάποιος χρήστης επιθυμεί να περιορίσει το πεδίο από το οποίο θα αντλήσει την πληροφορία που τον ενδιαφέρει, μπορεί να επιλέξει να αλληλεπιδρά μέσω του διαμεσολαβητή με συγκεκριμένες πηγές πληροφορίας, οπότε είναι πιθανό να μπορεί να αλληλεπιδράσει με τμήμα μόνο του μοντέλου πληροφορίας που φαίνεται στο Διάγραμμα 8-2⁵¹.

⁵¹ Αν, για παράδειγμα, κάποιος χρήστης επιλέξει να αλληλεπιδρά μόνο με τις πηγές πληροφορίας του Πολυτεχνείου Κρήτης και του Πανεπιστημίου Κρήτης, αυτός θα αλληλεπιδρά με το τμήμα εκείνο του μοντέλου πληροφορίας που δε σχετίζεται με την κλάση «Περιγραφή_Βιβλίου»



Διάγραμμα 8-2: Αφαιρετικό Διάγραμμα Κλάσεων που απεικονίζει το Μοντέλο Πληροφορίας που εξάγεται στους Χρήστες μέσω του Διαμεσολαβητή

Έστω τώρα κάποιος χρήστης που επιθυμεί να αναζητήσει στην πηγή πληροφορίας του Πολυτεχνείου Κρήτης τους τίτλους και τις ταυτότητες εγγράφων που το περιεχόμενό τους σχετίζεται με επικοινωνία ανθρώπων-υπολογιστών. Ο χρήστης αυτός κατασκευάζει την αντίστοιχη ερώτηση στο γραφικό επίπεδο διεπαφής που του παρέχεται μέσω του προγράμματος πλοήγησης του Παγκοσμίου Ιστού που χρησιμοποιεί, η οποία μετατρέπεται από το επίπεδο διεπαφής του διαμεσολαβητή στη μορφή που φαίνεται στο Παράδειγμα 8-8.

```
SELECT Τίτλος, Ταυτότητα FROM Έγγραφο*
WHERE Έγγραφο RELATES TO
  <ORTERMSET>
    <TERM WEIGHT=100%>Επικοινωνία Ανθρώπων-Υπολογιστών</TERM>
  </ORTERMSET>
AND Source="Πηγή Πληροφορίας Πολυτεχνείου Κρήτης"
```

Παράδειγμα 8-8: Ερώτηση για την ανάκτηση των Τίτλων και των Ταυτοτήτων Εγγράφων που το περιεχόμενό τους σχετίζεται με Επικοινωνία Ανθρώπων-Υπολογιστών και βρίσκονται στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης

Στη συνέχεια, αφού η ερώτηση μεταβιβαστεί στο περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, μετατρέπεται από αυτό για λογαριασμό της πηγής πληροφορίας του Πολυτεχνείου Κρήτης στη μορφή που φαίνεται στο Παράδειγμα 8-9.

```
Έγγραφο@Τίτλος, Έγγραφο@Ταυτότητα
[relates(include_subclasses(Έγγραφο), OR(Επικοινωνία Ανθρώπων-Υπολογιστών,
100%)]
```

Παράδειγμα 8-9: Η Ερώτηση που φαίνεται στο Παράδειγμα 8-8 μετά τη μετατροπή της από το περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα

Αν τώρα κάποιος χρήστης επιθυμεί να αναζητήσει στις πηγές πληροφορίας του Πολυτεχνείου Κρήτης και του Πανεπιστημίου Κρήτης τους τίτλους και τις ταυτότητες εγγράφων που το περιεχόμενό τους σχετίζεται με επικοινωνία ανθρώπων-υπολογιστών, η ερώτηση που αντιστοιχεί στην ερώτηση που φαίνεται στο Παράδειγμα 8-8 είναι αυτή που φαίνεται στο Παράδειγμα 8-10.

```
SELECT Τίτλος, Ταυτότητα FROM Έγγραφο*  
WHERE Έγγραφο RELATES TO  
    <ORTERMSET>  
        <TERM WEIGHT=100%>Επικοινωνία Ανθρώπων-Υπολογιστών</TERM>  
    </ORTERMSET>  
AND (Source="Πηγή Πληροφορίας Πολυτεχνείου Κρήτης" OR  
     Source="Πηγή Πληροφορίας Πανεπιστημίου Κρήτης")
```

Παράδειγμα 8-10: Ερώτηση για την ανάκτηση των Τίτλων και των Ταυτοτήτων Εγγράφων που το περιεχόμενό τους σχετίζεται με Επικοινωνία Ανθρώπων-Υπολογιστών και βρίσκονται στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης ή την Πηγή Πληροφορίας του Πανεπιστημίου Κρήτης

Στη συνέχεια, αφού η ερώτηση μεταβιβαστεί στο περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, μετατρέπεται από αυτό για λογαριασμό των πηγών πληροφορίας του Πολυτεχνείου Κρήτης και του Πανεπιστημίου Κρήτης στη μορφή που φαίνεται στο Παράδειγμα 8-9.

Αντίστοιχα, αν κάποιος χρήστης επιθυμεί να αναζητήσει στις πηγές πληροφορίας του Πολυτεχνείου Κρήτης, του Πανεπιστημίου Κρήτης και των Πανεπιστημιακών Εκδόσεων Κρήτης τους τίτλους και τις ταυτότητες αντικειμένων που το περιεχόμενό τους σχετίζεται με επικοινωνία ανθρώπων-υπολογιστών, η ερώτηση που αντιστοιχεί στην ερώτηση που φαίνεται στο Παράδειγμα 8-8 είναι αυτή που φαίνεται στο Παράδειγμα 8-11.

```
SELECT Τίτλος, Ταυτότητα FROM Αντικείμενο*  
WHERE Αντικείμενο RELATES TO  
    <ORTERMSET>  
        <TERM WEIGHT=100%>Επικοινωνία Ανθρώπων-Υπολογιστών</TERM>  
    </ORTERMSET>
```

Παράδειγμα 8-11: Ερώτηση για την ανάκτηση των Τίτλων και των Ταυτοτήτων Εγγράφων που το περιεχόμενό τους σχετίζεται με Επικοινωνία Ανθρώπων-Υπολογιστών και βρίσκονται σε οποιαδήποτε Πηγή Πληροφορίας γίνεται διαθέσιμη μέσω του Διαμεσολαβητή

Στη συνέχεια, αφού η ερώτηση μεταβιβαστεί στο περιτύλιγμα των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα, μετατρέπεται από αυτό για λογαριασμό των πηγών πληροφορίας του Πολυτεχνείου Κρήτης και του Πανεπιστημίου Κρήτης στη μορφή που φαίνεται στο Παράδειγμα 8-9. Στο σημείο αυτό πρέπει να σημειωθεί ότι το περιτύλιγμα των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα απλώς διακομίζει την ερώτηση στην Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης, καθώς οι

γλώσσες του διαμεσολαβητή και των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα είναι ταυτόσημες.

Έστω τώρα ότι τα κριτήρια της ερώτησης που φαίνεται στο Παράδειγμα 8-11 ικανοποιούνται από τα εξής αντικείμενα:

- Το Άρθρο που βρίσκεται στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης, το οποίο έχει τίτλο «Τεχνολογία Αιχμής στην Επικοινωνία Ανθρώπων-Υπολογιστών», ταυτότητα 125 και βαθμό συσχέτισης με την ερώτηση 100%.
- Την Αναφορά που βρίσκεται στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης, η οποία έχει τίτλο «Επικοινωνία Ανθρώπων-Υπολογιστών», ταυτότητα 126 και βαθμό συσχέτισης με την ερώτηση 100%.
- Το Άρθρο που βρίσκεται στην Πηγή Πληροφορίας του Πανεπιστημίου Κρήτης, το οποίο έχει τίτλο «Επικοινωνία Ανθρώπων-Υπολογιστών: Τεχνολογία Αιχμής», ταυτότητα 105 και βαθμό συσχέτισης με την ερώτηση 100%.
- Τη Διατριβή που βρίσκεται στην Πηγή Πληροφορίας του Πανεπιστημίου Κρήτης, το οποίο έχει τίτλο «Πειράματα Απόδοσης Τεχνικών Επικοινωνίας Ανθρώπων-Υπολογιστών», ταυτότητα 115 και βαθμό συσχέτισης με την ερώτηση 95%.
- Την Περιγραφή Βιβλίου που βρίσκεται στην Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης, η οποία έχει τίτλο «Τεχνικές Επικοινωνίας Ανθρώπων-Υπολογιστών», ταυτότητα 135 και βαθμό συσχέτισης με την ερώτηση 100%.

