

ΕΛΕΓΧΟΣ



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

186

ΨΗΦΙΑΚΕΣ ΒΙΒΛΙΟΘΗΚΕΣ ΒΑΣΙΣΜΕΝΕΣ ΣΤΟ  
WEB: ΑΡΧΙΤΕΚΤΟΝΙΚΕΣ, ΑΣΦΑΛΕΙΑ,  
ΑΝΑΚΤΗΣΗ ΠΛΗΡΟΦΟΡΙΩΝ

Αριστοβάθμια Διδακτορική Διατριβή  
Παπαδάκη Ιωάννη



00140689

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ	
ΑΡ. ΕΙΣ.	40689
ΣΟΜΠ.	4720 à 22384
ΤΑΞΙΝ.	025. 00285 ΠΑ
ΒΙΒΛΙΟΘΗΚΗ	

Πειραιάς, 2002



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΔΙΑΤΡΙΒΗ

για την απόκτηση Διδακτορικού Διπλώματος  
του Τμήματος Πληροφορικής

Ιωάννη Θ. Παπαδάκη

ΨΗΦΙΑΚΕΣ ΒΙΒΛΙΟΘΗΚΕΣ ΒΑΣΙΣΜΕΝΕΣ ΣΤΟ  
WEB: ΑΡΧΙΤΕΚΤΟΝΙΚΕΣ, ΑΣΦΑΛΕΙΑ,  
ΑΝΑΚΤΗΣΗ ΠΛΗΡΟΦΟΡΙΩΝ

*Τριμελής Συμβουλευτική Επιτροπή:*

*Επιβλέπων:*

**Αλεξάνδρης Νικόλαος**  
Καθηγητής Πανεπιστημίου  
Πειραιώς

*Μέλη:*

**Χρυσικόπουλος Βασίλειος**  
Καθηγητής Πανεπιστημίου Ιονίου

**Παναγιωτόπουλος Θεμιστοκλής**  
Αναπληρωτής Καθηγητής  
Πανεπιστημίου Πειραιώς

*Επταμελής Εξεταστική Επιτροπή:*

**Αλεξάνδρης Νικόλαος**  
Καθηγητής Πανεπιστημίου Πειραιώς

**Τσανάκας Παναγιώτης**  
Καθηγητής Ε.Μ.Π.

**Χρυσικόπουλος Βασίλειος**  
Καθηγητής Πανεπιστημίου Ιονίου

**Δουληγέρης Χρήστος**  
Αναπληρωτής Καθηγητής Πανεπιστημίου Πειραιώς

**Μαρτάκος Δρακούλης**  
Αναπληρωτής Καθηγητής Πανεπιστημίου Αθηνών

**Μπάκος Γεώργιος**  
Αναπληρωτής Καθηγητής Πανεπιστημίου Ιονίου

**Παναγιωτόπουλος Θεμιστοκλής**  
Αναπληρωτής Καθηγητής Πανεπιστημίου Πειραιώς



## Πίνακας Περιεχομένων

ΚΕΦΑΛΑΙΟ 1.....	1
1. Εισαγωγή.....	1
1.1 Οι ψηφιακές βιβλιοθήκες στο χώρο της κοινωνίας της πληροφορίας.....	1
1.2 Βασικές αρχές για την ανάπτυξη συστημάτων ψηφιακών βιβλιοθηκών.....	2
1.3 Αντικείμενο και γενικοί στόχοι της Διατριβής.....	7
1.4 Δομή της Διατριβής.....	8
1.5 Συνεισφορά στην ερευνητική περιοχή - Δημοσιεύσεις.....	10
1.6 Ευχαριστίες.....	12
ΚΕΦΑΛΑΙΟ 2.....	15
2. Ψηφιακές Βιβλιοθήκες βασισμένες στο Web και στην XML.....	15
2.1 Εισαγωγή.....	15
2.2 Μεταδεδομένα.....	17
2.3 XML και ψηφιακές βιβλιοθήκες.....	18
2.4 Απαιτήσεις για την ανάπτυξη ψηφιακών βιβλιοθηκών Πανεπιστημιακού υλικού.....	19
2.5 Ανασκόπηση της Διεθνούς βιβλιογραφίας.....	21
2.5.1 Αρχιτεκτονική βασισμένη στο Web και στην XML για αρχεία τηλεπισκόπησης.....	21
2.5.2 Αρχιτεκτονική βασισμένη σε αντικείμενα δεδομένων και συνδέσμους.....	23
2.5.3 Το σύστημα SOMLib.....	24
2.5.4 THEIPS: Μια ψηφιακή βιβλιοθήκη περιβαλλοντικών δεδομένων.....	24
2.5.5 FEDORA: Μια κατανεμημένη αρχιτεκτονική ψηφιακών βιβλιοθηκών.....	25
2.6 Πλαίσιο υποστήριξης ψηφιακών βιβλιοθηκών βασισμένων στο Web.....	26
2.6.1 Ενδιάμεσο στρώμα - το τμήμα XML Descriptor.....	28
2.6.2 Ενδιάμεσο στρώμα - το τμήμα Collection Constructor.....	30
2.6.3 Ενδιάμεσο στρώμα - το τμήμα Search Engine.....	32
2.6.4 Αποθηκευτικός χώρος.....	33
2.6.5 Διακινούμενα έγγραφα.....	34
2.6.5.1 Δομή εγγράφων.....	34

2.6.5.2	Τύποι δεδομένων .....	36
2.7	Η ψηφιακή βιβλιοθήκη DocML .....	37
2.7.1	Αποθηκευτικός χώρος .....	38
2.7.2	Ενδιάμεσο στρώμα .....	42
2.7.3	Στρώμα Πελάτη .....	44
2.7.4	Παρεχόμενες υπηρεσίες .....	44
2.7.4.1	Σύστημα διεπαφής .....	44
2.7.4.2	Αναζήτηση .....	45
2.7.4.3	Πλοήγηση .....	46
2.7.4.4	Μεταφορά δεδομένων .....	46
2.7.5	Διαχειριστικά θέματα .....	47
2.8	Συμπεράσματα .....	48
ΚΕΦΑΛΑΙΟ 3	.....	51
3. Μουσικές Ψηφιακές Βιβλιοθήκες βασισμένες στο Web	.....	51
3.1	Εισαγωγή .....	51
3.2	Απαιτήσεις μιας μουσικής ψηφιακής βιβλιοθήκης .....	52
3.3	Μεταδεδομένα στη μουσική πληροφορία .....	54
3.4	Ανασκόπηση της Διεθνούς βιβλιογραφίας .....	55
3.4.1	Η μουσική ψηφιακή βιβλιοθήκη Jukebox .....	55
3.4.2	Η μουσική ψηφιακή βιβλιοθήκη Patron .....	57
3.4.3	Η μουσική ψηφιακή βιβλιοθήκη VARIATIONS .....	58
3.4.4	Ανάκτηση πληροφοριών βασισμένη στο περιεχόμενο σε μουσικές ψηφιακές βιβλιοθήκες .....	59
3.4.5	Μουσικές ψηφιακές βιβλιοθήκες βασισμένες σε αρχιτεκτονικές Peer-to-Peer (P2P) .....	60
3.4.5.1	Η ψηφιακή βιβλιοθήκη Napster .....	61
3.4.5.2	Η ψηφιακή βιβλιοθήκη Gnutella .....	62
3.5	Η προτεινόμενη αρχιτεκτονική .....	63
3.5.1	Κίνητρο .....	63
3.5.2	Αρχιτεκτονική .....	63
3.5.2.1	Αποθηκευτικός χώρος .....	64

3.5.2.2	Ενδιάμεσο στρώμα .....	67
3.5.2.3	Στρώμα πελάτη .....	71
3.5.3	Πλάνο λειτουργίας ενός ραδιοφωνικού σταθμού στο Web .....	72
3.6	Συμπεράσματα και ανοιχτά θέματα .....	73
ΚΕΦΑΛΑΙΟ 4 .....		75
4	Ιατρικές Ψηφιακές Βιβλιοθήκες .....	75
4.1	Εισαγωγή .....	75
4.2	Ανασκόπηση της Διεθνούς βιβλιογραφίας .....	77
4.2.1	Ηλεκτρονικός Φάκελος Ασθενή (ΗΦΑ) βασισμένος σε συστήματα βάσεων δεδομένων πρόσβασιμα από το Web .....	77
4.2.1.1	Το σύστημα Ιπποκράτης .....	78
4.2.1.2	Το σύστημα PROMPT .....	78
4.2.2	Χρήση ηλεκτρονικών εγγράφων για την αναπαράσταση του ΗΦΑ .....	79
4.2.3	Roaming Electronic HealthCare Record (R-EHCR) .....	81
4.2.4	Η αρχιτεκτονική InterCare .....	82
4.2.5	Το σύστημα τηλεϊατρικής Euromed-Euromed ETS .....	84
4.3	Αρχιτεκτονική ασφαλούς ιατρικής ψηφιακής βιβλιοθήκης .....	85
4.3.1	Αποθηκευτικός Χώρος .....	86
4.3.2	Ενδιάμεσο στρώμα .....	86
4.3.3	Στρώμα πελάτη .....	87
4.4	Απειλές ασφάλειας .....	88
4.5	Υπηρεσίες Ασφάλειας .....	89
4.6	Επαγγελματικές Κάρτες Υγείας - ΕΚΥ και ΕΤΟ .....	92
4.7	Συμπεράσματα και ανοιχτά θέματα .....	93
ΚΕΦΑΛΑΙΟ 5 .....		95
5	Ανάκτηση πληροφοριών σε συστήματα ψηφιακών βιβλιοθηκών βασισμένων στο Web .....	95
5.1	Εισαγωγή .....	95
5.2	Ανασκόπηση της Διεθνούς βιβλιογραφίας .....	97
5.2.1	Ταξινομημένες λίστες με αποτελέσματα αναζήτησης σε ψηφιακές βιβλιοθήκες .....	98

5.2.2	Ομαδοποίηση αποτελεσμάτων αναζήτησης σε ψηφιακές βιβλιοθήκες .....	98
5.2.3	Οπτικοποίηση των αποτελεσμάτων .....	100
5.2.3.1	Δυσδιάστατες απεικονίσεις .....	100
5.2.3.2	Τρισδιάστατες απεικονίσεις .....	101
5.2.4	Συστήματα επεξεργασίας αποτελεσμάτων αναζήτησης στο Web .....	102
5.2.5	Μηχανές αναζήτησης στο Web και παρουσίαση αποτελεσμάτων .....	103
5.2.6	Αποτελέσματα αναζήτησης και ψηφιακές βιβλιοθήκες βασισμένες στο Web .....	104
5.3	Αλληλεπιδραστικά αποτελέσματα αναζήτησης .....	105
5.3.1	Αρχιτεκτονική .....	105
5.3.2	Μεταδεδομένα .....	106
5.3.3	Χώρος εκτέλεσης της εφαρμογής .....	107
5.3.4	Πρότυπη Υλοποίηση .....	108
5.3.4.1	Η Εφαρμογή .....	108
5.3.4.2	Θέματα υλοποίησης .....	112
5.3.4.3	Επέκταση στο Web .....	113
5.3.5	Αξιολόγηση .....	113
5.3.5.1	Η Μέθοδος .....	113
5.3.5.2	Δομή της μεθόδου αξιολόγησης .....	114
5.3.5.3	Το περιβάλλον εκτέλεσης (Testbed) .....	114
5.3.5.4	Οι συμμετέχοντες .....	115
5.3.5.5	Το ερωτηματολόγιο .....	115
5.3.5.6	Ανάλυση αποτελεσμάτων .....	116
5.4	Συμπεράσματα και ανοιχτά θέματα .....	118
Κεφάλαιο 6	.....	121
6. Συμπεράσματα της Διατριβής.....		121
6.1	Συμπεράσματα και ανοιχτά θέματα .....	121
6.1.1	Σχεδιασμός και αρχιτεκτονική ψηφιακών βιβλιοθηκών .....	122
6.1.2	Μουσικές ψηφιακές βιβλιοθήκες διαχείρισης αρχείων mp3 .....	123
6.1.3	Ιατρικές ψηφιακές βιβλιοθήκες .....	124
6.1.4	Ανάκτηση πληροφοριών .....	125

Κεφάλαιο 7 .....	127
Βιβλιογραφικές αναφορές .....	127
ΠΑΡΑΡΤΗΜΑ ΚΕΦΑΛΑΙΟΥ 5 .....	141
ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ (στην Αγγλική γλώσσα) .....	141

Πανεπιστήμιο Πειραιώς



## 1. Εισαγωγή

### 1.1 Οι ψηφιακές βιβλιοθήκες στο χώρο της κοινωνίας της πληροφορίας

Στην εποχή του Διαδικτύου και του Παγκόσμιου Ιστού (Web), ο χώρος των ψηφιακών βιβλιοθηκών βρίσκεται συνεχώς στο προσκήνιο. Η επέκταση του Διαδικτύου και ειδικότερα η τεράστια επιτυχία που έχει γνωρίσει το Web, οφείλεται σε μεγάλο ποσοστό σε τρία θεμελιώδη του χαρακτηριστικά. Εξαιτίας της *παγκοσμιότητας* του (universality), ο οποιοσδήποτε μπορεί να συμμετέχει στο Web είτε ως συγγραφέας, είτε ως εκδότης, είτε ως καταναλωτής με ελάχιστες επενδύσεις σε υλικό ή/και λογισμικό. Εξαιτίας της *ομοιομορφίας* του, πηγές πληροφορίας, υπηρεσίες και χρήστες συμμετέχουν στο Web ως ισότιμα μέλη σε ένα κοινό περιβάλλον πληροφορίας. Τέλος, εξαιτίας της *αποκέντρωμένης δομής* του, το Web είναι άναρχο, καθώς δεν υπάρχει απόλυτη συμφωνία σε τεχνικό επίπεδο όσον αφορά τις τεχνολογίες και τα πρωτόκολλα που επικρατούν.

Οι παραπάνω αρχές αν και αποτελούν το βασικό λόγο επιτυχίας του Web, εντούτοις αποτελούν και την αιτία των βασικών του προβλημάτων. Είναι κοινή λεπτοίηση ανάμεσα στους χρήστες του Web πως η παγκοσμιότητα συχνά μεταφράζεται σε “ποσότητα δίχως ποιότητα”. Επίσης, λόγω της ομοιομορφίας, είναι πολύ δύσκολο να εντοπίσει κανείς εργαλεία, τεχνολογίες και γενικότερα καθοδήγηση επικεντρωμένη σε ειδικευμένους επιστημονικούς χώρους (π.χ. γεωγραφικών δεδομένων, μουσικών, κ.ά). Τέλος, η αποκέντρωση καθιστά πολύ δύσκολη την

επιβολή οργανωτικών δομών τέτοιων που να διασφαλίζουν την ακεραιότητα της πληροφορίας, σε επίπεδο αξιοπιστίας, πρόσβασης, ασφάλειας και διατήρησης αυτής.

Προσπαθώντας να αντιμετωπίσει τα παραπάνω προβλήματα, ο τομέας των ψηφιακών βιβλιοθηκών φέρνει σε επιφάνεια επιστήμονες από διάφορες ερευνητικές επιστημονικές περιοχές. Ανάλογα με το επιστημονικό προφίλ του κάθε ερευνητή που ασχολείται με το χώρο αυτό, προκύπτουν διάφορες προσεγγίσεις. Έτσι, θα μπορούσε να υποστηριχθεί κανείς πως οι ψηφιακές βιβλιοθήκες προέκυψαν ως εξέλιξη των συμβατικών βιβλιοθηκών στην ψηφιακή εποχή, ή ακόμα πως ο τομέας αυτός δεν είναι τίποτα άλλο παρά ένα βήμα προς την αυτοματοποίηση των βιβλιοθηκονομικών εργασιών. Από μια άλλη οπτική γωνία, οι ψηφιακές βιβλιοθήκες μπορούν να θεωρηθούν ως μια μορφή συνέκωσης και ομογενοποίησης πολλών και διαφορετικών συστημάτων βάσεων δεδομένων. Μια άλλη θεωρία υποστηρίζει ότι οι ψηφιακές βιβλιοθήκες αποτελούν μια εξειδικευμένη χρήση του Web. Αν και όλες αυτές οι απόψεις αποτυπώνουν με επιτυχία διάφορες πτυχές του κλάδου αυτού, εντούτοις καμιά τους δεν καταφέρνει να αντιμετωπίσει τις ψηφιακές βιβλιοθήκες ως ένα ενιαίο και αυτόνομο χώρο έρευνας. Για να μπορέσει κανείς να καταλάβει ποιά είναι η θέση τους στη σύγχρονη κοινωνία της πληροφορίας, θα πρέπει να θεωρήσει τις ψηφιακές βιβλιοθήκες ως την τομή πολλών επιστημονικών περιοχών, εμπλουτισμένη με επιπρόσθετα χαρακτηριστικά που οδηγούν στη δημιουργία νέων ερευνητικών θεμάτων.

Έχοντας υπόψη τα παραπάνω δεδομένα, προκύπτει ότι οι ψηφιακές βιβλιοθήκες μπορούν να θεωρηθούν ως συλλογές ψηφιακών (αυγώς ή/και ψηφιοποιημένων) αντικειμένων, εμπλουτισμένες με τις υπηρεσίες εκείνες που είναι αναγκαίες για τη σωστή διαχείριση και χρήση αυτών. Ως ψηφιακό αντικείμενο μπορεί να θεωρηθεί κάτι απλό όπως ένα κείμενο χαρακτηριστών αλλά και κάτι πιο σύνθετο, όπως το τρισδιάστατο μοντέλο ενός ανθρώπινου οργάνου σε ψηφιακή μορφή.

## **1.2 Βασικές αρχές για την ανάπτυξη συστημάτων ψηφιακών βιβλιοθηκών**

Η ανάπτυξη μιας ψηφιακής βιβλιοθήκης προϋποθέτει την τήρηση κάποιων κανόνων όσον αφορά το σχεδιασμό, την υλοποίηση και τη διαχείριση αυτής. Έτσι, πρέπει πάντοτε να λαμβάνεται υπόψη το γεγονός ότι οι τεχνολογίες στο Διαδίκτυο διαρκώς αλλάζουν και εξελίσσονται. Το πρόβλημα της διαρκούς εναλλαγής τεχνολογιών είναι πολύ σημαντικό στο χώρο των ψηφιακών βιβλιοθηκών. Η διατήρηση των ψηφιακών αντικειμένων (π.χ. έγγραφα) σε



μια και μοναδική μορφοποίηση (format), τα πνευματικά και χρηστικά δικαιώματα της οποίας ανήκουν σε συγκεκριμένη εταιρία, περιέχει: πάντα τον κίνδυνο να μη μπορεί κανείς να προσπελάσει τα αντκειμένα αυτά τη στιγμή που η εταιρία αυτή διακόψει την υποστήριξη της συγκεκριμένης μορφοποίησης. Αν και σε αυτήν την περίπτωση τα ψηφιακά αντικείμενα είναι επαρκώς διατηρημένα, εντούτοις δε μπορεί κανείς να τα εκμεταλλευτεί!

Από την πλευρά των χρηστών, το πιο ενδιαφέρον και πολύτιμο συστατικό μιας ψηφιακής βιβλιοθήκης είναι το περιεχόμενό της. Κατά το σχεδιασμό και δημιουργία μιας ψηφιακής βιβλιοθήκης πρέπει να ληφθούν αποφάσεις σχετικές με την επιλογή των αντικειμένων εκείνων που θα αποτελέσουν το περιεχόμενό της.

Επίσης, πρέπει να αποφασιστεί ο τρόπος ψηφιοποίησης των αντικειμένων εκείνων που ενδεχομένως να υπάρχουν μόνο σε αναλογική μορφή, καθώς και η προσθήκη μεταδεδομένων σε αυτά χρησιμοποιώντας αντίστοιχες γλώσσες (π.χ. XML, SGML), βασισμένες σε ευρέως αποδεκτά πρότυπα (standards). Ειδικότερα, για τη χρήση των μεταδεδομένων, πρέπει να καθοριστούν εκ των προτέρων μέχρι ποιο βάθος θα υπάρχουν μεταδεδομένα (για παράδειγμα, μεταδεδομένα σε επίπεδο βιβλίου σε αντιδιαστολή με μεταδεδομένα σε επίπεδο κεφαλαίου). Η απόδοση μοναδικών προοδιοριστών (identifiers) στα ψηφιακά αντικείμενα μιας ψηφιακής βιβλιοθήκης είναι ένα ακόμα κρίσιμο σημείο στην ανάπτυξη μιας τέτοιας εφαρμογής, καθώς στο χώρο αυτό συναντάει κανείς αντικείμενα μέσα σε αντικείμενα που είναι με τη σειρά τους καταμεμημένα σε διάφορα και συνάμα διαφορετικά υπολογιστικά συστήματα.

Ενα ακόμα βασικό συστατικό για τον επιτυχή σχεδιασμό μιας ψηφιακής βιβλιοθήκης είναι και η σωστή επιλογή των εμπλεκόμενων προσώπων. Δεδομένου ότι ο χώρος αυτός συνδυάζει την τεχνολογία διαφόρων επιστημονικών κλάδων, στον κύκλο ανάπτυξης μιας ψηφιακής βιβλιοθήκης πρέπει να συμμετέχουν πρόσωπα από διάφορες ειδικότητες. Οι δυο κλάδοι που εμπλέκονται πιο άμεσα, είναι ο κλάδος της Πληροφορικής και ο κλάδος της Βιβλιοθηκονομίας. Οι επιστήμονες της Πληροφορικής γνωρίζουν τις προοπτικές, αλλά και τους περιορισμούς της τεχνολογίας αναφορικά με την επίτευξη ενός στόχου τέτοιου βεληγκεούς, ενώ οι βιβλιοθηκονόμοι προσφέρουν την εμπειρία τους πάνω σε θέματα καταλογογράφησης, αρχειοθέτησης και ευρετηρίασης πολύ μεγάλου όγκου δεδομένων. Επίσης, η προσωπική επαφή των βιβλιοθηκονόμων με πρόσωπα από διάφορες επιστημονικές ομάδες τους καθιστά ιδανικούς συμβούλους για την επιλογή των λειτουργιών που θα υποστηρίξει το τελικό προϊόν καθώς και για το σχεδιασμό του αντίστοιχου συστήματος διεπαφής. Η αρμονική συνεργασία προσώπων από διαφορετικούς επιστημονικούς χώρους είναι δυνατή μόνο αν υπάρχει σαφής καθορισμός

της ακτίνας δράσης όλων των εμπλεκόμενων προσώπων. Πρέπει ακόμα να τονιστεί πως η μακροπρόθεσμη διαθεσιμότητα των ψηφιακών αντικειμένων εξασφαλίζεται από την αντιμετώπιση της ψηφιακής βιβλιοθήκης ως ένα έργο με προοπτική στο χρόνο, όπως ακριβώς συμβαίνει με τις συμβατικές βιβλιοθήκες.

Οι περισσότερες ψηφιακές βιβλιοθήκες είναι διαθέσιμες στο Διαδίκτυο μέσω της τεχνολογίας του Web. Οι πολλές δυνατότητες που προσφέρει το Web στο σχεδιασμό εφαρμογών ευρείας χρήσης, έχουν ωθήσει τους σχεδιαστές ψηφιακών βιβλιοθηκών να παρέχουν συστήματα διεπαφής στο Web. Κατά τη δημιουργία λοιπόν τέτοιων εφαρμογών, πρέπει να δίνεται ιδιαίτερο βάρος στη χρηστικότητα (usability) του τελικού προϊόντος. Η χρηστικότητα εξαρτάται από ένα πλήθος παραγόντων, όπως είναι οι τεχνικές διαφορές μεταξύ των διαφόρων υπολογιστικών συστημάτων που κυκλοφορούν στην αγορά, καθώς και ο διαφορετικός τρόπος με τον οποίο ο εκάστοτε Web Browser (ακόμα και μεταξύ Web Browsers της ίδιας εταιρίας αλλά διαφορετικής έκδοσης) χειρίζεται και αποτυπώνει την πληροφορία που δέχεται από τον αντίστοιχο Web εξηρητητή. Γι αυτό λοιπόν, τα συστήματα διεπαφής των ψηφιακών βιβλιοθηκών που είναι βασισμένες στο Web πρέπει να είναι σχεδιασμένα όσο το δυνατόν πιο απλά, έτσι ώστε να διασφαλίζεται η ίδια συμπεριφορά από όλους τους Web Browsers που ενδεχομένως να τα φιλοξενήσουν.

Η κατοχύρωση ανοικτής πρόσβασης (open access) στις ψηφιακές βιβλιοθήκες σχετίζεται άμεσα με τη χρηστικότητά τους. Γενικότερα, ανοικτή πρόσβαση στην πληροφορία επιτυγχάνεται όταν υπάρχει πρόσβαση στο δίκτυο υπολογιστών που τη φιλοξενεί. Επίσης, η τεχνολογία που χρησιμοποιείται πρέπει να είναι εύκολη στη χρήση της και οι χρήστες να έχουν τα κατάλληλα προσόντα και τις απαραίτητες γνώσεις για να φτάσουν γρήγορα και με απλό τρόπο στην πηγή πληροφορίας που αναζητούν. Τέλος, η πληροφορία πρέπει να είναι σε τέτοια μορφή ώστε να γίνεται αμέσως αντιληπτή και κατανοητή. Ένας τρόπος για να εξασφαλιστεί ανοικτή πρόσβαση στο περιεχόμενο μιας ψηφιακής βιβλιοθήκης είναι η αποφυγή χρήσης υλικού και λογισμικού, τα δικαιώματα εκμετάλλευσης του οποίου μονοπωλούνται από συγκεκριμένη εταιρία (proprietary), όπου βέβαια αυτό είναι δυνατόν. Αν και είναι λογικό να χρησιμοποιούνται διάφορα εμπορικά πακέτα για τη δημιουργία του περιεχομένου, εντούτοις πρέπει να αποφεύγεται η απαίτηση εξειδικευμένου λογισμικού για την πρόσβαση σε αυτό. Την αρχή αυτή φαίνεται να την έχουν υιοθετήσει αρκετές εταιρίες που χρεώνουν την αγορά των εργαλείων δημιουργίας ψηφιακών αντικειμένων, αλλά παρέχουν δωρεάν τα συστήματα διεπαφής εκείνα που είναι απαραίτητα για την πρόσβαση σε αυτά. Με αυτόν τον τρόπο, το περιεχόμενο μιας

ψηφιακής βιβλιοθήκης θα είναι προσβάσιμο, όσο είναι διαθέσιμα τα αντίστοιχα συστήματα διεπαφής.

Ενα από τα πιο δυσπρόσιτα προβλήματα στο χώρο των ψηφιακών βιβλιοθηκών είναι και αυτό των πνευματικών δικαιωμάτων. Οι υπάρχοντες νόμοι ανά τον κόσμο αναφέρονται στην προστασία δικαιωμάτων περισσότερο πάνω σε υλικά αγαθά παρά σε ψηφιακά. Η κοινωνία της ψηφιακής πληροφορίας στην οποία είμαστε μέλη, έχει πολλά πλεονεκτήματα αλλά ταυτόχρονα εγκυμονεί και πολλούς κινδύνους. Τα πλεονεκτήματα έχουν να κάνουν με την ευκολία πρόσβασης στην πληροφορία που προσφέρεται, ενώ τα προβλήματα εντοπίζονται στις ευκαιρίες που υπάρχουν για κακόβουλη εκμετάλλευση της εύκολης αυτής πρόσβασης καθώς και στην παράνομη αναπαραγωγή αυτής. Δυστυχώς μέχρι σήμερα, δεν υπάρχουν άμεσες λύσεις όσον αφορά την επιβολή νόμων που κατοχυρώνουν τα πνευματικά δικαιώματα πληροφοριών που είναι διαθέσιμες σε ψηφιακή μορφή. Χωρίς αμφιβολία, στα επόμενα χρόνια, θα βρεθεί κάποια λύση σε αυτά τα προβλήματα. Μέχρι τότε, οι σχεδιαστές ψηφιακών βιβλιοθηκών πρέπει να υιοθετούν κατάλληλες τακτικές ώστε να προστατεύουν τις συλλογές τους όσο το δυνατόν καλύτερα.

Γενικότερα, σε μια ψηφιακή βιβλιοθήκη ισχύει πως όσο πιο πολλά αυτοματοποιημένα εργαλεία δημιουργούνται και χρησιμοποιούνται, τόσο μεγαλύτερη οικονομία εξασφαλίζεται σε ανθρώπινους πόρους. Τα εργαλεία αυτά πρέπει να είναι εύκολα στη χρήση τους, ώστε να επιτρέπουν στους αναλυτές να ασχολούνται μόνο με την ουσία της δουλειάς τους.

Μια από τις πιο σημαντικές αρχές στη σχεδίαση ψηφιακών βιβλιοθηκών είναι η υιοθέτηση και χρήση προτύπων (standards). Βασικά πλεονεκτήματα που προκύπτουν από τη χρήση προτύπων είναι η δημιουργία εφαρμογών που μπορούν εύκολα να επεκταθούν ανάλογα με τις απαιτήσεις και δυνατότητες που προσφέρει η εκάστοτε νέα τεχνολογία, να επικοινωνήσουν με παρόμοιες εφαρμογές και να μετατραπούν κατά τέτοιο τρόπο ώστε να μπορούν να εκτελεστούν σε περισσότερες από μια πλατφόρμες. Όλα αυτά τα χαρακτηριστικά είναι σημαντικά για τις ψηφιακές βιβλιοθήκες σε τομείς που απαιτούν σημαντική εργασία όπως στην ψηφιοποίηση, στην εισαγωγή μεταδομένων και στη σημείωση εγγράφων (markup).

Η αξιολόγηση της ποιότητας μιας ψηφιακής βιβλιοθήκης μπορεί να γίνει σε όλους τους επί μέρους τομείς αυτής, όπως στην επιλογή του περιεχομένου, στην εισαγωγή των αντίστοιχων μεταδομένων, στην ψηφιοποίηση και γενικότερα στη χρησιμότητά της. Κάποιες μέθοδοι αξιολόγησης μπορούν να ενταχθούν σε μια αυτοματοποιημένη διαδικασία, ενώ κάποιες άλλες απαιτούν σχολαστικότητα και κριτική σκέψη από έμπειρο προσωπικό. Είναι πολύ σημαντικό να



ενταχθεί η αξιολόγηση στον κύκλο λειτουργίας μιας ψηφιακής βιβλιοθήκης, έτσι ώστε να υπάρχει βεβαιότητα ανά πάσα στιγμή για την καλή ποιότητα του περιεχομένου και των υπηρεσιών που προσφέρει.

Τέλος, θεωρώντας την ψηφιακή βιβλιοθήκη ως μετεξέλιξη της συμβατικής βιβλιοθήκης, πρέπει να ληφθεί υπόψη ότι μια από τις πιο θεμελιώδεις απαιτήσεις μιας συμβατικής βιβλιοθήκης είναι και η αντοχή της στο χρόνο. Δυστυχώς, σύμφωνα με τους ειδικούς [Rothenberg, 1999], μέχρι σήμερα, δεν έχει βρεθεί οίγουρος τρόπος μακροπρόθεσμης διατήρησης της ψηφιακής πληροφορίας. Η πιο δημοφιλής προσέγγιση σε αυτό το θέμα είναι η τακτική της μετατροπής των ψηφιακών δεδομένων και προγραμμάτων από μια μορφή σε μια άλλη, ή η μεταφορά τους από ένα φυσικό μέσο αποθήκευσης σε κάποιο άλλο. Με αυτόν τον τρόπο, τα ψηφιακά αντικείμενα είναι πάντα διαθέσιμα μέσω των εργαλείων εκείνων και των μορφών που επικρατούν στην αγορά την κάθε στιγμή. Ωστόσο, για να έχει αποτέλεσμα αυτό του είδους η τακτική, πρέπει να ενταχθεί στον κύκλο λειτουργίας μιας ψηφιακής βιβλιοθήκης και να μην αποτελεί μια ενέργεια που θα εκτελεστεί μόνο μια φορά.

*Λαμβάνοντας υπόψη τα προαναφερθέντα, η παρούσα Διδακτορική Διατριβή προσεγγίζει τις ψηφιακές βιβλιοθήκες υπό το πρίσμα των αναγκών αυτού του χώρου για λύσεις που προέρχονται από την επιστήμη της Πληροφορικής. Συγκεκριμένα, παρουσιάζονται διάφορες αρχιτεκτονικές ψηφιακών βιβλιοθηκών, οι οποίες αντιμετωπίζουν προβλήματα που πηγάζουν από την ύπαρξη πολλών και διαφορετικών μορφών δεδομένων σε τέτοια συστήματα, όπως είναι τα δεδομένα κειμένου, εικόνας, ήχου καθώς και συνδυασμού αυτών. Επίσης, περιγράφεται ένα μοντέλο ασφάλειας σε επίπεδο αρχιτεκτονικής και σχεδιασμού, που σκοπό έχει να προστατέψει τις κρίσιμες και ευαίσθητες πληροφορίες που υπάρχουν στο μοντέλο μιας κατανεμημένης ιατρικής ψηφιακής βιβλιοθήκης. Τέλος, η παρούσα εργασία συνεισφέρει στο χώρο της ανάκτησης πληροφοριών, περιγράφοντας ένα δυναμικό τρόπο αλληλεπίδρασης των χρηστών με τα αποτελέσματα μιας αναζήτησής τους, όπως αυτά παρέχονται από μια μηχανή αναζήτησης (προερχόμενη είτε από το Διαδίκτυο ή από μια ψηφιακή βιβλιοθήκη), με σκοπό την εξαγωγή χρήσιμων και σαφών συμπερασμάτων όσον αφορά την χρηστικότητα των αποτελεσμάτων αναζήτησης.*

### 1.3 Αντικείμενο και γενικοί στόχοι της Διατριβής

Η γνωστή περιοχή και το αντικείμενο ενδιαφέροντος της παρούσας Διδακτορικής Διατριβής είναι θέματα αρχιτεκτονικής, ασφάλειας και ανώτακτης πληροφοριών σε ψηφιακές βιβλιοθήκες.

Στα πλαίσια της έρευνας στον τομέα της αρχιτεκτονικής ψηφιακών βιβλιοθηκών, παρουσιάζεται ένα πλαίσιο αρχιτεκτονικών ψηφιακών βιβλιοθηκών προσβάσιμων από το Web, καθώς και μια ψηφιακή βιβλιοθήκη - η DocML - για τη διαχείριση Πανεπιστημιακού υλικού, βασισμένη στο παραπάνω πλαίσιο. Τα δυο αυτά συστήματα βασίζονται στην κατανομημένη προσέγγιση και σε τεχνολογίες Διαδικτύου. Η επικοινωνία των διαφόρων αρχιτεκτονικών στρωμάτων μεταξύ τους καθώς και η επικοινωνία μεταξύ των τμημάτων λογισμικού στο ίδιο στρώμα επιτυγχάνεται μέσω της χρήσης της μεταγλώσσας eXtensible Markup Language (XML). Η XML χρησιμοποιείται και για τη δημιουργία ενός κατανομημένου συστήματος καταχώρισης μεταδεδομένων για την περιγραφή των αντικειμένων της ψηφιακής βιβλιοθήκης.

Ακολουθώντας τις παραπάνω αρχές, παρουσιάζεται ο σχεδιασμός και η αρχιτεκτονική μιας κατανομημένης ψηφιακής βιβλιοθήκης τριών στρωμάτων προσβάσιμης από το Web για τη διαχείριση μουσικών αρχείων που κωδικοποιούνται σε Mpeg Layer 3 (mp3). Η συγκεκριμένη ψηφιακή βιβλιοθήκη αξιοποιεί την πληροφορία που βρίσκεται ενθυλακωμένη μέσα στο κυρίως σώμα του αρχείου τύπου mp3 και αφορά μεταδεδομένα όπως "ερμηνευτής", "είδος μουσικής" και "τίτλος αναφερόμενου τραγουδιού". Η πληροφορία αυτή είναι οργανωμένη σε ένα κατανομημένο σύνολο μεταδεδομένων σε μορφή XML, για το οποίο υπάρχει δυνατότητα διαχείρισης και ελέγχου από ένα κεντρικό σημείο που βρίσκεται στο ενδιάμεσο στρώμα. Οι χρήστες του συστήματος επικοινωνούν με το ενδιάμεσο στρώμα το οποίο αναλαμβάνει να ικανοποιήσει τα αιτήματα των χρηστών, επικοινωνώντας με τη σειρά του με τον αποθηκευτικό χώρο στον οποίο βρίσκονται αποθηκευμένα τα αρχεία τύπου mp3.