Η πληροφορία που επιστρέφεται από την Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης έχει τη μορφή που φαίνεται στο Παράδειγμα 8-12, η πληροφορία που επιστρέφεται από την Πηγή Πληροφορίας του Πανεπιστημίου Κρήτης έχει τη μορφή που φαίνεται στο Παράδειγμα 8-13 και η πληροφορία που επιστρέφεται από την Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης έχει τη μορφή που φαίνεται στο Παράδειγμα 8-14. Τέλος, μετά το συνδυασμό της πληροφορίας που επιστρέφεται από τις διαφορετικές πηγές πληροφορίας, η πληροφορία που επιστρέφεται από τη γλώσσα του διαμεσολαβητή φαίνεται στο Παράδειγμα 8-15.

```
<Result>
  <Item Class="Άρθρο" Weight="100%">
    <Value Name="Τίτλος">
      Τεχνολογία Αιχμής στην Επικοινωνία Ανθρώπων-Υπολογιστών
    </Value>
    <Value Name="Ταυτότητα"> 125 </Value>
  </Item>
  <Item Class="Αναφορά" Weight="100%">
    <Value Name="Τίτλος"> Επικοινωνία Ανθρώπων-Υπολογιστών </Value>
    <Value Name="Ταυτότητα"> 126 </Value>
  </Item>
</Result>
```

Παράδειγμα 8-12: Η Πληροφορία που επιστρέφεται από την Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης ως Αποτέλεσμα της Ερώτησης που φαίνεται στο Παράδειγμα 8-11

```

<Result>
  <Item Class="Άρθρο" Weight="100%">
    <Value Name="Τίτλος">
      Επικοινωνία Ανθρώπων-Υπολογιστών: Τεχνολογία Αιχμής
    </Value>
    <Value Name="Ταυτότητα"> 105 </Value>
  </Item>
  <Item Class="Διατριβή" Weight="95%">
    <Value Name="Τίτλος">
      Πειράματα Απόδοσης Τεχνικών Επικοινωνίας Ανθρώπων-Υπολογιστών
    </Value>
    <Value Name="Ταυτότητα"> 115 </Value>
  </Item>
</Result>

```

Παράδειγμα 8-13: Η Πληροφορία που επιστρέφεται από την Πηγή Πληροφορίας του Πανεπιστημίου Κρήτης ως Αποτέλεσμα της Ερώτησης που φαίνεται στο Παράδειγμα 8-11

```

<Result>
  <Item Class="Περιγραφή_Βιβλίου" Weight="100%">
    <Value Name="Τίτλος">
      Τεχνικές Επικοινωνίας Ανθρώπων-Υπολογιστών
    </Value>
    <Value Name="Ταυτότητα"> 135 </Value>
  </Item>
</Result>

```

Παράδειγμα 8-14: Η Πληροφορία που επιστρέφεται από την Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης ως Αποτέλεσμα της Ερώτησης που φαίνεται στο Παράδειγμα 8-11

```

<Result>
  <Item Class="Αναφορά" Source="Πηγή Πληροφορίας Πολυτεχνείου Κρήτης"
    Weight="100%">
    <Value Name="Τίτλος"> Επικοινωνία Ανθρώπων-Υπολογιστών </Value>
    <Value Name="Ταυτότητα"> 126 </Value>
  </Item>
  <Item Class="Άρθρο" Source="Πηγή Πληροφορίας Πολυτεχνείου Κρήτης"
    Weight="100%">
    <Value Name="Τίτλος">
      Τεχνολογία Αιχμής στην Επικοινωνία Ανθρώπων-Υπολογιστών
    </Value>
    <Value Name="Ταυτότητα"> 125 </Value>
  </Item>
  <Item Class="Άρθρο" Source="Πηγή Πληροφορίας Πανεπιστημίου Κρήτης"
    Weight="100%">
    <Value Name="Τίτλος">
      Επικοινωνία Ανθρώπων-Υπολογιστών: Τεχνολογία Αιχμής
    </Value>
    <Value Name="Ταυτότητα"> 105 </Value>
  </Item>
</Result>

```

```

</Item>
<Item Class="Περιγραφή_Βιβλίου"
  Source="Πηγή Πληροφορίας των Πανεπιστημιακών Εκδόσεων Κρήτης"
  Weight="100%">
  <Value Name="Τίτλος">Τεχνικές Επικοινωνίας Ανθρώπων-Υπολογιστών</Value>
  <Value Name="Ταυτότητα"> 135 </Value>
</Item>
<Item Class="Διατριβή" Source="Πηγή Πληροφορίας Πανεπιστημίου Κρήτης"
  Weight="95%">
  <Value Name="Τίτλος">
    Πειράματα Απόδοσης Τεχνικών Επικοινωνίας Ανθρώπων-Υπολογιστών
  </Value>
  <Value Name="Ταυτότητα"> 115 </Value>
</Item>
</Result>

```

Παράδειγμα 8-15: Η Πληροφορία που επιστρέφεται από τη Γλώσσα του Διαμεσολαβητή ως Αποτέλεσμα της Ερώτησης που φαίνεται στο Παράδειγμα 8-11

Η πληροφορία που επιστρέφεται από τη γλώσσα του διαμεσολαβητή στη συνέχεια μορφοποιείται, με βάση τις προτιμήσεις του χρήστη και την υπολογιστική πλατφόρμα που αυτός χρησιμοποιεί από τον κατάλληλο παρουσιαστή και αποστέλλεται στο πρόγραμμα πλοήγησης του Παγκοσμίου Ιστού που αυτός χρησιμοποιεί για παρουσίαση. Έστω τώρα ότι ο χρήστης επιλέγει (πιθανώς με τη διάσχιση του αντίστοιχου συνδέσμου) να ανακτήσει το 2^ο αντικείμενο που του παρουσιάστηκε, το οποίο είναι το Άρθρο που βρίσκεται στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης και έχει ταυτότητα 125. Στην περίπτωση αυτή το επίπεδο διεπαφής του διαμεσολαβητή προκαλεί την εκτέλεση της ερώτησης που φαίνεται στο Παράδειγμα 8-16 και το αποτέλεσμά της, που φαίνεται στο Παράδειγμα 8-17 παρουσιάζεται στο χρήστη μετά τη μορφοποίησή του.

```

SELECT Content FROM Άρθρο
WHERE Άρθρο.Ταυτότητα="125" AND
      Source="Πηγή Πληροφορίας Πολυτεχνείου Κρήτης"

```

Παράδειγμα 8-16: Αίτηση Ανάκτησης του Περιεχομένου του Άρθρου που βρίσκεται στην Πηγή Πληροφορίας του Πολυτεχνείου Κρήτης και έχει ταυτότητα 125

```

<Result>
  <Item Class="Άρθρο" Source="Πηγή Πληροφορίας Πολυτεχνείου Κρήτης" >
    <Value Name="Content">
      <Άρθρο>
        <Τίτλος>
          Τεχνολογία Αιχμής στην Επικοινωνία Ανθρώπων-Υπολογιστών
        </Τίτλος>
        <Συγγραφέας>
          <Όνομα>X</Όνομα>
          <Επώνυμο>X</Επώνυμο>
        </Συγγραφέας>

```

```

<Σύνοψη>Σύνοψη του Άρθρου</Σύνοψη>
<Ενότητα>
  <Παράγραφος>1η Ενότητα</Παράγραφος>
</Ενότητα>
<Ενότητα>
  <Υποενότητα>
    <Παράγραφος>2η Ενότητα</Παράγραφος>
    <Παράγραφος>κείμενο...</Παράγραφος>
  </Υποενότητα>
</Ενότητα>
</Άρθρο>
</Value>
</Item>
</Result>

```

Παράδειγμα 8-17: Η Πληροφορία που επιστρέφεται ως Αποτέλεσμα της Αίτησης που φαίνεται στο Παράδειγμα 8-16

8.4. Ανακεφαλαίωση

Σ' αυτό το κεφάλαιο δόθηκε ένα παράδειγμα εφαρμογής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Πιο συγκεκριμένα, στο τμήμα 8.1 περιγράφηκε το *Περιβάλλον της Εφαρμογής*, στο τμήμα 8.2 περιγράφηκε η *Αρχιτεκτονική της Εφαρμογής* και στο τμήμα 8.3 έγινε αναφορά στη *Λειτουργία της Εφαρμογής*.

9^ο Κεφάλαιο

Ανακεφαλαίωση – Συνεισφορά – Μελλοντικές Επεκτάσεις

Στο κεφάλαιο αυτό, με το οποίο κλείνει η παρούσα διατριβή, αρχικά γίνεται μια σύντομη ανακεφαλαίωση της διατριβής (τμήμα 9.1). Στη συνέχεια, στο τμήμα 9.2 γίνεται μια αναφορά στη συνεισφορά της και στα συμπεράσματα που εξήχθησαν από τη διατριβή. Τέλος, μια προσπάθεια καταγραφής των πιθανών μελλοντικών επεκτάσεων της διατριβής γίνεται στο τμήμα 9.3.

9.1. Ανακεφαλαίωση

Στο τμήμα αυτό γίνεται μια σύντομη ανακεφαλαίωση της παρούσας διατριβής όπου περιγράφηκε ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Η παρούσα διατριβή επικέντρωσε στο σχεδιασμό της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» και των τμημάτων από τα οποία αυτή απαρτίζεται. Πιο συγκεκριμένα, εκτός από το σχεδιασμό της συνολικής αρχιτεκτονικής του συστήματος, στα πλαίσια της παρούσας διατριβής έλαβε χώρα ο σχεδιασμός του Διαμεσολαβητή και των Πηγών Πληροφορίας, με ιδιαίτερη έμφαση στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα.

Ο σχεδιασμός τόσο της συνολικής αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» όσο και των τμημάτων από τα οποία αυτή απαρτίζεται, βασίστηκε στο

συνδυασμό των τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό (XML, RDF, XQL κλπ.). Επιπλέον, ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» βασίστηκε στη χρήση της UML για την περιγραφή της αρχιτεκτονικής και της συμπεριφοράς του συστήματος.

Μετά από μια σύντομη εισαγωγή, η παρούσα διατριβή συνεχίστηκε με μια παρουσίαση της σχετικής με την παρούσα διατριβή έρευνας. Πιο συγκεκριμένα, έγινε αναφορά σε συστήματα που έρχονται να καλύψουν τις ίδιες ή παρόμοιες ανάγκες (ή μέρος των αναγκών) που καλύπτει το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» (Ψηφιακές Βιβλιοθήκες, Διαμεσολαβητές, Συστήματα Νέων κατ' Απαίτηση, Μηχανές Αναζήτησης και Πράκτορες) και στην έρευνα που γίνεται σε θέματα που πραγματεύεται η παρούσα διατριβή. Επιπλέον, δόθηκε μια σύντομη περιγραφή του συστήματος που αναπτύχθηκε στα πλαίσια του ευρωπαϊκού ερευνητικού προγράμματος VENIVA, καθώς ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» βασίστηκε στην εμπειρία που αποκτήθηκε από το πρόγραμμα αυτό και στον εντοπισμό των περιορισμών του συστήματος που αναπτύχθηκε στα πλαίσιά του.