Στα πλαίσια της έρευνας στον τομέα της ασφάλειας των ψηφιακών βιβλιοθηκών, παρουσιάζεται μια ασφαλής ψηφιακή βιβλιοθήκη βασισμένη στις προδιαγραφές της Common Object Request Broker Architecture (CORBA) για κατανομημένα συστήματα. Βασίζεται σε μια αρχιτεκτονική τριών στρωμάτων που έχει ως κεντρικό συστατικό τον Web εξυπηρετητή. Η ιδιόμορφη φύση των αιτηρών δεδομένων για τα οποία η ακεραιότητα και εμπιστευτικότητα πρέπει να είναι εξασφαλισμένη, υπαγορεύει τη χρήση τεχνολογίας Τριών Εμπιστων Οντοτήτων

(TPO, Trusted Third Parties-TTP) για την υποστήριξη των αντίστοιχων υπηρεσιών ασφάλειας. Ένα ακόμα βασικό χαρακτηριστικό του μηχανισμού ασφάλειας της προτεινόμενης ψηφιακής βιβλιοθήκης είναι η χρήση έξυπνης κάρτας για την αποθήκευση και διάδοση των απαραίτητων δεδομένων ασφάλειας (π.χ. μυστικό κλειδί κρυπτογράφησης).

Στα πλαίσια της έρευνας στον τομέα της ανάκτησης πληροφοριών, παρουσιάζεται μια μεθοδολογία διαχείρισης των αποτελεσμάτων όπως αυτά επιστρέφονται από μια μηχανή αναζήτησης. Η μεθοδολογία αυτή, απαιτεί αλληλεπίδραση με τους χρήστες για την εξαγωγή ασφαλών συμπερασμάτων όσον αφορά τη χρηστική αξία των επιστρεφόμενων αποτελεσμάτων. Επίσης, παρουσιάζεται ένας τρόπος ταξινόμησης των αποτελεσμάτων αυτών σύμφωνα με τη σπουδαιότητά τους, η οποία ορίζεται ως ένας σταθμισμένος συνδυασμός της σχετικότητάς τους με το αρχικό ερώτημα και της δημοτικότητάς τους στον πληθυσμό που έχει πρόσβαση σε αυτά. Η συγκεκριμένη έρευνα έχει αξιολογηθεί από χρήστες του Web, οι οποίοι κλήθηκαν να συμπληρώσουν ένα κατάλληλα διαμορφωμένο ερωτηματολόγιο, το οποίο αναφερόταν σε μια εφαρμογή που αναπτύχθηκε ειδικά για το σκοπό αυτό.

## **1.4 Δομή της Διατριβής**

Η παρούσα Διδακτορική Διατριβή περιλαμβάνει επτά (7) Κεφάλαια.

Στο Κεφάλαιο 1 “Εισαγωγή” περιλαμβάνεται το γνωστικό αντικείμενο, οι γενικότεροι στόχοι, η δομή της Διατριβής και παρουσιάζεται η συνεισφορά στην ερευνητική περιοχή.

Στο Κεφάλαιο 2 “Ψηφιακές Βιβλιοθήκες βασισμένες στο Web και στην XML” παρουσιάζεται ένα πλαίσιο στήριξης αρχιτεκτονικών βασισμένων στο Web και στο πρότυπο της XML. Το πλαίσιο αυτό βασίζεται στο μοντέλο των τριών στρωμάτων για καταναμημένες εφαρμογές στο Web, ενώ χρησιμοποιεί τις δυνατότητες που παρέχει το πρότυπο της XML για την υλοποίηση των παρεχόμενων υπηρεσιών προς τους χρήστες. Στη συνέχεια, παρουσιάζεται η αρχιτεκτονική της DocML, η οποία βασίζεται σε μεγάλο βαθμό στις κατευθύνσεις του παραπάνω πλαισίου. Το βασικό σημείο διαφοροποίησης των δυο προσεγγίσεων εστιάζεται στο μοντέλο ανάκτησης πληροφοριών. Ετσι, σε αντίθεση με το αρχικό πλαίσιο στήριξης όπου χρησιμοποιείται το καταναμημένο μοντέλο, η αρχιτεκτονική της DocML χρησιμοποιεί το συγκεντρωτικό μοντέλο



για την υποστήριξη των λειτουργιών ανάκτησης πληροφοριών που παρέχει στους χρήστες. Το Κεφάλαιο ολοκληρώνεται με την παράθεση των πλεονεκτημάτων και μειονεκτημάτων των δυο προσεγγίσεων.

Στο Κεφάλαιο 3 “*Μουσικές Ψηφιακές Βιβλιοθήκες στο Web*” παρουσιάζεται ο σχεδιασμός και η αρχιτεκτονική μιας κατανεμημένης μουσικής ψηφιακής βιβλιοθήκης στο Web. Η προτεινόμενη ψηφιακή βιβλιοθήκη διαχειρίζεται αρχεία ηχητικού περιεχομένου τύπου mp3, τα οποία είναι ευρέως διαδεδομένα στο Web. Επίσης, χρησιμοποιεί τεχνολογίες οι οποίες έχουν καθιερωθεί στο χώρο, όπως ο Web browser και ο Web εξυπηρετητής, έτσι ώστε να ευνοείται η γρήγορη διάδοση των αντίστοιχων εφαρμογών στην κοινωνία του Web. Η επικοινωνία των επί μέρους τμημάτων της επιτυγχάνεται μέσω της χρήσης του προτύπου της XML. Η παρουσίαση της αρχιτεκτονικής γίνεται μέσα από την περιγραφή ενός ραδιοφωνικού σταθμού, ο οποίος κατασκευάζει και εκπέμπει κανάλια ήχου στο Web.

Στο Κεφάλαιο 4 “*Ιατρικές Ψηφιακές Βιβλιοθήκες*” παρουσιάζεται η αρχιτεκτονική μιας κατανεμημένης ιατρικής ψηφιακής βιβλιοθήκης βασισμένη στην αρχιτεκτονική CORBA. Οι συλλογές που φιλοξενούνται στη συγκεκριμένη αρχιτεκτονική κυμαίνονται από απλά έγγραφα κειμένου μέχρι πολύπλοκα υπολογιστικά μοντέλα. Το κυρίως μέρος του Κεφαλαίου αυτού περιγράφει τη θωράκιση της ιατρικής ψηφιακής βιβλιοθήκης από κινδύνους ασφάλειας που συναντώνται σε ένα κατανεμημένο περιβάλλον. Κατόπιν, γίνεται ιδιαίτερη αναφορά στο ρόλο των Εμπιστών Τρίτων Οντοτήτων (ΕΤΟ) για τη διαχείριση των δεδομένων ασφάλειας που απαιτούνται από την αρχιτεκτονική, όπως επίσης και στις έξυπνες κάρτες για την αποθήκευση και μεταφεριμότητα των δεδομένων αυτών.

Στο Κεφάλαιο 5 “*Ανάκτηση πληροφοριών σε συστήματα ψηφιακών βιβλιοθηκών στο Web*” εισάγεται η έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης, όπως αυτά παρέχονται από συστήματα ανάκτησης πληροφοριών στο Web αλλά και από ψηφιακές βιβλιοθήκες βασισμένες στο Web ειδικότερα. Επίσης, ορίζεται η σημαντικότητα (importance) ως κριτήριο ταξινόμησης των αποτελεσμάτων αυτών, η οποία αποτελεί σταθμισμένο συνδυασμό της σχετικότητας (relevancy) και της δημοτικότητας (popularity) των αντίστοιχων πηγών στις οποίες αναφέρονται. Στη συνέχεια του Κεφαλαίου, παρουσιάζεται μια πρότυπη εφαρμογή αλληλεπιδραστικής διαχείρισης των αποτελεσμάτων αναζήτησης, όπως αυτά επιστρέφονται από



την αντίστοιχη μηχανή αναζήτησης της UPLN, μιας ψηφιακής βιβλιοθήκης Πανεπιστημιακού υλικού βασισμένης στο Web. Το κεφάλαιο ολοκληρώνεται με τα αποτελέσματα αξιολόγησης της πρότυπης αυτής εφαρμογής.

Στο Κεφάλαιο 6 “*Συμπεράσματα της Διατριβής*” ανακεφαλαιώνονται τα γενικότερα συμπεράσματα από την εκπόνηση της Διατριβής και επισημαίνονται τα ανοιχτά θέματα που προκύπτουν.

Στο Κεφάλαιο 7 “*Βιβλιογραφικές αναφορές*” περιλαμβάνεται πλήρης βιβλιογραφική αναφορά των δημοσιεύσεων που χρησιμοποιήθηκαν στην παρούσα Διατριβή.

### **1.5 Συνεισφορά στην ερευνητική περιοχή - Δημοσιεύσεις**

Από τη συνολική θεώρηση της Διατριβής, συνάγεται ότι η συμβολή στην ανάπτυξη του γνωστικού πεδίου των ψηφιακών βιβλιοθηκών εντοπίζεται στα εξής σημεία:

- Στη σύνθεση μιας κατακεμημένης αρχιτεκτονικής ψηφιακής βιβλιοθήκης βασισμένης στην XML προσβάσιμης από το Web. Η κύρια συνεισφορά στο αντίστοιχο ερευνητικό πεδίο έγκειται στα χρήσιμα συμπεράσματα που προέκυψαν από τη χρήση του νέου προτύπου της XML και των παρεμμερών τεχνολογιών σε συστήματα ψηφιακών βιβλιοθηκών στο Web.
- Στο σχεδιασμό μιας ψηφιακής βιβλιοθήκης τριών στρωμάτων για τη διαχείριση αρχείων ήχου κωδικοποιημένων σύμφωνα με το πρότυπο mp3. Η βιβλιοθήκη αυτή αξιοποιεί τα μεταδομένα που βρίσκονται ενθυλακωμένα μέσα στο κυρίως σώμα του αρχείου mp3. Η αρχιτεκτονική της βιβλιοθήκης ευνοεί την κατοχύρωση των πνευματικών δικαιωμάτων (copyright) των συντελεστών των ηχητικών αυτών δεδομένων.
- Στη χρήση του προτύπου της XML για τη διαχείριση μεταδομένων σε αρχιτεκτονικές κατακεμημένων ψηφιακών βιβλιοθηκών.
- Στον τομέα της ασφάλειας των ψηφιακών βιβλιοθηκών, περιγράφεται ένας μηχανισμός βασισμένος στην CORBA, ο οποίος διασφαλίζει τις βασικές απαιτήσεις σε θέματα ασφάλειας για μια κατακεμημένη ψηφιακή βιβλιοθήκη ιατρικών δεδομένων προσβάσιμης από το Web.

- Στον τομέα της ανάκτησης πληροφοριών εισάγεται η έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης, σύμφωνα με την οποία παρουσιάζεται ένας μηχανισμός αλληλεπίδρασης των χρηστών με τα αποτελέσματα, όπως αυτά επιστρέφονται από μια μηχανή αναζήτησης, γεγονός που διευκολύνει τη γρήγορη εξαγωγή ασφαλών συμπερασμάτων αναφορικά με τη χρησιμότητα των αποτελεσμάτων αναζήτησης από τους χρήστες.
- Στον τομέα της ανάκτησης πληροφοριών περιγράφεται μια μέθοδος επεξεργασίας των αποτελεσμάτων όπως αυτά παρέχονται από μια μηχανή αναζήτησης σύμφωνα με την οποία η αντιστοιχία εφαρμογή που υλοποιεί τη μέθοδο αυτή, εκτελείται στην πλευρά του πελάτη μέσω του αντιστοιχίου Web browser, έτσι ώστε να μην επιβαρύνεται το δίκτυο από άσκοπη ανταλλαγή δεδομένων μεταξύ Web εξυπηρετητή και Web Browser. Επίσης, η προσέγγιση αυτή συμβάλλει στην ταχύτερη εξυπηρέτηση των χρηστών, σε συνάρτηση με την υπολογιστική ισχύ του εκάστοτε σταθμού εργασίας του κάθε χρήστη.

Στα πλαίσια των παραπάνω συνεισφορών στον ευρύτερο χώρο των ψηφιακών βιβλιοθηκών, δημοσιεύτηκαν ερευνητικές εργασίες τόσο σε διεθνή επιστημονικά συνέδρια με έκδοση πρακτικών μετά από κρίση πλήρους κειμένου, όσο και σε έγχρωμα διεθνή επιστημονικά περιοδικά:

- Ι. Παπαδάκης, Β. Χρυσικόπουλος, Δ. Πολέμη, "Ασφάλεια στις Ιατρικές Ψηφιακές Βιβλιοθήκες", Πρακτικά 2ου Πανελληνίου Συνεδρίου Βιοιατρικής Τεχνολογίας, σελ. 153-157, Νοέμβριος 1999
- Ι. Papadakis, V. Chrissikopoulos, D. Polemi, "A Secure Web-based Medical Digital Library Architecture Based on TTPs", Proceedings of the 16th International Conference of Medical Infobahn in Europe - XVI MIE 2000, pp. 610-616, IOS Press, August 2000
- Ι. Papadakis, V. Chrissikopoulos, "A Digital Library Framework based on XML", Proceedings of the 3rd International Conference of Asian Digital Library - ICADL '00, pp. 81-88, Seoul, December 2000
- J. Papadakis, C. Douligenis, "Design and Architecture of a Digital Music Library", Proceedings of the 3rd International Conference on Information Integration and Web-based Applications & Services - IWAS 2001, pp. 445-447, Austria, September 2001
- Ι. Papadakis, V. Chrissikopoulos, D. Polemi, "Secure Medical Digital Libraries", International Journal of Medical Informatics, Vol. 64, No. 2-3, pp. 417-428, Elsevier Science, December 2001

- I. Papadakis, C. Douligeris, "Design and Architecture of a Digital Music Library on the Web", "New Review on Hypermedia and Multimedia -NRHM", Vol. 7, pp. 115-126, Taylor-Graham, 2001
- I. Papadakis, I. Andreou, V. Chrissikopoulos, "Interactive Search Results", Proceedings of 6th European Conference on Digital Libraries - ECDL 2002, pp. 448-462, Elsevier Science, Rome, September 2002
- I. Papadakis, V. Karakoidas, V. Chrissikopoulos, "DocMI: A Digital Library of University Data", Library Hi-tech, Vol. 20, No.3, pp. 378 – 387, Emerald Publications, 2002

Πανεπιστήμιο Πειραιώς

## 1.6 Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Βασίλειο Χρυσικόπουλο για την εμπιστοσύνη που έδειξε με την επιλογή του προσώπου μου ως υποψήφιο Διδάκτορα υπό την εποπτεία του, αλλά και για την καθοδήγησή του κατά τη διάρκεια εκπόνησης της Διδακτορικής μου Διατριβής. Επίσης, θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Νικόλαο Αλεξανδρή για την άριστη συνεργασία που είχαμε στην περίοδο όπου ανέλαβε ως επιβλέπων Καθηγητής μου.

Ιδιαίτερες ευχαριστίες οφείλονται και στον Αν. Καθ. κ. Χρήστο Δουληγέρη τόσο για τις πολύτιμες συμβουλές του κατά τη διάρκεια της ερευνητικής μου δράσης, όσο και για την έμπρακτη συμπαράσταση που μου πρόσφερε σε περιόδους όπου την είχα απόλυτη ανάγκη.

Θα ήθελα ακόμα να ευχαριστήσω τα υπόλοιπα μέλη της Συμβουλευτικής και της επταμελούς Εξεταστικής Επιτροπής: κ. Θεμιστοκλή Παναγιωτόπουλο (Αν. Καθ. Πανεπιστημίου Πειραιώς), κ. Παναγιώτη Τσανίκι (Καθ. Ε.Μ.Π), κ. Γεώργιο Μπίλιο (Αν. Καθ. Πανεπιστημίου Ιονίου) και κ. Μαρτίνο Δρακινόη (Αν. Καθ. Πανεπιστημίου Αθηνών) για τα εποικοδομητικά τους σχόλια κατά την τελική φάση μελέτης και παρουσίασης της παρούσας Διατριβής.

Επίσης, επιθυμώ να ευχαριστήσω τον Καθηγητή κ. Γεώργιο Βασιλακόπουλο για τις ιδιαίτερα ενδιαφέρουσες και εποικοδομητικές συζητήσεις που κάναμε στα πρώτα χρόνια της Διδακτορικής μου Διατριβής, αλλά και για τη γενικότερη συμπαράστασή του κατά τη διάρκεια της παρουσίας μου στο Πανεπιστήμιο Πειραιώς.

Παράλληλα, θα ήθελα να εκφράσω την ευγνωμοσύνη μου στους συναδέλφους κκ. Παναγιώτη Κοτζανικόλαου και Εμμανουήλ Μάγιο για το γεγονός ότι φρόντισαν ώστε να μη νιώσω ποτέ μόνος μου κατά τη διάρκεια της θητείας μου ως υποψήφιος Διδάκτορας.

Οφείλω ακόμα να ευχαριστήσω τους κκ. Ιωάννη Ανδρέου, Βασίλειο Καρακίδη και την Δρα. Δέσποινα Πολέμη για την ομαλή συνεργασία που είχαμε σε διάφορες ερευνητικές δραστηριότητες.

Τέλος, ευχαριστώ τους υποψήφιους Διδάκτορες του Τμήματος Πληροφορικής του Πανεπιστημίου Πειραιά Μαρία Μουντριδου, Βικτώρια Τσιρίγκα, Κατερίνα Καμπάση, Νικόλαο Αβραντινή, Μικέλα Πουλμενάκου και Ιωάννη Καλιγκάτσι και τους φίλους μου Γεώργιο Κατσιρέλο, Μαρία Κουτσούγερα, Ρόζα Μαυροπόδη και Σπύρο Βοσινάκη για τη βοήθεια που μου πρόσφεραν όποτε τους τη ζητούσα.



## 2. Ψηφιακές Βιβλιοθήκες βασισμένες στο Web και στην XML

### 2.1 Εισαγωγή

Ο τομέας των ψηφιακών βιβλιοθηκών αναπτύχθηκε για να παρέχει τρόπους οργάνωσης και αποδοτικής διαχείρισης μεγάλων συλλογών από έγγραφα διαφόρων ειδών. Σε αυτό το πλαίσιο, παρουσιάζονται δυο μοντέλα ψηφιακών βιβλιοθηκών βασισμένων στο Web για τη διαχείριση Πανεπιστημιακού υλικού. Τα αποτελέσματα που προέκυψαν από την έρευνα αυτή έρχονται να καλύψουν το κενό που υπάρχει στον τομέα της αποδοτικής διαχείρισης μεγάλου όγκου δεδομένων που διακινούνται στα Πανεπιστήμια. Πολύ συχνά, οι φοιτητές παραδίδουν ένα πλήθος εργασιών σε διάφορα μαθήματα κατά τη διάρκεια των σπουδών τους. Τα παραδοτέα κυμαίνονται από ένα απλό αρχείο κειμένου μέχρι πιο πολύπλοκες συνθέσεις, όπως ένα σύνολο αρχείων που αναπαριστούν ένα πολύπλοκο τρισδιάστατο περιβάλλον μοντελοποιημένο σύμφωνα με το πρότυπο της VRML (Virtual Reality Markup Language). Συνήθως, οι εργασίες των φοιτητών αποθηκεύονται με φυσικό τρόπο στο γραφείο του εκάστοτε Καθηγητή ή, σύμφωνα με το πιο αισιόδοξο σενάριο, σε ηλεκτρονική μορφή στον προσωπικό δικτυακό τόπο του Καθηγητή, για μικρό πάντως χρονικό διάστημα.

Αν και οι φοιτητές θα ωφελούνταν σε μεγάλο βαθμό αν είχαν πρόσβαση σε παλαιότερες εργασίες συναφούς περιεχομένου, κάτι τέτοιο είναι πρακτικά αδύνατο, λόγω της απουσίας ενός συστήματος που να είναι ικανό να αποθηκεύει, να ευρετηριάζει και να ανακτά τέτοιου είδους δεδομένα. Η δυνατότητα των ψηφιακών βιβλιοθηκών να διαχειρίζονται δεδομένα μεγάλης

πολυπλοκότητας και όγκου καθιστά την τεχνολογία αυτή ικανή να ανταποκριθεί στις απαιτήσεις που προαναφέρθηκαν.

Λαμβάνοντας υπόψη τις παραπάνω σχέσεις, στο κεφάλαιο αυτό παρουσιάζεται α) ένα πλαίσιο (framework) στήριξης αρχιτεκτονικών ψηφιακών βιβλιοθηκών βασισμένων στο Web και στο πρότυπο της XML (eXtensible Markup Language) και β) η DocML, μια ψηφιακή βιβλιοθήκη Πανεπιστημιακού υλικού, που βασίζεται σε μεγάλο ποσοστό στις αρχές του προαναφερθέντος πλαισίου.

Συγκεκριμένα, όπως θα φανεί στις επόμενες ενότητες, και τα δυο μοντέλα αναφέρονται σε μια αρχιτεκτονική τριών στρωμάτων ικανή να διαχειριστεί XML δεδομένα μέσω του Web. Το πρότυπο της XML χρησιμοποιείται για να περιγράψει τα έγγραφα που αποτελούν τον κατακεντημένο αποθηκευτικό χώρο (back-end) των αρχιτεκτονικών αυτών. Αν και τα έγγραφα αυτά έχουν όλο και λιγότερα κοινά στοιχεία καθώς γίνεται η μετάβαση από Τμήμα σε Τμήμα, εντούτοις στους χρήστες παρουσιάζεται μια ομοιογενής και καλά οργανωμένη αποθηκευτική δομή. Αυτό οφείλεται στη χρήση της XML, η οποία ευνώνει τη διαχείριση εγγράφων διαφόρων τύπων μορφοποίησης και περιεχομένου. Και στις δυο προσεγγίσεις, η βασική λειτουργικότητα των συστημάτων παρέχεται στο ενδιάμεσο στρώμα (middleware). Τέλος, το στρώμα του πελάτη (client layer) φιλοξενεί έναν απλό Web browser ο οποίος επικοινωνεί με τον κεντρικό Web εξυπηρετητή (server) του ενδιάμεσου στρώματος.

Πέρα από τις ομοιότητες αυτές, τα δυο περιγραφόμενα μοντέλα διαφοροποιούνται στον τρόπο που υποστηρίζουν την ανάκτηση πληροφοριών στον αποθηκευτικό τους χώρο, καθώς στη μια περίπτωση ακολουθείται η κατακεντημένη προσέγγιση, ενώ στην άλλη η συγκεντρωτική. Περισσότερες λεπτομέρειες για αυτό το θέμα θα δοθούν κατά τη διάρκεια της παρουσίασης των δυο μοντέλων.

Το υπόλοιπο του κεφαλαίου έχει διαρθρωθεί ως εξής: Αρχικά, γίνεται αναφορά στα μεταδεδομένα, τα οποία αποτελούν βασικό στοιχείο για την ανάπτυξη ψηφιακών βιβλιοθηκών. Ο ρόλος της XML στις ψηφιακές βιβλιοθήκες περιγράφεται στην αμέσως επόμενη ενότητα, ακολουθούμενη από μια καταγραφή των απαιτήσεων για την ανάπτυξη ψηφιακών βιβλιοθηκών Πανεπιστημιακού υλικού. Κατόπιν, γίνεται μια ανασκόπηση της διεθνούς βιβλιογραφίας πάνω σε θέματα παρεμφερών ψηφιακών βιβλιοθηκών βασισμένων στο Web και στην XML. Στη συνέχεια, παρουσιάζεται ένα πλαίσιο στήριξης αρχιτεκτονικών ψηφιακών βιβλιοθηκών, το οποίο υποστηρίζει αρχιτεκτονικές τριών στρωμάτων βασισμένων στο Web και στην τεχνολογία της XML. Κατά τη διάρκεια της παρουσίασης αυτής περιγράφονται αναλυτικά τα τρία στρώματα

καθώς και τα βασικότερα λειτουργικά τμήματα που τα αποτελούν. Η ενότητα αυτή ολοκληρώνεται με μια ανάλυση του τύπου των εγγράφων που διακινούνται στις ψηφιακές βιβλιοθήκες που προκύπτουν. Βασισμένη στα συμπεράσματα της προηγούμενης ενότητας, η επόμενη ενότητα παρουσιάζει την DocML, μια ψηφιακή βιβλιοθήκη Πανεπιστημιακού υλικού βασισμένη στο Web και στην τεχνολογία της XML. Στις ενότητες που αναφέρονται στην DocML, γίνεται μια λεπτομερειακή περιγραφή των συστατικών της μερών, τα οποία έχουν ομοιότητες αλλά και μερικές διαφορές από το αρχικό πλαίσιο. Ιδιαίτερη έμφαση δίνεται στις υπηρεσίες που παρέχονται μέσω της χρήσης της αρχιτεκτονικής αυτής καθώς και στον τρόπο με τον οποίο μεταφέρονται τα δεδομένα μεταξύ των διαφόρων στρωμάτων της DocML.

## **2.2 Μεταδεδομένα**

Στα πλαίσια των άρτια δομημένων αποθηκών δεδομένων, τα έγγραφα που είναι διαθέσιμα μέσω ψηφιακών βιβλιοθηκών συνοδεύονται πολύ συχνά από μεταδεδομένα. Ένα έγγραφο μιας ψηφιακής βιβλιοθήκης βασισμένης στο Web μπορεί να κυμαίνεται από ένα απλό αρχείο κειμένου τύπου ASCII μέχρι μια πολύπλοκη δομή αρχείων και φακέλων αποτελούμενη από αρχεία διαφόρων τύπων. Είναι συνεπώς αναγκαίο για μια ψηφιακή βιβλιοθήκη αποτελούμενη από τέτοιου είδους έγγραφα, να χρησιμοποιεί μεταδεδομένα για την ανάπτυξη τεχνικών που διευκολύνουν την ανάκτηση πληροφοριών καθώς και τη διαχείριση των εγγράφων αυτών.

Ο όρος “μεταδεδομένα” αναφέρεται στις πληροφορίες των δεδομένων. Στο χώρο των ψηφιακών βιβλιοθηκών βασισμένων στο Web, τα μεταδεδομένα περιγράφουν πηγές στο Διαδίκτυο, απαντώντας στα ερωτήματα: τί είναι η πηγή αυτή, περί τίνοσ πρόκειται, πού βρίσκεται, σε τί αναφέρεται, κλπ. Υπάρχουν τρία βασικά θέματα που αφορούν την ανάπτυξη μεταδεδομένων: α) η περιγραφή των πηγών που αναφέρονται, β) η διαδικασία παραγωγής των μεταδεδομένων και τέλος γ) η χρήση τους [Iannella, 1997].

Το πρώτο θέμα αφορά το είδος της πληροφορίας που καταγράφεται. Αυτό εξαρτάται από τον τύπο των αναφερόμενων πηγών και το σκοπό των μεταδεδομένων. Το χρησιμοποιούμενο σχήμα μεταδεδομένων πρέπει να είναι αρκετά ευέλικτο ώστε να μπορεί να αποτυπώσει χρήσιμη πληροφορία για μια μεγάλη ποικιλία πηγών καθώς και για ένα πλήθος σκοπών. Στην ιδανική περίπτωση, πρέπει να υπάρχει μόνο ένα σχήμα μεταδεδομένων για όλους



τους τύπους πληροφοριών. Δυστυχώς, κάτι τέτοιο δεν είναι εφικτό, καθώς απαιτείται συναίνεση από όλους τους εμπλεκόμενους στη διαδικασία δημοσίευσης δεδομένων στο Διαδίκτυο. Είναι λοιπόν, ένα μεγάλο μέρος της έρευνας γύρω από τα μεταδοδεμένα αφορά τη συνεργασία μεταξύ διαφορετικών σχημάτων μεταδοδεμένων και την εύρεση του σημείου ισορροπίας ανάμεσα στη χρησιμότητα και στο μέγεθος ενός συνόλου από στοιχεία μεταδοδεμένων.

Το δεύτερο θέμα αφορά την παραγωγή των μεταδοδεμένων. Θα μπορούσε να υποστηριχθεί ότι τα μεταδοδεμένα αποτελούν μια περιλήψη των δεδομένων σε διάφορα επίπεδα απλότητας. Η εμπλοκή του ανθρώπινου παράγοντα στην παραγωγή αυτών των περιλήψεων κοστίζει τόσο σε χρόνο όσο και κόστος. Ως εκ τούτου, τα συστήματα μεταδοδεμένων προσπαθούν να μειώσουν αυτό το κόστος αυτοματοποιώντας τη διαδικασία παραγωγής τους, όπου βεβαίως κάτι τέτοιο είναι εφικτό.

Τέλος, το τρίτο θέμα έχει να κάνει με την πρόσβαση στα μεταδοδεμένα καθώς και τον τρόπο χρήσης τους. Τα μεταδοδεμένα πρέπει να ανακτώνται με τέτοιο τρόπο ώστε να διατηρείται η πληροφορία και το νόημα που περιλαμβάνουν. Μια σημαντική χρήση των μεταδοδεμένων είναι ως ένας μηχανισμός εντοπισμού πηγών σε κατακεμημένα περιβάλλοντα όπως το Διαδίκτυο. Οι χρήστες μπορούν να ανιχθούν από αυτά πληροφορίες που αφορούν τη χρησιμότητα των πηγών που αντιπροσωπεύουν για αυτούς.

### **2.3 XML και ψηφιακές βιβλιοθήκες**

Όπως έχει ήδη αναφερθεί, απαραίτητη προϋπόθεση για τη μακροχρόνια διατήρηση των ψηφιακών βιβλιοθηκών είναι η συμμόρφωση τους με τα διάφορα πρότυπα που επικρατούν. Το πρότυπο της XML [Cole, 2000] μπορεί να προσφέρει πολλά στο χώρο των ψηφιακών βιβλιοθηκών.

Σε αυτό το πλαίσιο, η δυνατότητα της XML να αναπαριστά το απλό κείμενο ως μια ιεραρχική δομή περιεχομένου από αντίστοιχα αντικείμενα και ταυτόχρονα να ενθυλωνώνει μεταδοδεμένα μέσα σε αυτά τα αντικείμενα, καθιστά την τεχνολογία αυτή ως μια πολλά υποσχόμενη λύση για το σχεδιασμό ψηφιακών βιβλιοθηκών βασισμένων στο Web. Από την άλλη μεριά, το γεγονός ότι η XML είναι απόγονος της SGML (Standard Generalized Markup Language) εγκυμονεί κινδύνους για την τεχνολογία αυτή. Η SGML υπάρχει εδώ και αρκετό καιρό και παρολαυτά δεν έχει κατορθώσει ακόμα να καθιερωθεί στο Web. Η βασική χιτίλα

αποτυχίας της SGML όσον αφορά τη διάδοσή της στο Web οφείλεται στη μεγάλη της πολυπλοκότητα όπως επίσης και στη μεγάλη καμπύλη εκμάθησης που τη χαρακτηρίζει. Έχοντας υπόψη τα παραπάνω, οι εμπνευστές της XML αφαίρεσαν μεγάλο μέρος της πολυπλοκότητας της SGML ενώ κράτησαν τα χαρακτηριστικά εκείνα που ήταν απαραίτητα στο Web [Pardi, 1999].

Ακόμα ένα επιχείρημα υπέρ της χρήσης της XML στο σχεδιασμό ψηφιακών βιβλιοθηκών βασισμένων στο Web είναι το γεγονός ότι η τεχνολογία αυτή επιτρέπει την ύπαρξη διαφορετικών όψεων των ίδιων δεδομένων, λόγω της ικανότητάς της να διαχωρίζει τα δεδομένα από την παρουσίασή τους [Liechi, 1998]. Κάτι τέτοιο είναι πολύ χρήσιμο για τις ψηφιακές βιβλιοθήκες όπου, πολύ συχνά, διαφορετικές ομάδες χρηστών απαιτούν διαφορετικές όψεις των ίδιων εγγράφων. Επίσης, τα συστήματα διεπαφής που προκύπτουν, μπορούν πιο εύκολα να συντηρηθούν όταν είναι ανεξαρτημένα από τα κυρίως δεδομένα.

Η διαλειτουργικότητα (interoperability) μεταξύ των διαφόρων συστατικών μιας ψηφιακής βιβλιοθήκης επιτυγχάνεται πολύ πιο εύκολα μέσω της χρήσης της XML καθώς και των συναφών τεχνολογιών που προκύπτουν από αυτή. Ορίζοντας μια κοινή μορφοποίηση (format) για τις χρησιμοποιούμενες δομές δεδομένων που να είναι αναγνώσιμη τόσο από τον άνθρωπο όσο και από τους υπολογιστές, περιορίζεται δραστικά η πολυπλοκότητα στην αλληλεπίδραση μεταξύ υπολογιστών [Aloisiò, 1999].

Συνοψίζοντας, προκύπτει ότι η χρήση της XML παρέχει ένα απλό και συνάμα άμεσο τρόπο ανάπτυξης και διαχείρισης εύρωστων και λειτουργικών συστημάτων ψηφιακών βιβλιοθηκών.

## ***2.4 Απαιτήσεις για την ανάπτυξη ψηφιακών βιβλιοθηκών Πανεπιστημιακού υλικού***

Για τη σωστή ανάπτυξη μιας ψηφιακής βιβλιοθήκης Πανεπιστημιακού υλικού βασισμένης στο Web, πρέπει να ληφθεί υπόψη ένα σύνολο από παράγοντες. Έτσι, όπως κάθε εφαρμογή που εμπλέκει τεχνολογίες Διαδικτύου, μια τέτοια προσπάθεια πρέπει να υιοθετήσει συγκεκριμένες θεμελιώδεις αρχές σχεδιασμού που ισχύουν για το περιβάλλον αυτό.

Η υλοποίηση λοιπόν πρέπει να βασίζεται σε πρωτόκολλα και εργαλεία υλοποίησης που είναι καθιερωμένα στην κοινωνία του Web έτσι ώστε να επιτευχθεί διαλειτουργικότητα μεταξύ των διαφόρων τμημάτων της ψηφιακής βιβλιοθήκης. Βασικές αρχές σχεδιασμού όπως η

απλότητα, η τμηματικότητα (modularity) και η επεκτασιμότητα αυξάνουν τις πιθανότητες προσαρμογής της ψηφιακής βιβλιοθήκης στις καινούργιες τεχνολογίες που εμφανίζονται με ολόένα και πιο γρήγορο ρυθμό τα τελευταία χρόνια στο Web. Τέλος, η αποκέντρωση (decentralization) είναι μια απάντηση που μόλις πρόσφατα έχει προκύψει στον τομέα των μεγάλης κλίμακας εφαρμογών βασισμένων στο Web. Για να μπορεί μια ψηφιακή βιβλιοθήκη να αναπτυχθεί χωρίς περιορισμούς ενώ ταυτόχρονα να αποφεύγονται λάθη και καταρροές ουστημάτων, η αρχιτεκτονική της πρέπει να στηρίζεται σε όσο το δυνατόν λιγότερα συγκεντρωτικά μητρώα (registries).

Περαιτέρω απαιτήσεις προκύπτουν από το γεγονός ότι η συγκεκριμένη ψηφιακή βιβλιοθήκη βασίζεται σε Πανεπιστημιακό υλικό. Ετσι, η δομή του αποθηκευτικού της χώρου πρέπει να σχεδιαστεί κατά τέτοιο τρόπο ώστε να διευκολύνει τη διαχείριση μεγάλου όγκου δεδομένων που ενδεχομένως να υπάρχουν σε διάφορες μορφολογίες (formats), με αποδοτικό τρόπο [Aloisio, 1999]. Επίσης, η ταυτόχρονη πρόσβαση στην ψηφιακή βιβλιοθήκη από μεγάλο μέρος της Πανεπιστημιακής κοινότητας κατά τη διάρκεια συγκεκριμένων χρονικών περιόδων (π.χ. πριν από εξετάσεις) επιβάλλει την εξασφάλιση της ικανότητας από την πλευρά της ψηφιακής βιβλιοθήκης να εξυπηρετεί πολλές αιτήσεις πρόσβασης που συμβαίνουν ταυτόχρονα σε μικρό χρονικό διάστημα.

Πέρα από τις παραπάνω απαιτήσεις που προκύπτουν από τα ιδιαίτερα χαρακτηριστικά της συγκεκριμένης ψηφιακής βιβλιοθήκης (δηλ. προσβάσιμη από το Web, ειδικευμένη σε Πανεπιστημιακό υλικό), υπάρχουν κάποιες παρατηρήσεις που ισχύουν για κάθε εφαρμογή ψηφιακών βιβλιοθηκών. Ετσι, η αλληλεπίδραση με τους χρήστες πρέπει να είναι εύκολη και πλήρης, ενώ η παρεχόμενη ανάκτηση πληροφοριών πρέπει να είναι ακριβής. Για το σκοπό αυτό, η χρήση μεταδεδομένων αποτελεί μια λύση που έχει δοκιμαστεί στο παρελθόν με επιτυχία στο χώρο των ψηφιακών βιβλιοθηκών (Baldonado, 1997, Iannella 1997). Συγκεκριμένα, σε ένα περιβάλλον ψηφιακών βιβλιοθηκών, τα μεταδεδομένα ορίζονται ως δεδομένα που χρησιμοποιούνται για να περιγράψουν τα κυρίως δεδομένα της αναφερόμενης συλλογής. Επίσης, το γεγονός ότι μια ψηφιακή βιβλιοθήκη δεν αποτελεί απλά μια στατική εφαρμογή, υποδηλώνει ότι το περιγραφόμενο σύστημα δε μπορεί να διατηρηθεί μακροπρόθεσμα χωρίς την απαιτούμενη συντήρηση αλλά και χωρίς τακτική αναβάθμιση. Το μεγάλο μέγεθος και η ποικιλία μορφών στις οποίες είναι αποθηκευμένα τα κυρίως δεδομένα, υπαγορεύουν ότι τέτοιοι στόχοι πρέπει να επιτυγχάνονται μέσω αυτοματοποιημένων διαδικασιών. Ετσι, ο διαχωρισμός του περιεχομένου από την εμφάνιση αποτελεί θεμελιώδη κανόνα στην ανάπτυξη ψηφιακών



βιβλιοθηκών (Aarts, 1995). Τέλος, η δυνατότητα άμεσης επεκτασιμότητας (scalability) των εμπλεκόμενων εφαρμογών αποτελεί μείζον θέμα για το σχεδιασμό τέτοιου είδους συστημάτων. Η λειτουργικότητα μιας ψηφιακής βιβλιοθήκης πρέπει να προσαρμόζεται στις ιδιαίτερες ανάγκες, απαιτήσεις αλλά και παροχές σε θέματα Πληροφορικής της κάθε εποχής.