Στη συνέχεια έγινε αναφορά στη συσχετιζόμενη με την παρούσα διατριβή εργασία, στην οποία συμπεριλαμβάνονται τα πρότυπα που καθοδήγησαν το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Έτσι, έγινε αναφορά:

- Στη UML (Unified Modeling Language), που χρησιμοποιήθηκε για το σχεδιασμό και την περιγραφή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».
- Σε πρότυπα αναπαράστασης εγγράφων, που επηρέασαν το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» και ενσωματώθηκαν σ' αυτόν.
- Σε πρότυπα αναπαράστασης μεταδεδομένων, που χρησιμοποιήθηκαν στο σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση».
- Στο DOM (Document Object Model) που επηρέασε το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση».
- Σε γλώσσες αναζήτησης και διαχείρισης εγγράφων, που επηρέασαν το σχεδιασμό του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» και ενσωματώθηκαν, ολικά ή μερικά, σ' αυτόν.

Την αναφορά στη συσχετιζόμενη με την παρούσα διατριβή εργασία ακολούθησε η παρουσίαση της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Πιο συγκεκριμένα, δόθηκε μια γενική περιγραφή της αρχιτεκτονικής του συστήματος και έγιναν σύντομες αναφορές στα τμήματα που την απαρτίζουν. Ιδιαίτερη έμφαση δόθηκε στα τμήματα το σχεδιασμό των οποίων πραγματεύεται η παρούσα διατριβή.

Στη συνέχεια περιγράφηκε ο Διαμεσολαβητής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Πιο συγκεκριμένα, δόθηκε μια γενική περιγραφή του διαμεσολαβητή και καθορίστηκαν η λειτουργικότητά του, η αρχιτεκτονική του και τα συστήματα διεπαφής του διαμεσολαβητή με τις πηγές πληροφορίας και με τους χρήστες. Ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας και της βάσης δεδομένων του διαμεσολαβητή, που αποτελούν συνεισφορά της παρούσας διατριβής.

Την περιγραφή του Διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» ακολούθησε η περιγραφή των πηγών πληροφορίας. Πιο συγκεκριμένα, περιγράφηκαν η λειτουργικότητά και η αρχιτεκτονική των πηγών πληροφορίας, ενώ ιδιαίτερη έμφαση δόθηκε στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Τόσο στην περιγραφή των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα όσο και στην περιγραφή των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα, ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας και της βάσης δεδομένων τους, που αποτελούν συνεισφορά της παρούσας διατριβής. Ειδικά για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, έμφαση δόθηκε στην αναπαράσταση των εγγράφων στις σχεσιακές βάσεις δεδομένων όπου αποθηκεύονται τα περιεχόμενα των πηγών πληροφορίας που εντάσσονται σε αυτή την κλάση πηγών πληροφορίας.

Στη συνέχεια περιγράφηκε η διαχείριση των πηγών πληροφορίας, με ιδιαίτερη έμφαση στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και στις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Πιο συγκεκριμένα, περιγράφηκαν τα εργαλεία και οι μηχανισμοί που υποστηρίζουν τις λειτουργίες:

- Της δημιουργίας πηγών πληροφορίας.
- Της υποστήριξης κλάσεων και της δυνατότητα εξέλιξης του σχήματος και του μοντέλου πληροφορίας της βάσης δεδομένων των πηγών πληροφορίας ώστε να μπορούν να υποστηριχθούν από αυτές νέες κλάσεις.
- Της εισαγωγής αντικειμένων στις πηγές πληροφορίας.
- Της αναζήτησης αντικειμένων και τμημάτων αντικειμένων στις πηγές πληροφορίας.
- Της ανάκτησης αντικειμένων και τμημάτων αντικειμένων από τις πηγές πληροφορίας.
- Της δεικτοδότησης, μέσα στις πηγές πληροφορίας, απομακρυσμένης ως προς αυτές πληροφορίας.
- Της ενημέρωσης των πηγών πληροφορίας για πληροφορία που γίνεται διαθέσιμη και σχετίζεται με το περιεχόμενο των συγκεκριμένων πηγών πληροφορίας. Η ενημέρωση αυτή έχει ως συνέπεια είτε την εισαγωγή είτε τη δεικτοδότηση, μέσα στις πηγές πληροφορίας, των αντικειμένων αυτών.

Ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας ΚΑΔΔΑ, που περιγράφει τους κανόνες διαχείρισης, αναπαράστασης και δεικτοδότησης που ισχύουν για τα στιγμιότυπα κλάσεων που περιγράφονται από XML DTDs και επιτρέπει μ' αυτό τον τρόπο την εξέλιξη των σχημάτων και των μοντέλων πληροφορίας των βάσεων δεδομένων των πηγών πληροφορίας μέσω της υποστήριξης νέων κλάσεων.

Τέλος, δόθηκε ένα παράδειγμα εφαρμογής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» σε ακαδημαϊκό περιβάλλον, με στόχο να γίνουν εμφανείς η χρησιμότητα και η εφαρμοσιμότητα του συστήματος.

Η παρούσα διατριβή κλείνει με την τρέχουσα ανακεφαλαίωση, μια αναφορά στη συνεισφορά της και μια προσπάθεια καταγραφής των μελλοντικών επεκτάσεών της.

9.2. Συμπεράσματα – Συνεισφορά

Το θέμα αυτού του τμήματος είναι η καταγραφή τόσο των συμπερασμάτων που εξήχθησαν από την παρούσα διατριβή όσο και της συνεισφοράς της διατριβής.

Στα πλαίσια της παρούσας διατριβής και της διατριβής [Πουτ2000] έλαβε χώρα ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Τα στιγμιότυπα του συστήματος αυτού φιλοδοξούν να παίξουν, μέσα στον Παγκόσμιο Ιστό, το ρόλο της παγκόσμιας βιβλιοθήκης για κοινότητες χρηστών με κοινές απαιτήσεις. Για την κάλυψη των αναγκών αυτών, το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» διαθέτει χαρακτηριστικά τόσο των ψηφιακών βιβλιοθηκών, όσο και των διαμεσολαβητών και των συστημάτων νέων κατ' απαίτηση. Επιπλέον, στο σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» έχει ενσωματωθεί η αιχμή της τεχνολογίας των προτύπων για τον Παγκόσμιο Ιστό. Ο σχεδιασμός τόσο της συνολικής αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» όσο και των τμημάτων από τα οποία αυτή απαρτίζεται, βασίστηκε στο συνδυασμό των ώριμων τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό (XML, RDF, XQL κλπ.). Επιπλέον, ο σχεδιασμός του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» βασίστηκε στη χρήση της UML για την περιγραφή της αρχιτεκτονικής και της συμπεριφοράς του συστήματος.

Η παρούσα διατριβή επικέντρωσε στα εξής σημεία:

- Το σχεδιασμό με βάση το συνδυασμό των τεχνολογιών των σχεσιακών βάσεων δεδομένων και των τεχνικών ανάκτησης πληροφορίας με την αιχμή της τεχνολογίας των προτύπων που κυριαρχούν στον Παγκόσμιο Ιστό, της αρχιτεκτονικής του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση». Η αρχιτεκτονική του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση» απαρτίζεται από ένα *Διαμεσολαβητή (Mediator)* και ένα σύνολο *Πηγών Πληροφορίας (Information Sources)*, τα περιεχόμενα των οποίων γίνονται διαθέσιμα μέσω του Διαμεσολαβητή.
- Το σχεδιασμό του Διαμεσολαβητή του συστήματος «Παροχής Πληροφοριών κατ' Απαίτηση», με ιδιαίτερη έμφαση στα εξής σημεία:
 - Την περιγραφή του διαμεσολαβητή.
 - Τον καθορισμό της λειτουργικότητας του διαμεσολαβητή.
 - Τον καθορισμό της αρχιτεκτονικής του διαμεσολαβητή.
 - Τον καθορισμό των συστημάτων διεπαφής του διαμεσολαβητή με τις πηγές πληροφορίας και με τους χρήστες.
 - Τον καθορισμό, την περιγραφή και το σχεδιασμό των δομικών στοιχείων του διαμεσολαβητή. Ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας του διαμεσολαβητή και τη διαχείριση των δομημένων μεταδεδομένων που χρησιμοποιούνται στο διαμεσολαβητή για την περιγραφή της δομής και του περιεχομένου των πηγών πληροφορίας που γίνονται διαθέσιμες μέσω του διαμεσολαβητή αυτού.

- Το σχεδιασμό των Πηγών Πληροφορίας που περιέχουν δομημένα έγγραφα, με ιδιαίτερη έμφαση στα εξής σημεία:
 - Την περιγραφή των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Τον καθορισμό της λειτουργικότητας των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Τον καθορισμό της αρχιτεκτονικής των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Τον καθορισμό, την περιγραφή και το σχεδιασμό των δομικών στοιχείων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας και της βάσης δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Τον καθορισμό της αναπαράστασης δομημένων εγγράφων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα. Το μοντέλο αναπαράστασης δομημένων εγγράφων στη σχεσιακή βάση δεδομένων των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα καλύπτει τις απαιτήσεις που τίθενται για την αποδοτική και ευέλικτη αναπαράσταση XML εγγράφων.
 - Τον καθορισμό του τρόπου διαχείρισης των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Στον καθορισμό του τρόπου διαχείρισης των πηγών πληροφορίας εμπεριέχεται και ο σχεδιασμός των εργαλείων και των μηχανισμών που υποστηρίζουν τη διαχείριση των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα. Στον όρο «διαχείριση των πηγών πληροφορίας που περιέχουν δομημένα έγγραφα» περιλαμβάνονται οι εξής λειτουργίες:
 - Η δημιουργία πηγών πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Η υποστήριξη κλάσεων δομημένων εγγράφων από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα και η δυνατότητα εξέλιξης του σχήματος της βάσης δεδομένων τους ώστε να μπορούν να μπορούν να υποστηριχθούν νέες κλάσεις δομημένων εγγράφων.
 - Η εισαγωγή δομημένων εγγράφων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Η αναζήτηση δομημένων εγγράφων στις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Η ανάκτηση δομημένων εγγράφων από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα.
 - Η ενημέρωση των πηγών πληροφορίας για έγγραφα που γίνονται διαθέσιμα και το περιεχόμενό τους σχετίζεται με το περιεχόμενο των πηγών πληροφορίας. Η ενημέρωση αυτή έχει ως συνέπεια είτε την εισαγωγή είτε τη δεικτοδότηση, μέσα στις πηγές πληροφορίας, των εγγράφων αυτών.
- Το σχεδιασμό των Πηγών Πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Ο στόχος αυτός περιλαμβάνει:

- Την περιγραφή των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
- Τον καθορισμό της λειτουργικότητας των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
- Τον καθορισμό της αρχιτεκτονικής των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
- Τον καθορισμό, την περιγραφή και το σχεδιασμό των δομικών στοιχείων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας και της βάσης δεδομένων των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα.
- Τον καθορισμό του τρόπου διαχείρισης των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Στον καθορισμό του τρόπου διαχείρισης των πηγών πληροφορίας εμπεριέχεται και ο σχεδιασμός των εργαλείων και των μηχανισμών που υποστηρίζουν τη διαχείριση των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα. Στην έννοια «διαχείριση των πηγών πληροφορίας που περιέχουν παραδοσιακά δεδομένα» περιλαμβάνονται, όπως και για τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα, οι λειτουργίες της δημιουργίας των πηγών πληροφορίας, της εισαγωγής πληροφορίας σ' αυτές, της αναζήτησης πληροφορίας, της ανάκτησης πληροφορίας και της ενημέρωσης για αντικείμενα που γίνονται διαθέσιμα και σχετίζονται με το περιεχόμενο των συγκεκριμένων πηγών πληροφορίας.

Επιπλέον, ιδιαίτερη έμφαση δόθηκε στην περιγραφή της γλώσσας ΚΑΔΔΑ, που περιγράφει τους κανόνες διαχείρισης, αναπαράστασης και δεικτοδότησης που ισχύουν για τα στιγμιότυπα κλάσεων που περιγράφονται από XML DTDs και επιτρέπει μ' αυτό τον τρόπο την εξέλιξη των σχημάτων και των μοντέλων πληροφορίας των βάσεων δεδομένων των πηγών πληροφορίας μέσω της υποστήριξης νέων κλάσεων. Τόσο η γλώσσα ΚΑΔΔΑ όσο και ο σχεδιασμός των εργαλείων και των μηχανισμών που βασίζονται στη χρήση της (εργαλεία SupportDTD και ISAgent και μηχανισμοί εισαγωγής και δεικτοδότησης) αποτελούν συνεισφορά της παρούσας διατριβής.

Συμπερασματικά, το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» είναι ένα σύστημα που παρέχει στους χρήστες του ένα συνδυασμό των υπηρεσιών που παρέχουν τα συστήματα νέων κατ' απαίτηση, οι ψηφιακές βιβλιοθήκες και οι διαμεσολαβητές. Η παρούσα διατριβή επικέντρωσε στο σχεδιασμό τόσο της συνολικής αρχιτεκτονικής του συστήματος όσο και στο σχεδιασμό των τμημάτων από τα οποία αυτή απαρτίζεται με στόχο να προταθούν λύσεις για προβλήματα που απασχολούν όσους σχεδιάζουν συστήματα που καλύπτουν τις ίδιες (τουλάχιστον εν μέρει) απαιτήσεις με το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση». Επιπλέον, ο σχεδιασμός του συστήματος επηρεάστηκε από τα νέα πρότυπα που προτείνονται στην περιοχή του Παγκοσμίου Ιστού και βασίστηκε στη χρήση της UML για την περιγραφή της αρχιτεκτονικής και της συμπεριφοράς του συστήματος.

9.3. Μελλοντικές Επεκτάσεις

Σ' αυτό το τμήμα γίνεται μια προσπάθεια καταγραφής πιθανών μελλοντικών επεκτάσεων της παρούσας διατριβής. Στις επεκτάσεις αυτές περιλαμβάνονται οι εξής:

- Η περιγραφή, ο σχεδιασμός και η υποστήριξη από το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» νέων κλάσεων πηγών πληροφορίας, όπως είναι οι πηγές πληροφορίας που περιέχουν γεωγραφικά δεδομένα, οι πηγές πληροφορίας που περιέχουν ημιδομημένα δεδομένα, οι πηγές πληροφορίας που περιέχουν αντικείμενα πολυμέσων κλπ.
- Η περιγραφή, ο σχεδιασμός και η υποστήριξη από το σύστημα «Παροχής Πληροφοριών κατ' Απαίτηση» της διαδικασίας ενσωμάτωσης ανεξάρτητων υπαρχουσών πηγών πληροφορίας.
- Η υποστήριξη από τις πηγές πληροφορίας που περιέχουν δομημένα έγγραφα XML εγγράφων που η δομή τους περιγράφεται από σχήματα εκφρασμένα στην XML Schema Language. Καθώς η XML Schema Language είναι αυτό-τεκμηριούμενη, το αντίστοιχο με το SupportDTD εργαλείο, το οποίο θα παρέχει τη δυνατότητα υποστήριξης νέων κλάσεων εγγράφων που η δομή τους περιγράφεται από XML Schema Language σχήματα, είναι πιθανό να απαιτεί λιγότερη αλληλεπίδραση με τους διαχειριστές των πηγών πληροφορίας για την υποστήριξη μιας νέας κλάσης εγγράφων.
- Η υποστήριξη από τις πηγές πληροφορίας που περιέχουν παραδοσιακά δεδομένα αντικειμένων που η δομή τους περιγράφεται από σχήματα εκφρασμένα στην XML Schema Language. Καθώς η XML Schema Language είναι αυτό-τεκμηριούμενη, το αντίστοιχο με το SupportDTD εργαλείο, το οποίο θα παρέχει τη δυνατότητα υποστήριξης νέων κλάσεων αντικειμένων που η δομή τους περιγράφεται από XML Schema Language σχήματα, είναι πιθανό να απαιτεί λιγότερη αλληλεπίδραση με τους διαχειριστές των πηγών πληροφορίας για την υποστήριξη μιας νέας κλάσης αντικειμένων.
- Η επέκταση της ΚΑΔΔΑ, έτσι ώστε να μπορεί να υποστηρίξει τους κανόνες αναπαράστασης και διαχείρισης αντικειμένων και εγγράφων που η δομή τους περιγράφεται από σχήματα εκφρασμένα στην XML Schema Language.
- Η υποστήριξη από τις πηγές πληροφορίας μηχανισμών αυτόματης κατηγοριοποίησης εγγράφων και αντικειμένων, βάσει του περιεχομένου τους.
- Η ανάπτυξη τεχνικών αξιολόγησης των διευθύνσεων που επισκέπτονται οι πράκτορες και ο καθορισμός, βάσει των συμπερασμάτων που προκύπτουν από την αξιολόγηση, της προτεραιότητας και της συχνότητας των επισκέψεων κάθε διεύθυνσης.

Τα κριτήρια που χρησιμοποιούνται κατά την αξιολόγηση θα είναι ο βαθμός συσχέτισης των περιεχομένων των διευθύνσεων αυτών με το περιεχόμενο των πηγών πληροφορίας για την ενημέρωση των οποίων είναι υπεύθυνος κάθε πράκτορας.

- Η αντιμετώπιση θεμάτων που έχουν να κάνουν με τη συνεργασία πρακτόρων (collaborative agents) και πολυπρακτορικά περιβάλλοντα (multi-agent environments).