## **2.5 Ανασκόπηση της Διεθνούς βιβλιογραφίας**

Η σημασία της τεχνολογίας των ψηφιακών βιβλιοθηκών έχει ήδη αναγνωριστεί από πολλούς οργανισμούς και κυβερνήσεις ανά τον κόσμο [Schäuble 1998]. Πρωτοπόροι σε αυτή την προσπάθεια ήταν οι ΗΠΑ οι οποίες το 1994 ανακοίνωσαν την πρώτη επιδοτούμενη πρόσκληση ενδιαφέροντος για προγράμματα ψηφιακών βιβλιοθηκών (Digital Library Initiative-DLI). Από τότε έχουν αναπτυχθεί διάφορα συστήματα ψηφιακών βιβλιοθηκών τόσο στις ΗΠΑ όσο και στον υπόλοιπο κόσμο. Σε αυτήν την ενότητα θα περιγραφούν συστήματα ψηφιακών βιβλιοθηκών που είναι βασισμένα στο Web.

### **2.5.1 Αρχιτεκτονική βασισμένη στο Web και στην XML για αρχεία τηλεπισκόπησης**

Ο Aloisio et al. [Aloisio, 1999] πρότειναν μια αρχιτεκτονική τριών στρωμάτων βασισμένη στην τεχνολογία CGI (Common Gateway Interface) για τη διαχείριση αρχείων τηλεπισκόπησης. Στη συγκεκριμένη αρχιτεκτονική ορίζονται το στρώμα αποθήκευσης δεδομένων (back-end), το ενδιάμεσο στρώμα λειτουργικότητας (middleware) και το στρώμα του πελάτη (client layer).

Στο στρώμα αποθήκευσης δεδομένων μπορεί να βρισκονται διάφορα εμπορικά συστήματα διαχείρισης δεδομένων τηλεπισκόπησης τα οποία αντιμετωπίζονται από την ψηφιακή βιβλιοθήκη ως “μαύρα κουτιά”, με την έννοια ότι καλούνται από το σύστημα να απαντήσουν σε διάφορες αιτήσεις επιστρέφοντας την αντίστοιχη πληροφορία χωρίς να είναι αναγκαία η γνώση του τρόπου λειτουργίας τους.

Το ενδιάμεσο στρώμα λειτουργικότητας παρέχει ένα πλούσιο και ταυτόχρονα επεκτάσιμο σύνολο από ολοκληρωμένες υπηρεσίες διαχείρισης της ψηφιακής βιβλιοθήκης διαμεσολάβησης για την ανταλλαγή πληροφοριών μεταξύ των συστατικών της μερών και

μετατροπής των δεδομένων της από μια μορφή σε κάποια άλλη ανάλογα με τις εκάστοτε απαιτήσεις. Ένας διαμεσολαβητής για παράδειγμα, δέχεται τις διάφορες αιτήσεις του συστήματος για δεδομένα και αποφασίζει σύμφωνα με κάποια κριτήρια να απευθύνει την αίτηση στους κατάλληλους εξυπηρετητές. Στην επιλογή αυτή λαμβάνεται υπόψη και η απόδοση των υποψήφιων εξυπηρετητών. Ένας μετατροπέας αναλαμβάνει τη μετατροπή πολύπλοκων εντολών προς κάποιο εμπορικό σύστημα που χρησιμοποιείται από την ψηφιακή βιβλιοθήκη σε μια σειρά από πιο απλές εντολές, τις οποίες μπορεί να “καταλάβει” το εμπορικό αυτό σύστημα, επεκτείνοντας με αυτόν τον τρόπο την προσαρμοστικότητα της ψηφιακής βιβλιοθήκης. Τέλος, τα αποτελέσματα από κάποια αναζήτηση σε μια βάση δεδομένων θα μπορούσαν να χρησιμοποιηθούν ως ερωτήματα για περαιτέρω ερωτήσεις στη βάση ή ακόμη και ως εντολές σε άλλες εφαρμογές του συστήματος, παρέχοντας με αυτόν τον τρόπο ένα ολοκληρωμένο περιβάλλον εργασίας.

Το στρώμα του πελάτη είναι υπεύθυνο για την αλληλεπίδραση με τους χρήστες του συστήματος καθώς και για την υποβολή ερωτημάτων στις υπηρεσίες που παρέχονται από το ενδιάμεσο στρώμα. Το στρώμα αυτό θα μπορούσε να αποτελείται μόνο από ένα συμβατικό Web browser, ενώ στην πιο πλούσια έκδοσή του θα μπορούσε να αναβαθμιστεί με τμήματα κώδικα που μεταποθηκεύονται (download) εκεί από το ενδιάμεσο στρώμα (applets). Ένα άλλο σενάριο υπαγορεύει την ανάγκη εγκατάστασης εξειδικευμένου λογισμικού στο στρώμα του πελάτη ανεξάρτητο από τον Web browser.

Συγκριτικά με παρόμοιες αρχιτεκτονικές βασισμένες στην τεχνολογία των CGI, οι συγγραφείς της συγκεκριμένης δουλειάς υποστηρίζουν ότι η ιδιότητα των παρεχόμενων υπηρεσιών να στηρίζονται στο πρότυπο της XML, διευκολύνει την επικοινωνία μεταξύ χρηστών και συστήματος, καθώς φυλάσσεται το ιστορικό των προηγούμενων καταστάσεων του εκάστοτε διαύλου επικοινωνίας με τους χρήστες. Από την άλλη μεριά, η προσέγγιση αυτή είναι ευέλικτη στα γενικότερα προβλήματα των τμημάτων κώδικα CGI που προκύπτουν από το γεγονός ότι δεν έχουν τρόπο να πληροφορηθούν την προηγούμενη κατάσταση από την οποία έχουν προσέλθει, οπότε και δυσχεραίνεται η ανάπτυξη ευέλικτων εφαρμογών, ικανών να αξιολογήσουν τη γενικότερη συμπεριφορά των χρηστών παρέχοντάς τους έτσι πιο ολοκληρωμένη εξυπηρέτηση.

## 2.5.2 Αρχιτεκτονική βασισμένη σε αντικείμενα δεδομένων και συνδέσμους

Σύμφωνα με τον Gurta et al. στο [Gurta, 1999], υπάρχουν τέσσερα βασικά στοιχεία που ξεχωρίζουν την εμπειρική πληροφορία από την επιστημονική πληροφορία.

- Το πρώτο στοιχείο αναφέρεται στο γεγονός ότι η επιστημονική πληροφορία είναι πολυσχιδής εκ φύσεως. Μαζί με τα συνηθισμένα, πλήρως δομημένα δεδομένα, η επιστημονική πληροφορία ενδέχεται να αποτελείται από έγγραφα, εικόνες, video, τρισδιάστατες απεικονίσεις, πειράματα προσομοίωσης και οποιοδήποτε συνδυασμό των προαναφερθέντων.
- Το δεύτερο χαρακτηριστικό της επιστημονικής πληροφορίας συνίσταται στην παρατήρηση ότι πολλές φορές, εκτός από δεδομένο-κεντρική, η πληροφορία μπορεί να χρειάζεται να αντιμετωπιστεί ως διεργασιο-κεντρική. Με άλλα λόγια, πολλές φορές δεν αρκεί μόνο η πρόσβαση στην πληροφορία. Εξίσου σημαντική είναι και η γνώση του πώς αναπαράγεται.
- Ένα τρίτο χαρακτηριστικό της επιστημονικής πληροφορίας είναι η ετερογένειά της. Αυτό συμβαίνει όταν ένα κομμάτι πληροφορίας σχετίζεται λογικά με πολλά και διαφορετικά από αυτό είδη πληροφορίας. Σε αυτήν την περίπτωση, μια τέτοια σύνθεση αποτελεί ένα είδος ετερογενούς πληροφορίας.
- Τέλος, οι χρήστες της επιστημονικής πληροφορίας απαιτούν ολοκληρωμένη πρόσβαση στην πληροφορία αυτή, που συνίσταται σε ανάκτηση, επεξεργασία και οπτικοποίηση (visualization) ενός ή περισσότερων συνόλων από δεδομένα.

Βασισμένο στις παραπάνω παρατηρήσεις, το πληροφοριακό μοντέλο που προτάθηκε από τον Gurta et al. [Gurta, 1999] βασίζεται σε αντικείμενα δεδομένων (data objects) και συνδέσμους (links). Τα αντικείμενα δεδομένων αποτελούν τον υποκείμενο αποθηκευτικό χώρο, ενώ μπορούν να συσχετισθούν με άλλα αντικείμενα δεδομένων μέσω στατικών συνδέσμων. Ο τρόπος πλοήγησης στο αντικείμενοστραφές αυτό σύστημα, σύμφωνα με τους συγγραφείς αυτής της εργασίας, ενοεί τη μοντελοποίηση επιστημονικής πληροφορίας, στην οποία και αναφέρεται.



### 2.5.3 Το σύστημα SOMLib

Σύμφωνα με την προσέγγιση που ακολουθείται στο σύστημα SOMLib [Raubert, 1999], παρέχεται πρόσβαση σε κατανομημένους αποθηκευτικούς χώρους σύμφωνα με τις προδιαγραφές ενός νευρωνικού δικτύου. Στο σύστημα βρίσκονται ενσωματωμένα τυπικά συστήματα διεπαφής βασισμένα στο Web, ενώ παρέχονται επίσης και "ευρύη υπολογιστικά εργαλεία" που συντελούν στην αύξηση της λειτουργικότητας ολόκληρου του συστήματος.

Οι ψηφιακές βιβλιοθήκες που προκύπτουν, βασίζονται σε ένα δημοφιλέστερο νευρωνικό δίκτυο χωρίς επίβλεψη (unsupervised), στο SOM (Self-Organizing Map) [Kohonen, 1995], το οποίο ενσωματώνει και συνδυάζει τοπικά αρχεία κειμένου αλλά και απομακρυσμένα συστήματα βιβλιοθηκών που χρησιμοποιούν συμβατές διανυσματικές (vector space) αναπαράστασεις. Σύμφωνα με την αρχιτεκτονική SOMLib, πολλές τοπικές ψηφιακές βιβλιοθήκες που βασίζονται στο SOM μπορούν να ολοκληρωθούν έτσι ώστε να δημιουργήσουν χάρτες βιβλιοθηκών. Η χρησιμοποίηση συγκεκριμένης διανυσματικής αναπαράστασης χαρακτηριστικών επιτρέπει το συνδυασμό διαφόρων αποθηκευτικών χώρων πληροφορίας χωρίς την ανάγκη δημιουργίας ειδικών εργαλείων μετατροπής για κάθε μια συλλογή. Αυτό επιτρέπει την τοπική συντήρηση των επί μέρους βιβλιοθηκών, η οποία απαιτεί σαφώς λιγότερη εκπαίδευση, λόγω του σχετικά μικρού όγκου δεδομένων. Οι χάρτες βιβλιοθηκών που βρίσκονται σε υψηλότερο επίπεδο χρησιμοποιούν τους ήδη εκπαιδευμένους τοπικούς χάρτες ως βάση για τα δεδομένα εκπαίδευσής τους, μειώνοντας έτσι το μέγεθος αλλά και το χρόνο εκπαίδευσης. Αυτή η διαδικασία συντελεί στη δημιουργία μιας "υπερ-βιβλιοθήκης". Πολλές υπερ-βιβλιοθήκες SOMLib μπορούν να συσχετιστούν (όχι απαραίτητα ιεραρχικά) δημιουργώντας έτσι ένα ιστό βιβλιοθηκών. Οι ψηφιακές βιβλιοθήκες που βασίζονται σε αυτήν την προσέγγιση, χαρακτηρίζονται από επεκτασιμότητα (scalability), από λειτουργικότητα βασισμένη στο πρόβλημα του εκάστοτε χρήστη και από ιδιαίτερα αποδοτικούς μηχανισμούς αναζήτησης.

### 2.5.4 THETIS: Μια ψηφιακή βιβλιοθήκη περιβαλλοντικών δεδομένων

Η αρχιτεκτονική που προτείνεται στο [Housis, 1997] είναι μια κατανομημένη αρχιτεκτονική που αποτελείται από τρία βασικά συστατικά στοιχεία: τη μηχανή αναζήτησης, τη μηχανή ανάκτησης πληροφοριών και ένα σύστημα διεπαφής βασισμένο στο Web που



ενσωματώνει τεχνολογία GIS (Geographic Information System). Είναι ένα καταναμημένο περιβάλλον που αποτελείται από έναν ή περισσότερους κόμβους-εξυπηρετητές (server nodes), κάθε ένας εκ των οποίων διαχειρίζεται έναν ή περισσότερους αποθηκευτικούς χώρους. Ένας κόμβος-εξυπηρετητής περιέχει μια μηχανή αναζήτησης και μια μηχανή ανάκτησης πληροφοριών. Αυτά τα συστατικά της ψηφιακής βιβλιοθήκης λειτουργούν αυτόνομα για τον εντοπισμό και την ολοκλήρωση πληροφοριών που είναι διαθέσιμες στον εκάστοτε κόμβο που αναφέρονται. Οι κόμβοι-εξυπηρετητές του συστήματος THETIS είναι όμοιοι, καθώς οι χρήστες λαμβάνουν την ίδια πληροφορία ανεξάρτητα από τον κόμβο από τον οποίο έστειλαν το αρχικό ερώτημα. Η ανακτούμενη πληροφορία παρουσιάζεται στους χρήστες μέσω ενός συστήματος διεπαφής, το οποίο παρέχει λειτουργικότητα GIS, επιτρέποντας με αυτόν τον τρόπο την οπτικοποίηση (visualization) των δεδομένων μέσω χαρτών.

Κύρια χαρακτηριστικά της αρχιτεκτονικής του συστήματος THETIS είναι η δυνατότητα πρόσβασης από ένα σημείο, η ολοκλήρωση της υποκείμενης πληροφορίας και η οπτικοποίηση των καταναμημένων γεωγραφικών δεδομένων στο Διαδίκτυο.

#### **2.5.5 FEDORA: Μια καταναμημένη αρχιτεκτονική ψηφιακών βιβλιοθηκών**

Η μεθοδολογία που ακολουθείται στο σύστημα FEDORA [Payette, 1998] βασίζεται στις τεχνολογίες CORBA (Common Object Request Broker Architecture) και Java για την κατασκευή μιας καταναμημένης υποδομής για ψηφιακές βιβλιοθήκες. Η βασική απαίτηση που επιδιώκει να ικανοποιήσει αυτό το σύστημα είναι η εξασφάλιση ενός αξιόπιστου και ασφαλούς τρόπου αποθήκευσης αλλά και πρόσβασης σε ψηφιακή πληροφορία. Επίσης, έχει δοθεί ιδιαίτερη βαρύτητα στην υποστήριξη κατάλληλων μηχανισμών που θα παρέχουν επεκτασιμότητα και διαλειτουργικότητα μεταξύ των διαφόρων τύπων δεδομένων καθώς και των συστημάτων διαχείρισης πνευματικών δικαιωμάτων που εμπλέκονται στην ψηφιακή βιβλιοθήκη FEDORA. Συνολικά, οι προκλήσεις που καλείται να αντιμετωπίσει αυτό το σύστημα είναι οι εξής:

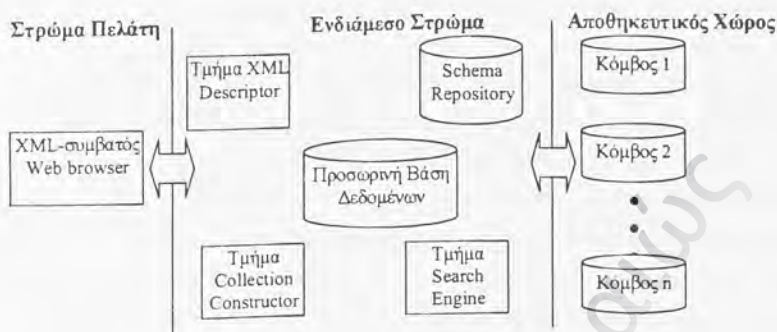
- Η δημιουργία πολύπλοκων αντικειμένων που περιλαμβάνουν και συνδυάζουν ετερογενείς τύπους δεδομένων, πιθανότατα καταναμημένους σε διάφορες τοποθεσίες.

- Η ικανότητα κληροδότησης κατά βούληση σε αυτά τα αντικείμενα μοντέλων συμπεριφοράς που χαρακτηρίζουν διάφορες συλλήψεις περιεχομένου γενικού ή/και ειδικού τύπου (όπως βιβλίο, περιοδικό, άρθρο, ταινία κ.ά.) ανεξάρτητα από την υποκείμενη δομή των αντικειμένων αυτών.
- Η πρόσβαση των χρηστών σε πολλαπλές θεωρήσεις (views) ή κοινοποιήσεις (disseminations) των δεδομένων κάθε αντικειμένου μέσω της διαρκούς ενεργοποίησης εξωτερικών μηχανισμών που εκτελούν αυτές τις συμπεριφορές τύπου περιεχομένου.
- Η συσχέτιση των εξωτερικών συστημάτων διαχείρισης πνευματικών δικαιωμάτων με οποιαδήποτε κοινοποίηση (dissemination) του περιεχομένου κάθε αντικειμένου.
- Η επεκτασιμότητα τύπων δεδομένων και συστημάτων διαχείρισης πνευματικών δικαιωμάτων.

## **2.6 Πλαίσιο υποστήριξης ψηφιακών βιβλιοθηκών βασισμένων στο Web**

Το περιγραφόμενο μοντέλο στο κεφάλαιο αυτό της Διατριβής αποτελεί ένα αφαιρετικό πλαίσιο υποστήριξης αρχιτεκτονικών τριών στρωμάτων βασισμένων στο Web και στην τεχνολογία της XML. Πρόκειται για ένα σύστημα που σε αντιδιαστολή με τα υπόλοιπα συστήματα ψηφιακών βιβλιοθηκών που έχουν αναφερθεί μέχρι τώρα βασίζεται εξ' ολοκλήρου στις παραπάνω τεχνολογίες. Για την καλύτερη κατανόηση του πλαισίου αυτού, η παρουσίασή του θα γίνει μέσω της αρχιτεκτονικής μας ψηφιακής βιβλιοθήκης Πανεπιστημιακού υλικού.

Ο σχεδιασμός της ψηφιακής βιβλιοθήκης είναι βασισμένος στην οργανωτική δομή ενός Πανεπιστημίου. Έτσι, ο αποθηκευτικός της χώρος αποτελείται από ένα πλήθος κόμβων που αντιστοιχούν στα Πανεπιστημιακά Τμήματα. Κάθε Τμήμα διατηρεί ένα σύνολο από Πανεπιστημιακά έγγραφα. Αν και μεταβαίνοντας από Τμήμα σε Τμήμα, τέτοιου είδους έγγραφα απαντώντας έχουν κοινά στοιχεία ως προς το περιεχόμενο και τη δομή τους, οι χρήστες του συστήματος έρχονται σε επαφή με μια ομοιογενή και καλά οργανωμένη δομή. Αυτό οφείλεται στο γεγονός ότι η χρήση της XML διευκολύνει τη διαχείριση ανομοιογενών εγγράφων.



Σχήμα 1: Γενική περιγραφή των τμημάτων της ψηφιακής βιβλιοθήκης

Όπως φαίνεται και στο σχ. 1, το ενδιάμεσο στρώμα αποτελείται από τρία τμήματα που υποστηρίζουν τη βασική λειτουργικότητα της ψηφιακής βιβλιοθήκης:

- Το τμήμα *XML Descriptor*
- Το τμήμα *Collection Constructor*
- Το τμήμα *Search Engine*

Το τμήμα *XML Descriptor* περιγράφει τα είδη εγγράφων που βρίσκονται αποθηκευμένα σε κάθε κόμβο του αποθηκευτικού χώρου. Η βασική λειτουργία του τμήματος αυτού είναι να παρέχει μια συνοπτική περιγραφή του περιεχομένου του κάθε κόμβου στα υπόλοιπα τμήματα της ψηφιακής βιβλιοθήκης. Για παράδειγμα, το *XML Descriptor* τμήμα ενός κόμβου που αναφέρεται στο Τμήμα Ιστορίας ενός Πανεπιστημίου, ενημερώνει την ψηφιακή βιβλιοθήκη ότι ο εν λόγω κόμβος περιέχει έγγραφα με ιστορικό περιεχόμενο.

Το τμήμα *Search Engine* αναλαμβάνει τον εντοπισμό εγγράφων που ικανοποιούν τα κριτήρια αναζήτησης των χρηστών. Αποτελεί τη μηχανή αναζήτησης της βιβλιοθήκης και βασίζεται στην τεχνολογία της XML για τη μορφοποίηση τόσο των ερωτημάτων των χρηστών όσο και των αποτελεσμάτων που επιστρέφονται σε αυτούς [W3C, 1998].

Το τμήμα *Collection Constructor* ταξιάζει τις περιγραφές των εγγράφων που βρίσκονται αποθηκευμένα στους διάφορους κόμβους, με τα κριτήρια αναζήτησης των χρηστών. Κατόπιν,

συγχωνώνει τις περιγραφές των εγγράφων μαζί με τους συνδέσμους πρόσβασης σε αυτά σε μια λίστα απαρίθμησης αποτελεσμάτων την οποία και προωθεί στο χρήστη. Το τμήμα αυτό αναλαμβάνει επίσης την αλληλεπίδραση με τους χρήστες και παρέχει αυξημένη λειτουργικότητα συγκριτικά με παραμετρικά συστήματα διεπαφών μηχανών αναζήτησης.

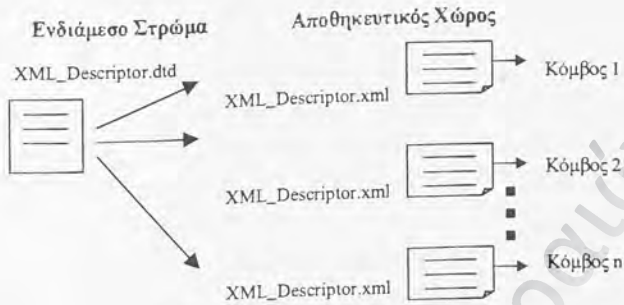
Εκτός από τα τρία αυτά βασικά τμήματα, το ενδιάμεσο στρώμα διατηρεί και κάποιες άλλες εφαρμογές, οι οποίες λειτουργούν επικουρικά στο σύστημα. Έτσι, για λόγους ταχύτερης εξυπηρέτησης των αιτήσεων των χρηστών, η ψηφιακή βιβλιοθήκη συντηρεί μια προσωρινή (cache) βάση δεδομένων στο ενδιάμεσο στρώμα. Σύμφωνα με τη λογική λειτουργίας του τμήματος αυτού, η προσωρινή βάση δεδομένων φιλοξενεί τα πιο δημοφιλή έγγραφα της ψηφιακής βιβλιοθήκης, έτσι ώστε να περιοριστούν, όσο αυτό είναι δυνατό, οι χρονοβόρες επισκέψεις στον αποθηκευτικό χώρο. Επίσης, στο ενδιάμεσο στρώμα συντηρείται και το *Schema Repository*, το οποίο αποτελεί μια κωθήκη για τα διάφορα σχήματα (schemas) που αντιστοιχούν σε πρότυπα για τις κατηγορίες εγγράφων από τις οποίες οι συγγραφείς της ψηφιακής βιβλιοθήκης μπορούν να επιλέξουν την κατηγορία εκείνη, η οποία περιγράφει καλύτερα τις απαιτήσεις τους. Τέλος, το στρώμα του πελάτη στην ουσία δεν είναι τίποτα άλλο παρά ένας τυπικός Web browser ικανός να διαχειριστεί αρχεία τύπου XML (π.χ. Internet Explorer 5). Τα κυριότερα συστατικά του περιγραφόμενου πλαισίου αναλύονται στις προσεχείς ενότητες.

### **2.6.1 Ενδιάμεσο στρώμα - το τμήμα XML Descriptor**

Το τμήμα *XML Descriptor* υποστηρίζει την ανάκτηση πληροφοριών από τους χρήστες της ψηφιακής βιβλιοθήκης. Όπως αναφέρεται στην εργασία του Jannella et al. [Ward, 1998], "Η ανάκτηση πληροφοριών αναφέρεται στον εντοπισμό, στην πρόαβαση, στην ανάκτηση και στη διαχείριση συναφών πληροφοριών για τους χρήστες". Σε αυτό το πλαίσιο, κάθε συγμύπο του τμήματος αυτού, περιγράφει στο τμήμα που έκανε την αίτηση τα περιεχόμενα του κόμβου στον οποίο ανήκει. Μια παρόμοια προσέγγιση ακολουθείται στο σύστημα NDLTD [Powel, 1998 - Fox, 1997], όπου εισάγεται η γλώσσα SearchDB-ML (Searchable Database Markup Language). Η δομή του τμήματος *XML Descriptor* ορίζεται στο αντίστοιχο XML σχήμα (το οποίο με ορολογία XML αντιστοιχεί σε ένα αρχείο .dtd - document type definition), το οποίο ονομάζεται *XML\_Descriptor.dtd*. Το αρχείο αυτό μετακινάται σε όλους τους κόμβους της ψηφιακής βιβλιοθήκης με τη μορφή αρχείων υπό το όνομα *XML\_Descriptor.xml*, όπως



φαίνεται στο σχ. 2. Τμήματα των αρχείων XML\_Descriptor.dtd και XML\_Descriptor.xml παρουσιάζονται στα σχ. 3α και 3β.



Σχήμα 2: Το τμήμα XML Descriptor

```
<!ELEMENT XML_Descriptor (University)>
<!ELEMENT University (Department)>
<!ATTLIST University name CDATA #REQUIRED>
<!ELEMENT Department (Lecture*, ThematicArea*)>
<!ATTLIST Department name CDATA #REQUIRED>
<!ELEMENT Lecture (#PCDATA)>
<!ELEMENT ThematicArea (#PCDATA)>
```

Σχήμα 3α: Το αρχείο XML\_Descriptor.dtd

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE XML_Descriptor SYSTEM "XML_Descriptor.dtd">
<XML_Descriptor> <!--Το συστατικό-ρίζα-->
<University name = "ΥπiPi" <!--Το Πανεπιστήμιο στο οποίο
ανήκει ο κόμβος-->
<Department name = "Informatics"> <!--Το τμήμα στο οποίο
αναφέρεται-->
<Lecture> Compilers </Lecture> <!--Τα μαθήματα -->
<Lecture> Computer Architecture </Lecture>
<ThematicArea> Computers </ThematicArea> <!--Οι θεματικές
περιοχές στις οποίες ανήκουν τα έγγραφα του κόμβου-->
</Department>
</University>
</XML_Descriptor>
```

Σχήμα 3β: Το αρχείο XML\_Descriptor.xml

Τα συστατικά μέρη του τμήματος XML *Descriptor* είναι στην ουσία μεταδεδομένα [Baldonado, 1997, Iannella, 1997]. Ειδικότερα, το συστατικό (element) “University” αναφέρεται στην επωνυμία του Πανεπιστημιακού ιδρύματος, ενώ περιέχει το υποσυστατικό (sub element) “Department”, που με τη σειρά του αναφέρεται στα αντίστοιχα Πανεπιστημιακά Τμήματα. Κάθε Τμήμα μπορεί να περιέχει από ένα έως περισσότερα μαθήματα (Lectures) και από μια έως περισσότερες επιστημονικές περιοχές (Thematic Areas) στις οποίες να ανήκει. Η ιεραρχική δομή που ορίζεται σε αυτό το αρχείο έχει ως συστατικό-ρίζα (root element) το συστατικό XML\_Descriptor.

Η ανάγκη χρησιμοποίησης μεταδεδομένων στις ψηφιακές βιβλιοθήκες και ειδικότερα σε ψηφιακές βιβλιοθήκες Πανεπιστημιακού υλικού, όπου υπάρχει μεγάλη ποικιλία μορφών και ειδών περιεχομένου, έχει αναγνωριστεί ευρέως από την επιστημονική κοινότητα. Το μεγάλο πλεονέκτημα των μεταδεδομένων έγκειται στο γεγονός ότι διευκολύνουν τη διαθεσιμότητα μεγάλου όγκου πληροφορίας τόσο στους χρήστες όσο και στις εμπλεκόμενες εφαρμογές [Liechā, 1998].

Η χρήση της τεχνολογίας XML παρέχει την υποδομή για την υιοθέτηση ενός απλού αλλά και εύχρηστου τρόπου εντοπισμού των κόμβων εκείνων που διαθέτουν χρήσιμες για τους χρήστες πληροφορίες. Μέσω της λειτουργικότητας που παρέχεται στο ενδιαμέσο στρώμα, οι χρήστες της ψηφιακής βιβλιοθήκης έχουν τη δυνατότητα να επιλέξουν τους κόμβους που θα συμμετέχουν στην αναζήτηση σύμφωνα με τα εξής κριτήρια:

- Επωνυμία Πανεπιστημιακού Τμήματος
- Περιεχόμενο επιστημονικής περιοχής
- Μαθήματα που διδάσκονται

Συγκεκριμένα, όπως θα φανεί στις επόμενες ενότητες, το τμήμα *Search Engine* ακριβώς τις τιμές των παραπάνω ετικετών (tags), έτσι ώστε να απορρίψει/επιλέξει τα αντίστοιχα έγγραφα που θα συμμετέχουν στη διαδικασία της αναζήτησης.

## 2.6.2 Ενδιάμεσο στρώμα - το τμήμα *Collection Constructor*

Μόλις κάποιος χρήστης συνδεθεί με στο σύστημα, το τμήμα XML *Descriptor* κάθε κόμβου καλείται να μεταφέρει τα XML δεδομένα που περιέχει στο τμήμα *Collection Constructor*. Όπως έχει ήδη αναφερθεί, το τμήμα αυτό βρίσκεται στο ενδιαμέσο στρώμα και λειτουργεί σε

αυτή την περίπτωση ως “μεσίτης” (broker). Κατόπιν, το τμήμα *Collection Constructor* συγχωνεύει όλα τα XML δεδομένα σε μια συγκεντρωτική XML δενδρική δομή, η οποία περιλαμβάνει τις περιγραφές κάθε κόμβου. Αυτό το μοντέλο λήψης (pull model) [Marschal, 1999] που χρησιμοποιείται για την κατασκευή του παραγόμενου XML δένδρου, προτιμάται από το αντίστοιχο μοντέλο προώθησης (push model), λόγω της συνέπειας (consistency) της συλλογής των αποθηκευμένων εγγράφων που εγγράφεται η χρήση του. Από την άλλη μεριά, αν κάποιος κόμβος της ψηφιακής βιβλιοθήκης δεν είναι διαθέσιμος κατά τη διάρκεια μιας δοσολήψης με κάποιον χρήστη, τα έγγραφα που είναι αποθηκευμένα σε αυτόν τον κόμβο δεν πρόκειται να ληφθούν υπόψη. Επίσης, ανάλογα με τον πληθυσμό των εγγράφων κάθε ενός κόμβου, το ακολουθούμενο μοντέλο θα μπορούσε να αποδειχτεί αρκετά χρονοβόρο. Παρολαυτά, υπάρχουν σαφείς ενδείξεις πως στο κοντινό μέλλον, θα είναι διαθέσιμο λογισμικό ικανό να διαχειριστεί αποδοτικά μεγάλο όγκο XML δεδομένων.

Σε δεύτερη φάση, ο χρήστης καλείται να αποφασίσει πώς θα κατηγοριοποιήσει τη συλλογή που του παρουσιάζεται. Η παρούσα αρχιτεκτονική παρέχει τρεις δυνατούς τρόπους κατηγοριοποίησης (όπως παρέχονται από το τμήμα *XML Descriptor*): α) μέσω μιας λίστας από τα μαθήματα που διδάσκονται στο Πανεπιστήμιο, β) μέσω μιας λίστας από τα Τμήματα του Πανεπιστημίου ή γ) μέσω μιας λίστας από τις προκαθορισμένες επιστημονικές περιοχές στις οποίες ανήκουν τα μαθήματα που διδάσκονται. Επιλέγοντας ένα ή περισσότερα κριτήρια κατηγοριοποίησης (π.χ. τα Τμήματα Ψυχολογίας και Κοινωνιολογίας της κατηγοριοποίησης ανά Τμήμα), στην ουσία περιορίζεται η αναζήτηση στους κόμβους εκείνους που σχετίζονται με τις τιμές των κριτηρίων αυτών. Μ' αυτόν τον τρόπο, η διαδικασία της αναζήτησης εκτελείται ταχύτερα, ενώ έχει και πιο ακριβή αποτελέσματα. Εναλλακτικά, οι χρήστες μπορούν να πλοηγηθούν στα έγγραφα που είναι αποθηκευμένα στους επιλεγμένους κόμβους της ψηφιακής βιβλιοθήκης.

Η δυνατότητα ύπαρξης πολλών θεωρήσεων (views) των ίδιων δεδομένων, προκύπτει από το γεγονός ότι κάθε XML αρχείο αντιστοιχεί σε ένα XML δένδρο, το οποίο μπορεί να παρουσιαστεί με διάφορους τρόπους, ανάλογα με το XSL (Extensible Stylesheet Language) αρχείο στο οποίο αντιστοιχίζεται [W3C, 1999].

Ακόμα ένα χαρακτηριστικό της προτεινόμενης ψηφιακής βιβλιοθήκης είναι η επεκτασιμότητά (scalability) της. Μόλις εντοπιστούν περαιτέρω επεκτάσεις που πρέπει να γίνουν, το σχήμα του τμήματος *XML Descriptor* μπορεί εύκολα να ανανεωθεί με τη βοήθεια



ενός απλού κειμενογράφου, καθώς η μορφοποίηση των δεδομένων του τμήματος αυτού είναι σε κωδικοποίηση ASCII.

### 2.6.3 Ενδιάμεσο στρώμα - το τμήμα *Search Engine*

Μετά την επίταξη σε ένα συγκεκριμένο υποσύνολο ολόκληρης της συλλογής εγγράφων της ψηφιακής βιβλιοθήκης, οι χρήστες μπορούν να απευθύνουν τα κριτήρια αναζήτησης που επιθυμούν στο τμήμα *Search Engine*, το οποίο αποτελεί τη μηχανή αναζήτησης της προτεινόμενης αρχιτεκτονικής. Το τμήμα αυτό κωδικοποιεί τα κριτήρια σε μορφή XML και κατασκευάζει το XML δένδρο που προκύπτει στους κόμβους που έχουν επιλεγεί να συμμετέχουν στην αναζήτηση, μέσω του πρώτου σταδίου της αναζήτησης. Οι δενδρικές δομές XML που αντιστοιχούν στα έγγραφα που ικανοποιούν τα κριτήρια που έχουν τεθεί, επιστρέφουν το τμήμα εκείνο του XML δένδρου τους που περιέχει κάποια μεταδεδομένα σχετικά με το εν λόγω έγγραφο (π.χ. κατηγορία, συγγραφείς, κλπ), με βασικότερα ειδικά τα μεταδεδομένα που έχουν να κάνουν με την περιγραφή του περιεχομένου του εγγράφου. Για παράδειγμα, στην περίπτωση που κάποιος χρήστης θέσει τη λέξη-κλειδί "Java" ως κριτήριο αναζήτησης, ένα εγχειρίδιο χρήσης της Java θα επιστρέφει στη μηχανή αναζήτησης το εννοιολογικό πλαίσιο στο οποίο υπάγεται η λέξη Java για το έγγραφο αυτό:

- Κατηγορία: Εγχειρίδιο χρήσης. Συγγραφείς: Γιώργος Παπαδόπουλος. Τίτλος: Ηλεκτρονικό βιβλιοπωλείο. Ημερομηνία δημοσίευσης: 19/02/2001. Γλώσσα υλοποίησης: Java

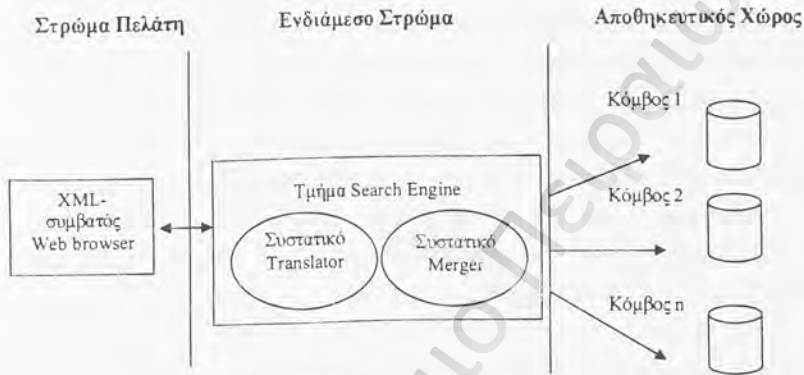
Σε μια άλλη περίπτωση ενός τεχνικού εγγράφου σχετικού με τεχνολογίες του Web (άρα και με τη γλώσσα Java), το αντίστοιχο XML υποδένδρο που θα επιστραφεί στη μηχανή αναζήτησης θα έχει τη μορφή:

- Κατηγορία: τεχνικό έγγραφο. Συγγραφείς: Γιώργος Μίνος. Τίτλος: Τεχνολογίες του Web. Ημερομηνία δημοσίευσης: 11/12/1999. Επικρατούσες τεχνολογίες: Java.

Όλες οι αναφορές των εγγράφων που ικανοποιούν τα κριτήρια αναζήτησης, συγχωνεύονται σε ένα συγκεντρωτικό XML δένδρο από το συστατικό *Merge*. Αυτό το συστατικό δέχεται διάφορα XML δένδρα από τους κόμβους του αποθηκευτικού χώρου και συνθέτει ένα ενιαίο δένδρο στο ενδιάμεσο στρώμα, το οποίο αντιστοιχεί στη ζητούμενη λίστα με τα αποτελέσματα της αναζήτησης. Το τελικό δένδρο αποστέλλεται στον χρήστη που ξεκίνησε τη διαδικασία



αναζήτησης. Χρησιμοποιώντας κατάλληλο λογισμικό διαχείρισης XML δεδομένων, το XML δένδρο μετατρέπεται σε αρχείο τύπου XML, το οποίο και παρουσιάζεται στο χρήστη. Κάθε σύνδεσμος που υπάρχει σε αυτό το αρχείο οδηγεί (χρησιμοποιώντας τεχνολογία XPath [W3C, 1999]) στα αντίστοιχα έγγραφα που βρίσκονται αποθηκευμένα στον αποθηκευτικό χώρο. Το σενάριο αυτό παρουσιάζεται στο σχ. 4. Δεν υπάρχει ανάγκη επέκτασης της αναζήτησης πέρα των XML αρχείων, καθώς κάθε αρχείο σχετίζεται άμεσα ή έμμεσα με κάποιο XML αρχείο.



Σχήμα 4: Το τμήμα Search Engine

Η γλώσσα προγραμματισμού που αποτελεί την ιδανικότερη λύση για τη διαχείριση XML δεδομένων είναι η Java. Συνεπώς, η Java προτείνεται ως η γλώσσα με την οποία θα υλοποιηθούν οι διάφορες εφαρμογές που θα παρέχουν τη λειτουργικότητα της ψηφιακής βιβλιοθήκης. Η ανεξαρτησία υπολογιστικής πλατφόρμας που παρέχει η Java, όπως και η ειδίκευσή της σε Web εφαρμογές, καθιστούν τη γλώσσα αυτή ως την πιο κατάλληλη για την ανάπτυξη της περιγραφόμενης ψηφιακής βιβλιοθήκης.

#### 2.6.4 Αποθηκευτικός χώρος

Ενας Web εξυπηρετητής φιλοξενείται σε κάθε κόμβο της ψηφιακής βιβλιοθήκης. Κατοχυρώνει τον κόμβο στο ενδιάμεσο στρώμα χρησιμοποιώντας τη λειτουργικότητα του τμήματος XML Descriptor. Θεωρώντας την ψηφιακή βιβλιοθήκη ως τρεις γενικότερες κλάσεις

συστατικών (δεδομένα, μεταδεδομένα και διεργασίες [Nürnberg, 1995]), κάθε στιγμιότυπο της κάθε κλάσης πρέπει να προσδιορίζεται με συνέπεια μέσω μόνιμων ονομάτων και προσδιοριστών [McGrath, 1999]. Το γεγονός ότι κάθε κόμβος της συλλογής στηρίζεται σε έναν προσωιωμένο Web εξυπηρετητή, παρέχει το απαραίτητο εύρος ονομάτων (namespace) που εξασφαλίζει τη μοναδικότητα των ονομάτων για τα έγγραφα της ψηφιακής βιβλιοθήκης.

Επίσης, η χρήση της τεχνολογίας XML namespaces [W3C, 1999] παρέχει τις εγγυήσεις εκείνες που απαιτούνται για τη σωστή διαχείριση των ονομάτων των διαφόρων συστατικών της ψηφιακής βιβλιοθήκης που σχετίζονται με την XML (ονόματα συστατικών (elements), ονόματα χαρακτηριστικών (attributes), κλπ). Κάθε δομή XML σχετίζεται λογικά με ένα εύρος ονομάτων (namespace), το οποίο περιλαμβάνει τους ορισμούς των περιεχομένων συστατικών της δομής αυτής. Είναι βεβαίως αυτονόητο πως η ανάπτυξη κάποιου εύρους ονομάτων (namespace) απαιτεί τη συμμετοχή έμπειρων ατόμων μέσα από την επιστημονική κοινότητα, έτσι ώστε να μην είναι δυσχερής η χρήση τους από τους συγγραφείς των εγγράφων της ψηφιακής βιβλιοθήκης.

Όπως θα φανεί στις επόμενες ενότητες, το περιγραφόμενο σύστημα ψηφιακής βιβλιοθήκης υποστηρίζει έγγραφα διαφόρων τύπων δεδομένων. Παρολαυτά, το γεγονός ότι κάθε έγγραφο στον αποθηκευτικό χώρο σχετίζεται με κάποιο XML αρχείο, υποχρεώνει πως οι εφαρμογές του ενδιαμέσου στρώματος οφείλουν να γνωρίζουν μόνο την τοποθεσία των αρχείων αυτών. Συνεπώς, πρέπει να υιοθετηθεί μια κοινή δομή αρχείων και φακέλων σε κάθε κόμβο της ψηφιακής βιβλιοθήκης, έτσι ώστε να επιτευχθεί σωστή και διαφανής ανάκτηση πληροφοριών. Σε μια πιλοτική υλοποίηση του συστήματος αυτού, επιλέγεται μια δομή παρόμοια με αυτή που ακολουθείται και στο πρωτόκολλο dienst [Logozz, 1995]. Αν και βασίζεται σε μια μάλλον αρχή επίπεδη οργάνωση αρχείων και φακέλων, εντούτοις παραμένει απλή στη χρήση της και ανεξάρτητη από εμπορικά συστήματα διαχείρισης βάσεων δεδομένων.

## **2.6.5 Διακινούμενα έγγραφα**

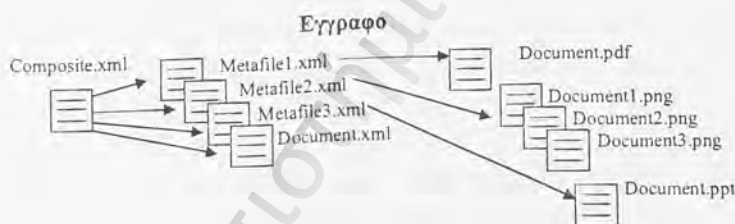
### **2.6.5.1 Δομή εγγράφων**

Όπως έχει προαναφερθεί, κάθε έγγραφο της συλλογής μπορεί να αποτελείται από ένα ή περισσότερα αρχεία διάφορων μορφοποιήσεων. Το πιθανότερο σενάριο υποχρεώνει ότι πολλές

από τις χρησιμοποιούμενες μορφοποιήσεις δεν περιγράφουν το περιεχόμενο του αρχείου στο οποίο αναφέρονται. Για αυτόν τον λόγο, όλα τα έγγραφα συνοδεύονται από ένα μετα-αρχείο (metafile.xml), το περιεχόμενο του οποίου δεν είναι στην ουσία τίποτα άλλο παρά μεταδεδομένα που παρέχουν μια περιγραφή του περιεχομένου του εν λόγω εγγράφου (π.χ. pdf, ppt, doc, κλπ).

Χρησιμοποιώντας τεχνολογία XPath [W3C, 1999], τα συστατικά αρχεία των εγγράφων καθίστανται προσπελάσιμα μέσω συνδέσμων που ανήκουν σε ένα συνοδευτικό αρχείο με το όνομα composite.xml. Η δομή του αρχείου αυτού παρουσιάζεται στο σχ. 5. Οι χρήστες έχουν τη δυνατότητα να πλοηγηθούν στην πληροφορία που περιλαμβάνει τα έγγραφα της συλλογής ξεκινώντας τη διαδικασία της πλοήγησης από τους συνδέσμους που βρίσκονται στο αρχείο composite.xml.

Η δομή που παρουσιάζεται στο σχ. 5 είναι κοινή για κάθε έγγραφο της συλλογής και ακολουθείται από κάθε κόμβο της ψηφιακής βιβλιοθήκης. Συγκεκριμένα, τα XML αρχεία υπόκεινται σε επεξεργασία από κατάλληλα εργαλεία (parsers) και τελικά παράγουν δενδρικές δομές όπου κάθε κόμβος τους αντιστοιχεί σε κάποιο συστατικό του αρχικού αρχείου.



Σχήμα 5: Η δομή ενός εγγράφου της ψηφιακής βιβλιοθήκης

Χρησιμοποιώντας κατάλληλο λογισμικό διαχείρισης XML δεδομένων μέσω τεχνολογίας βάσεων δεδομένων, οι δομές αυτές θα μπορούσαν να αποθηκευτούν ως έχουν (δηλ. ως δένδρα), παρέχοντας με αυτόν τον τρόπο πιο γρήγορη πρόσβαση. Το τμήμα *Search Engine* του ενδιάμεσου στρώματος χρειάζεται να ψάξει μόνο τα XML δεδομένα, επειδή η πληροφορία για την τοποθεσία αλλά και το περιεχόμενο των υπόλοιπων δεδομένων βρίσκεται ενθυλακωμένη στα XML δεδομένα. Τα δεδομένα αυτά μπορεί να βρίσκονται αποθηκευμένα οπουδήποτε μέσα στον κόμβο, αρκεί να είναι “ορατά” από τον Web εξυπηρετητή του κόμβου αυτού και να είναι προσβάσιμα μέσω κάποιας ηλεκτρονικής διεύθυνσης (URL). Σύμφωνα με την περιγραφόμενη

αρχιτεκτονική, κάθε κόμβος της ψηφιακής βιβλιοθήκης διατηρεί τα έγγραφα του αποθηκευμένα σε μια τοπική βάση δεδομένων.

Η χρήση της τεχνολογίας XPath δίνει μεγάλη ώθηση στον τομέα της αλληλεπίδρασης της ψηφιακής βιβλιοθήκης με τους χρήστες. Μερικά από τα χαρακτηριστικά που παρέχει η τεχνολογία αυτή στις εφαρμογές που την υιοθετούν είναι σύνδεσμοι που οδηγούν σε πολλαπλούς προορισμούς, κριτήριομη συνδεσμολογία, επαναληπόμενοι ρόλοι, περιγραφικοί τίτλοι κ.ά. Το μειονέκτημα αυτής της προσέγγισης είναι η έλλειψη των κατάλληλων εργαλείων στην αγορά που θα αυτοματοποιούσαν τη διαδικασία δημιουργίας και διαχείρισης των αλληλεπιδραστικών αρχείων. Παρολαυτά, πρέπει να ληφθεί υπόψη ότι η ευρεία αποδοχή της XML από όλες σχεδόν τις εταιρίες που ασχολούνται με την ανάπτυξη εφαρμογών στο Web, προεξοφλεί την εξάλειψη των προβλημάτων αυτών στο κοντινό μέλλον.

#### **2.6.5.2 Τύποι δεδομένων**

Η προτεινόμενη αρχιτεκτονική της ψηφιακής βιβλιοθήκης παρέχει τη δυνατότητα υποστήριξης οποιουδήποτε τύπου δεδομένων ικανού να προσπελαστεί από κάποια ηλεκτρονική διεύθυνση. Παρολαυτά, για να αξιοποιηθούν οι δυνατότητες που προσφέρει η τεχνολογία της XML, οι συγγραφείς προτείνονται να δημιουργήσουν τα έγγραφά τους σε μορφή XML. Για το σκοπό αυτό, διατίθεται ένα σύνολο από αρχεία σχημάτων (schemas-document definition files), τα οποία βρίσκονται αποθηκευμένα στο *Schema Repository* του ενδιαμέσου στρώματος. Κάθε ένα αρχείο-σχήμα ορίζει μια κατηγορία εγγράφων που προκύπτουν από αυτό, καθώς το χρησιμοποιούν ως πρότυπο (template). Για την περιγραφόμενη ψηφιακή βιβλιοθήκη Πανεπιστημιακού υλικού, έχουν εντοπιστεί οι ακόλουθες κατηγορίες εγγράφων:

- Εκπαιδευτικό υλικό (παρασιάσεις, case studies)
- Τεχνικά έγγραφα
- Διπλωματικές εργασίες (Μεταπτυχιακά Διπλώματα Ειδίκευσης, Διδακτορικές Διατριβές)
- Εγχειρίδια εργασιών



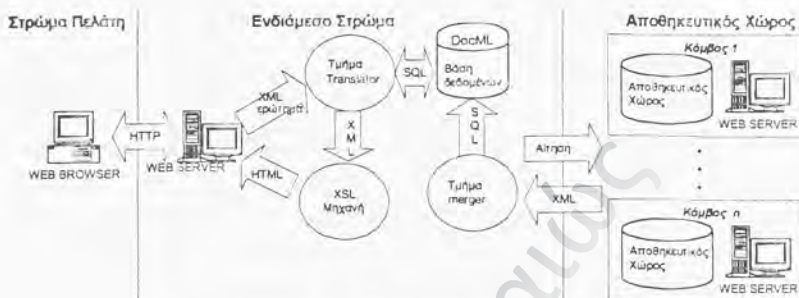
Για την ολοκληρωμένη υποστήριξη και το σωστό σχεδιασμό αυτών των σχημάτων, απαιτείται η εμπλοκή έμπειρων μελών της επιστημονικής κοινότητας στη διαδικασία ανάπτυξης τους. Μια ακόμα παρατήρηση σχετικά με τα δικαιουόμενα έγγραφα υπογορεύει ότι μια κατηγορία εγγράφων μπορεί να περιέχει παραπάνω από ένα σχήματα, αν αποφασιστεί ότι η ύπαρξη ενός γενικευμένου και ταυτόχρονα πολύπλοκου σχήματος ενδεχομένως να αποθαρρύνει τους υποψήφιους συγγραφείς από τη χρησιμοποίησή του. Τα XML έγγραφα που θα προκύψουν πρέπει να δημιουργούνται εύκολα. Η απλότητα συνεπώς είναι ένα χαρακτηριστικό που προβλέπεται να επηρεάσει σε μεγάλο βαθμό τη διάδοση της προτεινόμενης αρχιτεκτονικής. Για το σκοπό αυτό, προτείνεται ανεπιφύλακτα η χρήση εξειδικευμένων εργαλείων συγγραφής XML εγγράφων. Αν και η XML είναι μια σχετικά καινούργια τεχνολογία, υπάρχουν διαθέσιμα αρκετά εργαλεία συγγραφής XML, τα οποία ευνοούν τη δημιουργία XML εγγράφων από ανειδίκευτους χρήστες.

## 2.7 Η ψηφιακή βιβλιοθήκη DocML

Στην ενότητα αυτή περιγράφεται η αρχιτεκτονική και ο σχεδιασμός της DocML, μιας ψηφιακής βιβλιοθήκης Πανεπιστημιακού υλικού που βασίζεται πάνω στα συμπεράσματα και τις παρατηρήσεις που προέκυψαν από την προαναφερθείσα μελέτη του πλαισίου υποστήριξης αρχιτεκτονικών τριών στρωμάτων βασισμένων στο Web και στην τεχνολογία της XML.

Η DocML, όπως και η προηγούμενη προσέγγιση, βασίζεται σε μια αρχιτεκτονική τριών στρωμάτων ικανή να διαχειριστεί XML δεδομένα στο Web. Ο ρόλος της XML είναι να περιγράψει τα έγγραφα που συνθέτουν τον κατανεμημένο αποθηκευτικό χώρο της εν λόγω βιβλιοθήκης. Τα έγγραφα αυτά είναι στην ουσία οι εργασίες των φοιτητών. Με μικρές αλλαγές, η αρχιτεκτονική της DocML μπορεί να διαχειριστεί Πανεπιστημιακά έγγραφα διαφόρων ειδών.

Ο αποθηκευτικός χώρος αποτελείται από ένα πλήθος κόμβων που αντιστοιχούν στα Πανεπιστήμια-μέλη της ψηφιακής βιβλιοθήκης. Κάθε κόμβος συντηρεί μια αποθήκη ηλεκτρονικών εγγράφων που ανήκουν στο Πανεπιστήμιο. Ο σχεδιασμός του τοπικού αποθηκευτικού χώρου κάθε κόμβου βασίζεται στην οργανωτική δομή του Πανεπιστημίου. Όπως έχει αναφερθεί και στις προηγούμενες ενότητες, η χρήση της XML διευκολύνει τη διαχείριση των ποικιλόμορφων αρχείων και εγγράφων εντός ενός Πανεπιστημίου, τόσο από άποψη δομής όσο και περιεχομένου.



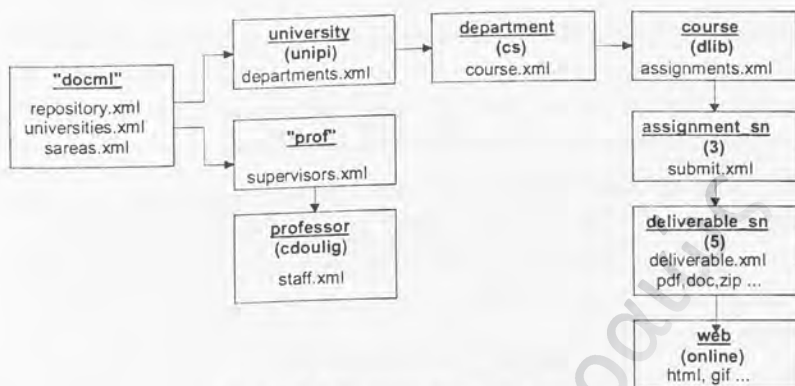
Σχήμα 6: Η αρχιτεκτονική της DocML.

Το ενδιάμεσο στρώμα υλοποιεί τη βασική λειτουργικότητα της προτεινόμενης αρχιτεκτονικής. Τα τμήματα *Translator* και *Merger* είναι υπεύθυνα για την επικοινωνία (μέσω της χρήσης ενός αφοσιωμένου (dedicated) Web εξυπηρετητή) του ενδιάμεσου στρώματος με το στρώμα του πελάτη και τον αποθηκευτικό χώρο αντίστοιχα. Συγκεκριμένα, το τμήμα *Translator* δέχεται και ικανοποιεί τις αιτήσεις των χρηστών, ενώ το τμήμα *Merger* συλλέγει τα XML δεδομένα που περιγράφουν τα διάφορα έγγραφα του αποθηκευτικού χώρου. Κατόπιν, τα δεδομένα αυτά (στην ουσία μεταδεδομένα) καταλήγουν στη βάση δεδομένων DocML.

Το στρώμα του πελάτη αποτελείται από έναν τυπικό Web browser, ο οποίος επικοινωνεί με τον κεντρικό Web εξυπηρετητή του ενδιάμεσου στρώματος. Μια γενική εικόνα της προτεινόμενης αρχιτεκτονικής παρουσιάζεται στο σχ. 6, παρέχοντας παράλληλα πληροφορίες για τα διάφορα τμήματα της ψηφιακής βιβλιοθήκης.

### 2.7.1 Αποθηκευτικός χώρος

Ο αποθηκευτικός χώρος αποτελείται από πολλούς κόμβους που αντιστοιχούν στα Πανεπιστήμια-μέλη της DocML. Όπως φαίνεται και στο σχ. 6, κάθε κόμβος συντηρεί έναν Web εξυπηρετητή και μια συλλογή εγγράφων, η οποία αποτυπώνεται σε ένα ιεραρχικό σύστημα αρχείων XML και άλλων μορφών. Η γενικότερη δομή της ιεραρχίας κάθε κόμβου παρουσιάζεται στο σχ. 7.



Σχήμα 7: Η δομή του αποθηκευτικού χώρου

Ο φάκελος-ρίζα της ιεραρχίας σε κάθε κόμβο της DocML έχει το όνομα "docml" και περιέχει τρία XML αρχεία: α) το αρχείο repository.xml, β) το αρχείο universities.xml και γ) το αρχείο sareas.xml. Το αρχείο repository.xml περιέχει γενικές πληροφορίες για το Πανεπιστήμιο. Ένα παράδειγμα τέτοιου αρχείου παρουσιάζεται στο σχ. 8.

```

<?xml version="1.0" encoding="UTF-8"?>
<repository>
  <id value="unipi"/>
  <url value="http://rainbow.cs.unipi.gr"/>
  <description value="University Of Piraeus"/>
  <comment>
    This is the university of piraeus
  </comment>
  <image src="logo.gif"/>
  <contact value="webmaster@unipi.gr"/>
  <address value="80 Karaoli & Dimitriou Str"/>
  <tel value="4142231"/>
</repository>
    
```

Σχήμα 8: Ένα αρχείο repository.xml

Τα υπόλοιπα δυο αρχεία, το αρχείο Universities.xml και sareas.xml, τα οποία μεταφέρονται αυτούσια από το ενδιαμέσο στρώμα όπου διατηρούνται, περιέχουν πληροφορίες σχετικά με τα

Πανεπιστήμια που συμμετέχουν στην DocML και τις επιστημονικές περιοχές αντίστοιχα. Ένα αρχείο sareas.xml παρουσιάζεται στο σχ. 9.

```
<?xml version="1.0" encoding="UTF-8"?>
<sareas>
  <area name = "Science" description = "Description"/>
  <area name= "Physics" description="Description"/>
  <area name= "Chemistry" description="Description"/>
  <area name="Informatics" description="Description"/>
</sareas>
```

Σχήμα 9: Ένα αρχείο sareas.xml

Ο φάκελος-ρίζα "docml" περιέχει δυο υποφάκελους, τον φάκελο "prof" και τον φάκελο "university". Ο πρώτος περιέχει το αρχείο staff.xml, το οποίο με τη σειρά του αποτελείται από πληροφορίες σχετικές με το προσωπικό του Πανεπιστημίου, ενώ το δεύτερο, το departments.xml περιγράφει τα διάφορα Τμήματα ενός Πανεπιστημίου. Ένα αρχείο departments.xml παρουσιάζεται στο σχ. 10.

```
<?xml version="1.0" encoding="UTF-8"?>
<departments>
  <department id="cs" name="Dept. of
  Informatics" description="Description"
  url="http://www.cs.unipi.gr/" />
  <department id="stat" name="Dept. of
  Statistics" description="Description"
  url="http://www.ode.unipi.gr/" />
</departments>
```

Σχήμα 10: Ένα αρχείο departments.xml

Το χαρακτηριστικό "id" στην ετικέτα (tag) "department" σχετίζεται με το όνομα του επόμενου φακέλου στην ιεραρχία. Έτσι, σύμφωνα με τη δομή XML που παρουσιάζεται στο σχ. 10, υπάρχουν δυο υποφάκελοι, ο υποφάκελος "cs" (δηλ. Τμήμα Πληροφορικής-Computer Science) και ο υποφάκελος "stat" (δηλ. Τμήμα Στατιστικής-Statistics). Κάθε φάκελος που αντιστοιχεί σε Τμήμα, περιέχει ένα αρχείο course.xml, το οποίο περιγράφει τα διάφορα μαθήματα που παραδίδονται σε κάθε Τμήμα. Ομοίως, κάθε Τμήμα περιέχει τις εργασίες (αντιστοιχούν στο αρχείο assignments.xml) που ανατίθενται στους φοιτητές, ενώ σε κάθε



εργασία αντιστοιχεί μια λίστα από τα παραδοτέα των φοιτητών (αντιστοιχούν στο αρχείο submit.xml) αναφορικά με την εκάστοτε εργασία. Τα έγγραφα που έχουν παραδώσει οι φοιτητές βρίσκονται αποθηκευμένα σε φακέλους με ονόματα που αντιστοιχούν σε αύξοντες αριθμούς. Με τον όρο "έγγραφο", ορίζεται το σύνολο των φακέλων ή/και αρχείων που μπορεί να παραδώσει ένας φοιτητής στα πλαίσια της εργασίας κάποιου μαθήματος. Τα έγγραφα περιέχουν τα πραγματικά δεδομένα της ψηφιακής βιβλιοθήκης. Κάθε έγγραφο της ψηφιακής βιβλιοθήκης μπορεί να περιέχει πολλά αρχεία διάφορων μορφών. Η πιθανότητα ύπαρξης αρχείων σε κωδικοποιήσεις που δεν περιγράφουν το περιεχόμενό τους, επιβάλλει την υιοθέτηση ενός επιπλέον αρχείου τύπου XML (το αρχείο deliverable.xml), το οποίο συνοδεύει κάθε έγγραφο. Το αρχείο αυτό αποτελείται από μεταδεδομένα που παρέχουν πληροφορίες σχετικές με το περιεχόμενο αλλά και αναφορές στα υπόλοιπα αρχεία που συνθέτουν το εν λόγω έγγραφο (π.χ. αρχεία τύπου pdf, ppt, doc, κλπ). Ένα αρχείο deliverable.xml παρουσιάζεται στο σχ. 11.

```
<?xml version="1.0" encoding="iso-8859-7"?>
<deliverables>
  <deliverable id="1">
    <university id="unipi"/>
    <department id="cs"/>
    <supervisor id="cdbulig"/>
    <course id="dlib"/>
    <assignment id="1"/>
    <abstract>a brief abstract</abstract>
    <comment>general comment</comment>
    <author date="15-03-2001" name="I.E" id="p95053"
    homepage="http://thalis.cs.unipi.gr/~jpap/"
    email="jpap@unipi.gr"/>
    <author date="15-03-2001" name="B.K" id="p97053"
    homepage="http://thalis.cs.unipi.gr/~jpap/"
    email="bkarak@unipi.gr"/>
    <installation os="Windows 95/98/2000/NT">
      <comment>Comments on
      installation</comment>
      <file name="mplah.doc"
      mime="application/doc"/>
    </installation>
    <web value="online/index.html"/>
  </deliverable>
</deliverables>
```

Σχήμα 11: Ένα αρχείο deliverable.xml

Σημειώνεται ότι το συστατικό "web" χρησιμοποιείται για να παρέχει την ηλεκτρονική διεύθυνση (URL) του διαθέσιμου εγγράφου.

Λαμβάνοντας υπόψη την κατανομημένη φύση του αποθηκευτικού χώρου, κάθε έγγραφο της ψηφιακής βιβλιοθήκης πρέπει να έχει μοναδικό όνομα (δηλ. προσδιοριστή). Έτσι, σύμφωνα με την προτεινόμενη προσέγγιση, το όνομα κάθε εγγράφου χωρίζεται στους ακόλουθους πέντε τομείς:

- Πανεπιστήμιο
- Τμήμα
- Μάθημα
- Σειριακός αριθμός εργασίας
- Σειριακός αριθμός παραδοτέου

Για παράδειγμα, ένα έγγραφο του Τμήματος Πληροφορικής του Πανεπιστημίου Πειραιά για το μάθημα των ψηφιακών βιβλιοθηκών, θα μπορούσε να έχει το εξής όνομα-προσδιοριστή: "unipi.cs.dlib.4.5".

### 2.7.2 Ενδιάμεσο στρώμα

Το ενδιάμεσο στρώμα ενυψλώνει τη βασική λειτουργικότητα της ψηφιακής βιβλιοθήκης. Η λειτουργικότητα αυτή παρέχεται μέσω της χρήσης των ακόλουθων συστατικών στοιχείων:

- Τμήμα *Merger*
- Τμήμα *Collector*
- Βάση δεδομένων DocML
- Μηχανή ανάλυσης δεδομένων XSL
- Αρχείο universities.xml
- Αρχείο sareas.xml

Τα διάφορα XML αρχεία που περιέχουν μεταδεδομένα για κάθε έγγραφο της ψηφιακής βιβλιοθήκης, συλλέγονται και αποθηκεύονται από το τμήμα *Merger* στη βάση δεδομένων DocML, όπως φαίνεται και από το σχ. 6. Το τμήμα *Merger* μπορεί να προγραμματιστεί να

επισκέπτεται κάθε κόμβο της ψηφιακής βιβλιοθήκης όταν η κίνηση στο σύστημα είναι μέτρια, μειώνοντας έτσι την πιθανότητα να επηρεαστεί αρνητικά η απόδοση του συστήματος.

Το τμήμα *Translator* υποστηρίζει την ανάκτηση πληροφοριών για τους χρήστες της ψηφιακής βιβλιοθήκης. Έτσι, όπως φαίνεται και στο σχ. 13, τα ερωτήματα των χρηστών λαμβάνονται από το στρώμα του πελάτη σε μορφή XML.