Αναφορές

- [ACPS96] Adali S., Candan K., Papakonstantinou Y., Subrahmanian V.S., *Query caching and optimization in distributed mediator systems*, Proceedings of SIGMOD-96, 1996
- [AFLM99] Abiteboul S., Florescu D., Levy A., Moerkotte G., *Foundations for Information Integration*, Dagstuhl-Seminar-Report 244, 1999
- [AKS96] Arens Y., Knoblock C., Shen W-M., *Query reformulation for dynamic information integration*, International Journal on Intelligent and Cooperative Information Systems (6) 2/3:99-130, June 1996
- [AK97] Ashish N., Knoblock C., *Wrapper Generation for Semi-structured Internet Sources*, SIGMOD Record, Vol. 26, No. 4, December 1997
- [AMM98] Atzeni P., Mecca G., Merialdo P., *Design and Maintenance of data-intensive Web sites*, In EDBT '98, Valencia, Spain, March 1998
- [AMM97] Atzeni P., Mecca G., Merialdo P., *To Weave the Web*, Proceedings of the 23rd VLDB Conference Athens, Greece, 1997, pp 206-215.
- [Bar98] Barja M.L., Bratvold T.A., Myllymaki J., Sonnenberger G., *Informia: a Mediator for Integrated Access to Heterogeneous Information Sources*. In Gardarin, French, Pissinou, Makki, Bouganim (editors), Proceedings of the Conference on Information and Knowledge Management CIKM'98, Washington DC, USA, November 3-7, 1998, pages 234-241, ACM Press, 1998
- [Bea99] Bearman D., Rust G., Weibel S., Miller E., Trant J., *A Common Model to Support Interoperable Metadata*, <http://www.dlib.org/dlib/january99/bearman/01bearman.html>, D-Lib Magazine, January 1999
- [BJN97] Baader F., Jeusfeld M., Nutt W., *Intelligent Access to Heterogeneous Information Sources, Report on the 4th Workshop on Knowledge Representation Meets Databases*, SIGMOD Record, Vol. 26, No. 4, December 1997
- [Bha98] Bharat K., Kamba T., Albers M., *Personalized, interactive news on the Web*, Multimedia Systems (1998) 6: 349-358, Springer – Verlag, 1998
- [Bos98] Bosak J., *Media Independent Publishing: Four Myths about XML*, IEEE Computer, October 1998
- [BMC97] Barret R., Maglio R., Kellem D., *How to Personalize the Web*, ACM CHI 97, 1997
- [Bour99] Bouret R., *XML-DBMS: Java Packages for Transferring Data between XML Documents and Relational Databases*, <http://www.informatik.tu-darmstadt.de/DVSI/staff/bouret/xmldbms/readme.html>
- [Bour99] Bourbakis N., Meng W., Zhang C., Wu Z., Salerno N.J., Borek S., *Retrieval of Multimedia Web Documents and Removal of Redundant Informtaion*, International Journal on Artificial Intelligence Tools, Vol. 8, No 1 (1999)
- [BRJ98] Booch G., Raumbaugh J., Jacobson I., *The Unified Modeling Language User Guide*, Addison-Wesley, 1998
- [Buf94] Buford J., Rutledge L., Rutledge J., Keskin C., *HyOctane: a HyTime engine for an MMIS*, Multimedia Systems 1994, pp. 173-194
- [Bun97] Buneman P., *Tutorial: Semistructured Data*, Proceedings of the ACM SIGMOD Symposium on Principles of Database Systems, 1997
- [CanXML] Bray T., Clark J., Tauber J., *Canonical XML*, W3C Working Draft, <http://www.w3.org/TR/xml-c14n/>, November 1999
-

- [CC/PP] Reynolds F., Hjelm J., Dawkins S., Singhal S., *Composite Capability/Preference Profiles (CC/PP): A user side framework for content negotiation*, W3C, <http://www.w3.org/TR/NOTE-CCPP/>, July 1999
- [CGMH+94] Chawathe S., Garcia-Molina H., Hammer J., Ireland K., Papakonstantinou Y., Ullman J., Widom J., *The TSIMMIS project: Integration of heterogeneous information sources*, In Proceedings of IPSJ, Tokyo, Japan, October 1994
- [CG98] Crescenzi V., Mecca G., *Grammars have exceptions*, Information Systems 1998, pp 1-25, Elsevier Science, 1998
- [Chan98] Chan D., *A Document-driven Approach to Database Report Generation*, INRIA, 1998
- [Cho98] Cho Y., Garcia – Molina H., Page L., *Efficient Crawling Through URL Ordering*, 7th International Web Conference (WWW '98), Brisbane, Australia, 1998
- [Chris97] Christodoulakis S., Pappas N., Kyriakaki G., Maragoudakis Y., Mavraganis Y., C. Tsinaraki C., *The KYDONIA Multimedia Information Server*, Ed. Serge Fdida and Michele Morganti "Multimedia Applications, Services and Techniques - ECMAST'97", Proceedings of Second European Conference Milan, Italy, May 1997 Springer 1997
- [Chu96] Chu W.W., Yang H., Chiang K., Minock M., Chow G., Larson C., *CoBase: A Scalable and Extensible Cooperative Information System*, Journal of Intelligent Information Systems, vol. 6, 2/3, pp. 223-259, June 1996
- [Coh95] Cohen W. W., *Text Categorization and Relational Learning*, In Proceedings of the 12th International Conference on Machine Learning (MK95), 1995
- [Coh96] Cohen W. W., Singer Y., *Context-sensitive learning methods for text categorization*, In Proceedings of the 9th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 307-315, 1996
- [Cra97] Crawley S., Davis S., Indulska J., McBride S., Raymond K., *Meta Information Management*, Proc. Second IFIP International conference on Formal Methods for Open Object-Based Distributed Systems (FMOODS '97), Canterbury, United Kingdom, 21st-23rd July, 1997
- [Cra98] Crabtree B., Soltysiak S., *Identifying and tracking changing interests*, International Journal on Digital Libraries (1998) 2: 38-53, , Springer – Verlag, 1998
- [Cra98] Crawley S., Davis S., Indulska J., McBride S., Raymond K., *Meta-meta is better-better!*, <http://www.dstc.edu.au/MOF>
- [CSS1] H. W. Lie and B. Bos, *Cascading Style Sheets, level 1*, <http://www.w3.org/TR/REC-CSS1-961217>, December 1996
- [CSS2] B. Bos, H. W. Lie, C. Lilley, I. Jacobs, *Cascading Style Sheets, level 2 (CSS2) Specification*, <http://www.w3.org/TR/REC-CSS2/>, May 1998
- [CORBA] Object Management Group, *The common Object Request Broker: Architecture and Specification*, Revision 2.0, <http://www.omg.org/>, April 1995
- [Date] M. Wolf, Wicksteed C., *Date and Time Formats (based on ISO8601)*, <http://www.w3.org/TR/NOTE-datetime>
- [DC97] S. Weibel, J. Kunze, C. Lagoze, M. Wolf, *Dublin Core Metadata for Resource Discovery*, September 1998, RFC 2413
- [DCD] Bray T., Frankston C., Malhotra A., *Document Content Description for XML*, W3C, <http://www.w3.org/TR/NOTE-dcd>
- [DDML] *Document Definition Markup Language (DDML) Specification, Version 1.0*, October 1998, <http://www.simonstl.com/xschema/spec/xscspecv5.htm>
- [Dels96] T. Delsey, *The Evolution of MARC Formats*, National Library of Canada, <http://www.acctbief.org/avenir/evmarc.html>, October 1996

- [Dem96] Dempsey L., Weibel S., *The Warwick Metadata Workshop: A Framework for the Deployment of Resource Description*, D-Lib Magazine, August 1996, <http://www.dlib.org/dlib/july96/07weibel.html>
- [DLP98] Daniel R., Lagoze C., Payette S., *A Metadata Architecture for Digital Libraries*, ACM Digital Libraries Conference (ADL 98), <http://www2.cs.cornell.edu/lagoze/papers/ADL98/dar-adl.html>, 1998
- [DOA] The Corporation for National Research Initiative, *Digital Object Architecture Project*, <http://www.cnri.reston.va.us/doa.html>, July 1998
- [DOM1] Apparao V., Byrne S., Champion M., Isaacs S., Le Hors A., Nicol G., Robie J., Sharpe P., Smith B., Sorensen J., Sutor R., Whitmer R., Wilson C., *Document Object Model (DOM) Level 1 Specification*, Version 1.0, <http://www.w3.org/TR/REC-DOM-Level-1>, October 1998
- [DOM2] Apparao V., Champion M., Kesselman J., Le Hors A., Le Hegaret P., Pixley T., Robie J., Sharpe P., Wilson C., Wood L., *Document Object Model (DOM) Level 2 Specification*, Version 1.0, <http://www.w3.org/TR/WD-DOM-Level-2>, July 1999
- [DP96] Dao S., Perry B., *Information Mediation in Cyberspace: Scalable Methods for Declarative Information Networks*, Journal of Intelligent Information Systems, vol. 6, 2/3, pp. 131-150, 1996
- [DS83] Davis R., Smith R., *Negotiation as a Metaphor for Distributed Problem Solving*, Journal on Artificial Intelligence, vol. 20, pp. 63-109, 1983
- [DSig1] *Digital Signature Label Architecture*, W3C Working Draft, <http://www.w3.org/TR/WD-DSIG-label-arch>
- [DSig2] Reagle J., *XML-Signature Requirements*, W3C Working Draft, <http://www.w3.org/TR/xmlsig-requirements>, October 1999
- [DSig3] Reagle J., Solo D., Bartel M., Boyer J., Eastlake D., Fox B., *XML-Signature Core Syntax*, W3C Working Draft, <http://www.w3.org/TR/xmlsig-core>, October 1999
- [DSSSL96] ISO/IEC 10179:1996, *Information technology -- Processing languages -- Document Style Semantics and Specification Language (DSSSL)*, <http://occam.sjf.novell.com:8080/dsssl/dsssl96>, 1996
- [DSW97] Decker K., Sycara K., Williamson M., *Intelligent adaptive information agents*, Journal of Intelligent Information Systems, vol. 9:239-260, 1997
- [DuCha98] B. DuCharme, *Links That Are More Valuable Than the Information They Link?*, <http://www.xml.com/xml/pub/98/07/xlink/index.html>
- [EW94] Etzion O., Weld D., *A softbot-based interface to the Internet*, CACM, 37(7): 72-76, 1994
- [FGDC] *FGDC Content Standards for Digital Geospatial Metadata*, <http://www.dwr.ehnr.state.nc.us/metadata/generate/overview.html>
- [FFFH+98] Frew J., Freeston M., Freitas N., Hill L., Janee G., Lovette K., Nideffer R., Smith T., Zheng Q., *The Alexandria Digital Library Architecture*, In Proceedings of 2nd European Conference on Digital Libraries (ECDL 98), Heraclion, Greece, 1998, Springer – Verlag
- [FIPA1] *FIPA: Foundation for Intelligent Physical Agents*, <http://drogo.csel.it/fipa>
- [FIPA2] *FIPA Agent Communication Language*, <http://www.fipa.org/spec/fipa97/fipa97.html>
- [FKL97] Florescu D., Koller D., Levy A., *Using Probabilistic Information in Data Integration*, Proceedings of the 23rd VLDB Conference Athens, Greece, 1997, pp 216-225.
- [FLM98] Florescu D., Levy A., Mendelzon A., *Database Techniques for the World-Wide Web: A Survey*, SIGMOD Record, Vol. 27, No 3, September 1998
- [Fow97] Fowler M., Scott K., *UML Distilled: Applying the Standard Modeling Language*, Addison-Wesley 1997