```
<?xml version="1.0" encoding="UTF-8"?>
<query>
  <essay>
    <tag name="author" value="Liakopoulos"/>
    <tag name="author" value="Koliopoulos"/>
    <tag name="installation" value="win"/>
  </essay>
</query>
```

Σχήμα 13: Το ερώτημα ενός χρήστη σε μορφή XML.

Το μήνυμα στο σχ. 13 περιγράφει ένα ερώτημα για κάποια παραδοτέα εργασία των φοιτητών "Λιακόπουλος" και "Κολιόπουλος". Το τμήμα *Translator* μετατρέπει τα ερωτήματα των χρηστών σε SQL και απευθύνει το τελικό μήνυμα στη βάση δεδομένων DocML. Τα αποτελέσματα που προκύπτουν ενθυλακώνονται σε αρχεία τύπου XML και συσχετίζονται με κατάλληλα αρχεία μορφοποίησης XSL. Μια μηχανή XSL μετατρέπει τα XML δεδομένα σε HTML σύμφωνα με τις εντολές που υπάρχουν στο XSL αρχείο. Τα HTML δεδομένα αποστέλλονται στον Web browser του χρήστη μέσω της χρήσης ενός αφροσιωμένου Web εξυπηρετητή.

Όπως έχει ήδη αναφερθεί, τα αρχεία *universities.xml* και *sareas.xml* περιέχουν πληροφορία σχετική με τα συμμετέχοντα Πανεπιστήμια και τις αναρχωρισμένες επιστημονικές περιοχές αντίστοιχα. Έτσι, το αρχείο *universities.xml* αποτελείται κυρίως από διαχειριστικές πληροφορίες για κάθε Πανεπιστήμιο, ενώ το αρχείο *sareas.xml* αποτελείται από μια λίστα με τις επιστημονικές περιοχές που μπορούν να συσχετιστούν με κάθε Πανεπιστημιακό Τμήμα που περιγράφεται στη δομή του αποθηκευτικού χώρου. Τέλος, η βάση δεδομένων DocML περιέχει τα μεταδεδομένα που βρίσκονται στον κατανομημένο αποθηκευτικό χώρο. Το περιεχόμενο της βάσης δεδομένων DocML ενημερώνεται ανά τακτά χρονικά διαστήματα για τις αλλαγές που έχουν πραγματοποιηθεί στα αντίστοιχα XML αρχεία του αποθηκευτικού χώρου, μέσω της λειτουργικότητας που παρέχεται από το τμήμα *Merge*.

### 2.7.3 Στρώμα Πελάτη

Η αλληλεπίδραση με τους χρήστες της ψηφιακής βιβλιοθήκης επιτυγχάνεται μέσω του Web. Έτσι, ο Web εξυπηρετητής στο ενδιαμέσο στρώμα δέχεται και προωθεί αιτήσεις που γίνονται από τους χρήστες προς τα υπόλοιπα τμήματα του ενδιαμέσου στρώματος. Οι χρήστες αλληλεπιδρούν με το σύστημα μέσω σταθμών εργασίας που συμπεριφέρονται ως “λεπτοί πελάτες” (thin clients). Οι σταθμοί αυτοί είναι εξοπλισμένοι με έναν τυπικό Web browser ικανό να διαχειριστεί XML δεδομένα. Σύμφωνα με την ακολουθούμενη στρατηγική, η DocML μπορεί να προσπελάσει από οποιοδήποτε υπολογιστή που είναι συνδεδεμένος στο Διαδίκτυο και διαθέτει έναν Web browser, ανεξάρτητα από το λειτουργικό σύστημα που τον διέπει.

### 2.7.4 Παρεχόμενες υπηρεσίες

#### 2.7.4.1 Σύστημα διεπαφής

Η ανομοιογένεια των υπολογιστικών συστημάτων από τα οποία μπορεί κάποιος χρήστης να προσπελάσει την DocML, υπαγορεύει ότι πρέπει να δοθεί ιδιαίτερη βεβαιότητα σε θέματα συμβατότητας που προκύπτουν από τη χρήση Web browsers που προέρχονται από διαφορετικούς κατασκευαστές. Έτσι, το σύστημα διεπαφής είναι υλοποιημένο σε γλώσσα σημείωσης (mark-up) HTML 3.2, η οποία υποστηρίζεται από όλους τους Web browsers της αγοράς. Το μεγαλύτερο μέρος της παρεχόμενης λειτουργικότητας του συστήματος διεπαφής είναι υλοποιημένο με την τεχνολογία Java servlets και την τεχνολογία JavaScript, οι οποίες επίσης υλοποιούν την ανάκτηση πληροφοριών της DocML, όπως θα περιγραφεί στις επόμενες ενότητες. Ο συνδυασμός χρήσης προγραμμάτων που εκτελούνται τόσο στη μεριά του εξυπηρετητή όσο και στη μεριά του πελάτη, επιτυγχάνει την καλύτερη δυνατή απόδοση για το σύστημα διεπαφής.



#### 2.7.4.2 Αναζήτηση

Η διαδικασία της αναζήτησης πραγματοποιείται μόλις κάποιος χρήστης θέσει ένα ερώτημα που ακολουθεί συγκεκριμένα κριτήρια όπως όνομα συγγραφέα, ημερομηνία εισαγωγής κ.α. Οι όροι του ερωτήματος συγκρίνονται με πεδία της βάσης δεδομένων DocML που περιγράφουν το περιεχόμενο των αντίστοιχων εγγράφων στα οποία αναφέρονται, όπως τα πεδία “επιστημονική περιοχή”, “συγγραφέας” και “αντικείμενο” (subject). Η ικανότητα να διατηρείται τέτοιου είδους πληροφορία στη βάση δεδομένων προκύπτει από το γεγονός ότι η δομή του αποθηκευτικού χώρου βασίζεται σε έγγραφα τύπου XML. Αυτού του είδους τα έγγραφα περιέχουν ενθυλακωμένα μεταδεδομένα που αναφέρονται στο περιεχόμενο του εγγράφου. Η πληροφορία αυτή μεταφέρεται αυτούσια στη βάση δεδομένων DocML, διευκολύνοντας έτσι την ανάκτηση όσο το δυνατόν πιο σχετικών με το ερώτημα του χρήστη εγγράφων.

Στη συνέχεια, το τμήμα *Translator* μετατρέπει τη λίστα με τα αποτελέσματα της αναζήτησης σε μορφή XML. Η λίστα αυτή μετατρέπεται στον εξυπηρετητή σε μορφή HTML χρησιμοποιώντας ένα κατάλληλο XSL stylesheet και τελικά, μεταφέρεται στον Web browser του χρήστη. Η τεχνολογία της XSL χρησιμοποιείται επίσης για την επέκταση της λειτουργικότητας αναφορικά με τη διαχείριση της λίστας των αποτελεσμάτων. Έτσι, οι χρήστες της ψηφιακής βιβλιοθήκης έχουν τη δυνατότητα να αλληλεπιδρούν με τα αποτελέσματα της αναζήτησης και να ορίσουν τον προσωπικό τους τρόπο παρουσίασης των αποτελεσμάτων αυτών, επιλέγοντας την ταξινόμησή ή/και το φιλτράρισμά τους σύμφωνα με τα παρεχόμενα κριτήρια. Για παράδειγμα, κάποιος χρήστης μπορεί να αποφασίσει να τοποθετήσει στην κορυφή της λίστας τα πιο πρόσφατα έγγραφα αντί των πιο σχετικών. Περισσότερες λεπτομέρειες για τη λειτουργικότητα που παρέχεται στη λίστα με τα αποτελέσματα, αναφέρονται στο πέμπτο κεφάλαιο, “Αλληλεπιδραστικά αποτελέσματα αναζήτησης”. Τέλος, κάθε αναφορά σε έγγραφο που παρουσιάζεται ως σύνδεσμος στην εξερχόμενη HTML ιστοσελίδα, μπορεί να οδηγήσει στα διάφορα συστατικά ενός εγγράφου που βρίσκονται στον αποθηκευτικό χώρο (χρησιμοποιώντας τεχνολογία XPath [W3C, 1999]), όπως φαίνεται στο σχ. 7.

### **2.7.4.3** Πλοήγηση

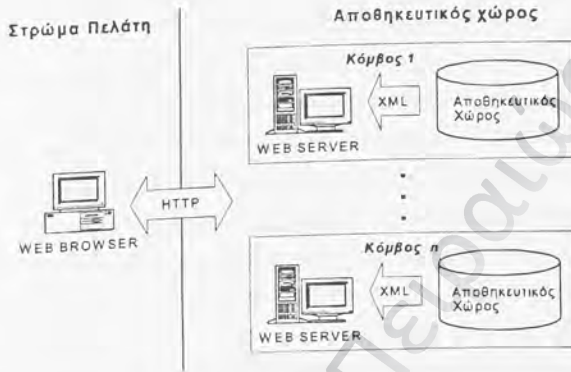
Οι χρήστες της DocML έχουν τη δυνατότητα να πλοηγηθούν στην υποκείμενη συλλογή εγγράφων με δυο τρόπους. Σύμφωνα με την πρώτη επιλογή, ακολουθώντας τους συνδέσμους που αντιστοιχούν στα διάφορα επίπεδα της ιεραρχίας του δομημένου αποθηκευτικού χώρου της ψηφιακής βιβλιοθήκης (βλ. σχ. 7). Για παράδειγμα, η πρόσβαση στα έγγραφα που αντιστοιχούν στο μάθημα “Μεταγλωττιστές” γίνεται μέσω της επίσκεψης των συνδέσμων “Πανεπιστήμιο Πειραιά”, “Τμήμα Πληροφορικής” και τελικά, του συνδέσμου “Μεταγλωττιστές”, ο οποίος οδηγεί σε μια ιστοσελίδα με συνδέσμους στα έγγραφα που αναφέρονται στο εν λόγω μάθημα.

Εναλλακτικά, οι χρήστες μπορούν να ταξινομήσουν την υποκείμενη συλλογή εγγράφων πριν την έναρξη της διαδικασίας της πλοήγησης, σύμφωνα με τις επιστημονικές περιοχές στις οποίες ανήκουν τα μαθήματα που διδάσκονται. Αυτή η “κατά περιεχόμενο” ταξινόμηση, υλοποιείται μέσω της χρήσης του αρχείου `sareas.xml`, το οποίο παρουσιάστηκε στο σχ. 9. Το αρχείο αυτό περιέχει τις επιστημονικές περιοχές που συνθέτουν τον επιστημονικό ορίζοντα ενός Πανεπιστημίου. Έτσι, τα Τμήματα εντός ενός Πανεπιστημίου σχετίζονται με μια ή περισσότερες επιστημονικές περιοχές. Επιλέγοντας μια συγκεκριμένη επιστημονική περιοχή, οι χρήστες στην ουσία εστιάζουν την ανάκτηση πληροφοριών σε ένα υποσύνολο του αρχικού συνόλου των εγγράφων του κατανεμημένου αποθηκευτικού χώρου. Μ’ αυτόν τον τρόπο, η διαδικασία της πλοήγησης γίνεται απλούστερη και πιο γρήγορη, ενώ ταυτόχρονα η διαδικασία της αναζήτησης καταλήγει σε πιο ακριβή αποτελέσματα. Η δυνατότητα κατηγοριοποίησης των εγγράφων με διαφορετικούς τρόπους από αυτόν που αντιστοιχεί στον τρόπο με τον οποίο έχουν αποθηκευτεί φυσικά στα διάφορα αποθηκευτικά μέσα, προκύπτει από το γεγονός ότι η χρησιμοποιούμενη XML επιτρέπει την ύπαρξη πολλαπλών όψεων (views) του ίδιου συνόλου δεδομένων. Συγκεκριμένα, κάθε XML αρχείο αντιστοιχεί σε μια δενδρική δομή, η οποία μπορεί να απεικονιστεί με διάφορους τρόπους, χρησιμοποιώντας κάθε φορά το κατάλληλο αρχείο XSL stylesheet.

### **2.7.4.4** Μεταφορά δεδομένων

Ανάλογα με τη φύση των αρχείων που αντιστοιχούν στα έγγραφα που παρουσιάζονται στη λίστα με τα αποτελέσματα από μια αναζήτηση ή πλοήγηση ενός χρήστη, υπάρχει η

δυνατότητα παρουσίασης των εγγράφων αυτών από την τοποθεσία αποθήκευσής τους ή της μεταποθήκευσής τους (download) τοπικά στον υπολογιστή του χρήστη για περαιτέρω επεξεργασία (βλ. σχ. 14).



Σχήμα 14: Μεταφορά δεδομένων από τον Πελάτη στον Αποθηκευτικό χώρο και αντίστροφα

Συγκεκριμένα, αρχεία σε μορφή XML, μετατρέπονται σε αρχεία μορφής HTML από αντίστοιχα αρχεία XSL stylesheets και αποθηκεύονται άμεσα με το που εισάγονται στην ψηφιακή βιβλιοθήκη. Με αυτόν τον τρόπο, αποφεύγεται η υπερβολική χρήση των πηγών (resources) του αποθηκευτικού χώρου. Τέτοιου είδους προβλήματα θα μπορούσαν να προκύψουν στην περίπτωση όπου τα XML αρχεία μετατρέπονταν σε αρχεία τύπου HTML τη στιγμή που γινόταν κάποια αίτηση από τους χρήστες (on the fly).

## 2.7.5 Διαχειριστικά θέματα

Μια Κεντρική Υπηρεσία - ΚΥ είναι υπεύθυνη για την εγγραφή καινούργιων μελών στην DocML. Αρχικά, το Πανεπιστήμιο που επιθυμεί να συμμετάσχει στην ψηφιακή βιβλιοθήκη στέλνει μια αίτηση εγγραφής στην ΚΥ. Μόλις εγκριθεί η αίτηση, η ΚΥ επιστρέφει μια εφαρμογή στο Πανεπιστήμιο, η οποία, μόλις εγκατασταθεί στον τοπικό εξυπηρετητή, αναλαμβάνει την εγκατάσταση του νέου κόμβου. Η ΚΥ κατοχυρώνει το νέο μέλος της DocML προσθέτοντας την αντίστοιχη εγγραφή στο αρχείο universities.xml.

Σε αυτό το σημείο, πρέπει να σημειωθεί ότι τα αρχεία `universities.xml` και `sareas.xml` διατηρούνται στο ενδιάμεσο στρώμα. Για λόγους βελτισποίησης της απόδοσης του όλου συστήματος, αντίγραφα των αρχείων αυτών υπάρχουν σε κάθε ένα από τους κόμβους του κατακευματισμένου αποθηκευτικού χώρου ξεχωριστά.

Το τμήμα *Merger* αναλαμβάνει τη συνέπεια (consistency) των δεδομένων που υπάρχουν στη βάση δεδομένων DocMI και στους κόμβους της ψηφιακής βιβλιοθήκης. Έτσι, το τμήμα αυτό ενημερώνει τη βάση δεδομένων DocMI για τις αλλαγές που γίνονται τοπικά σε κάθε Πανεπιστήμιο επικοινωνώντας περιοδικά με κάθε κόμβο και συλλέγοντας πληροφορίες σχετικές με καινούρια ή/και ενημερωμένα τοπικά έγγραφα. Έτσι, η διαδικασία της ενημέρωσης μπορεί να προγραμματιστεί να πραγματοποιείται σε περιόδους όπου η κίνηση στο δίκτυο της ψηφιακής βιβλιοθήκης είναι μειωμένη.

Τέλος, η συντήρηση του κάθε κόμβου πραγματοποιείται τοπικά από τον διαχειριστή του Πανεπιστημίου με τη βοήθεια αντίστοιχων εργαλείων που παρέχονται από την ΚΥ. Ο διαχειριστής μπορεί να προσθέσει, να αφαιρέσει και να ενημερώσει έγγραφα του τοπικού αποθηκευτικού χώρου.

## **2.8 Συμπεράσματα**

Το κεφάλαιο αυτό ασχολήθηκε με τις ψηφιακές βιβλιοθήκες βασισμένες στο Web και την XML. Η κύρια συνεισφορά του κεφαλαίου έγκειται στα χρήσιμα συμπεράσματα που προέκυψαν από τη χρήση του νέου προτύπου της XML και των παρεμφερών τεχνολογιών σε συστήματα ψηφιακών βιβλιοθηκών στο Web. Συγκεκριμένα, στο πρώτο μέρος του κεφαλαίου έγινε η παρουσίαση ενός πλαισίου στήριξης αρχιτεκτονικών αυτού του τύπου μέσα από μια ψηφιακή βιβλιοθήκη Πανεπιστημιακού υλικού. Το προτεινόμενο πλαίσιο στήριξης διακρίνεται για την απλότητα και ευκολία με την οποία υποστηρίζει βασικές απαιτήσεις των ψηφιακών βιβλιοθηκών όπως επεκτασιμότητα (scalability) και υποστήριξη ονοματολογίας (namespaces). Επίσης, η χρήση της XML διευκολύνει τη διαχείριση των μεταδεδομένων με απλότητα και ομοιογένεια, ενώ ταυτόχρονα παρέχει την κατάλληλη υποδομή για την ανταλλαγή μηνυμάτων μεταξύ των συστατικών των αντίστοιχων συστημάτων, βασισμένων σε μια κοινή μορφοποίηση. Το γεγονός ότι η XML ξεχωρίζει τα δεδομένα από την παρουσίασή τους, ευνοεί την ανάπτυξη ψηφιακών βιβλιοθηκών όπου οι χρήστες μπορούν να προσαρμόζουν την οπτικοποίηση



(visualization) των δεδομένων ανάλογα με τις προσωπικές τους προτιμήσεις. Ακόμα ένα χαρακτηριστικό του προτεινόμενου πλαισίου είναι το γεγονός ότι η ανάκτηση πληροφοριών πραγματοποιείται σε δυο στάδια. Στο πρώτο στάδιο, οι χρήστες εστιάζουν την αναζήτηση στους κόμβους της ψηφιακής βιβλιοθήκης όπου βρίσκονται τα πιο σχετικά προς τις ανάγκες τους έγγραφα. Στο δεύτερο στάδιο, πραγματοποιείται η αναζήτηση στους προεπιλεγμένους κόμβους, βελτιώνοντας με αυτόν τον τρόπο την απόδοση της ανάκτησης πληροφοριών. Επίσης, επιβαρύνεται σε μικρότερο βαθμό το δίκτυο από άσκοπες μεταφορές δεδομένων.

Έχοντας υπόψη τα συμπεράσματα που προέκυψαν από την πρώτη προσέγγιση, το υπόλοιπο του κεφαλαίου περιγράφει την αρχιτεκτονική της DocML, μιας ψηφιακής βιβλιοθήκης βασισμένης στο Web και στην XML για τη διαχείριση Πανεπιστημιακού υλικού. Η βασική διαφορά της DocML από την πρώτη προσέγγιση αναφέρεται στον τρόπο με τον οποίο πραγματοποιείται η ανάκτηση πληροφοριών. Είτσι, στην DocML, τα μεταδεδομένα των εγγράφων που βρίσκονται αποθηκευμένα τοπικά σε κάθε κόμβο, αντιγράφονται (replicate) σε μια κεντρική βάση δεδομένων, όπου και απευθύνονται τα ερωτήματα των χρηστών. Είτσι, αντίθετα με την πρώτη προσέγγιση, η αναζήτηση πραγματοποιείται αποκλειστικά και μόνο στο ενδιάμεσο στρώμα της αρχιτεκτονικής. Το βασικό πλεονέκτημα αυτής της προσέγγισης είναι η ταχύτητα εξυπηρέτησης των ερωτημάτων των χρηστών, καθώς δεν απαιτείται επικοινωνία με τον αποθηκευτικό χώρο. Από την άλλη μεριά, πρέπει να δοθεί ιδιαίτερη έμφαση στη διατήρηση της συνέπειας (consistency) των δεδομένων στον καταμετρημένο αποθηκευτικό χώρο και στη βάση δεδομένων, έτσι ώστε να αποφευχθούν τυχόν ανακρίβειες στις λίστες με τα αποτελέσματα αναζήτησης. Μια ακόμη διαφορά των δυο προσεγγίσεων έγκειται στο γεγονός ότι η πρώτη διαχειρίζεται συλλογές εγγράφων ενός Πανεπιστημίου, ενώ η δεύτερη διαχειρίζεται συλλογές εγγράφων πολλών Πανεπιστημίων, τα οποία και αντιστοιχούν στους κόμβους της ψηφιακής βιβλιοθήκης.

Τέλος, πρέπει να τονιστεί ότι η επιτυχία εφαρμογών αυτού του βεληκεούς, εξαρτάται σε μεγάλο βαθμό από τη συμμετοχή κατά τη φάση της ανάπτυξής της, των ομάδων εκείνων που θα επωφεληθούν από τη χρήση της. Πρέπει λοιπόν να δοθούν τα κατάλληλα κίνητρα στους φοιτητές, ερευνητές και καθηγητές έτσι ώστε να παρέχουν τα απαραίτητα σχήματα (schemas) που θα χρησιμοποιηθούν από την υπόλοιπη ακαδημαϊκή κοινότητα για τη δημιουργία επιστημονικών εγγράφων και όχι απλών ψηφιακών δεδομένων.

Πανεπιστήμιο Πειραιώς

### 3. Μουσικές Ψηφιακές Βιβλιοθήκες βασισμένες στο Web

#### 3.1 Εισαγωγή

Ο βασικός στόχος μιας ψηφιακής βιβλιοθήκης είναι να παρέχει τρόπους οργάνωσης και σωστής διαχείρισης μεγάλων συλλογών από ψηφιακά δεδομένα. Κατά τη σχεδίαση μιας μουσικής ψηφιακής βιβλιοθήκης βασισμένης στο Web, πρέπει να ληφθούν υπόψη τα ιδιαίτερα χαρακτηριστικά όχι μόνο του περιβάλλοντος αυτού, αλλά και του τρόπου αποθήκευσης και διάδοσης της μουσικής σε ένα περιβάλλον με τόσες ιδιαιτερότητες όπως είναι το Web. Επει, το γεγονός ότι δεν υπάρχει κάποιο κυρίαρχο λειτουργικό σύστημα στο Web, υποδηλώνει ότι, τουλάχιστον στην πλευρά του χρήστη, η προτεινόμενη προσπάθεια πρέπει να βασίζεται σε λογισμικό όπως ο Web browser, το οποίο είναι ανεξάρτητο πλατφόρμας. Είναι επίσης σημαντικό να ληφθεί υπόψη το γεγονός ότι το MPEG-1 layer 3 format (mp3) είναι ο καθιερωμένος τρόπος αποθήκευσης και διανομής μουσικής πληροφορίας στο Web.

Η δημοτικότητα των αρχείων mp3 ώθησε πολλές εταιρίες λογισμικού στη δημιουργία εφαρμογών οι οποίες να είναι ικανές να μετατρέψουν σχεδόν οποιοδήποτε τύπο αρχείων ήχου σύμφωνα με το πρότυπο mp3. Συνεπώς, είναι λογικό να θεωρηθούν τα αρχεία τύπου mp3 ως ο πυρήνας της προτεινόμενης μουσικής ψηφιακής βιβλιοθήκης. Οι εφαρμογές που διαπραγματεύονται τη διαχείριση ηχητικών δεδομένων στο Web, πρέπει επίσης να αντιμετωπίσουν δυο βασικά θέματα: τους περιορισμούς που επιβάλλει το δίκτυο και τα προβλήματα της πνευματικής ιδιοκτησίας (copyright). Οι περιορισμοί λόγω δικτύου πηγάζουν από το γεγονός ότι το Web βασίζεται στο πρωτόκολλο TCP/IP, το οποίο δεν είναι σχεδιασμένο για να διαχειρίζεται αποδοτικά δεδομένα πολυμέσων. Ευτυχώς, σύμφωνα με τις ενδείξεις που υπάρχουν στο χώρο των δικτυακών πρωτοκόλλων, δε θα αργήσουν να βρεθούν λύσεις σε αυτό το πρόβλημα [Lawton, 1999]. Το θέμα των πνευματικών δικαιωμάτων

αναφέρεται στο γεγονός ότι τα αρχεία μουσικού περιεχομένου δεν μπορούν να προσπελαστούν και να διαδοθούν στα πλαίσια μιας μουσικής ψηφιακής βιβλιοθήκης, χωρίς να αντιμετωπίσουν τους νομικούς περιορισμούς που αναφέρονται σε υλικό που υπόκειται στο καθεστώς της πνευματικής ιδιοκτησίας [Crews, 1997].

Με βάση τους παραπάνω προβληματισμούς, στο κεφάλαιο αυτό παρουσιάζεται ο σχεδιασμός και η αρχιτεκτονική μιας κατανεμημένης μουσικής ψηφιακής βιβλιοθήκης στο Web. Η προτεινόμενη ψηφιακή βιβλιοθήκη χρησιμοποιεί τεχνολογίες οι οποίες έχουν καθιερωθεί στο χώρο, όπως ο Web browser και ο Web εξυπηρετητής. Επίσης, η επικοινωνία των επί μέρους τμημάτων της, επιτυγχάνεται μέσω της χρήσης του προτύπου της XML. Η παρουσίαση γίνεται μέσα από την περιγραφή ενός ραδιοφωνικού σταθμού στο Web, ο οποίος βασίζεται πάνω στην προτεινόμενη ψηφιακή βιβλιοθήκη.

Το υπόλοιπο του κεφαλαίου έχει διαρθρωθεί ως εξής: Αρχικά, αναφέρονται οι απαιτήσεις μιας μουσικής ψηφιακής βιβλιοθήκης στο Web. Κατόπιν, περιγράφεται ο ρόλος των μεταδεδομένων στη μουσική πληροφορία και ειδικότερα στα αρχεία τύπου mp3. Στην επόμενη ενότητα γίνεται μια ανασκόπηση της διεθνούς βιβλιογραφίας πάνω σε θέματα παρεμφερών μουσικών ψηφιακών βιβλιοθηκών βασισμένων στο Web και στα αρχεία τύπου mp3. Στη συνέχεια, περιγράφεται αναλυτικά η προτεινόμενη κατανεμημένη αρχιτεκτονική διαχείρισης αρχείων τύπου mp3 στο Web. Η περιγραφή γίνεται μέσα από την παρουσίαση ενός ραδιοφωνικού σταθμού που επέμπει στο Web, ο οποίος χρησιμοποιεί την εν λόγω αρχιτεκτονική. Ιδιαίτερη έμφαση δίνεται στις υπηρεσίες που παρέχονται καθώς και στον τρόπο με τον οποίο μεταφέρονται τα δεδομένα (ή/και μεταδεδομένα) μεταξύ των διάφορων στρωμάτων της. Το κεφάλαιο ολοκληρώνεται με την παράθεση των συμπερασμάτων που προκύπτουν από την έρευνα πάνω στις μουσικές ψηφιακές βιβλιοθήκες στο Web.

### **3.2 Απαιτήσεις μιας μουσικής ψηφιακής βιβλιοθήκης**

Μια κατανεμημένη μουσική ψηφιακή βιβλιοθήκη πρέπει να σέβεται τα νόμιμα πνευματικά δικαιώματα των ανθρώπων που εμπιστεύονται τα έργα τους σε αυτή, καθώς περιέχει προϊόντα με κατοχυρωμένα πνευματικά δικαιώματα δημιουργού. Ένας τρόπος για να επιτευχθεί ο στόχος αυτός είναι μέσω της δημιουργίας εφαρμογών αναμετάδοσης (broadcasting) που στηρίζονται σε εξυπηρετητές συνεχούς ροής (streaming servers). Με αυτόν τον τρόπο, τα μουσικά αρχεία δεν είναι δυνατό να αντιγραφούν σε ανεπιθύμητους προορισμούς χωρίς έλεγχο. Αντί αυτού, ένα ή περισσότερα κανάλια ήχου (audio streams) αναμεταδίδονται στους



συνδεδεμένους σταθμούς εργασίας. Μια παρόμοια διαδικασία, η οποία μάλιστα ακολουθείται από τις ραδιοφωνικές εφαρμογές στο Web, είναι αποδεκτή από τη μουσική βιομηχανία καθώς κινείται στα πλαίσια λειτουργίας του παραδοσιακού ραδιοφώνου, για το οποίο υπάρχει ήδη κατοχυρωμένο ένα ισχυρό νομικό πλαίσιο. Επίσης, οι εφαρμογές αναμετάδοσης παρέχουν στους χρήστες τους μειωμένο έλεγχο των μουσικών δεδομένων, καθώς τα μουσικά αρχεία παραμένουν σε κάθε περίπτωση στην πλευρά του εξυπηρετητή.

Εκτός από τα προαναφερθέντα νομικά θέματα, πρέπει να ληφθούν υπόψη οι τεχνικές απαιτήσεις για την ανάπτυξη μιας μουσικής ψηφιακής βιβλιοθήκης. Η δομή του αποθηκευτικού της χώρου πρέπει να είναι ικανή να διαχειριστεί μεγάλο αριθμό αρχείων τύπου mp3 με συνεπή, δομημένο και αποτελεσματικό τρόπο. Αυτού του είδους τα αρχεία μπορεί να είναι κωδικοποιημένα σε διάφορους ρυθμούς δειγματοληψίας (bit gates). Επίσης, τυχόν επεκτάσεις στο λογισμικό, πρέπει να έχουν ελάχιστες επιπτώσεις στην επίδοση του συστήματος.

Για την παροχή αποτελεσματικών υπηρεσιών ανάκτησης πληροφοριών από την ψηφιακή βιβλιοθήκη, η εφαρμογή κατανεμημένων λύσεων στον τομέα αυτό φαίνεται να έχει περισσότερα πλεονεκτήματα σε σχέση με την εφαρμογή αντίστοιχων συγκεντρωτικών. Οι κατανεμημένες μηχανές αναζήτησης έχουν καλύτερες επιδόσεις από τις συγκεντρωτικές καθώς προσφέρουν τοπική αυτονομία. Επίσης, η ευρετηρίαση (indexing) του περιεχομένου της ψηφιακής βιβλιοθήκης καταναλώνει λιγότερο εύρος ζώνης δικτύου, καθώς η διαδικασία αυτή πραγματοποιείται ξεχωριστά σε κάθε κόμβο, χωρίς να απαιτείται ανταλλαγή δεδομένων με τους υπόλοιπους κόμβους [Wells, 1997].

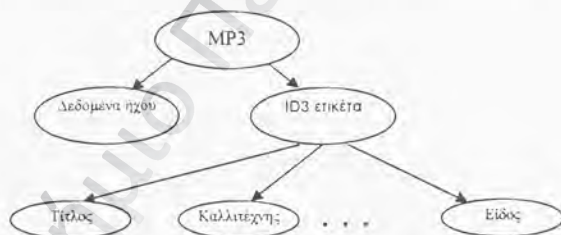
Μια ακόμα σημαντική απαίτηση μιας ψηφιακής βιβλιοθήκης που παρέχει η κατανεμημένη αποθήκευση και ανάκτηση είναι η ανεξαρτησία πλατφόρμας. Οι χρήστες πρέπει να έχουν τη δυνατότητα προσπέλασης της ψηφιακής βιβλιοθήκης μέσω δικτύου ανεξάρτητα από το λειτουργικό σύστημα που διέπει το σταθμό εργασίας τους. Ομοια, οι κατανεμημένοι αποθηκευτικοί χώροι πρέπει να είναι ικανοί να λειτουργούν κάτω από διαφορετικά λειτουργικά συστήματα.

Η θεώρηση του Web ως ένα παγκόσμιο λειτουργικό σύστημα και η χρησιμοποίηση γλωσσών προγραμματισμού όπως η Java (η οποία είναι ανεξάρτητη λειτουργικού συστήματος) και προτύπων όπως αυτό της XML για τη διάδοση της πληροφορίας μεταξύ των διάφορων αρχιτεκτονικών τμημάτων της ψηφιακής βιβλιοθήκης, προσφέρουν μια λύση που ικανοποιεί όλες τις παραπάνω απαιτήσεις [Roland, 2000].

### 3.3 Μεταδεδομένα στη μουσική πληροφορία

Ενα από τα χαρακτηριστικά που διακρίνει το περιβάλλον των ψηφιακών βιβλιοθηκών στο Web από τον ευρύτερο χώρο του Web, είναι η ικανότητά τους να αποθηκεύουν τα δεδομένα με συνέπεια σε οργανωμένες δομές [Lyon, 1997].

Στο πλαίσιο των καλώς δομημένων αποθηκών δεδομένων, τα έγγραφα μιας ψηφιακής βιβλιοθήκης συνοδεύονται συχνά από μεταδεδομένα. Όπως έχει ήδη αναφερθεί, τα μεταδεδομένα ορίζονται ως δεδομένα τα οποία χρησιμοποιούνται για να περιγράψουν τέτοια έγγραφα. Στην περίπτωση μιας μουσικής ψηφιακής βιβλιοθήκης, τα έγγραφα αντιστοιχούν σε μουσικά αρχεία κωδικοποιημένα σύμφωνα με το πρότυπο mp3. Τα αρχεία αυτά χαρακτηρίζονται από το γεγονός ότι τα μεταδεδομένα είναι ενσωματωμένα στο σώμα (body) κάθε αρχείου.



Σχήμα 1: Δομή ενός αρχείου τύπου mp3 με ετικέτα ID3

Σύμφωνα με τις προδιαγραφές του προτύπου mp3, μεταδεδομένα όπως “καλλιτέχνης”, “είδος” και “συλλογή” υπάρχουν στην ετικέτα ID3 καθενός από τα αρχεία τύπου mp3. Συγκεκριμένα, το ID3 είναι ένα μπλοκ δεδομένων ενσωματωμένο στην ηχητική ψηφιακή πληροφορία. Κάθε ετικέτα ID3 περιέχει ένα ή περισσότερα μπλοκ πληροφορίας τα οποία ονομάζονται πλαίσια (frames). Αυτά τα πλαίσια περιέχουν διάφορες πληροφορίες για το αντίστοιχο τραγούδι όπως “τίτλος”, “συλλογή”, “εκτελεστής”, “ιστοσελίδα”, “στίχοι” και άλλα. Το σχ. 1 παρουσιάζει τη δομή ενός αρχείου τύπου mp3. Εχοντας υπόψη τα παραπάνω, προκύπτει πως μια μουσική ψηφιακή βιβλιοθήκη που αποτελείται από αρχεία τύπου mp3, οφείλει να εκμεταλλευτεί τις ετικέτες ID3, έτσι ώστε να παρέχει αποτελεσματικές μεθόδους ανάκτησης πληροφοριών.

Υπάρχουσες μουσικές ψηφιακές βιβλιοθήκες στο Web βασίζονται μόνο στο όνομα αρχείων τύπου mp3 για την εξόχωση συμπερασμάτων όσον αφορά το περιεχόμενο των αρχείων αυτών. Συγκεκριμένα, έχει καθιερωθεί να ονομάζονται τα αρχεία τύπου mp3 σύμφωνα με το

όνομα του καλλιτέχνη ακολουθούμενο από τον τίτλο του τραγουδιού. Κατόπιν, τα διάφορα προγράμματα διαχείρισης των αρχείων αυτών επεξεργάζονται τα ονόματα έτσι ώστε να δημιουργήσουν αντίστοιχα ευρετήρια. Ωστόσο, μια τέτοια προσέγγιση έχει ως αποτέλεσμα τη δημιουργία ευρετηρίων αμφιβόλου ακρίβειας, καθώς δεν υπάρχει αυστηρά καθορισμένος τρόπος ονοματολογίας των αρχείων τύπου mp3, ενώ επίσης υπάρχει ελάχιστος χώρος για την αποθήκευση τέτοιου είδους πληροφορίας στο όνομα αυτών των αρχείων.

Από την άλλη πλευρά, επιβάλλεται να υπάρχει μια έγκυρη ετικέτα ID3 σε κάθε ένα από τα αρχεία της συλλογής, όταν η ανάκτηση μεταδεδομένων από τα αρχεία τύπου mp3 βασίζεται στις ετικέτες αυτές. Σε αντίθετη περίπτωση, ολόκληρο το σύστημα ανάκτησης πληροφοριών θα είναι ανακριβές. Για το λόγο αυτό, βασική απαίτηση μιας τέτοιας εφαρμογής είναι η διασφάλιση της ακεραιότητας των δεδομένων και των μεταδεδομένων της μουσικής ψηφιακής βιβλιοθήκης.

### 3.4 Ανασκόπηση της Διεθνούς βιβλιογραφίας

#### 3.4.1 Η μουσική ψηφιακή βιβλιοθήκη Jukebox

Ο στόχος του συστήματος Jukebox [Fonss-Jorgensen, 1997] είναι η εγκατάσταση αλλά και η αξιολόγηση μιας νέας υπηρεσίας βιβλιοθήκης, όπου οι χρήστες που βρίσκονται σε απομακρυσμένες περιοχές μπορούν να αποκτήσουν απευθείας πρόσβαση μέσω του Web, σε ηχογραφημένες συλλογές. Οι συλλογές αυτές είναι αποθηκευμένες σε ιδρύματα συγκεκριμένων Ευρωπαϊκών κρατών (Ρώμη, Ααρχούς, Λονδίνο).

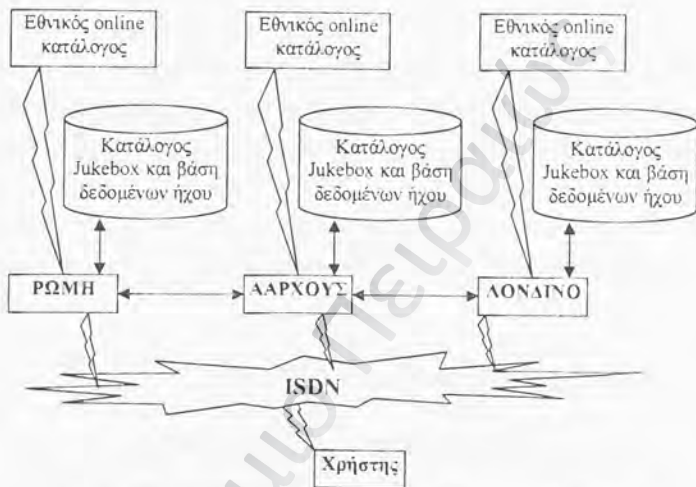
Μόλις ένας χρήστης εισέλθει στην ψηφιακή βιβλιοθήκη, συνδέεται αυτόματα με τον κόμβο που αναφέρεται στη χώρα του. Ο κόμβος αυτός περιέχει τα ηχητικά δεδομένα της συγκεκριμένης χώρας, όπως επίσης και τη βιβλιογραφική περιγραφή των αντίστοιχων δεδομένων που είναι αποθηκευμένα στους υπόλοιπους κόμβους της ψηφιακής βιβλιοθήκης, όπως φαίνεται στο σχ. 2.

Ο κατάλογος Jukebox είναι ένα ιδιαίτερο χαρακτηριστικό του συστήματος. Έτσι λοιπόν, μέσω της λειτουργικότητας που παρέχει ο κατάλογος αυτός, οι χρήστες έχουν τη δυνατότητα να εντοπίζουν ηχητικά δεδομένα που βρίσκονται κατανεμημένα στους υπόλοιπους κόμβους του συστήματος.

Αρχικά, οι χρήστες απευθύνουν τα ερωτήματά τους στον κόμβο της χώρας τους. Αν το ηχητικό ντοκουμέντο που ικανοποιεί κάποιο ερώτημα ανήκει σε διαφορετικό κόμβο από αυτόν από τον οποίο προήλθε το εν λόγω ερώτημα, το σύστημα συνδέεται μέσω δικτύου Integrated



Services Digital Network - ISDN με τον απομακρυσμένο κόμβο και μεταφέρει το ζητούμενο ηχητικό υλικό τοπικά στο σταθμό εργασίας του χρήστη. Η ανάκτηση πληροφοριών όπως και η μεταφορά δεδομένων γίνεται με τέτοιο τρόπο, ώστε οι χρήστες να μην ασχολούνται με την προέλευση των αρχείων που παραλαμβάνουν (αν είναι δηλαδή τοπικά ή απομακρυσμένα



Σχήμα 2: Η αρχιτεκτονική της ψηφιακής βιβλιοθήκης Jukebox

αρχεία). Τα αποτελέσματα των αναζητήσεων είναι εμπλουτισμένα με μεταδεδομένα, τα οποία είναι οργανωμένα κατά στήλες, όπου κάθε στήλη αναφέρεται σε κάποιο μεταδεδομένο του αντίστοιχου ηχητικού υλικού, όπως “τίτλος”, “δημιουργός”, “συνθέτης”, “έτος ηχογράφησης” και “χώρα προέλευσης”. Τα αποτελέσματα μπορούν να ταξινομηθούν με οποιαδήποτε σειρά επιλέξουν οι χρήστες. Τέλος, εκτενέστερες πληροφορίες για το κάθε μέλος της λίστας εμφανίζονται με την επιλογή της αντίστοιχης εγγραφής.

Αν και η ψηφιακή βιβλιοθήκη Jukebox βασίζεται στο πρότυπο mp3 για την κωδικοποίηση των αρχείων ήχου, εντούτοις η πρόσβαση στη συλλογή αυτή είναι περιορισμένη εξαιτίας της εξάρτησης από συγκεκριμένο (εμπορικό) αποκωδικοποιητή υλικού στην πλευρά του χρήστη.



### 3.4.2 Η μουσική ψηφιακή βιβλιοθήκη Patron

Το σύστημα Patron [Lyon, 1997] αναφέρεται σε μια ψηφιακή βιβλιοθήκη ήχου και video βασισμένη σε μια αρχιτεκτονική πελάτη-εξυπηρετητή (client-server), σχεδιασμένη κατά τέτοιο τρόπο ώστε να παρέχει ήχο και video κατά απαίτηση (on demand) για τους χρήστες του Πανεπιστημίου του Surrey. Η βασική σχεδιαστική απαίτηση της αρχιτεκτονικής του συστήματος είναι η υιοθέτηση μιας τμηματικής (modular) προσέγγισης, έτσι ώστε να δημιουργηθεί ένα ανοικτό σύστημα, στο οποίο να μπορούν να ανανεώνονται και να αντικαθιστώνται τα διάφορα συστατικά του μέρη κατά βούληση. Το σύστημα αυτό αποτελείται από τρία κύρια τμήματα:

- Ένα εξυπηρετητή ικανό να διαχειρίζεται δεδομένα συνεχούς ροής (streaming data)
- Δίκτυο υψηλών ταχυτήτων
- Σταθμούς εργασίας με τις λιγότερες απαιτήσεις όσον αφορά τη συντήρησή τους και την εξάρτησή τους από το υλικό (hardware).

Ο εξυπηρετητής αποτελείται από έναν Web εξυπηρετητή, μέσω του οποίου ελέγχεται και η πρόσβαση στην ψηφιακή βιβλιοθήκη. Η παρεχόμενη λειτουργικότητα υποστηρίζεται από ένα σύνολο από τμήματα κώδικα σε Java εκτελούμενα στον εξυπηρετητή (Java servlets), τα οποία με τη σειρά τους παρέχουν ένα επίπλεον επίπεδο αφαίρεσης μεταξύ του συστήματος διεπαφής και της υποκείμενης βάσης δεδομένων. Συγκεκριμένα, τα τμήματα κώδικα χωρίζουν μια ηλεκτρονική διεύθυνση (URL) στα επί μέρους της συστατικά, τα οποία και υπόκεινται σε ξεχωριστή επεξεργασία.

Η εφαρμογή, που βρίσκεται στη μεριά του πελάτη, εκτελείται ολοκληρωτικά μέσα από έναν Web browser. Τα αντικείμενα που παραλαμβάνονται από τον εξυπηρετητή, εμφανίζονται σε μορφή HTML με τη βοήθεια κατάλληλων ενθυλακωμένων εφαρμογών αναπαραγωγής ήχου ή/και video. Το πλεονέκτημα αυτής της προσέγγισης έγκειται στο γεγονός ότι ο εξυπηρετητής αναλαμβάνει μόνο την παράδοση των δεδομένων και όχι την επεξεργασία τους. Έτσι, αυτή η κατανομή αρμοδιοτήτων συντελεί στη δημιουργία ενός αρκετά αποδοτικού συστήματος.

Η καλή ποιότητα των ηχητικών δεδομένων και των δεδομένων τύπου video που ανταλλάσσονται μεταξύ πελατών και εξυπηρετητή οφείλεται στη δικτυακή υποστήριξη του συστήματος από ένα δίκτυο Asynchronous Transfer Mode - ATM, το οποίο είναι εγκατεστημένο στο Πανεπιστήμιο του Surrey.

Το γεγονός ότι η συγκεκριμένη ψηφιακή βιβλιοθήκη βασίζεται σε παγιωμένες διαδικτυακές τεχνολογίες όπως ο Web εξυπηρετητής και ο Web browser για τη διαχείριση των δεδομένων της αλλά και την αλληλεπίδραση με τους χρήστες, προσδίδει στο σύστημα τα

χαρακτηριστικά της επεκτασιμότητας καθώς και της μεγάλης διαθεσιμότητας, η οποία περιορίζεται μόνο από τις δυνατότητες του δικτύου στο οποίο φιλοξενείται.

### 3.4.3 Η μουσική ψηφιακή βιβλιοθήκη VARIATIONS

Η μουσική ψηφιακή βιβλιοθήκη VARIATIONS [Dunn, 1999] παρέχει πρόσβαση σε περισσότερες από 5.000 ηχογραφήσεις της βιβλιοθήκης του William και Gayle Cook στη Μ. Βρετανία. Το σύστημα αυτό αποτελεί στην ουσία την ολοκλήρωση διαφόρων αυτόνομων πακέτων λογισμικού σε μια εφαρμογή που προσφέρει λειτουργικότητα, η οποία συναντάται συνήθως σε συστήματα ψηφιακών βιβλιοθηκών.

Τα αντικείμενα που βρίσκονται αποθηκευμένα στην ψηφιακή βιβλιοθήκη VARIATIONS αποτελούνται από ένα αρχείο μεταδεδομένων και από ένα ή περισσότερα αρχεία ήχου, όπως φαίνεται στο σχ. 3.

Τα μεταδεδομένα του συστήματος χωρίζονται σε τρεις τουλάχιστον κατηγορίες: διαχειριστικά, δομικά και περιγραφικά [Fleischhauer, 1996]. Τα μεταδεδομένα αυτά



Σχήμα 3: Ένα αντικείμενο της ψηφιακής βιβλιοθήκης VARIATIONS

ακολουθούν το πρότυπο USMARC, ενώ όπου κρίνεται απαραίτητο, το πρότυπο αυτό έχει επεκταθεί. Τα περιγραφικά μεταδεδομένα χρησιμοποιούνται από τους χρήστες κυρίως για τις διαδικασίες αναζήτησης και προσδιορισμού των αντικειμένων στα οποία αναφέρονται.

Αναφορικά με τον τρόπο αποθήκευσης των δεδομένων, χρησιμοποιείται ένα σύστημα αποθήκευσης δυο στρωμάτων: ένας εξυπηρετητής αρχειοθέτησης (archive server) διατηρεί τα πρωτότυπα αρχεία όλων των δεδομένων ήχου σε κασέτες (tape cartridges) που είναι αποθηκευμένες σε μια αυτοματοποιημένη βιβλιοθήκη κασετών. Παράλληλα, ένας εξυπηρετητής αναπαραγωγής (playback server) διατηρεί τα αρχεία ήχου, τα οποία έχουν ανακτηθεί πιο πρόσφατα από την ψηφιακή βιβλιοθήκη, ευνοώντας έτσι τη γρήγορη ικανοποίηση των

αιτημάτων των χρηστών, καθώς και τη δυνατότητα για αποδοτική αναμετάδοση ηχητικών δεδομένων συνεχούς ροής στους συνδεδεμένους χρήστες.

Η πρόσβαση στην ψηφιακή βιβλιοθήκη, παρέχεται από περίπου εξήντα υπολογιστές βασισμένους στο λειτουργικό σύστημα Windows NT. Οι υπολογιστές αυτοί είναι εξοπλισμένοι με κάρτες ήχου υψηλής ποιότητας, όπως επίσης και με ακουστικά και μικρόφωνα. Από πλευράς λογισμικού, στους σταθμούς εργασίας των χρηστών είναι εγκατεστημένοι: ένας εμπορικός αποκωδικοποιητής αρχείων τύπου MPEG της Microsoft, μια εμπορική εφαρμογή αναπαραγωγής video της IBM και μια εξειδικευμένη εφαρμογή της ψηφιακής βιβλιοθήκης VARIATIONS.

### 3.4.4 Ανάκτηση πληροφοριών βασισμένη στο περιεχόμενο σε μουσικές ψηφιακές βιβλιοθήκες

Μια πρωτότυπη προσέγγιση, όσον αφορά τη διάθεση μουσικών συλλογών στο Web ακολουθείται από την ψηφιακή βιβλιοθήκη μουσικών και ακουστικών δεδομένων της Νέας Ζηλανδίας [McNab, 1996]. Σύμφωνα με την αρχιτεκτονική του συστήματος αυτού, δεν απαιτείται η ύπαρξη βιβλιογραφικών δεδομένων για την ανάκτηση των ζητούμενων μουσικών αρχείων. Αντίθετα, ένα σύστημα "ευρετηρίου μελωδιών" ανακτά μουσικά δεδομένα χρησιμοποιώντας ως κριτήρια αναζήτησης μια σειρά από νότες, οι οποίες εισάγονται στο σύστημα μέσω ενός απλού μικροφώνου. Η εισαγωγή αυτή πραγματοποιείται από τους χρήστες με διάφορους τρόπους. Μπορεί κανείς να σφυρίξει ένα μέρος του σκοπού, ή να τραγουδήσει, ή ακόμα και να χρησιμοποιήσει κάποιο μουσικό όργανο.

Η ανάκτηση πληροφοριών στη συγκεκριμένη ψηφιακή βιβλιοθήκη, βασίζεται στην αναζήτηση μιας βάσης δεδομένων από παρτιτούρες. Αυτό βέβαια απαιτεί την ύπαρξη αποδοτικών αλγορίθμων διαχείρισης αλφαριθμητικών (pattern matching), καθώς οι παρτιτούρες αυτές αναπαρίστανται ως αλφαριθμητικά στο μέσο αποθήκευσης. Τα αποτελέσματα μιας έρευνας πάνω στην ακρίβεια με την οποία οι άνθρωποι τραγουδάνε μια γνωστή μελωδία, έδειξαν ότι πρέπει να γίνεται προσεγγιστική σύγκριση των δεδομένων εισόδου με τα δεδομένα του συστήματος. Επίσης, η ανάλυση δυο βάσεων δεδομένων που αποτελούνταν από λαϊκά τραγούδια, έδειξε ότι πέντε ή και λιγότερες νότες είναι αρκετές για αποδοτική ανάκτηση πληροφοριών, με την προϋπόθεση ότι οι νότες αυτές εισάγονται με το σωστό ρυθμό. Αν δεν συμπεριλαμβάνεται ο ρυθμός με τον οποίο συνοδεύονται οι νότες, είναι απαραίτητες μια με δυο νότες παραπάνω. Τέλος, η προσεγγιστική αναζήτηση απαιτεί δώδεκα ή περισσότερες



νότες, έτσι ώστε τα αποτελέσματα της αναζήτησης να μπορούν εύκολα να υποστούν διαχείριση από τους χρήστες.

Μια ακόμα προσπάθεια για ανάκτηση μουσικών δεδομένων με βάση το περιεχόμενό τους, ακολουθείται από το σύστημα SMILE [Melucci, 2000]. Αν και το σύστημα αυτό δεν αναφέρεται σε δεδομένα που είναι αποθηκευμένα σε μια ψηφιακή βιβλιοθήκη αλλά σε μια βάση δεδομένων, εντούτοις παρουσιάζει ενδιαφέρον, καθώς είναι σχεδιασμένο κατά τέτοιο τρόπο ώστε να μπορεί εύκολα να προσαρμοστεί σε μια ψηφιακή βιβλιοθήκη ακουστικού υλικού.

Το σύστημα SMILE ικανοποιεί τρεις απαιτήσεις μιας μουσικής ψηφιακής βιβλιοθήκης [Melucci, 2000]:

- Ανάκτηση πληροφοριών με βάση το περιεχόμενο μουσικών αρχείων, όπου η αναπαραστάση περιεχομένου έχει μελετηθεί τόσο στη γλώσσα της ευρετηρίασης, όσο και στη γλώσσα της ανάκτησης.
- Αυτοματοποίηση των λειτουργιών της ευρετηρίασης, ανάκτησης καθώς και της κατασκευής συνδέσμων (links).
- Ολοκλήρωση των διαδικασιών πλοήγησης και αναζήτησης σε μουσικές συλλογές, σύμφωνα με ένα συνεπές μεθοδολογικό και τεχνολογικό πλαίσιο.

Οι λειτουργίες που προκύπτουν στηρίζονται στη μεθοδολογία του συστήματος, η οποία με τη σειρά της αποτελείται από τα ακόλουθα βήματα:

- Την αυτόματη κατάταξη και κανονικοποίηση μουσικών εγγράφων, έτσι ώστε να προκύψουν μουσικές φράσεις. Τα μουσικά έγγραφα αναπαριστώνται χρησιμοποιώντας μορφοποιήσεις βασισμένες σε σημειολογία συμβόλων (symbolic notation).
- Την αυτόματη ευρετηρίαση και ανάκτηση για την περιγραφή του περιεχομένου των εγγράφων, έτσι ώστε να ανακτώνται τα πιο σχετικά έγγραφα με τις πληροφοριακές ανάγκες των χρηστών.
- Την αυτόματη κατασκευή υπέρ-μουσικής (hyper-music), που δίνει τη δυνατότητα στους χρήστες να εφαρμόζουν ολοκληρωμένες αναζητήσεις και πλοηγήσεις στο υποκείμενο σύστημα.

### **3.4.5 Μουσικές ψηφιακές βιβλιοθήκες βασισμένες σε αρχιτεκτονικές Peer-to-Peer (P2P)**

Σε ένα σύστημα Peer to Peer - P2P η πληροφορία αλλά και οι λειτουργίες διαχείρισης και επεξεργασίας της πληροφορίας είναι κατανεμημένες στους κόμβους αντί να είναι



συγκεντρωμένες σε κάποιον κεντρικό εξυπηρετητή. Οι χρήστες ανταλλάσσουν δεδομένα με απευθείας συνδέσεις μεταξύ τους. Ένα P2P σύστημα ορίζεται ως ένα κατακεντρωμένο σύστημα χωρίς καθόλου κεντρικό έλεγχο, όπου το λογισμικό που εκτελείται σε κάθε κόμβο είναι ισοδύναμο όσον αφορά τη λειτουργικότητά που προσφέρει. Μερικά από τα πλεονεκτήματα της P2P αρχιτεκτονικής σε σχέση με τις αρχιτεκτονικές πελάτη-εξυπηρετητή (client-server) και τις αρχιτεκτονικές τριών στρωμάτων (3-tier) είναι τα ακόλουθα:

- **Εμφυτη επεκτασιμότητα (inherent scalability).** Είναι πολύ πιο εύκολο να προστεθεί ή/και να αφαιρεθεί κάποιος κόμβος από το σύστημα P2P, καθώς το μόνο που χρειάζεται είναι η τυποποιημένη εγκατάσταση κάποιας εφαρμογής στον υποψήφιο κόμβο.
- **Διαθεσιμότητα μεγάλου όγκου πληροφορίας.** Η ευκολία συμμετοχής στις ψηφιακές βιβλιοθήκες που βασίζονται σε αυτού του είδους τις αρχιτεκτονικές ως παροχείς πληροφορίας (content providers), έχει οδηγήσει στη διαθεσιμότητα πολύ μεγάλου όγκου πληροφορίας στο Διαδίκτυο. Ο τύπος δεδομένων που ευνοείται περισσότερο από το καθεστώς αυτό είναι: τα αρχεία ήχου και ειδικότερα τα αρχεία τύπου mp3, τα οποία και αποτελούν το πιο διαδεδομένο μέσο ανταλλαγής ηχητικής πληροφορίας στο χώρο του Διαδικτύου.

Πέρα από τα πολύ σημαντικά αυτά πλεονεκτήματα, τα συστήματα ψηφιακών βιβλιοθηκών που είναι βασισμένα σε αρχιτεκτονικές P2P, παρουσιάζουν το εξίσου σημαντικό μειονέκτημα που έγκειται στο γεγονός ότι δεν προσφέρουν καμία εγγύηση όσον αφορά την ποιότητα των υπηρεσιών που παρέχουν. Έτσι, τα αρχεία που μεταποθηκεύουν (download) οι χρήστες μπορεί να μην αντιστοιχούν στα κριτήρια αναζητήσης που είχαν θέσει, ή ακόμα και να μην ολοκληρώνεται η μεταποθήκευσή τους. Επίσης, η ανεξέλεγκτη μεταποθήκευση αρχείων που αντιστοιχούν σε δεδομένα με κατοχυρωμένα πνευματικά δικαιώματα, έχει δημιουργήσει αμφιβολίες, όσον αφορά τη νομιμότητα των αντίστοιχων συστημάτων P2P που τα διαχειρίζονται. Στη συνέχεια, παρουσιάζονται δυο ψηφιακές βιβλιοθήκες βασισμένες σε παραλλαγές της αρχιτεκτονικής P2P, για τη διαχείριση αρχείων τύπου mp3.

#### 3.4.5.1 Η ψηφιακή βιβλιοθήκη Napster

Στο σύστημα Napster [Napster, 2000] υπάρχει ένας κεντρικός εξυπηρετητής καταλόγου (directory server) ο οποίος διατηρεί ευρετήριο με τα μεταδεδομένα όλων των αρχείων των κόμβων που βρίσκονται στο δίκτυο. Τα μεταδεδομένα μπορεί να περιέχουν στοιχεία όπως: ονόματα αρχείων, ημερομηνίες δημιουργίας και πληροφορίες πνευματικών

δικαιωμάτων. Επίσης ο εξυπηρετητής διατηρεί ένα πίνακα με πληροφορίες σύνδεσης των κόμβων όπως η διεύθυνση IP του χρήστη και η χωρητικότητα της γραμμής του.

Στο σύστημα Napster, η λειτουργία της αναζήτησης χρησιμοποιεί ως δεδομένο εισόδου μια λίστα που αποτελείται από λέξεις-κλειδιά, όπως αυτές τίθενται από τους χρήστες. Όταν ο κεντρικός εξυπηρετητής λάβει μια ερώτηση, ψάχνει στο ευρετήριο του για να βρει αρχεία που να ταιριάζουν με τα κριτήρια που έθεσε ο χρήστης. Τα αποτελέσματα της αναζήτησης περιλαμβάνουν μια λίστα από κόμβους οι οποίοι διατηρούν τα ζητούμενα αρχεία και αποστέλλονται πίσω στον χρήστη. Στη συνέχεια, ο χρήστης ανοίγει μια απευθείας (direct) σύνδεση με τον κόμβο ο οποίος έχει το ζητούμενο αρχείο, με σκοπό να το μεταποληθείσει τοπικά στο σταθμό εργασίας του.

Τα πλεονεκτήματα που προκύπτουν από αυτή την συγκεντρωτική προσέγγιση όσον αφορά την ευρετηρίαση, είναι τα εξής:

- Γρήγορος και αποδοτικός εντοπισμός των αρχείων
- Οι αναζητήσεις είναι όσο το δυνατόν πιο περιεκτικές
- Όλοι οι καταχωρημένοι χρήστες συμμετέχουν στο δίκτυο.

Αντίστοιχα, υπάρχουν και τα εξής μειονεκτήματα:

- Τα δεδομένα που διακινούνται είναι ευαίσθητα στη λογοκρισία και στις τεχνικές βλάβες
- Το φαινόμενο Slashdot (Slashdot effect): Τα δημοφιλή δεδομένα γίνονται λιγότερο προσβάσιμα εξαιτίας του φόρτου των αιτήσεων σε έναν κεντρικό εξυπηρετητή
- Το κεντρικό ευρετήριο δεν είναι συγχρονισμένο με τα επί μέρους ευρετήρια, επειδή η βάση δεδομένων του κεντρικού εξυπηρετητή ενημερώνεται περιοδικά.

#### **3.4.5.2 Η ψηφιακή βιβλιοθήκη Gnutella**

Σύμφωνα με την αρχιτεκτονική του συστήματος Gnutella [Gnutella, 2000], δεν υπάρχει κεντρικός εξυπηρετητής (directory server), αλλά οι συμμετέχοντες κόμβοι έχουν την ίδια δυνατότητες και ευθύνες. Έτσι, η επικοινωνία μεταξύ των κόμβων είναι συμμετρική, ενώ το ευρετήριο με τα μεταδεδωμένα των διαμοιραζόμενων αρχείων βρίσκεται αποθηκευμένο τοπικά σε κάθε κόμβο.

Τα πλεονεκτήματα που προκύπτουν από αυτήν την κατανομημένη προσέγγιση όσον αφορά την ευρετηρίαση, είναι τα εξής:

- Εμφυτη επεκτασιμότητα (inherent scalability)
- Αποφυγή του φαινομένου Slashdot εξαιτίας της έλλειψης κεντρικού σημείου πρόσβασης

- Ανθεκτικότητα του συστήματος στα σφάλματα

Αντίστοιχα, υπάρχουν και τα εξής μειονεκτήματα:

- Χρονοβόρα διαδικασία ανάκτησης της ζητούμενης πληροφορίας
- Αυξημένη κίνηση στο δίκτυο μόνο για αναζήτηση

### 3.5 Η προτεινόμενη αρχιτεκτονική

#### 3.5.1 Κίνητρο

Κίνητρο για το σχεδιασμό της προτεινόμενης αρχιτεκτονικής αποτελεί το γεγονός ότι δεν υπάρχουν κατάλληλες εφαρμογές στο Web που να χειρίζονται αποτελεσματικά μουσικά δεδομένα τα οποία να είναι κωδικοποιημένα με βάση το δημοφιλές πρότυπο mp3. Προσπάθειες όπως αυτές που ακολουθούνται από τις εταιρίες Napster και Gnutella δεν εκμεταλλεύονται το γεγονός της ύπαρξης περιγραφικής πληροφορίας ενθυλακωμένης στο σώμα (body) τέτοιων αρχείων. Συγκεκριμένα, κάθε αρχείο τύπου mp3 έχει η δυνατότητα να περιέχει μεταδεδομένα, τα οποία να είναι αποθηκευμένα στην ετικέτα ID3 αυτού.

Ένα άλλο επιχείρημα υπέρ της προτεινόμενης αρχιτεκτονικής είναι η ανάγκη παροχής ενός πλαισίου διαχείρισης μουσικής πληροφορίας στο Web που να διασφαλίζει τα κατοχυρωμένα πνευματικά δικαιώματα των δημιουργών των μουσικών αυτών δεδομένων. Μια εφαρμογή που διαχειρίζεται αρχεία τύπου mp3 πρέπει να συμμορφώνεται με τους νόμους που αφορούν την πνευματική ιδιοκτησία, ενώ ταυτόχρονα πρέπει να παρέχει όσο το δυνατό μεγαλύτερη λειτουργικότητα.

Ακολουθώντας τις παραπάνω επισημάνσεις, στο κεφάλαιο αυτό παρουσιάζεται μια μουσική ψηφιακή βιβλιοθήκη ικανή να διαχειρίζεται αποδοτικά μουσικά αρχεία, τα οποία είναι κωδικοποιημένα σύμφωνα με το πρότυπο mp3. Παράλληλα, δίνεται ιδιαίτερη έμφαση στο σεβασμό των πνευματικών δικαιωμάτων των δημιουργών τους. Η παρεχόμενη λειτουργικότητα βασίζεται στην εκμετάλλευση των ετικετών ID3 των αρχείων τύπου mp3 για την ανάπτυξη τεχνικών ανάκτησης πληροφοριών επικεντρωμένων στο περιεχόμενο των αρχείων αυτών.

#### 3.5.2 Αρχιτεκτονική

Η περιγραφόμενη μουσική ψηφιακή βιβλιοθήκη βασίζεται σε μια αρχιτεκτονική τριών στρωμάτων, τυπική για καταναμημένες εφαρμογές όπως αυτή που παρουσιάζεται στο σχ. 4.



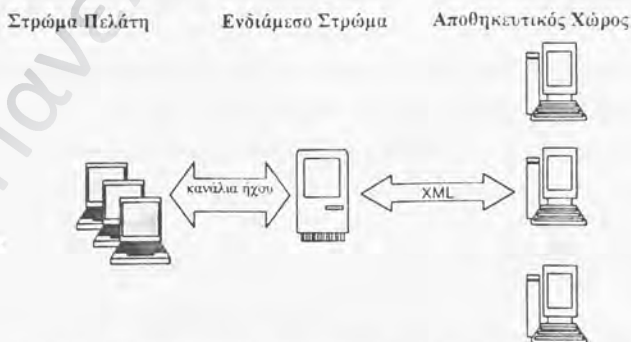
Βασίζεται στο πρότυπο της XML για τη διαχείριση συλλογών αρχείων τύπου mp3, τα οποία διατηρούνται σε έναν κατανεμημένο αποθηκευτικό χώρο που αποτελείται από λειτουργικά αυτόνομους κόμβους. Τα δεδομένα που ζητούνται και ανακτώνται από τον αποθηκευτικό χώρο υπόκεινται σε επεξεργασία με διάφορους τρόπους στο ενδιάμεσο στρώμα, το οποίο θέτει και τους κανόνες λειτουργίας του συστήματος.

Τα ζητούμενα δεδομένα (αρχεία τύπου mp3) μεταφέρονται από τον αποθηκευτικό χώρο στο ενδιάμεσο στρώμα και στη συνέχεια αναμεταδίδονται στους Web browsers που βρίσκονται στο στρώμα του πελάτη, αφού πρώτα μετατραπούν σε κανάλια ήχου (audio streams) από εξυπηρετητές συνεχούς ροής (streaming servers). Σύμφωνα με αυτό το σενάριο, οι Web browsers συμπεριφέρονται ως "λεπτοί πελάτες" (thin clients), με την έννοια ότι δεν υπάρχει προεγκατεστημένο ειδικευμένο λογισμικό για την επεξεργασία και διαχείριση των πολυμεσικών αυτών δεδομένων. Είναι ευθύνη των Web browsers να διαχειριστούν κατάλληλα τα κανάλια ήχου με ανάλογο λογισμικό (με την μορφή plug-ins ή built-in λειτουργικότητας).

Η προσέγγιση που ακολουθείται ευνοεί την κατοχύρωση των πνευματικών δικαιωμάτων των συντελεστών των δεδομένων που περιέχουν τα αρχεία τύπου mp3, καθώς τα αρχεία αυτά μετατρέπονται σε ηχητικά κανάλια συνεχούς ροής, τα οποία και αναμεταδίδονται στο Διαδίκτυο, χωρίς να ευνοείται η ανεξέλεγκτη μεταποθήκευσή τους στους σταθμούς εργασίας των χρηστών.

### 3.5.2.1 Αποθηκευτικός χώρος

Ο αποθηκευτικός χώρος αποτελείται από κατανεμημένες συλλογές αρχείων τύπου mp3, οι οποίες αντιστοιχούν στους διάφορους κόμβους της ψηφιακής βιβλιοθήκης. Λαμβάνοντας



Σχήμα 4: Η αρχιτεκτονική τριών στρωμάτων της ψηφιακής βιβλιοθήκης



υπόψη την κατανεμημένη φύση του αποθηκευτικού χώρου, κάθε αρχείο της ψηφιακής βιβλιοθήκης έχει ένα μοναδικό όνομα/προσδιοριστή. Οι κόμβοι της ψηφιακής βιβλιοθήκης διατηρούν από έναν Web εξυπηρετητή, ο οποίος παρέχει μοναδικά ονόματα στα αρχεία που φιλοξενεί (σύμφωνα με το πρωτόκολλο URL), ενώ επίσης τα καθιστά διαθέσιμα στο ενδιαμέσο στρώμα. Όλα τα αρχεία τύπου mp3 κάθε κόμβου της αρχιτεκτονικής εντάσσονται σε ένα προκαθορισμένο ιεραρχικό σύστημα αρχείων. Ο φάκελος-ρίζα μιας τέτοιας ιεραρχίας αντιστοιχεί στο φάκελο-ρίζα εγγράφων (document root folder) του Web εξυπηρετητή. Τα βασικά δεδομένα κάθε κόμβου (αρχεία τύπου mp3) βρίσκονται στο φάκελο "Collection", ο οποίος με τη σειρά του ανήκει στο φάκελο-ρίζα εγγράφων του Web εξυπηρετητή. Το σχ. 5 παρουσιάζει το ιεραρχικό σύστημα φακέλων/ αρχείων κάθε κόμβου.