- [FRV96] Florescu D., Rashid L., Valduriez P., *A methodology for query reformulation in cis using semantic knowledge*, International Journal on Intelligent and Cooperative Information Systems 5(4), 1996
- [Gai97] Gaines B.R., *Knowledge Management in Societies of Intelligent Adaptive Agents*, Journal of Intelligent Information Systems, vol. 9, pp.277-298, 1997
- [GM95] Garcia-Molina et. al. *The TSIMMIS Approach to Mediation: Data Models and Languages*, In Proceedings of Workshop NGITS, 1995
- [GMHW99] Goldman R., McHugh J., Widom J., *From Semistructured Data to XML: Migrating the Lore Data Model and Query Language*, <http://www-db.stnford.edu/lore>, 1999
- [Gil97] Gilbert D., *Intelligent Agents: The Right Information at the Right Time*, <http://www.networking.ibm.com/iag/iaghome.html>, 1997
- [Gott98] B. Gottesmann, *Why XML matters*, PC Magazine, October 1998
- [GGM97] L. Gravano, H. Garcia-Molina, *Merging Ranks from Heterogeneous Internet Sources*, Proceedings of the 23rd VLDB Conference Athens, Greece, 1997, pp 196-205.
- [Gue97] Guenther R., *Dublin Core Qualifiers/Substructure*, October 1997, <http://www.loc.gov/marc/dcqualif.html>
- [HNBG+98] Hamard K., Nie J-Y., Bochmann G., Godin R., Kerherve B., Radhakrishnan T., Shinghal R., Berouti F., Ferrie F.P., *A Multi-Level Agent System for Digital Libraries*, 1998
- [HGN+97] Hammer J., Garcia-Molina H., Nestorov S., Yermení R., Breuning M., Vassalos V., *Template-Based Wrappers*, Proceedings of SIGMOD 1997
- [HKWY97] L. Haas, D. Kossmann, E. Wimmers, J. Yang, *Optimizing Queries across Diverse Data Sources*, Proceedings of the 23rd VLDB Conference Athens, Greece, 1997, pp 276-285.
- [HTML] Berners-Lee T., *Hypertext Markup Language - A Representation of Textual Information and Metainformation for Retrieval and Interchange*, Internet Draft, CERN 1993
- [HTML4] Raggett D., Le Hors A., Jacobs I., *HTML 4.0 Specification*, W3C, <http://www.w3.org/TR/REC-html40/>, April 1998
- [Hynode] IBM, HyNoDe, *Specification of an HTML-based News Document structure*, 1996.
- [HyTime92] ISO/IEC 10744:1992, *Information Technology -- Hypermedia/Time-based Structuring Language (HyTime)*, 1992
- [IK97] Imam I., Kerschberg L., *Adaptive Intelligent Agents*, Journal of Intelligent Information Systems, vol. 9, pp. 211-214, 1997
- [IMS] *IMS Meta-Data Specification*, http://www.imsproject.org/md_overview.html
- [ISAD] Ad Hoc Commission on Descriptive Standards, *ISAD(G): General International Standard Archival Description*, 1995
- [ISBD] Working Group on the General International Standard Bibliographic Description set by IFLA, *International Standard Bibliographic Description*, 1977
- [ISO11404] *ISO 11404 -- Information Technology -- Programming Languages, their environments and system software interfaces -- Language-independent datatypes*, ISO/IEC 11404:1996, 1996
- [Ker97.1] Kerschberg L., *Knowledge Rovers: Cooperative Intelligent Agent Support for Enterprise Information Architectures*, P. Kandzia & Klush M. (editors), Cooperative Information Agents, Proceedings of 1st International Workshop CIA-97, Lecture Notes in Artificial Intelligence, Vol. 1202, Springer – Verlag, 1997

- [Ker97.2] Kerschberg L., *The Role of Intelligent Agents in Advanced Information Systems*, Small C., Douglas P., Johnson R., King P. & Martin N. (editors), *Advances in Databases*, Lecture Notes in Computer Science, Vol. 1271, Springer – Verlag, 1997
- [KH96] Kuokka D., Harada L., *Integrating Information via Matchmaking*, *Journal of Intelligent Information Systems*, vol. 6(2/3), pp. 261-279, 1996
- [Klu98] Klush M. (editor), *Intelligent Information Agents – Agent-Based Information Discovery and Management on the Internet*, ISBN 3-540-65112-8, Springer – Verlag, 1998
- [KLMW94] Kemper A., Lockemann P.C., Moerkotte G., Walter H-D., *Autonomous Objects: A Natural Model for Complex Applications*, *Journal of Intelligent Information Systems*, vol. 3, pp. 113-150, 1994
- [Kos95] Koster M., *Robots in the Web: threat or treat?*, *ConneXions*, Volume 9, No. 4, <http://info.webcrawler.com/mak/projects/robots/threat-or-treat.html>, April 1995
- [KQML1] Finin T., Labroux Y., Matfeld J., *KQML as an agent communication language*, Bradshaw J. (editor), *Software Agents*, MIT Press, Menlo Park, 1995
- [KQML2] Finin T., McKay D., Fritzson R., McEntire R., *KQML: An Information and Knowledge Exchange Protocol*, Fuchi K. & Yokoi T. (editors), *Knowledge Building and Knowledge Sharing*, Ohmsha and IOS Press, 1994
- [KWD97] Kushmerick N, Weld D.S., Doorenbos R., *Wrapper Induction for Information Extraction*, In *Proceedings of IJCAI 1997*
- [KW96] Kwok C., Weld D., *Planning to gather Information*, In *Proceedings of the AAAI 13th Conference on Artificial Intelligence*, 1996
- [Lag96] Carl Lagoze, Clifford A. Lunch, Ron Daniel Jr., *The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata*, June 1996, <http://cs-tr.cs.cornell.edu/Dienst/UI/2.0/Describe/ncstrl.cornell/TR96-1593>
- [LPH97] Laird J.E., Pearson D.J., Huffman S.B., *Knowledge-Directed Adaptation in Multi-Level Agents*, *Journal of Intelligent Information Systems*, vol. 9, pp.261-276, 1997
- [LRO96] Levy A., Rajaraman A., Ordille J., *Querying heterogeneous information sources using source descriptions*, In *Proceedings of VLDB '96*, Bombay, India 1996
- [MathML] Buswell S., Devitt S., Diaz A., Ion P., Miner R., Poppelier N., Smith B., Soiffer N., Sutor R., Watt S., *Mathematical Markup Language (MathML™) 1.01 Specification*, W3C, <http://www.w3.org/TR/REC-MathML/>, July 1999
- [MBP98] McBrien P., Poulouvassilis A., *A formalization of semantic schema integration*, *Information Systems* Vol. 23, No 5, pp. 307-334, Elsevier Science 1998
- [McGra98] McGrath S., *Rendering XML Documents Using XSL*, *Dr Dobbs Journal*, July 1998
- [MFDG98] Mace S., Flohr U., Dobson R., Graham T., *Weaving a Better Web*, *Byte*, March 1998
- [MHEG] ISO/IEC DIS 13522-1, Part 1: *MHEG Object Oriented Representation -- Base Notation (ASN.1)*, ISO/IEC DIS 13522-2, Part 2: *MHEG Object Oriented Representation -- Base Notation (SGML)*, ISO/IEC DIS 13522-3, Part 3: *MHEG Extensions for Scripting Language Support*, ISO/IEC DIS 13522-4, Part 4: *Registration Procedure for Format Identifiers*, ISO/IEC DIS 13522-5, Part5: *MHEG Subset for Base Level Implementation*
- [Mic98] Michard A., Christophides V., Scholl M., Stapleton M., Sutcliffe D., Vercoustre A-M., *The Aquarelle Resource Discovery System*, In *Journal of Computer Networks and ISDN Systems* 30(13), pages 1185-1200, Elsevier Science, August 1998
- [Mil98] Miller E., *An Introduction to the Resource Description Framework*, *D-Lib Magazine*, May 1998, ISSN 1082-9873, <http://www.dlib.org/dlib/may98/05miller.html>
- [Mins99] Hauswirth M., *A Component and Communication System for Push Systems*, *ESEC/FSE '99*, Toulouse, France