```
Document root folder/Collection/song1.mp3
song2.mp3
...
songn.mp3
```

Σχήμα 5: Το ιεραρχικό σύστημα φακέλων της μουσικής ψηφιακής βιβλιοθήκης

Ετσι, σύμφωνα με την παραπάνω διαδικασία ονοματολογίας, το μοναδικό όνομα/προσδιοριστής για ένα αρχείο τύπου mp3, το οποίο βρίσκεται στον Web εξυπηρετητή με το όνομα "thalis.com", θα είναι "thalis.com/Collection/mysong.mp3".

Για κάθε κόμβο της ψηφιακής βιβλιοθήκης καθορίζεται μια ιεραρχική δομή που περιγράφει τον εκάστοτε αποθηκευτικό χώρο. Είναι αποτυπωμένη σε μορφή XML και αποτελείται από υποδένδρα συστατικών που αντιστοιχούν σε πληροφορία, η οποία έχει εξαχθεί από τις ετικέτες ID3. Ανάλογα με το μέγεθος των συλλογών αρχείων τύπου mp3, η ιεραρχία αυτή μπορεί να απεικονιστεί σε δομές δεδομένων που κυμαίνονται από ένα απλό αρχείο τύπου XML έως ένα σύνθετο Σύστημα Διαχείρισης Βάσεων Δεδομένων (RDBMS). Στο σχ. 6 παρουσιάζεται η ιεραρχική δομή ενός αρχείου τύπου XML (repository.xml).

```
<?xml version="1.0" encoding="ISO-8859-7"?>
<mp3files>
  <mp3file entry="November 13, 1999"
    counter="23" title="Lotus"
    id="www.thalis.com/Collection/mysong.mp3"
    artist="REM" genre="Rock" album="The
    Greatest '99" •••/>
  <mp3file entry="November 23, 1998"
    counter="564" title="catch"
    id="www.thalis.com/Collection/mysong2.mp3"
    artist="the cure" genre="Rock" album="the
    cure" •••/>
  <mp3file entry="November 13, 1999"
    counter="566" title="with or without you"
    id="www.thalis.com/Collection/mysong4.mp3"
    artist="U2" genre="Rock" album="U2" •••/>
  •
  •
  •
  •
</mp3files >
```

Σχήμα 6: Ένα αρχείο repository.xml

Το χαρακτηριστικό "counter" μετράει πόσες φορές το συγκεκριμένο αρχείο τύπου mp3 έχει προσπελαστεί από τους χρήστες της ψηφιακής βιβλιοθήκης. Σκοπός του είναι να ποσοτικοποιήσει τη δημοτικότητα του συγκεκριμένου αρχείου για τους χρήστες της ψηφιακής βιβλιοθήκης, ενώ αυξάνει κατά μία μονάδα κάθε φορά που επιλέγεται από κάποιον χρήστη. Η πληροφορία αυτή χρησιμοποιείται από το λογισμικό τμήμα "Mediator", το οποίο βρίσκεται στο ενδιάμεσο στρώμα και μεταξύ άλλων κατατάσσει τα αποτελέσματα μιας αναζήτησης σύμφωνα με τη δημοτικότητά τους. Το χαρακτηριστικό "entry" υποδηλώνει την ημερομηνία στην οποία το αναφερόμενο αρχείο εισήχθη για πρώτη φορά στη μουσική ψηφιακή βιβλιοθήκη. Τα υπόλοιπα χαρακτηριστικά που φαίνονται στο σχ. 6, αποτελούν στην ουσία μεταδεδομένα που περιέχονται στην ετικέτα ID3 του αρχείου τύπου mp3.

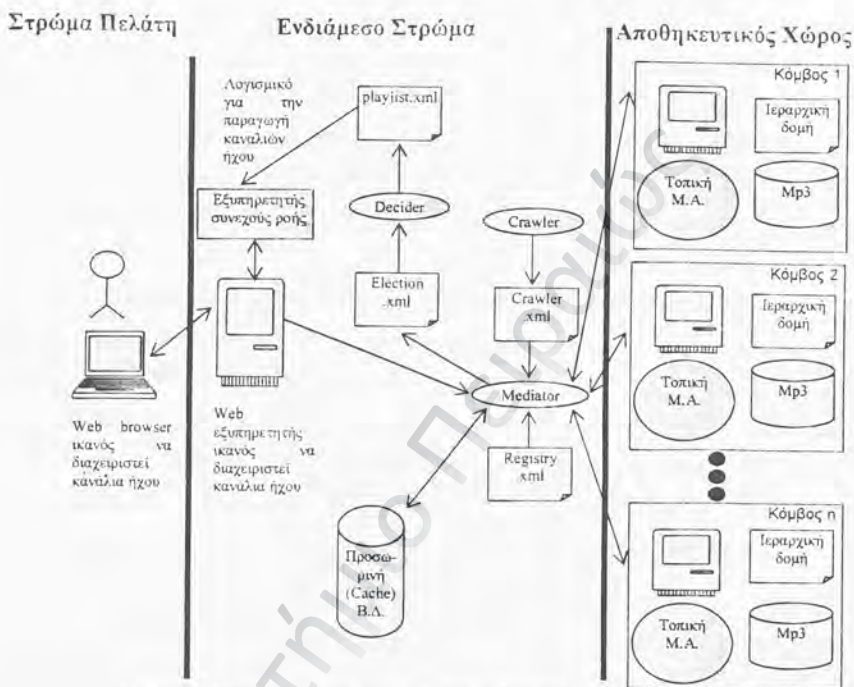
Η συνέπεια των δεδομένων της ιεραρχικής δομής του εκάστοτε αποθηκευτικού χώρου και των πραγματικών περιεχομένων του εξασφαλίζεται από ένα σύνολο υποπρογραμμάτων που εκτελούνται περιοδικά και τοπικά σε κάθε κόμβο. Τέλος, κάθε κόμβος της μουσικής ψηφιακής βιβλιοθήκης διατηρεί μια τοπική μηχανή αναζήτησης, η οποία εξυπηρετεί ερωτήματα τα οποία απευθύνονται στην τοπική ιεραρχική δομή του κόμβου. Οι μηχανές αναζήτησης στο σύνολό τους κατευθύνονται από το ενδιάμεσο στρώμα και συγκεκριμένα από το λογισμικό τμήμα "Mediator", το οποίο και περιγράφεται αναλυτικά στην επόμενη ενότητα.

### 3.5.2.2 Ενδιάμεσο στρώμα

Η βασική λειτουργικότητα της προτεινόμενης αρχιτεκτονικής παρέχεται από το ενδιάμεσο στρώμα. Είναι ευθύνη του ενδιάμεσου στρώματος να παράγει τα κανάλια ήχου, τα οποία θα αναμεταδοθούν τελικά στους συνδεδεμένους χρήστες. Σύμφωνα με την επιχειρησιακή λογική λειτουργίας της εφαρμογής που υιοθετεί την προτεινόμενη μουσική ψηφιακή βιβλιοθήκη, μπορούν να ακολουθηθούν διάφορα σενάρια για την κατασκευή των καναλιών αυτών. Έτσι, θα μπορούσε κανείς να θεωρήσει τους ραδιοφωνικούς σταθμούς που εκπέμπουν στο Web καθώς και τις Πανεπιστημιακές μουσικές βιβλιοθήκες, ως δυο πιθανές περιπτώσεις όπου βρίσκει εφαρμογή το προτεινόμενο μοντέλο.

Στην περίπτωση ενός ραδιοφωνικού σταθμού που εκπέμπει στο Web, το ενδιάμεσο στρώμα δέχεται τις αιτήσεις των χρηστών για κάποιο μουσικό κομμάτι και ανάλογα απευθύνει ερωτήματα στον αποθηκευτικό χώρο, έτσι ώστε να αποκτήσει τα ζητούμενα μουσικά αρχεία τύπου mp3. Κατόπιν, χρησιμοποιώντας κατάλληλο λογισμικό (εξυπηρετητής συνεχούς ροής), συνθέτει τα κανάλια ήχου που θα αναμεταδώσει στο στρώμα του πελάτη. Στο σχ. 7 παρουσιάζεται η μουσική ψηφιακή βιβλιοθήκη, όπως αυτή βρίσκει εφαρμογή σε ένα ραδιοφωνικό σταθμό που εκπέμπει στο Web.

Σύμφωνα με τη λογική λειτουργίας του σταθμού, οι χρήστες καλούνται αρχικά να ψηφίσουν τα τραγούδια εκείνα που θέλουν να ακούσουν από το σταθμό. Αν τα τραγούδια που



Σχήμα 7: Η αρχιτεκτονική της μουσικής ψηφιακής βιβλιοθήκης

ψηφίζουν αντιστοιχούν σε αρχεία τύπου mp3 που βρίσκονται στον αποθηκευτικό χώρο της ψηφιακής βιβλιοθήκης, προστίθενται εγγραφές στο αρχείο election.xml. Το αρχείο αυτό είναι δομημένο σε μορφή XML και περιέχει το όνομα και την τοποθεσία των αρχείων τύπου mp3 που ψήφισαν οι χρήστες. Η ψηφιογραφία συνεχίζεται κατά τη διάρκεια λειτουργίας του σταθμού. Ανά τακτά χρονικά διαστήματα, το λογισμικό τμήμα "Decider" αποφασίζει ποια αρχεία από αυτά που αναφέρονται στο αρχείο election.xml (βλ. σχ. 8) θα μεταφερθούν στο αρχείο playlist.xml (βλ. σχ. 9), έτσι ώστε να συμμετέχουν στο κανάλι ήχου που θα αναμεταδοθεί στο στρώμα του πελάτη. Το κανάλι αυτό αποτελείται από αρχεία τύπου mp3, τα οποία μετατρέπονται σε ένα ενιαίο κανάλι ήχου από τον εξυπηρετητή συνεχούς ροής, το οποίο στη συνέχεια αναμεταδίδεται στους συνδεδεμένους χρήστες.



```
<?xml version="1.0" encoding="UTF-8" ?>
<election>
  <mp3file TITLE="Tis Kyriakes apo paidi tis
sixainomai"
LOCATION="www.thalis.com/Collection/mysong1.
mp3" COUNTER="6" ARTIST_NAME="Xaris kai
Panos Katsimixas" GENRE="5" />
  <mp3file TITLE="Dark Side of the Moon"
LOCATION="www.thalis.com/Collection/mysong2.
mp3" COUNTER="10" ARTIST_NAME="Pink Floyd"
GENRE="2" />
  <mp3file TITLE="Leei"
LOCATION="www.thalis.com/Collection/mysong3.
mp3" COUNTER="11" ARTIST_NAME="Kotsiras"
GENRE="4" />
  .
  .
  .
</election>
```

Σχήμα 8: Ένα αρχείο ψηφοφορίας: election.xml

```
<?xml version="1.0" encoding="UTF-8" ?>
<playlist>
  <mp3file TITLE="Tis Kyriakes apo paidi tis
sixainomai"
LOCATION="www.thalis.com/Collection/mysong1.
mp3" ARTIST_NAME="Xaris &Panos Katsimixas"
GENRE="5" />
  <mp3file TITLE="Leei"
LOCATION="www.thalis.com/Collection/mysong3.
mp3" ARTIST_NAME="Kotsiras" GENRE="4" />
  .
  .
  .
</playlist>
```

Σχήμα 9: Ένα αρχείο playlist.xml

Για τον εντοπισμό αρχείων τύπου mp3 στη μουσική ψηφιακή βιβλιοθήκη, ακολουθείται μια παρόμοια προσέγγιση, με αυτή που προτάθηκε από τον Dushay et al. [Dushay, 1999]. Συγκεκριμένα, οι χρήστες έχουν τη δυνατότητα να εισάγουν σε μια φόρμα αναζήτησης διάφορα μεταδεδομένα σχετικά με το τραγούδι που επιθυμούν να ακούσουν από το ραδιοφωνικό σταθμό στο Web, όπως όνομα καλλιτέχνη και τίτλο τραγουδιού. Κατόπιν, το λογισμικό τμήμα "Mediator" δέχεται τις αιτήσεις των χρηστών και τις μεταφράζει σε μορφή XML, έτσι ώστε να

είναι κατανοητές από τον υπολογιστή. Σύμφωνα με τις πληροφορίες που παρέχουν τα αρχεία crawler.xml και registry.xml, το λογισμικό τμήμα "Mediator" αποφασίζει σε ποιές μηχανές αναζήτησης του καταμετρημένου αποθηκευτικού χώρου θα απευθύνει τις αιτήσεις των χρηστών για τον εντοπισμό των επιθυμητών αρχείων τύπου mp3. Συγκεκριμένα, το αρχείο registry.xml περιέχει πληροφορίες σχετικές με την τοποθεσία των κόμβων που συμμετέχουν στην ψηφιακή βιβλιοθήκη, όπως φαίνεται και στο σχ. 10.

```
<?xml version="1.0" encoding="ISO-8859-7"?>
<nodes>
  <node code="pcd-03" URL=http://pcD-
    03.cs.unipi.gr/ port="3245" />
  <node code="pcd-05" URL=http://pcD-
    05.cs.unipi.gr/ port="3245" />
</nodes>
```

Σχήμα 10: Ένα αρχείο registry.xml

Αντίστοιχα, το αρχείο crawler.xml περιέχει πληροφορίες σχετικές με το περιεχόμενο κάθε κόμβου της ψηφιακής βιβλιοθήκης. Οι πληροφορίες αυτές αναφέρονται στα είδη μουσικής (συστατικό "category") των αρχείων τύπου mp3 του εκάστοτε κόμβου καθώς και στους καλλιτέχνες που έχουν δημιουργήσει τα τραγούδια στα οποία αναφέρονται τα αρχεία αυτά (βλ. σχ. 11). Το αρχείο crawler.xml ενημερώνεται από το λογισμικό τμήμα "Crawler", το οποίο εκτελείται περιοδικά και συλλέγει τα αντίστοιχα μεταδεδομένα από τους κόμβους της ψηφιακής βιβλιοθήκης.