- [Mot95] Motro A., *Multiplex: A Formal Model for Multidatabases and Its Implementation*, ISSE Department, George Mason University, Fairfax, VA, Technical Report ISSE-TR-95-10, 1995
- [Mor96] Morin J.-H., *HyperNews, a Hypermedia Electronic – NewsPaper Environment Based on Agents*, in Object Applications, Ed. Tsichritzis D., CUI, University of Geneva, 1996
- [ModXHTML] Altheim M., Austin D., Boumphrey F., Dooley S., McCarron S., Wugofski T., *Modularization of XHTML™*, W3C, <http://www.w3.org/TR/xhtml-modularization>, April 1999
- [MOF] DSTC, *Meta-Object Facility*, OMG TC document cf/97-01-01, Linnaeus Project, January 1997
- [MPP98] Marazakis M., Papadakis D., Papadakis S., *A Framework for the Encapsulation of Value-Added services in Digital Objects*, In Proceedings of 2nd European Conference on Digital Libraries (ECDL 98), Heraclion, Greece, 1998, Springer – Verlag
- [MS96] Motro A., Smets P., *Uncertainty Management in Information Systems: form Needs to Solutions*, Norwal, MA: Kluwer Academic Publishers, 1996
- [NCSTRL] Davis J., Lagoze C., *NCSTRL: Design and Deployment of a Globally Distributed Digital Library*, IEEE Computer, February 1999
- [New91] Newcomb S., Kipp N., Newcomb V., *The 'HyTime' Hypermedia/Time based Document Structuring Language*, Communications of the ACM, November 1991, Vol. 34, No. 11
- [Norg98] Norgard B., *Entry Vocabulary Modules and Agents*, <http://www.sims.berkeley.edu/research/metadata/agents.html>, 1998
- [NVML] Sekiguchi M., Takayama K., Naito H., Maeda Y., Horai H., Toriumi M., *NaVigation Markup Language (NVML)*, W3C, <http://www.w3.org/TR/NVML>, August 1999
- [ODA89] ISO 8613:1989, *Information Processing -- Text and Office Systems -- Office Document Architecture (ODA) and Interchange Format*, 1989
- [OQL] Atwood T., Dubl J., Ferran G., Loomis M., *Object Database Standard: ODMG-93*, ISBN 1-55860-302-6, Gattell R.G.G., Morgan Kaufmann Publishers Inc.
- [P3P] Marchiori M., Reagle J., *Platform for Privacy Preferences (P3P) Specification*, <http://www.w3.org/TR/WD-P3P>, August 1999
- [Page98] Page L., Brin S., *The Anatomy of a Search Engine*, 7th International Web Conference (WWW '98), Brisbane, Australia, 1998
- [Pae98] Paepke A., Chang C.C., Garcia-Molina H., Winograd T., *Interoperability for Digital Libraries Worldwide*, Communications of the ACM, pages 33-43, April 1998
- [PAGM96] Papakonstantinou Y., Abiteboul S., Garcia-Molina H., *Object Fusion in mediator systems*, In Proceedings of VLDB '96, pages 413-424, Bombay, India 1996
- [PGMU96] Papakonstantinou Y., Garcia-Molina H., Ullman J., *MedMaker: A mediationsystem based on declarative specifications*, ICDE 96, 1996
- [PIDL] Koike Y., Kamba T., Langheinrich M., *PIDL – Personalized Information Description Language*, W3C Note, <http://www.w3.org/TR/NOTE-PIDL>, February 1999
- [PICS1] *Rating Services and Rating Systems and Their Machine Readable Descriptions Version 1.1*, W3C Recommendation, <http://www.w3.org/TR/REC-PICS-services>
- [PICS2] *PICS Label Distribution Label Syntax and Communication Protocols Version 1.1*, W3C Recommendation, <http://www.w3.org/TR/REC-PICS-labels>
- [PICS3] Y. Chu, P. DesAutels, B. LaMacchia, P. Lipp, *PICS Signed Labels (DSig) 1.0 Specification*, <http://www.w3.org/TR/REC-DSig-label/>, May 1998
- [PICS4] J. Reagle, D. Weitzner, *Statement on the Intent and Use of PICS: Using PICS Well*, <http://www.w3.org/TR/NOTE-PICS-Statement>, June 1998

- [PL98] Payette S., Lagoze C., *Flexible and Extensible Digital Object and Repository Architecture (FEDORA)*, In Proceedings of 2nd European Conference on Digital Libraries (ECDL 98), Heracleion, Greece, 1998, Springer – Verlag
- [PML] Ram A., Catrambone R., Guzdial M., Kehoe C., McCrickard S., Stasko J., *PML: Adding Flexibility to Multimedia Presentations*, IEEE Multimedia Summer 1999, pp. 40-52
- [POIX] Kanemitsu H., Kamada T., *POIX: Point Of Interest eXchange Language Specification*, W3C, <http://www.w3.org/TR/poix>, June 1999
- [Pri93] Price R., *MHEG: An Introduction to the future International Standard for Hypermedia Object Interchange*, ACM Multimedia 93
- [RDF] Lassila O., Swick R., *Resource Description Framework (RDF) Model and Syntax Specification*, <http://www.w3.org/TR/REC-rdf-syntax.htm>, February 1999
- [RDFS] Brickley D., Guha R.V., Layman A., *Resource Description Framework (RDF) Schema Specification*, <http://www.w3.org/TR/PR-rdf-schema.htm>, March 1999
- [RFC1766] IETF (Internet Engineering Task Force), *RFC 1766: Tags for the Identification of Languages*, 1995, <http://ds.internic.net/rfc/rfc1766.txt>
- [RGK98] Rus D., Gray R., Kotz D., *Transportable Information Agents*, Huhns M & Singh M.P. (editors), Readings in Agents, Morgan Kaufmann, 1998
- [RJB98] Raumbaugh J., Jacobson I., Booch G., *The Unified Modeling Language Reference Manual*, Addison-Wesley, 1998
- [Rob98] Robie J., *The Design of XQL*, <http://www.texcel.no/whitepapers/xql-design.html>
- [Rob99] Robie J., Derksen E., Fankhauser P., Howland E., Huck G., Macherius I., Murata M., Resnick M., Schöning H., *XQL (XML Query Language)*, <http://metalab.unc.edu/xql/xql-proposal.html>, August 1999
- [Ros91] Rosenberg J., Sherman M.- Ann Marks - Jaap Akkerhuis, *Multi-media Document Translation - ODA and the EXPRES Project*, Springer-Verlag 1991
- [RS97] Roth M. T., Scharz P., *Don't Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources*, Proceedings of the 23rd VLDB Conference Athens, Greece, 1997, pp 266-275.
- [Rus98] Rusch-Feja D., *Metadata: Standards for Retrieving WWW Documents (and Other Digitized and Non-Digitized Resources)*, Library and Information Services in Astronomy III, ASP Conference Series, Vol 153, 1998, <http://www.stsci.edu/meetings/lisa3/ruschfejad.html>
- [Sah98] Sahuguet A., Azavant F., *Taming Web sources with "minute-made" wrappers*, Technical Report, University of Pennsylvania, 1998
- [SGML86] ISO 8879:1986, *Information processing – Text and office systems – Standard Generalized Markup Language (SGML)*, 1986
- [SGML90] Goldfarb C., *THE SGML HANDBOOK*, Oxford University Press, 1990
- [SK97] Seligman L., Kerschberg L., *A Mediator for Approximate Consistency: Supporting "Good Enough" Materialized Views*, Journal of Intelligent Information Systems, vol. 8, pp. 203-225, 1997
- [SMIL] Ayars J., Cohen A., Day K., Hodge E., Hoschka P., Lanphier R., Layaida N., van Ossenbruggen J., Rutledge L., Saccocio B., Schmitz P., ten Kate W., Wugofski T., Yu J., *Synchronized Multimedia Integration Language (SMIL) Boston Specification*, W3C, <http://www.w3.org/TR/smil-boston>, August 1999
- [SOX] Davidson A., Fuchs M., Hedin M., Jain M., Koistinen J., Lloyd C., Maloney M., Schwarzhof K., *Schema for Object-oriented XML 2.0*, W3C, <http://www.w3.org/TR/NOTE-SOX/>, July 1999

- [SPDL] ISO/IEC DIS 10190:1991, *Information Technology -- Text Communication -- Standard Page Description Language (SPDL)*
- [ST98] Salminen A., Tompa F., *Grammars++ for Modeling Information in Text*, 1998
- [Ste95] Steinmetz R., Nahrstedt K., *Multimedia: Computing, Communications and Applications*, Prentice Hall, 1995
- [SVG99] Bowler J., Cohn R., Capsimalis M., Donoho A., Duce D., Evans J., Ferraiolo J., Furman S., Graffagnino P., Henderson L., Hester A., Hopgood B., Lawrence K., Lilley C., Mansfield P., McCluskey K., Nguyen T., Sandal T., Santangeli P., Sheikh H., State G., Stevahn R., Zhou S., *Scalable Vector Graphics (SVG) 1.0 Specification*, <http://www.w3.org/TR/WD-SVG>, July 1999
- [TEI95] DeRose S., Durand D., *The TEI Guidelines*, In "Computing and Humanities" 29 (3). Reprinted in *Text Encoding Initiative: Background and Context*, ISBN 0-7923-3704-2
- [TRV97] Tomasic A., Raschid L., Valduriez P., *A data model and query processing techniques for scaling access to distributed heterogeneous databases in disco*, IEEE Transactions on Computers, special issue on Distributed Computing Systems, 1997
- [Tsin98] Tsinaraki C., Anestis G., Moumoutzis N., Christodoulakis S., *Implementing Powerful Retrieval Capabilities in a Distributed Environment for Libraries and Archives*, In Proceedings of 2nd European Conference on Digital Libraries (ECDL 98), Heraclion, Greece, 1998, Springer – Verlag
- [UCLP] SAIC/Bellcore, *Universal Commerce Language and Protocol*, <http://www.w3.org/TR/NOTE-uclp>, January 1999
- [UML] Object Management Group (OMG), *Unified Modeling Language Specification*, <http://www.omg.org>, July 1998
- [URI97] T. Berners-Lee, R. Fielding, L. Masinter, *Uniform Resource Identifiers (URI): Generic Syntax and Semantics*, <http://www.ics.uci.edu/pub/ietf/uri/draft-fielding-uri-syntax-01.txt>, November 1997
- [URL94] T Berners-Lee, L Masinter, M McCahill, *Uniform Resource Locators (URL)*, RFC 1738, CERN, Xerox Corporation, University of Minnesota, December 1994
- [Vel99] Velegrakis G, Christophides V., Constantopoulos P., *Declarative Specification of Z39.50 Wrappers*, In Proceedings of the ECDL '99, Paris, France, September 1999
- [Wal98] Walsh N., *A Technical Introduction to XML*, <http://nwalsh.com/articles/XML/index.html>, February 1998
- [WAP98b] Wireless Application Protocol Forum (WAP), *WAP WML*, <http://www.wapforum.org/>, April 98
- [WBXML] Martin B., Jano B., *WAP Binary XML Content Format*, <http://www.w3.org/TR/wbxml>, June 1999
- [WeiHa98] Weibel S., Hakala J., *DC-5: The Helsinki Metadata Workshop*, D-Lib Magazine, February 1998, ISSN 1082-9873, <http://www.dlib.org/dlib/february98/02weibel.html>
- [Wein98] Weinstein P., Birmingham W., *Creating ontological metadata for digital library content and services*, International Journal on Digital Libraries (1998) 2: 20-37, 1998
- [Wie96] Wiederhold G., *Foreword: Intelligent integration of information*, Journal of Intelligent Information Systems, vol. 6(2/3):93-97, May 1996
- [Wu99] Wu x., Ngu D., *User-Centered Agents for Structured Information Location*, 1999
- [WWW92] Berners-Lee T., Cailliau R., Groff J.F., Pollenmann B., *World-Wide Web: The Information Universe, Electronic Networking: Research, Applications and Policy*, Vol 1 No 2, Meckler, Westport CT, Spring 1992

- [XFA-FC] Tardif M., Devitt R., McDougall R., McKenzie G., Miller A., Neilson A., *XFA-FormCalc Version 1.0*, JetForm, <http://www.w3.org/1999/05/XFA/xf-formcalc.html>, June 1999
- [XFA-Templ] McKenzie G., Devitt R., McDougall R., Miller A., Neilson A., Tardif M., *XFA-Template Version 1.0*, JetForm, <http://www.w3.org/1999/05/XFA/xf-template>, June 1999
- [XHTML] Pemberton S., Altheim M., Austin D., Boumphrey F., Burger J., Donoho A., Dooley S., Hofrichter K., Hoschka P., Ishikawa M., Kate W., King P., Klante P., Matsui S., McCarron S., Navaro A., Nies Z., Ragget D., Schmitz P., Wilson C., Wugofski T., Zigmond D., *XHTML™ 1.0: The Extensible HyperText Markup Language – A Reformulation of HTML 4.0 in XML 1.0*, <http://www.w3.org/TR/WD-html-in-xml/>, March 1999
- [XIncl] Marsh J., Orchard D., *XML Inclusion Proposal*, W3C Note, <http://www.w3.org/TR/xinclude>, November 1999
- [XLink] DeRose S., Orchard D., Trafford B., *XML Linking Language (XLink)*, W3C, <http://www.w3.org/TR/xlink>, July 1999
- [XMI] Unisys Corporation, International Machines Corporation, Cooperative Research Centre for Distributed Systems Technology (DSTC), Oracle Corporation, Platinum Technology Inc., Fujitsu, softbeam, Recerca Informatica, Daimler-Benz, *XML Metadata Interchange (XMI)*, OMG Document ad/98-10-05, October 1998
- [XML] Bray T., Paoli J., Sperberg-McQueen C. M., *Extensible Markup Language (XML) 1.0 Specification*, W3C Recommendation, <http://www.w3.org/TR/REC-xml>, February 1998
- [XML-Data] Layman A., Jung E., Maler E., Thompson H., Paoli J., Tigue J., Mikula N., DeRose S., *XML Data*, <http://www.w3.org/TR/NOTE-XML-data>, January 1998
- [XML-Frag] Grosso P., Veilard D., *XML Fragment Interchange*, W3C Working Draft, <http://www.w3.org/TR/WD-xml-fragment>, June 1999
- [XMLISRQ] Megginson D., *XML Information Set Requirements*, W3C Note, <http://www.w3.org/TR/NOTE-xml-info-set-req/>, February 1999
- [XMLInfSet] Cowan J., Megginson D., *XML Information Set*, W3C Working Draft, <http://www.w3.org/TR/xml-info-set/>, May 1999
- [XML-Name] Bray T., Hollander D., Layman A., *Namespaces in XML*, W3C Recommendation, <http://www.w3.org/TR/REC-xml-names/>, January 1999
- [XML-QL] Deutsch A., Fernandez M., Florescu D., Levy A., *XML-QL: A Query Language for XML*, <http://www.w3.org/TR/NOTE-xml-ql/>, August 1998
- [XML-RPC] Winer D., *XML-RPC Specification*, [http://www.xmlrpc.com/stories/storyReader\\$7](http://www.xmlrpc.com/stories/storyReader$7)
- [XMLSche] Malhotra A., Maloney M., *XML Schema Requirements*, <http://www.w3.org/TR/NOTE-xml-schema-req/>, February 1999
- [XML-Sche1] Beech D., Maloney M., Mendelsohn N., Thompson H., *XML Schema Part 1: Structures*, W3C Working Draft, <http://www.w3.org/TR/xmlschema-1/>, September 1999
- [XML-Sche2] Biron P., Malhotra A., *XML Schema Part 2: Datatypes*, W3C Working Draft, W3C Working Draft, <http://www.w3.org/TR/xmlschema-2/>, September 1999
- [XML-SS] Clark J., *Associating Style Sheets with XML Documents – Version 1.0*, W3C Recommendation, <http://www.w3.org/TR/xml-style-sheet>, June 1999
- [XMTP98] Borden J., *eXtensible Mail Transport Protocol (XMTP)*, 1999
- [XPath] Clark J., DeRose S., *XML Path Language (XPath) Version 1.0*, W3C Recommendation, <http://www.w3.org/TR/xpath>, November 1999
- [XPointer] Daniel R., DeRose S., *XML Pointer Language (XPointer)*, W3C Working Draft, <http://www.w3.org/TR/WD-xptr>, July 1999

- [XQL] Robie J., Lapp J., *XML Query Language*, <http://www.w3.org/TandS/QL/QL98/pp/xql.html>, September 1998
- [XSL] Deach S., *Extensible Stylesheet Language (XSL)*, W3C Working Draft, <http://www.w3.org/TR/WD-xsl>, April 1999
- [XSLT] Clark J., *XSL Transformations (XSLT) Specification Version 1.0*, W3C Recommendation, <http://www.w3.org/TR/xslt>, November 1999
- [VP97] V. Vassalos, Y. Papakonstantinou: *Describing and Using Query Capabilities of Heterogeneous Sources*, Proceedings of the 23rd VLDB Conference Athens, Greece, 1997, pp 256-265.
- [YNA] F Yergeau, G Nicol, G Adams, M Dürst, *Internationalization of the Hypertext Markup Language*, <ftp://ietf.org/internet-drafts/draft-ietf-html-i18n-05.txt>
- [Ανεσ97] Ανέστης Γ., *Σχεδιασμός και Υλοποίηση Συστήματος Λογικών Ερωτήσεων και Ερωτήσεων Ομοιότητας σε Σχεσιακά Συστήματα Διαχείρισης Βάσεων Δεδομένων*, Διπλωματική Εργασία, Πολυτεχνείο Κρήτης, 1997
- [Καπετ95] Καπετανάκης Γ, *Δημιουργία και Ανάκτηση Εγγράφων στο Αντικειμενοστρεφές Σύστημα KYDONIA*, Διατριβή για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης, Πολυτεχνείο Κρήτης, 1995
- [Καρασ98] Καρασαββίδης Σ., *Σχεδιασμός και Υλοποίηση των Υποσυστημάτων Κατηγοριοποίησης & Αναζήτησης Εγγράφων και Διαχείρισης Διαγραμμάτων σε Εξατομικευμένα Συστήματα Νέων Κατ' Απαίτηση*, Διπλωματική Εργασία, Πολυτεχνείο Κρήτης, 1998
- [Κοπαν96] Κοπανάκης Α., *Σχεδιασμός και Υλοποίηση Συστήματος Διαχείρισης Δομημένων Εγγράφων και Ορισμού Εναλλακτικών Τρόπων για τη Διαχείριση και Παρουσίαση Δεδομένων που αποθηκεύονται σε Σχεσιακές Βάσεις Δεδομένων*, Διπλωματική Εργασία, Πολυτεχνείο Κρήτης, 1996
- [Μουμ98] Μουμουτζής Ν., *Σχεδιασμός ενός Συστήματος που υποστηρίζει την Ανάπτυξη Αλληλεπιδραστικών Γεωγραφικών Εφαρμογών*, Διατριβή για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης, Πολυτεχνείο Κρήτης, 1998
- [Πετρ98] Πετρόπουλος Μ., *Συγγραφή και Παρουσίαση Εγγράφων σε Εξατομικευμένα Συστήματα Νέων Κατ' Απαίτηση*, Διπλωματική Εργασία, Πολυτεχνείο Κρήτης, 1998
- [Πουτ2000] Πούτος Κ., *Ανάκτηση Πληροφορίας στο Σύστημα Παροχής Πληροφοριών κατ' Απαίτηση*, Διατριβή για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης, Πολυτεχνείο Κρήτης, 2000
- [Σκον98] Σκόνδρας Π., *Σχεδιασμός και Υλοποίηση Μηχανισμών Προώθησης και Ανέλκυσης (Push & Pull) σε Σύστημα Νέων Κατ' Απαίτηση*, Διπλωματική Εργασία, Πολυτεχνείο Κρήτης, 1998
- [Τσιν95] Τσινιράκη Χ., *Σχεδιασμός και Υλοποίηση Συστήματος Ομοιόμορφης και Διαφανούς Αναπαράστασης Εγγράφων και Συνδέσμων στο Αντικειμενοστρεφές Σύστημα Διαχείρισης Βάσεων Δεδομένων "Kydonia" και στον Παγκόσμιο Ιστό*, Διπλωματική Εργασία, Πολυτεχνείο Κρήτης, Οκτώβριος 1995