```
<?xml version="1.0" encoding="ISO-8859-7"?>
<nodes>
  <node code= "pcd-03">
    <category>Childrens</category>
    <category>Rock</category>
    <artist>REM</artist>
    . . .
  </node>
  <node code= "pcd-05">
    . . .
  </node>
  . . .
</nodes>
```

Σχήμα 11: Ένα αρχείο crawler.xml

Το επόμενο βήμα είναι να απευθυνθούν τα ερωτήματα των χρηστών στους επιλεγμένους κόμβους. Σε κάθε κόμβο εκτελείται (μέσω της αντίστοιχης μηχανής αναζήτησης) τοπικά μια αναζήτηση, έτσι ώστε να εντοπιστούν τα ζητούμενα αρχεία. Στην περίπτωση, όπου η αναζήτηση φέρει αποτελέσματα, οι αναφορές των αρχείων που εντοπίστηκαν, συγκεντρώνονται σε μια λίστα μορφής XML και επιστρέφονται στο λογισμικό τμήμα "Mediator". Το τμήμα αυτό δέχεται λίστες με αποτελέσματα από διάφορους κόμβους, τα οποία και συγχωνεύει σε μια ενιαία λίστα σε μορφή XML, διαγράφοντας παράλληλα τις διπλοεγγραφές. Τέλος, το τμήμα "Mediator" ταξινομεί τη λίστα που προκύπτει με βάση τη σχετικότητα (relevancy) καθώς και τη δημοτικότητα (popularity) των συμμετεχόντων αρχείων τύπου mp3.

Στη συνέχεια, η λίστα με τα συγχωνευμένα αποτελέσματα αναζήτησης μετατρέπεται με τη βοήθεια μιας μηχανής XSLT (XSLT processor) σε ένα αρχείο τύπου HTML, το οποίο και αποστέλλεται στους συνδεδεμένους χρήστες. Στην περίπτωση, όπου η λίστα περιέχει αναφορές σε αρχεία που αντιστοιχούν στο τραγούδι που είχε αρχικά ζητηθεί, υπάρχει η δυνατότητα επιλογής του συγκεκριμένου αρχείου και αποστολή της επιλογής αυτής στο ενδιάμεσο στρώμα. Αυτόματα, προστίθεται μια νέα εγγραφή στο αρχείο election.xml, το οποίο περιέχει ουσιαστικά την ψηφοφορία των χρηστών.

Τέλος, για την ομαλή λειτουργία καθώς και για την καλή απόδοση του συστήματος, το ενδιάμεσο στρώμα διατηρεί μια προσωρινή (cache) βάση δεδομένων με τα αρχεία που παρουσιάζουν τη μεγαλύτερη ζήτηση. Μ' αυτόν τον τρόπο, περιορίζεται η ανάγκη επικοινωνίας με τον αποθηκευτικό χώρο, γεγονός που ελαχιστοποιεί το χρόνο απόκρισης των αιτημάτων των χρηστών από τη ψηφιακή βιβλιοθήκη.

### 3.5.2.3 Στρώμα πελάτη

Το στρώμα του πελάτη αποτελείται από ένα τυπικό Web browser, ο οποίος κατά κανόνα έχει δυνατότητες αναπαραγωγής πολυμεσικών δεδομένων (ήχος). Το προτεινόμενο μοντέλο εκμεταλλεύεται το γεγονός της ευρείας διαφθεσιμότητας του λογισμικού αυτού σε όλα σχεδόν τα λειτουργικά συστήματα που υπάρχουν στην αγορά. Έτσι, δυνατότητα σύνδεσης στις εφαρμογές που στηρίζονται σε αυτήν την ψηφιακή βιβλιοθήκη, έχουν ουσιαστικά όλοι οι χρήστες με πρόσβαση στο Web. Περιορισμοί στην πρόσβαση τίθενται μόνο από την ικανότητα του δικτύου να διαχειριστεί μεγάλες ποσότητες δεδομένων ήχου. Το δικτυακό πρωτόκολλο πάνω στο οποίο βασίζεται το Web (TCP/IP), είναι προσαρμοσμένο στη διαχείριση δεδομένων κειμένου και όχι στη διαχείριση πολυμεσικών δεδομένων. Ευτυχώς, ορισμένα πρωτοεμφανιζόμενα πρωτόκολλα δικτύων όπως το ATM over IP και το Ipv6, υπόσχονται να

παρέχουν καλύτερη υποστήριξη σε πολυμεσικές εφαρμογές στο Web, όπως αυτή που παρουσιάζεται σε αυτό το κεφάλαιο.

### 3.5.3 Πλάνο λειτουργίας ενός ραδιοφωνικού σταθμού στο Web

Στην περίπτωση του προτεινόμενου ραδιοφωνικού σταθμού στο Web, η κατασκευή των καναλιών ήχου που αναμεταδίδονται στο στρώμα του πελάτη είναι ευθύνη του λογισμικού τμήματος "Decider".

Το λογισμικό τμήμα "Decider" κατασκευάζει το αρχείο playlist.xml, το οποίο περιέχει τα αρχεία τύπου mp3 που αναμεταδίδονται στο στρώμα του πελάτη. Σύμφωνα με τη λογική λειτουργίας του σταθμού, το πρόγραμμα φτιάχνεται αυτόματα από τους ακροατές του. Αποτελείται από δίωρες εκπομπές, κατά τη διάρκεια των οποίων οι ακροατές έχουν τη δυνατότητα να ψηφίσουν τα τραγούδια εκείνα που θα ήθελαν να ακούσουν από το σταθμό.

Συγκεκριμένα, το τμήμα "Decider" αναζητά στο αρχείο ψηφοφορίας election.xml την κατηγορία μουσικής εκείνη, στην οποία αναφέρονται πιο συχνά τα αρχεία τύπου mp3 που βρίσκονται στο αρχείο αυτό. Σε κάθε ραδιοφωνική εκπομπή μεταδίδονται τραγούδια που ανήκουν στην πιο "δημοφιλή" κατηγορία, σύμφωνα με την ψηφοφορία που πραγματοποιήθηκε κατά τη διάρκεια της προηγούμενης εκπομπής.

Τα τραγούδια που θα ακουστούν από το σταθμό κατά τη διάρκεια μιας εκπομπής, επιλέγονται από το τμήμα "Decider" στο τέλος κάθε "χρονικού παραθύρου". Κάθε εκπομπή αποτελείται από ένα πλήθος σταθερών χρονικών περιόδων (π.χ. δέκα λεπτών), οι οποίες ονομάζονται "χρονικά παράθυρα".

Το τμήμα "Decider" βασίζεται σε τρία κριτήρια για να επιλέξει τα τραγούδια του αρχείου election.xml που θα στελεχώσουν το αρχείο playlist.xml. Η επιλογή γίνεται κατά τέτοιο τρόπο ώστε να προσομοιωθεί η διαδικασία επιλογής τραγουδιών μέσω αριερώσεων που ακολουθείται κατά κανόνα στους συμβατικούς ραδιοφωνικούς σταθμούς. Κατά συνέπεια, το πρώτο κριτήριο επιλογής αναφέρεται στο γεγονός ότι πρέπει να αναμεταδίδονται τα τραγούδια εκείνα που συγκεντρώνουν τις περισσότερες προτιμήσεις των χρηστών σε κάθε εκπομπή.

Επίσης, το τμήμα "Decider" φροντίζει ώστε κατά τη διάρκεια μιας εκπομπής να αναμεταδίδονται τραγούδια μιας κατηγορίας. Η κατηγορία αυτή καθορίζεται, όπως έχει ήδη αναφερθεί, από τη συνολική ψηφοφορία της προηγούμενης εκπομπής. Η κατηγορία των τραγουδιών που επιλέγονται προς αναμετάδοση μπορεί να αλλάξει μόνο στο τέλος κάθε εκπομπής.



Τέλος, το τμήμα "Decider" ελέγχει το αρχείο election.xml για διπλές εγγραφές στο τέλος κάθε χρονικού παράθυρου έτσι ώστε να μην επιλέγονται για αναμετάδοση τα ίδια τραγούδια περισσότερες από μια φορά στην ίδια εκπομπή.

Ακολουθώντας τις παραπάνω αρχές λειτουργίας, ο προτεινόμενος ραδιοφωνικός σταθμός στο Web παράγει μια συνεχή ροή τμημάτων δεδομένων που αντιστοιχεί στις απαιτήσεις που θέτουν οι ακροατές, με στόχο να προσεγγιστεί η αντίστοιχη διαδικασία που ακολουθείται σε ένα συμβατικό ραδιοφωνικό σταθμό.

### 3.6 Συμπεράσματα και ανοιχτά θέματα

Στο κεφάλαιο αυτό παρουσιάστηκε μια κατανομημένη μουσική ψηφιακή βιβλιοθήκη βασισμένη στο Web, ικανή να διαχειρίζεται αρχεία τύπου mp3. Το σύστημα αυτό βασίζεται σε ένα μοντέλο τριών στρωμάτων, για την αποθήκευση και ανάκτηση του υποκείμενου περιεχομένου. Το κίνητρο δημιουργίας της ψηφιακής βιβλιοθήκης είναι η έλλειψη κατάλληλων εργαλείων και εφαρμογών στο Web για την οργάνωση, διαχείριση και διάδοση αρχείων τύπου mp3 με έναν απλό και αποδοτικό τρόπο. Ταυτόχρονα, απαραίτητη απαίτηση που ικανοποιεί το προτεινόμενο σύστημα είναι ο σεβασμός των πνευματικών δικαιωμάτων των συντελεστών των δεδομένων που βρίσκονται αποθηκευμένα στην ψηφιακή βιβλιοθήκη.

Ετσι, λοιπόν, σε αντίθεση με τις υπάρχουσες εφαρμογές διαχείρισης αρχείων τύπου mp3 στο Web, όπου οι χρήστες έχουν τη δυνατότητα να μεταποθηκεύουν τα αρχεία αυτά στον προσωπικό τους σταθμό εργασίας, στο κεφάλαιο αυτό παρουσιάστηκε μια μέθοδος μετατροπής των αρχείων τύπου mp3 σε κανάλια ήχου συνεχούς ροής, τα οποία και αναμεταδίδονται στο Διαδίκτυο. Με αυτόν τον τρόπο, τα αρχεία αυτά παραμένουν σε ένα σημείο (δηλ. στον αποθηκευτικό χώρο), χωρίς να δημιουργούνται αντίγραφα τους στους σταθμούς εργασίας των χρηστών της ψηφιακής βιβλιοθήκης. Ετσι, περιορίζεται αισθητά η δυνατότητα μη εξουσιοδοτημένης εκμετάλλευσης των αντίστοιχων δεδομένων, ενώ ταυτόχρονα εξασφαλίζεται σε μεγάλο βαθμό η κατοχύρωση των πνευματικών δικαιωμάτων στα οποία υπόκεινται τα δεδομένα αυτά.

Η προτεινόμενη ψηφιακή βιβλιοθήκη μπορεί να χρησιμοποιηθεί από διάφορες εφαρμογές μουσικού περιεχομένου, κυμαινόμενες από ραδιοφωνικούς σταθμούς στο Web μέχρι πολύπλοκα βιβλιοθηκονομικά συστήματα, τα οποία διαχειρίζονται πολύ μεγάλο όγκο ηχητικού υλικού.

Σύμφωνα με τα συμπεράσματα που προκύπτουν από το σενάριο χρήσης της ψηφιακής βιβλιοθήκης στα πλαίσια ενός ραδιοφωνικού σταθμού στο Web, το προτεινόμενο μοντέλο έχει

τη δυνατότητα να διαχειρίζεται αποδοτικά μεγάλο αριθμό αρχείων τύπου mp3. Επίσης, η ανεξαρτησία πλατφόρμας καθώς και η εμμετάλλευση των δυνατοτήτων παρωμένων εφαρμογών όπως ο Web browser που χαρακτηρίζουν την προτεινόμενη μουσική ψηφιακή βιβλιοθήκη, προσδίδουν στο ραδιοφωνικό σταθμό το χαρακτηριστικό της μεγάλης διαθεσιμότητας για τους χρήστες του Web.

Αν και η κατανεμημένη φύση της προτεινόμενης αρχιτεκτονικής την καθιστά ικανή να διαχειρίζεται αποδοτικά μεγάλο αριθμό αρχείων τύπου mp3, εντούτοις προκύπτουν κάποια προβλήματα που έχουν να κάνουν με τους περιορισμούς που τίθενται από τη διαθεσιμότητα του δικτύου. Έτσι, ο αριθμός των χρηστών που μπορούν να είναι ταυτόχρονα συνδεδεμένοι στην ψηφιακή βιβλιοθήκη, εξαρτάται από τη χωρητικότητα του δικτύου πάνω στο οποίο στηρίζεται η εκάστοτε εφαρμογή. Η ραγδαία εξέλιξη δικτυακών πρωτοκόλλων στο Web όπως το πρωτόκολλο ATM over IP και το πρωτόκολλο IPv6, υπόσχονται καλύτερη υποστήριξη στην προτεινόμενη μουσική ψηφιακή βιβλιοθήκη καθώς και γενικότερα σε εφαρμογές που διαχειρίζονται πολυμεσικά δεδομένα στο Web.

## 4. Ιατρικές Ψηφιακές Βιβλιοθήκες

### 4.1 Εισαγωγή

Τα τελευταία χρόνια, τα Πληροφοριακά Συστήματα Υγείας - ΠΣΥ (Health Information System - HIS) που βασίζονται αποκλειστικά στον υπολογιστή κυριαρχούν στον Ευρωπαϊκό χώρο, καθώς ευνοούν την αποδοτική ανταλλαγή ιατρικών πληροφοριών. Έτσι, η παροχή υπηρεσιών υγείας συντελείται με μεγαλύτερη ευκολία, σε λιγότερο χρόνο και με λιγότερα έξοδα. Επίσης, όπως όλα σχεδόν τα νεότερα συστήματα, τα ΠΣΥ επηρεάζονται άμεσα από νέες τεχνολογίες όπως το Web. Οι σημερινοί επαγγελματίες του χώρου της Υγείας, επιθυμούν να εργάζονται σε ένα ομοιογενές πληροφοριακό περιβάλλον και να έχουν πρόσβαση σε ιατρικά δεδομένα από διαφορετικές γεωγραφικές τοποθεσίες με έναν προτυποποιημένο τρόπο. Ένα σύστημα αυτού του είδους πρέπει να είναι ικανό να εξυπηρετεί τους χρήστες, ακόμα και αν αυτοί βρίσκονται διασκορπισμένοι σε διαφορετικές τοποθεσίες. Από την άλλη μεριά, η επικοινωνία μέσω υπολογιστή εισάγει ορισμένα σημαντικά θέματα ασφάλειας, όπως εμπιστευτικότητα, ακεραιότητα δεδομένων και καταλογισμός ευθύνης, τα οποία, λόγω της πολυπλοκότητας αλλά και της ευαισθησίας που χαρακτηρίζουν τα ιατρικά δεδομένα, είναι απαραίτητα να αντιμετωπιστούν με τον πιο ασφαλή και κατάλληλο τρόπο.

Τα υπάρχοντα ΠΣΥ βασίζονται σε κλειστές αρχιτεκτονικές που χρησιμοποιούν συστήματα διαχείρισης βάσεων δεδομένων ειδικά σχεδιασμένα για την ελάχιστη εφαρμογή. Αν και τα συστήματα αυτά έχουν αποδειχτεί λειτουργικά και κατά συνέπεια έχουν κατακλύσει την αγορά, εντούτοις δε φαίνονται ικανά να προσαρμοστούν στις νέες απαιτήσεις που έχουν τεθεί

από τη ραγδαία εξάπλωση του Διαδικτύου. Οι χρήστες θέλουν πλέον να δουλεύουν κάτω από ένα ομοιογενές περιβάλλον με τρόπο σαφώς καθορισμένο και ασφαλή.

Τα κατακεντημένα συστήματα που βασίζονται στην τεχνολογία του Web ανταποκρίνονται σε αυτές τις απαιτήσεις και στηρίζουν εφαρμογές που διακρίνονται για την υψηλή ποιότητα εξυπηρέτησης των χρηστών τους (Quality of Service-QoS). Επιπλέον, η δεδομένη εξοικείωση των χρηστών με το Διαδίκτυο άρα και με τις εφαρμογές στο Web, εγγυάται τη γρήγορη προσαρμογή τους στο καινούργιο αυτό περιβάλλον εργασίας.

Ο μεγάλος και ανομοιογενής όγκος πληροφορίας που διακινείται στις εφαρμογές της Υγείας, καθιστά την τεχνολογία των ψηφιακών βιβλιοθηκών ως την πιο βιώσιμη επιλογή για την ανάπτυξη μιας αρχιτεκτονικής ικανής να διαχειρίζεται αποδοτικά αλλά και με ασφάλεια την πληροφορία αυτή. Πράγματι, οι ψηφιακές βιβλιοθήκες αποτελούν οργανωμένες συλλογές ψηφιακών δεδομένων πολλών και διαφορετικών μεταξύ τους μορφών. Έτσι, μέσα από μια ψηφιακή βιβλιοθήκη είναι δυνατή η διαχείριση των πολυδιάστατων ιατρικών δεδομένων, που μπορούν να κυμαίνονται από μια απλή ιατρική συνταγή μέχρι μια σύνθετη απεικόνιση ενός ανθρωπίνου οργάνου μέσα από το πρότυπο της Virtual Reality Markup Language (VRML).

Σε αυτό το κεφάλαιο, παρουσιάζεται η αρχιτεκτονική μιας ασφαλούς κατακεντημένης ιατρικής ψηφιακής βιβλιοθήκης. Η αρχιτεκτονική αυτή υλοποιεί την έννοια του Ηλεκτρονικού Φακέλλου Ασθενή - ΗΦΑ (Computer Patient Record - CPR) [OMG, 1998, Polemi, 1998, OMG, 1997a] που διαχειρίζεται προϋπάρχοντα ιατρικά συστήματα. Παρέχει επίσης στο χρήστη-επαγγελματία του χώρου, έναν ομοιογενή τρόπο αναζήτησης και επεξεργασίας ιατρικών πληροφοριών που είναι αποθηκευμένες σε διαφορετικά μεταξύ τους συστήματα. Χαρακτηριστικό γνώρισμα της αρχιτεκτονικής αποτελεί το γεγονός ότι βασίζεται στις προδιαγραφές CORBA (Common Object Request Broker Architecture) για κατακεντημένες εφαρμογές στο Web. Αυτή ακριβώς η κατακεντημένη φιλοσοφία της ψηφιακής βιβλιοθήκης απαιτεί την προστασία των απόρρητων πληροφοριών για κάθε ασθενή από απειλές που συναντώνται συχνά σε ένα κατακεντημένο περιβάλλον. Για το σκοπό αυτό, η περιγραφόμενη αρχιτεκτονική εμπλουτίζεται με ένα σύνολο από υπηρεσίες ασφάλειας που υποστηρίζονται από την CORBA και βασίζονται στην τεχνολογία των Εμπιστων Τρίτων Οντοτήτων - ΕΤΟ (Trusted Third Parties - TTPs) [Minton, 1997, Wreder, 1998, Golden, 1998, Prior, 1998, Iona, 1999, Blobel, 1998].

Το υπόλοιπο του κεφαλαίου έχει διαρθρωθεί ως εξής: Αρχικά, γίνεται μια ανασκόπηση της διεθνούς βιβλιογραφίας πάνω σε θέματα αρχιτεκτονικής παρεμφερών Πληροφοριακών Συστημάτων Υγείας. Ιδιαίτερη έμφαση δίνεται σε συστήματα που εστιάζονται στην ασφάλεια της παρεχόμενης λειτουργικότητας. Κατόπιν, περιγράφεται αναλυτικά η κατακεντημένη αρχιτεκτονική μιας ασφαλούς ιατρικής ψηφιακής βιβλιοθήκης. Η αρχιτεκτονική αυτή



αποτελείται από τρία στρώματα, τα οποία επικοινωνούν μεταξύ τους με βάση τις προδιαγραφές της CORBA για κατανεμημένες αρχιτεκτονικές. Στη συνέχεια, παρουσιάζονται οι απειλές ασφαλείας που ενδέχεται να προκύψουν στην ψηφιακή βιβλιοθήκη, καθώς και οι υπηρεσίες ασφαλείας που τις αντιμετωπίζουν. Οι υπηρεσίες αυτές βασίζονται σε μια ΕΤΟ για τη διαχείριση των δεδομένων ασφαλείας των χρηστών, όπως επίσης και στην τεχνολογία των έξυπνων καρτών, οι οποίες παρουσιάζονται σε επόμενη ενότητα. Τέλος, το κεφάλαιο ολοκληρώνεται με την παράθεση των συμπερασμάτων που προκύπτουν από την έρευνα πάνω στις ασφαλείς ιατρικές ψηφιακές βιβλιοθήκες.

## 4.2 Ανασκόπηση της Διεθνούς Βιβλιογραφίας

### 4.2.1 Ηλεκτρονικός Φάκελος Ασθενή (ΗΦΑ) βασισμένος σε συστήματα βάσεων δεδομένων προσβάσιμα από το Web

Υπάρχουν διάφορες προσεγγίσεις αναφορικά με την έννοια του ΗΦΑ που βασίζεται στο Web. Η έννοια αυτή έχει οριστεί από τον Prier et al. ως “*μία συλλογή από ετερογενή ηλεκτρονικά έγγραφα, τα οποία συνθέτουν το ιατρικό ιστορικό του κάθε ατόμου*”. Οι περισσότερες από τις προσεγγίσεις αυτές στηρίζονται σε συστήματα βάσεων δεδομένων, τα οποία είναι προσβάσιμα από το Web [Makris, 1998, Labreze, 1998]. Τέτοιου είδους προσεγγίσεις υποφέρουν από το γεγονός ότι το πρωτόκολλο HTTP, πάνω στο οποίο βασίζεται το Web, είναι ένα πρωτόκολλο που δεν έχει τη δυνατότητα καταγραφής της κατάστασής του (stateless). Κάθε αίτηση από ένα Web browser ακολουθείται από μια απάντηση από τον αντίστοιχο Web εξυπηρετητή. Αντίθετα, στις εφαρμογές βάσεων δεδομένων ένα σύνολο από αιτήσεις ορίζουν μια δοσοληψία (transaction), η οποία ικανοποιείται συνολικά από το αντίστοιχο σύστημα βάσεων δεδομένων [Lu, 1998].

Ακόμα ένα χαρακτηριστικό των συστημάτων που υλοποιούν τον ΗΦΑ μέσα από μια βάση δεδομένων, αποτελεί το γεγονός ότι το οπτικό αποτέλεσμα που παράγεται στην οθόνη του τελικού χρήστη είναι συνήθως δυσνόητο [Γρεγοί, 1999, Hamalka, 1998]. Αυτό οφείλεται στο ότι σε μια βάση δεδομένων αποθηκεύεται πληροφορία που έχει τη μορφή εγγράφου. Η αποθήκευση γίνεται όμως με πλήρως δομημένο τρόπο σύμφωνα με το σχεσιακό μοντέλο, με αποτέλεσμα να υπάρχει κίνδυνος απώλειας πληροφορίας όσον αφορά το νοηματικό περιεχόμενο της αρχικής πληροφορίας (ΗΦΑ) κατά την ανάκτηση και οπτικοποίησή (visualization) της.

#### **4.2.1.1 Το σύστημα Ιπποκράτης**

Χαρακτηριστικό παράδειγμα της προσέγγισης αυτής αποτελεί το σύστημα Ιπποκράτης [Varvatsiotis, 1998] το οποίο είναι μια πλατφόρμα ολοκλήρωσης εφαρμογών τηλεϊατρικής που καταλήγει σε ένα αυτόνομο σύστημα Τηλεϊατρικής. Η καρδιά του συστήματος αυτού αποτελείται από μια κεντρική βάση δεδομένων, η οποία διατηρεί πληροφορίες των χρηστών, διαγνωστικό υλικό (οπτικό υλικό τύπου DICOM (Digital Imaging and COmmunications in Medicine)) και δημογραφικά στοιχεία των ασθενών. Τα δεδομένα αυτά είναι προσβάσιμα και επιλέξιμα μέσω ενός συστήματος διεπαφής με το Web (Web interface). Συγκεκριμένα, μια εφαρμογή στον εξυπηρετητή αναλαμβάνει τη μετάφραση των πληροφοριών της βάσης δεδομένων σε μορφή αρχείων τύπου HTML, τα οποία εμφανίζονται στις οθόνες των χρηστών μέσα από κατάλληλο λογισμικό (Web browser).

Επίσης, το σύστημα Ιπποκράτης παρέχει τις εξής CSCW (Computer-Supported Cooperative Work) υπηρεσίες:

- Πολυμεσική (audio-video) συνομιλία
- Αλληλεπιδραστική ανταλλαγή πληροφοριών, η οποία περιλαμβάνει ανταλλαγή εικόνων, αρχείων και εγγράφων.
- Συνεργασία των χρηστών μέσω διαμοιραζόμενων κειμενογράφων, σχεδιαστικών εργαλείων και πινάκων ανακοινώσεων (bulletin boards).

Σύμφωνα με τους δημιουργούς του, ένα αρκετά σημαντικό πλεονέκτημα του συστήματος αυτού, αποτελεί η απομακρυσμένη γνωμάτευση (consulting): Οι ειδικευμένοι ιατροί που χρησιμοποιούν το σύστημα, μπορούν να στέλνουν τη διάγνωσή τους σε απομακρυσμένους δικτυακούς τόπους (Web sites). Με αυτόν τον τρόπο, ευνοείται η ακριβής διάγνωση και η ολοκληρωτική εκμετάλλευση των ιατρικών δεδομένων. Τέλος, μια κατάλληλη εφαρμογή ασφάλειας διαχειρίζεται τα κρυπτογραφικά κλειδιά που απαιτούνται για την κρυπτογράφηση των καναλιών επικοινωνίας του συστήματος με τους χρήστες του. Η εφαρμογή αυτή εξασφαλίζει επίσης την εμπιστευτικότητα των επικοινωνιών, αλλά και την ακεραιότητα των δεδομένων που ανταλλάσσονται.

#### **4.2.1.2 Το σύστημα PROMPT**

Ακόμα ένα σύστημα το οποίο παρέχει πρόσβαση στα δεδομένα του μέσω του Web, είναι το PROMPT [Labreze, 1998]. Σκοπός του συστήματος αυτού είναι η συντονισμένη επικοινωνία και συνεργασία όλων των επαγγελματιών του χώρου της Υγείας, από ιατρούς του

ιδιωτικού τομέα, δημόσιους ιατρούς και ιατρικό προσωπικό νοσοκομείων, μέχρι γενικότερο προσωπικό περίθαλψης, όπως νοσοκόμοι και τεχνικοί εργαστηρίων (ακτινολογικό, αξονικό, τομογράφος), έτσι ώστε να επιτευχθεί συνεχής και συνεπής περίθαλψη των ασθενών, σε όλα τα στάδια της διαδικασίας αυτής.

Κεντρικό σημείο αναφοράς του PROMPT, αποτελεί ο Ηλεκτρονικός Φάκελος Πολυμέσων Ασθενή (ΗΦΠΑ), ο οποίος βασίζεται σε ένα πληροφοριακό δίκτυο (Intranet) που ολοκληρώνει έναν μεγάλο αριθμό από εξυπηρετητές ιατρικών δεδομένων ασθενών, βάσεων γνώσεων (knowledge bases) και εξυπηρετητών τεκμηρίωσης (documentation servers). Οι βάσεις γνώσης είναι υπεύθυνες για τη δόμηση των δεδομένων, έτσι ώστε να είναι δυνατή η συσχέτιση των εγγράφων των ασθενών (ακτινογραφίες, πληρωμές, κ.α) με εξωτερικές πληροφορίες (οδηγίες χρήσης, χάρτες, κ.α), δημιουργώντας με αυτόν τον τρόπο ένα ολοκληρωμένο ΗΦΑ. Όλες οι πληροφορίες είναι προσβάσιμες από ένα κεντρικό σημείο πρόσβασης (single point of access), το οποίο με τη σειρά του είναι προσβάσιμο από το σταθμό εργασίας κάθε χρήστη. Έτσι, ο σταθμός εργασίας ενός χρήστη του PROMPT μπορεί να συνδεθεί σε ένα Intranet που του παρέχει τις κατάλληλες πληροφορίες, οι οποίες συνθέτουν τον ΗΦΠΑ του κάθε ασθενή.

Το σύστημα PROMPT δεν επιδιώκει να επηρεάσει τις δραστηριότητες της έρευνας στο χώρο της Υγείας, αλλά να αναγνωρίσει και να επικυρώσει τις τεχνολογίες που χρειάζεται να χρησιμοποιηθούν για την υλοποίηση των δραστηριοτήτων αυτών. Ιδιαίτερο βάρος έχει δοθεί στην ενθάρρυνση της συμμετοχής των χρηστών στον κύκλο ζωής του συστήματος αυτού, με σκοπό τη δημιουργία ενός συστήματος περίθαλψης προσανατολισμένο στους επαγγελματίες του χώρου της Υγείας.

#### **4.2.2 Χρήση ηλεκτρονικών εγγράφων για την αναπαράσταση του ΗΦΑ**

Για την αντιμετώπιση των προβλημάτων που προκύπτουν από την αναπαράσταση του ΗΦΑ μέσα από τη χρήση συστημάτων βάσεων δεδομένων, αναπτύχθηκαν προηγμένα συστήματα, τα οποία θέτουν ως βασική τους προτεραιότητα την εύρεση τρόπου αποδοτικής αποθήκευσης και διάδοσης των ιατρικών δεδομένων, καθώς και τον ορισμό των πολιτικών πρόσβασης στα δεδομένα αυτά.

Στο [Laforest, 2000] παρουσιάζεται ένας υβριδικός τρόπος υλοποίησης του ΗΦΑ μέσα από ένα σύστημα βάσεων δεδομένων για την αποθήκευση/ανάκτηση της αντίστοιχης πληροφορίας και του προτύπου της XML για τη δημιουργία των κατάλληλων συστημάτων διεπαφής. Σύμφωνα με τους συγγραφείς της εργασίας αυτής, η κύρια αιτία αποτυχίας καθιέρωσης των διαφόρων συστημάτων ΗΦΑ που έχουν εμφανιστεί κατά καιρούς στην αγορά, είναι η



επιμονή τους στη χρησιμοποίηση κάποιας φόρμας (form) για τη σύλληψη (capture) μορφοποιημένης πληροφορίας, η οποία βρίσκεται αποθηκευμένη στην υποκείμενη βάση δεδομένων. Η χρήση μιας βάσης δεδομένων είναι σημαντική, καθώς αποτελεί τον πιο αποδοτικό τρόπο ανάκτησης πληροφοριών. Παρολαυτά, ο χώρος της Υγείας δεν ευνοεί τη χρήση φόρμών, καθώς κάθε περίπτωση ασθενή είναι διαφορετική από τις υπόλοιπες και κάθε επαγγελματίας ιατρός έχει τη δικιά του μέθοδο εργασίας.

Η προτεινόμενη λύση υιοθετεί τη χρήση ημιδομημένων εγγράφων ως μια πιο βιώσιμη λύση για την αναπαράσταση της ιατρικής πληροφορίας. Βέβαια, τα εργαλεία που χρησιμοποιούνται για την ανάκτηση πληροφοριών από έγγραφα είναι λιγότερο αποδοτικά από τα αντίστοιχα εργαλεία που αναφέρονται σε βάσεις δεδομένων. Γι' αυτό λοιπόν, σύμφωνα με την περιγραφόμενη προσέγγιση, προτείνεται ένα σύστημα που συνδυάζει τα πλεονεκτήματα μιας βάσης δεδομένων και ενός συστήματος διαχείρισης εγγράφων. Συγκεκριμένα, τα έγγραφα χρησιμοποιούνται για να παράγουν τη ζητούμενη πληροφορία, ενώ η βάση δεδομένων χρησιμοποιείται απλώς για την αποθήκευση μέρους των δεδομένων που αναφέρονται στα έγγραφα και αφετέρου για την ανάκτηση των δεδομένων αυτών ή/και ολόκληρων των εγγράφων. Η εξόρυξη πληροφορίας δεν αφορά το σύνολο της πληροφορίας στο έγγραφο, αλλά την πληροφορία εκείνη που βρίσκεται αποθηκευμένη και στη βάση.

Η αρχιτεκτονική του προτεινόμενου συστήματος αποτελείται από τρία βασικά τμήματα:

- Το πρώτο τμήμα αναλαμβάνει τη διαχείριση και διακίνηση των εγγράφων που χρησιμοποιούνται. Έτσι λοιπόν, τα συστατικά του τμήματος αυτού είναι τα τελικά έγγραφα που παράγονται, τα αρχεία ορισμού (document definition files - dtd) πάνω στα οποία βασίζονται τα τελικά έγγραφα και οι ετικέτες (tags), οι οποίες συνθέτουν τα αρχεία ορισμού.
- Το δεύτερο τμήμα αναφέρεται σε ένα απλοποιημένο λεξικό της σχεσιακής βάσης δεδομένων που χρησιμοποιείται από το εν λόγω σύστημα.
- Το τρίτο τμήμα ονομάζεται τμήμα διασύνδεσης και αναλαμβάνει να φέρει σε επαφή τα άλλα δυο τμήματα. Περιέχει τα απαιτούμενα δεδομένα για τον καθορισμό της πληροφορίας προς αναζήτηση, τον καθορισμό των σημειωμένων (από ετικέτες) παραγράφων των εγγράφων μέσα στις οποίες θα πραγματοποιηθεί η αναζήτηση και τέλος, τον καθορισμό των πεδίων της βάσης που θα ενημερωθούν από την πληροφορία που θα ανακτηθεί.

Το κύριο προτέρημα του συστήματος αυτού, έγκειται στο γεγονός ότι οι επαγγελματίες του χώρου της Υγείας έχουν τη δυνατότητα να κατανοήσουν την παρεχόμενη πληροφορία μέσα από έγγραφα, τα οποία αποτελούν μια καλύτερη αναπαράσταση των ιατρικών δεδομένων σε



σχέση με το συμβατικό τρόπο που βασίζεται στη χρήση φόρμας. Επίσης, η ταχύτητα εξυπηρέτησης των αιτημάτων των χρηστών βρίσκεται σε υψηλά επίπεδα, λόγω της χρήσης ενός συστήματος βάσεων δεδομένων. Από την άλλη μεριά, το βασικό πρόβλημα που προκύπτει από τον υβριδικό αυτό τρόπο διαχείρισης ιατρικών δεδομένων εντοπίζεται στην πλεονάζουσα πληροφορία που πρέπει να συνηρθείται τόσο στη βάση δεδομένων όσο και στα έγγραφα. Έτσι, η ενημέρωση του συστήματος μπορεί να αποδειχτεί μια αρκετά επίπονη και χρονοβόρα διαδικασία, ενώ επίσης, τυχόν αναβάθμισή του πρέπει να γίνει πολύ προσεκτικά, έτσι ώστε να μη χαθεί ο συγχρονισμός μεταξύ του σχήματος της βάσης δεδομένων και των εγγράφων που αποτελούν το περιβάλλον διεπαφής των χρηστών με το σύστημα.

#### 4.2.3 Roaming Electronic HealthCare Record (R-EHCR)

Το σύστημα R-EHCR που προτάθηκε από τη Grammatikou et al. στο [Grammatikou, 2000], είναι μια ακόμα κατανεμημένη αρχιτεκτονική για παροχή υπηρεσιών πρωτεύουσας ιατρικής περίθαλψης, η οποία βασίζεται στην CORBA και στο πρότυπο της XML. Συγκεκριμένα, το σύστημα αυτό υλοποιεί την έννοια του ΗΦΑ σε ένα κατανεμημένο περιβάλλον ιατρικής περίθαλψης. Το πρόβλημα της κατανομής των χρηστών του συστήματος επιλύεται μέσα από την εισαγωγή της έννοιας του R-EHCR, της χρήσης τεχνικών προσωρινής τοπικής αποθήκευσης των δεδομένων και της βαθμωτής (incremental) ενημέρωσης ενός κεντρικού ευρετηρίου. Η επιλογή της κατανεμημένης προσέγγισης για τη δημιουργία ενός συστήματος παροχής ιατρικών υπηρεσιών, αποτελεί, σύμφωνα με τους συγγραφείς της εργασίας αυτής, την πιο κατάλληλη λύση. Έτσι, διευκολύνεται η διαχείριση μεγάλου όγκου ιατρικών δεδομένων μέσα από δίκτυα Wide Area Networks - WAN με βελτιωμένη απόδοση, ανοχή στα σφάλματα (fault tolerance) και ασφάλεια. Οι ανάγκες που καλύπτονται από την προτεινόμενη αρχιτεκτονική, είναι οι εξής:

- Διαφάνεια τοποθεσίας και ευρείας κλίμακας αποδοτική πρόσβαση στον ΗΦΑ
- Ασφαλής διακίνησης δεδομένων σε ιατρικούς φακέλους.

Σύμφωνα με την αρχιτεκτονική του R-EHCR, κάθε κέντρο παροχής υπηρεσιών πρωτεύουσας ιατρικής περίθαλψης (Primary HealthCare Center), είναι εφοδιασμένο με ένα τοπικό ΠΣΥ και συνδέεται με τα υπόλοιπα κέντρα μέσω ενός WAN δικτύου που βασίζεται στο πρωτόκολλο IP (Intranet). Κάθε ΠΣΥ υλοποιεί το συγκεκριμένο μοντέλο ροής εργασιών που του αντιστοιχεί και διατηρεί μια βάση δεδομένων για την αποθήκευση και διαχείριση τοπικά κατοχυρωμένων ΗΦΑ. Το βασικό χαρακτηριστικό της κατανεμημένης αυτής προσέγγισης, είναι η λειτουργικότητα που εισάγεται από την έννοια του R-EHCR: ο ΗΦΑ μεταναστεύει στο

κατάλληλο ΠΣΥ, έτσι ώστε να ακολουθεί τις επισκέψεις των ασθενών στα διαφορετικά μέρη που πιθανώς να επισκέπτονται για την ιατρική τους περίθαλψη.

Κάθε τοπικό ΠΣΥ φιλοξενεί μια βάση δεδομένων για τη διατήρηση των εικόνων που συνοδεύουν τον ΗΦΑ κάθε ασθενή. Επίσης, υπάρχει η οντότητα *Roaming Manager*, η οποία αφενός διατηρεί ένα κεντρικό ευρετήριο όλων των ΗΦΑ, αφετέρου λειτουργεί επικουρικά στα κατανεμημένα ΠΣΥ κατά τη διάρκεια εντοπισμού των διαφόρων ΗΦΑ των ασθενών. Το ευρετήριο αυτό διατηρεί την τοποθεσία όλων των πρωτευόντων αλλά και δευτερευόντων (σ.α. αντγράφων) ΗΦΑ, συμπεριλαμβανομένων και διαφόρων διαχειριστικών δεδομένων που τους αφορούν. Για λόγους βελτίωσης της απόδοσης του συστήματος, διατηρούνται αντίγραφα του κεντρικού ευρετηρίου σε κάθε τοπικό ΠΣΥ. Τέλος, ο συγχρονισμός των ευρετηρίων επιτυγχάνεται μέσω της βαθμωτής ενημέρωσης των διασκορπισμένων ευρετηρίων.

Η ασφάλεια του συστήματος εξασφαλίζεται από μια υπηρεσία πιστοποιητικών (*Certification Authority - CA*). Η υπηρεσία αυτή διατηρείται σε κεντρικό σημείο της αρχιτεκτονικής και αναλαμβάνει την έκδοση πιστοποιητικών αλλά και την αυθεντικοποίηση των χρηστών του συστήματος. Κάθε οντότητα της αρχιτεκτονικής αντιστοιχεί σε ένα ψηφιακό πιστοποιητικό, το οποίο συμβάλλει στην ασφάλή τους επικοινωνία. Η χρήση των ψηφιακών πιστοποιητικών εγγυάται επίσης τον καταλογοισμό ευθών των χρηστών (*non repudiation*), την ασφάλή πιστοποίηση των εμπλεκόμενων εξυπηρετητών και την αποφυγή της πλαστοπροσωπείας (*masquerading*).

#### **4.2.4 Η αρχιτεκτονική InterCare**

Το σύστημα InterCare [Lodder, 2000, Verderio, 2000] έχει ως στόχο την πραγματοποίηση και αξιολόγηση του κατανεμημένου ιδεατού Ηλεκτρονικού Χαρτοφύλακα Ασθενή - ΗΧΑ (*distributed virtual Electronic Patient Dossier - EPD*), χρησιμοποιώντας νέες τεχνολογίες. Το σύστημα αυτό υποστηρίζει την κοινή χρήση ιατρικών δεδομένων μεταξύ των διαφορετικών επαγγελματιών του χώρου της Υγείας που είναι υπεύθυνοι για κάθε ασθενή. Η παρεχόμενη λειτουργικότητα αποτελεί αναγκαία προϋπόθεση για την πραγματοποίηση του στόχου της διαμοιραζόμενης περίθαλψης (*shared care*).

Το κεντρικό τμήμα της αρχιτεκτονικής InterCare είναι ο εξυπηρετητής InterCare, ο οποίος λειτουργεί ως ενδιάμεσος μεσολαβητής μεταξύ των χρηστών του συστήματος και των διαφόρων τμημάτων στα οποία είναι αποθηκευμένα τα ιατρικά δεδομένα των ασθενών. Τα τμήματα αυτά έχουν προκύψει από ένα σύνολο συστημάτων στο χώρο της ιατρικής περίθαλψης, τα οποία προϋπήρχαν του συστήματος InterCare. Στην ουσία λοιπόν, το σύστημα αυτό

ολοκληρώνει τις υπηρεσίες που παρέχονται από τα προηγούμενα συστήματα. Ο εξυπηρετητής InterCare αποτελείται από τα εξής επί μέρους τμήματα:

- Εναν εξυπηρετητή ΗΧΑ
- Εναν εξυπηρετητή διαχείρισης έξυπνων καρτών
- Εναν διαχειριστή αναγνώρισης και αναφοράς ασθενών
- Εναν εξυπηρετητή ασφάλειας

Σε ένα πληροφοριακό περιβάλλον που αποτελείται από πολλούς και διαφορετικούς παροχείς υπηρεσιών και πληροφοριών, ο ΗΧΑ ενδέχεται να είναι κατανομημένος μεταξύ διαφορετικών πληροφοριακών συστημάτων, τα οποία να λειτουργούν σε διαφορετικές φυσικές τοποθεσίες. Για την ανάκτηση των ιατρικών δεδομένων από τους επαγγελματίες του χώρου της Υγείας καθώς και για την παρουσίασή τους με έναν ολοκληρωμένο τρόπο, πρέπει να πληρούνται οι ακόλουθες προδιαγραφές:

- Οι αιτήσεις για ανάκτηση ιατρικών δεδομένων πρέπει να ακολουθούνται από άμεσες αποκρίσεις του εμπλεκόμενου συστήματος, χωρίς σημαντικές καθυστερήσεις
- Η ποιότητα των ανακτούμενων ιατρικών δεδομένων πρέπει να είναι όσο το δυνατόν πιο υψηλή. Έτσι λοιπόν, τα δεδομένα πρέπει να ανακτώνται απευθείας από τον πρωτογενή χώρο φύλαξής τους και όχι από ενδιάμεσες βάσεις δεδομένων, που ενδεχομένως να συντηρούνται για λόγους καλύτερης απόδοσης του συστήματος. Σε περιπτώσεις όπου αυτό δεν είναι δυνατό, η ιδιότητα των δεδομένων πρέπει να μένει αναλλοίωτη
- Η δυνατότητα παρουσίασης των διαφορετικών τμημάτων του ΗΧΑ στην πλευρά των χρηστών με έναν ολοκληρωμένο τρόπο, υποδηλώνει την ύπαρξη ενός υποκείμενου μοντέλου αντικειμένων (object model), στο οποίο η δομή του ΗΧΑ είναι σαφώς καθορισμένη
- Οι κανόνες ασφάλειας του συστήματος πρέπει να τηρούνται κατά γράμμα. Ένας λόγος παραπάνω για την αυστηρή τήρηση των κανόνων αυτών, αποτελεί η ύπαρξη πολλών διαφορετικών συστημάτων που επικοινωνούν μεταξύ τους
- Για τη διατήρηση του αυτόνομου χαρακτήρα των υποκείμενων συστημάτων, δεν πρέπει να απαιτούνται σημαντικές τροποποιήσεις στα συστήματα αυτά.



#### 4.2.5 Το σύστημα τηλεϊατρικής Euromed-Euromed ETS

Το σύστημα τηλεϊατρικής Euromed [Polemi, 1998] έχει ως στόχο τη δημιουργία ενός ιδεατού ιατρικού κόσμου όπου όλη η ιατρική πληροφορία είναι προσπελάσιμη από οποιοδήποτε σημείο.

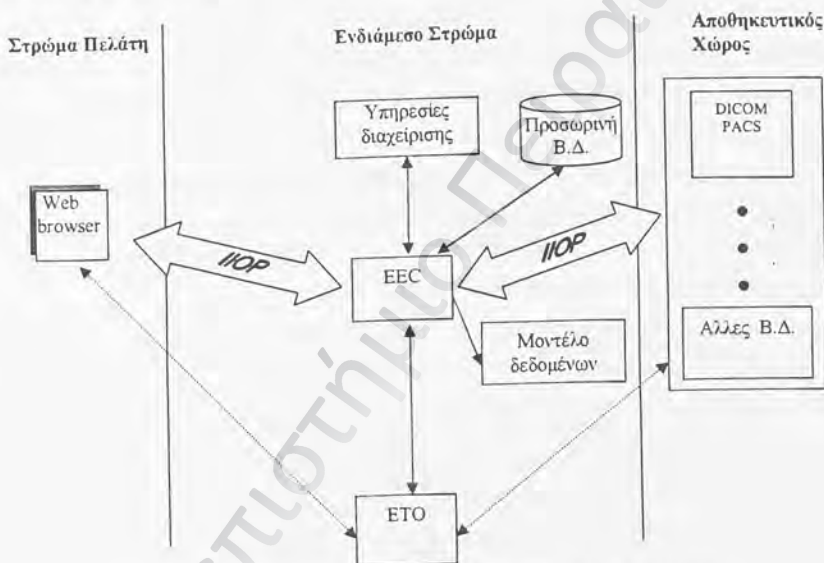
Το σύστημα Euromed βασίζεται σε διαδεδομένες τεχνολογίες του Διαδικτύου, έτσι ώστε να είναι απλή και εύκολη η επικοινωνία μεταξύ των δικτυακών τόπων που συμμετέχουν. Το δίκτυο του περιλαμβάνει ένα πλήθος δικτυακών τόπων, στους οποίους είναι αποθηκευμένη ιατρική πληροφορία σχετική με ασθενείς καθώς και εφαρμογές επεξεργασίας εικόνας όπως επίσης και αποθήκευσης δεδομένων. Ένας χρήστης του συστήματος που αναζητεί πληροφορίες με σκοπό να κάνει κάποια διάγνωση, εισάγεται στο σύστημα με τη βοήθεια ενός Web browser. Τα ιατρικά δεδομένα αποτελούνται από δισδιάστατες ή/και τρισδιάστατες εικόνες προερχόμενες από διάφορες ιατρικές συσκευές, βιοσήματα (biosignals), καρδιογραφήματα, αποτελέσματα εξετάσεων και αναφορές κειμένου. Τα διαθέσιμα δεδομένα για κάθε ασθενή δεν είναι υποχρεωτικό να βρίσκονται συγκεντρωμένα σε ένα δικτυακό τόπο. Σύμφωνα με τη λογική του Euromed, κάθε πρωτογενής πληροφορία αρχειοθετείται στο σημείο όπου αρχικά δημιουργήθηκε. Οι πληροφορίες αυτές σχολιάζονται με δείκτες που δείχνουν ο ένας στον άλλο και συνοψίζονται σε μια σελίδα τύπου HTML, μοναδική για κάθε ασθενή. Ο επαγγελματίας του χώρου της Υγείας που χειρίζεται το πρόγραμμα, εισάγει τα δεδομένα του αναφερόμενου ασθενή στην κατάλληλη εφαρμογή και τα επεξεργάζεται. Η γλώσσα προγραμματισμού Java έχει επιλεγεί από το σύστημα Euromed ως η τεχνολογία πάνω στην οποία βασίζεται η υλοποίηση των εμπλεκόμενων εφαρμογών στο Web.

Το σύστημα Euromed-ETS [Varvitsiotis, 1998, Spinellis 1999, Polemi, 1997] συμπληρώνει το σύστημα Euromed πάνω σε θέματα ασφάλειας για την κοινωνία της ιατρικής πληροφορίας. Η βασική απαίτηση που ικανοποιεί το σύστημα Euromed-ETS είναι η κατοχύρωση της ασφαλούς επικοινωνίας όλων των χρηστών του συστήματος μέσω του Web. Η υποδομή ασφάλειας του συστήματος βασίζεται στην τεχνολογία των ETO. Οι ETO φροντίζουν για την ασφαλή διακίνηση της κρίσιμης (όσον αφορά την ασφάλεια) ιατρικής πληροφορίας στο σύστημα. Έχουν μελετηθεί όλα τα θέματα που αφορούν τη χρήση των ETO σε εφαρμογές τηλεϊατρικής στο Web (διαχειριστικά, επιχειρησιακά, τεχνικά, λειτουργικά, οργανωτικά και νομικά). Τέλος, το δικτυακό πρωτόκολλο ασφάλειας που χρησιμοποιείται από το εν λόγω σύστημα είναι το Secure Socket Layer - SSL.



### 4.3 Αρχιτεκτονική ασφαλούς ιατρικής ψηφιακής βιβλιοθήκης

Η αρχιτεκτονική της ασφαλούς ιατρικής ψηφιακής βιβλιοθήκης που παρουσιάζεται σε αυτό το κεφάλαιο, βασίζεται σε μια αρχιτεκτονική CORBA τριών στρωμάτων, τυπική για κατανεμημένα συστήματα. Όπως φαίνεται και από το σχ. 1, ο κύριος όγκος των δεδομένων είναι αποθηκευμένος στο τρίτο στρώμα, που ονομάζεται αποθηκευτικός χώρος. Οι αιτήσεις για την ανάκτηση των δεδομένων από τους χρήστες του συστήματος που βρίσκονται στο στρώμα του πελάτη, φιλτράρονται και υπόκεινται σε επεξεργασία από το ενδιάμεσο στρώμα, το οποίο ενθυλακώνει και τους επιχειρησιακούς κανόνες (business rules) της εφαρμογής. Αξίζει να σημειωθεί ότι το λογισμικό που είναι εγκατεστημένο στο στρώμα του πελάτη δεν υπερβαίνει τις



Σχήμα 1: Η αρχιτεκτονική της ιατρικής ψηφιακής βιβλιοθήκης

προδιαγραφές ενός κοινού Web browser [OMG, 1997b], καθιστώντας έτσι τους σταθμούς εργασίας των χρηστών "λεπτούς πελάτες" (thin clients) [Golden, 1998, Dewire, 1998].

Το γεγονός ότι η προτεινόμενη ιατρική ψηφιακή βιβλιοθήκη ολοκληρώνει υπηρεσίες και δεδομένα που προέρχονται από προϋπάρχοντα ΠΣΥ, υποδηλώνει πως υπάρχει η πιθανότητα (ανάλογα με την πολιτική που ακολουθεί κάθε Τμήμα των εμπλεκόμενων οργανισμών) να χρειαστούν παραπάνω από ένα Object Request Broker - ORB [Iona, 1999, OMG, 1997a] για την υποστήριξη των διαφόρων δεξαμενών δεδομένων του αποθηκευτικού χώρου. Το ORB είναι στην ουσία ένας διάυλος λογισμικού (software bus), ο οποίος διαχειρίζεται αιτήσεις και

αποκρίσεις σε μια εφαρμογή βασισμένη στην CORBA. Κάθε ένα από τα διαφορετικά ORBs, επικοινωνεί με το κεντρικό ORB που βρίσκεται στο ενδιαμέσο στρώμα μέσω του πρωτοκόλλου Internet Inter-ORB Protocol- ΙΙΟΡ [Brando, 1995, Minton, 1997, OMG, 1998]. Το ΙΙΟΡ είναι μια συλλογή από αιτήσεις μηνυμάτων, τις οποίες μπορούν να απευθύνουν τα ORBs στο πρωτόκολλο TCP/IP [Brando, 1995].

#### **4.3.1 Αποθηκευτικός Χώρος**

Ο αποθηκευτικός χώρος αποτελείται από διάφορα αποθηκευτικά συστήματα τα οποία αντιστοιχίζονται σε αντικείμενα της CORBA μέσω της γλώσσας ορισμού συστήματος διεπαφής της (Interface Definition Language - IDL [Vinoski, 1996]). Για παράδειγμα, στην περίπτωση ενός συστήματος DICOM/PACS (Digital Imaging & COmmunications in Medicine/ Picture Archiving & Communication System), τα απαιτούμενα αντικείμενα της CORBA κατασκευάζονται από τα αντίστοιχα SOPs (Service Object Pairs) [Revet, 1997] και αποστέλλονται κατόπιν στο ενδιαμέσο στρώμα μέσω του ΙΙΟΡ πρωτοκόλλου επικοινωνίας της CORBA.

#### **4.3.2 Ενδιάμεσο στρώμα**

Το ενδιαμέσο στρώμα παρέχει την κύρια λειτουργικότητα της αρχιτεκτονικής αυτής. Η βασική αρμοδιότητα του ενδιαμέσου στρώματος είναι η ασφαλής διαχείριση των αιτήσεων των χρηστών προς τα ιατρικά δεδομένα, τα οποία είναι αποθηκευμένα σε διάφορα ΠΣΥ που συνθέτουν τον αποθηκευτικό χώρο. Ένας Εξυπηρετητής Εφαρμογών CORBA - EEC είναι υπεύθυνος για την παράδοση αντικειμένων ΗΦΑ, τα οποία βρίσκονται στον αποθηκευτικό χώρο, μέσω του ΙΙΟΡ, στους τελικούς χρήστες. Η νομοιογένεια των δεδομένων που οφείλεται αφενός στην ύπαρξη πολλών και διαφορετικών ειδών μορφοποιήσεων και αφετέρου στην εξάχτησή τους από διαφορετικές σημειολογίες (semantics), επιβάλλει την ύπαρξη ενός μοντέλου δεδομένων του χώρου (Domain Data Model - DDM). Το μοντέλο αυτό παρέχει στις εισερχόμενες αιτήσεις μια ενιαία περιγραφή των δεδομένων. Η ιατρική ψηφιακή βιβλιοθήκη που παρουσιάζεται σε αυτό το κεφάλαιο είναι σχεδιασμένη κατά τέτοιο τρόπο, ώστε να ικανοποιεί πολλές αιτήσεις, οι οποίες απευθύνονται σε πληροφορία που βρίσκεται στον αποθηκευτικό χώρο. Για την καλή απόδοση του συστήματος, μια βάση δεδομένων προσωρινών πληροφοριών (cache) συντηρείται στο ενδιαμέσο στρώμα. Έτσι, οι αιτήσεις πρόσβασης στα δεδομένα του αποθηκευτικού χώρου ελαχιστοποιούνται, με αποτέλεσμα να ικανοποιούνται πιο γρήγορα οι

αρχικές αιτήσεις των χρηστών. Το ενδιάμεσο στρώμα παρέχει: επίσης υπηρεσίες διαχείρισης των διαφόρων αντικειμένων του στρώματος αυτού, όπως η *υπηρεσία καταγραφής γεγονότων* (event monitoring service), η οποία καταγράφει την κατάσταση εκτέλεσης κάθε αντικειμένου και η *υπηρεσία διαχείρισης συνόδων* (session management service), η οποία επιβλέπει τις συνόδους μεταξύ των αντικειμένων-πελάτων και των αντικειμένων-στόχων (client and target objects).

Το γεγονός ότι η προτεινόμενη αρχιτεκτονική διαχειρίζεται δεδομένα που προέρχονται από συστήματα διαφορετικής φύσεως μεταξύ τους, υπαγορεύει πως τα ανακτούμενα από τον αποθηκευτικό χώρο δεδομένα δεν είναι κατάλληλα για απευθείας αποστολή στο στρώμα του πελάτη. Έτσι, ο εξυπηρετητής EEC χρησιμοποιείται για την προετοιμασία αποστολής των εικόνων υψηλής ανάλυσης του αποθηκευτικού χώρου προς τους τελικούς χρήστες. Πολλές φορές, για τη μείωση του φόρτου της δικτυακής κυκλοφορίας, οι εικόνες υψηλής ανάλυσης του αποθηκευτικού χώρου μετατρέπονται σε εικόνες μέτριας ανάλυσης, έτσι ώστε να μειωθεί το μέγεθός τους και να είναι πιο γρήγορη η μεταφορά τους στο στρώμα του πελάτη. Σε αυτήν την περίπτωση, ο EEC χρησιμεύει ως εξυπηρετητής οπτικοποίησης (visualization server).

Τα τμήματα του ενδιάμεσου στρώματος που έχουν περιγραφεί μέχρι τώρα, εξασφαλίζουν την καλή απόδοση και λειτουργικότητα του συστήματος. Η κατανομημένη φύση της ιατρικής ψηφιακής βιβλιοθήκης απαιτεί την τήρηση κάποιων κανόνων ασφαλείας για την προστασία των δεδομένων αυτών από πιθανούς κινδύνους που γενικότερα υπάρχουν στα κατανομημένα συστήματα [Council of Europe, 1997]. Στο χώρο της Υγείας, οι κανόνες αυτοί υλοποιούνται μέσα από την παροχή συγκεκριμένων υπηρεσιών ασφαλείας, οι οποίες ακολουθούν τις προδιαγραφές ISHTAR/SEISMED [Council of Europe, 1997, Barber, 1998, ISHTAR, 1997] και παρέχονται από έναν οργανισμό ΕΤΟ. Αν και τα θέματα που αφορούν την υποδομή και νομική υπόσταση των ΕΤΟ είναι πολύ σημαντικά, εντούτοις περαιτέρω αναφορά σε τέτοιου είδους θέματα κρίνεται μη σχετική με τη φιλοσοφία του παρόντος κεφαλαίου. Χρήσιμες πληροφορίες σχετικές με τέτοιου είδους θέματα υπάρχουν στο [Trusthealth I-II, 1999]. Στις επόμενες ενότητες γίνεται εκτενής αναφορά στις παρεχόμενες υπηρεσίες ασφαλείας, όπως επίσης και στις τεχνικές προδιαγραφές πάνω στις οποίες βασίζεται μια ΕΤΟ.

#### **4.3.3 Στρώμα πελάτη**

Η αλληλεπίδραση με τους τελικούς χρήστες του συστήματος επιτυγχάνεται μέσω της χρήσης ενός κεντρικού Web εξυπηρετητή (ενσωματωμένου στον EEC), ο οποίος είναι εγκατεστημένος στο ενδιάμεσο στρώμα. Ο Web εξυπηρετητής δέχεται και προωθεί τις αιτήσεις των χρηστών προς τα υπόλοιπα τμήματα του ενδιάμεσου στρώματος. Οι αιτήσεις αυτές



αποστέλλονται από το στρώμα του πελάτη μέσω των Web browsers, οι οποίοι είναι εφοδιασμένοι με δυνατότητα επεξεργασίας και εκτέλεσης τμημάτων κώδικα σε Java (Java applets), παρέχοντας έτσι αυξημένη λειτουργικότητα στους τελικούς χρήστες. Αντίστοιχα, ο κεντρικός Web εξυπηρετητής αναλαμβάνει την αποστολή των αποτελεσμάτων των αιτήσεων των χρηστών, όπως αυτά διαμορφώθηκαν από το υπόλοιπο σύστημα, προς τους Web browsers. Τέλος, η επικοινωνία με το ενδιάμεσο στρώμα επιτυγχάνεται μέσω του πρωτοκόλλου ΠΙΟΡ.

#### **4.4 Απειλές ασφάλειας**

Η αρχιτεκτονική της περιγραφόμενης ιατρικής ψηφιακής βιβλιοθήκης καθιστά διαθέσιμη την ιατρική πληροφορία στο σύνολο των επαγγελματιών του χώρου της Υγείας. Αυτό το μεγάλο προτέρημα αποτελεί ταυτόχρονα το σημαντικότερο σημείο προβληματισμού της αρχιτεκτονικής αυτής. Τα ιατρικά δεδομένα περιέχουν συνήθως ευαίσθητες πληροφορίες των αντίστοιχων ασθενών στους οποίους αναφέρονται. Έτσι, προκειμένου να προστατευτούν οι πληροφορίες αυτές, περιγράφεται ένα μοντέλο ασφάλειας βασισμένο στο πρωτόκολλο CSI-ECMA [Wilson, 1997], το οποίο υποστηρίζεται από την CORBA και εγγυάται την ικανοποίηση των προσδοκιών των χρηστών σε θέματα ασφάλειας. Το μοντέλο αυτό παρέχει ορισμένες υπηρεσίες ασφάλειας, οι οποίες αντιμετωπίζουν απειλές που συναντώνται συνήθως σε ένα κατανεμημένο πληροφοριακό σύστημα [Blobel, 1998, Choi, 1997, Wilson, 1997]. Οι απειλές αυτές είναι:

- μη εξουσιοδοτημένη αποκάλυψη μηνυμάτων (unauthorized disclosure of messages)
- πλαστοπροσωπία (impersonation)
- μη εξουσιοδοτημένη πρόσβαση (unauthorized access)
- μη εξουσιοδοτημένη μεταβολή της μεταδιδόμενης πληροφορίας (unauthorized modification of information in transit)
- καταλογισμός ευθύνης (non repudiation)
- άρνηση υπηρεσιών (denial of service)

*Μη εξουσιοδοτημένη αποκάλυψη μηνυμάτων:* Η παρακολούθηση των διαύλων επικοινωνίας αποτελεί συνηθισμένη απειλή στο Web, καθώς το περιβάλλον αυτό είναι ευάλωτο σε πολλά είδη παρακλούθησεων. Η κατάληψη συνόδων (session hijacking), η αποκάλυψη δρομολογήσεων δικτύου (network routing) και η παγίδευση γραμμών (wiretapping) βρίσκονται μεταξύ των πιο βασικών απειλών της επικοινωνίας στο Web.

*Πλαστοπροσωπία:* Μια από τις πιο σημαντικές απειλές ασφάλειας σε ένα περιβάλλον πολλών χρηστών είναι και η πλαστοπροσωπία. Σε αυτήν την περίπτωση, ο χρήστης που



εισέρχεται στο σύστημα, παριστάνει ότι είναι κάποιος άλλος. Αυτό επιτυγχάνεται κλέβοντας κωδικούς ή άλλου είδους δεδομένα ασφαλείας.

*Μη εξουσιοδοτημένη πρόσβαση:* Η ύπαρξη πολλών και διαφορετικών κατηγοριών χρηστών σε ένα περιβάλλον ιατρικής περιθαλψής επιβάλλει την ανάθεση διαφορετικών δικαιωμάτων πρόσβασης στα ιατρικά δεδομένα για τους διάφορους χρήστες. Χωρίς αυτόν τον περιορισμό, δεν μπορεί να αποφευχθεί η μη εξουσιοδοτημένη πρόσβαση σε ευαίσθητα δεδομένα, τα οποία δεν πρέπει να αποκαλυφθούν σε ορισμένους χρήστες.

*Μη εξουσιοδοτημένη μεταβολή της μεταδιδόμενης πληροφορίας:* Η προσάρτηση, διαγραφή, μεταβολή δεδομένων είναι εφικτή όταν τα κανάλια επικοινωνίας παρακολουθούνται ή όταν οι δικτυακές διευθύνσεις ελέγχονται.

*Καταλογισμός ευθύνης:* Η απειλή αυτή λαμβάνει χώρα όταν ο δημιουργός ή/και ο παραλήπτης ενός μηνύματος έχει τη δυνατότητα να αρνηθεί την πράξη της δημιουργίας ή/και παραλαβής του μηνύματος αυτού.

*Αρνηση υπηρεσιών:* Η δυνατότητα ενός συστήματος να χρησιμοποιεί αλγόριθμους κρυπτογράφησης για την προστασία της ανταλλασόμενης πληροφορίας, δεν είναι αρκετή για να το καταστήσει ασφαλές. Συχνά, τα καταμετρημένα συστήματα υπόκεινται σε επιθέσεις άρνησης υπηρεσιών. Σε μια τέτοια περίπτωση, οι πόροι του συστήματος δεν είναι διαθέσιμοι σε εξουσιοδοτημένους χρήστες εξαιτίας κακόβουλων επιθέσεων (malicious attacks), ακατάλληλου υλικού (hardware), ή λανθασμένης χρήσης του συστήματος από άλλους χρήστες.

## **4.5 Υπηρεσίες Ασφάλειας**

Για την αντιμετώπιση των απειλών που συναντώνται σε ένα καταμετρημένο περιβάλλον όπως είναι αυτό του χώρου της ιατρικής περιθαλψής, παρέχονται οι εξής υπηρεσίες ασφαλείας:

- Εμπιστευτικότητα (confidentiality)
- Αυθεντικοποίηση οντοτήτων (entity authentication)
- Έλεγχος πρόσβασης βασισμένος σε ρόλους χρηστών (role-based access control)
- Ακεραιότητα (integrity)
- Καταλογισμός ευθύνης (non repudiation)
- Διαθεσιμότητα (availability)

Η υπηρεσία εμπιστευτικότητας που εφαρμόζεται σε αυτήν την αρχιτεκτονική, φροντίζει έτσι ώστε τα δεδομένα που αποστέλλονται μέσω του ΠΟΡ να είναι ορατά μόνο σε εκείνους που έχουν εξουσιοδοτημένη πρόσβαση σε αυτά. Αυτό επιτυγχάνεται χρησιμοποιώντας ένα σύστημα κρυπτογράφησης δυο κλειδιών, του βασικού κλειδιού (basic key) και του κλειδιού διαλόγου (dialogue

key). Το βασικό κλειδί διανέμεται στα αντικείμενα που θέλουν να ανταλλάξουν πληροφορίες και χρησιμοποιείται για τη δημιουργία του κλειδιού διαλόγου, το οποίο δημιουργείται τοπικά σε κάθε υπολογιστή. Το κλειδί διαλόγου λειτουργεί ως *κλειδί συνόδου* (session key) για την ασφαλή ανταλλαγή μηνυμάτων μεταξύ των αντικειμένων. Το όφελος που προκύπτει από τη χρήση του αλγορίθμου αυτού, είναι πως το κλειδί συνόδου που κρυπτογραφεί τα μηνύματα, δεν κυκλοφορεί στο δίκτυο, οπότε δεν είναι διαθέσιμο προς υποκλοπή.

Για την αντιμετώπιση της απειλής της πλαστοπροσωπείας, οι χρήστες του συστήματος αυθεντικοποιούνται μέσω της υπηρεσίας *αυθεντικοποίησης οντοτήτων* βασισμένης στα *διαπιστευτήριά τους* (credentials). Τα διαπιστευτήρια είναι αντικείμενα που περιέχουν δεδομένα που χρησιμοποιούνται για την αυθεντικοποίηση των χρηστών, όπως επίσης και των διαφόρων οντοτήτων του συστήματος (principals). Συγκεκριμένα, μια ειδική οντότητα, ο *μεσολαβητής χρήστη* (user sponsor) αναλαμβάνει την απόκτηση των διαπιστευτηρίων ενός χρήστη ή μιας οντότητας του συστήματος, τα οποία και παρέχει σε μια άλλη, ειδική οντότητα, τον *αυθεντικοποιητή οντοτήτων* (principal authenticator), η οποία με τη σειρά της αυθεντικοποιεί τους χρήστες ή/και τις οντότητες. Ο αυθεντικοποιητής οντοτήτων επικοινωνεί με μια ΕΤΟ, για να διασταυρώσει τις πληροφορίες που έχει και τελικά να διαπιστώσει αν το αιτούμενο αντικείμενο δικαιούται να αυθεντικοποιηθεί στο σύστημα.

Η παραπάνω υπηρεσία αυθεντικοποίησης δεν είναι αρκετή για να εμποδίσει κάποιον αυθεντικοποιημένο χρήστη του συστήματος να αποκτήσει πρόσβαση σε δεδομένα που δεν πρέπει να οικειοποιηθεί. Για την αντιμετώπιση του προβλήματος αυτού, υποστηρίζεται μια *υπηρεσία ελέγχου πρόσβασης βασισμένη σε ρόλους*, η οποία ακολουθεί τις προδιαγραφές της CORBA για έλεγχο πρόσβασης επιπέδου 2 (Common Secure Interoperability - CSI - level 2) [OMG, 1997a]. Οι πληροφορίες ασφάλειας του κάθε χρήστη, συμπεριλαμβανομένου και του επιπέδου πρόσβασης (δηλ. ο ρόλος του) στο οποίο ανήκει, ενθυλακώνονται σε ένα *πιστοποιητικό δικαιωμάτων* (Privilege Attribute Certificate - PAC) το οποίο χρησιμοποιείται ως πληροφορία αυθεντικοποίησης και παρέχεται από μια ΕΤΟ. Η *ακεραιότητα* των δεδομένων εξασφαλίζεται μέσα από την εφαρμογή ενός αλγορίθμου ψηφιακής υπογραφής των αντίστοιχων μηνυμάτων. Σύμφωνα με τον αλγόριθμο αυτό, χρησιμοποιείται ένα σύστημα *κρυπτογράφησης δημοσίου κλειδιού* (public key encryption) για την ψηφιακή υπογραφή των μηνυμάτων αυτών. Η διαχείριση των απαραίτητων κλειδιών (ιδιωτικών-δημοσίων) που χρησιμοποιεί ο αλγόριθμος γίνεται από μια ΕΤΟ.

Για τον καταλογισμό ευθύνης στις οντότητες που έχουν παράγει ή αποδεχθεί κάποια μηνύματα, εφαρμόζεται η υπηρεσία *καταλογισμού ευθύνης*. Η υπηρεσία αυτή, όπως και η υπηρεσία που εξασφαλίζει την ακεραιότητα των δεδομένων, βασίζεται στην τεχνολογία των ψηφιακών

υπογραφών. Κάθε μήνυμα που ανταλλάσσεται μεταξύ δυο *c* ποτήτων, υπογράφεται ηλεκτρονικά με το *ιδιωτικό κλειδί* (private key) του αποστολέα. Επίσης, το μήνυμα συνοδεύει και μια *χρονική ετικέτα* (timestamp). Τόσο το κλειδί όσο και η ετικέτα, παρέχονται από μια ΕΤΟ.

Η διασφάλιση της συνεχούς διαθεσιμότητας ενός κατανεμημένου συστήματος όπως αυτό που περιγράφεται σε αυτό το κεφάλαιο, εξαρτάται σε μεγάλο βαθμό από τη σωστή διαχείριση των διαφόρων λειτουργικών συστημάτων (operating systems) του αποθηκευτικού χώρου και του Web εξυπηρετητή (ο οποίος χρησιμοποιείται από τον EEC) που διατηρείται στο ενδιάμεσο στρώμα. Ετσι, λοιπόν, οι εξουσιοδοτημένοι χρήστες πρέπει να έχουν *χωροχρονικά όρια* πρόσβασης στους πόρους της ιατρικής ψηφιακής βιβλιοθήκης.

Οι διάφορες υπηρεσίες ασφάλειας που έχουν περιγραφεί μέχρι τώρα, εφαρμόζονται σε συγκεκριμένα τμήματα της αρχιτεκτονικής της ιατρικής ψηφιακής βιβλιοθήκης, όπως φαίνεται στο σχ. 2. Οι χρήστες αυθεντικοποιούνται αρχικά στο σύστημα, ενώ ο χώρος πρόσβασής τους



Σχήμα 2: Οι υπηρεσίες ασφάλειας της ιατρικής ψηφιακής βιβλιοθήκης

καθορίζεται από τα αντίστοιχα πιστοποιητικά τους, τα οποία διαχειρίζεται μια ΕΤΟ. Κάθε κίνηση των χρηστών καταγράφεται από το σύστημα και δεν μπορεί να αμφισβητηθεί, καθώς συνοδεύεται από την ψηφιακή υπογραφή του δημιουργού της. Η ακεραιότητα και εμπιστευτικότητα των μηνυμάτων που ανταλλάσσονται μεταξύ των τριών στρωμάτων της αρχιτεκτονικής, διασφαλίζεται από τις παρεχόμενες υπηρεσίες ασφάλειας. Το κεντρικό τμήμα του ενδιάμεσου στρώματος, ο EEC, λειτουργεί ως πελάτης όταν επικοινωνεί με τα συστήματα του αποθηκευτικού χώρου, τα οποία σε αυτήν την περίπτωση λειτουργούν ως εξυπηρετητές. Οι ρόλοι ανλλάσσονται όταν τα δεδομένα του αποθηκευτικού χώρου αποστέλλονται στον EEC. Τα αντικείμενα της CORBA στο ενδιάμεσο στρώμα λειτουργούν ως πελάτες και αυθεντικοποιούνται



στα αντικείμενα στα οποία αναφέρονται, σύμφωνα με τις πληροφορίες που υπάρχουν στα πιστοποιητικά των χρηστών που τα επικαλέστηκαν.

Τέλος, για την ασφαλή διακίνηση των πληροφοριών μεταξύ των διαφόρων στρωμάτων της αρχιτεκτονικής αυτής, χρησιμοποιείται το πρωτόκολλο Secure Inter-ORB Protocol (SEC-IOP), το οποίο αναλαμβάνει την προστασία των μηνυμάτων που διακινούνται μέσω του ΠΙΟΡ. Το πρωτόκολλο αυτό ενεργοποιείται αυτόματα μόλις παραστεί ανάγκη, χωρίς να χρειάζεται να ενεργοποιηθεί άμεσα από κάποια υπηρεσία.

#### **4.6 Επαγγελματικές Κάρτες Υγείας - ΕΚΥ και ΕΤΟ**

Οι επαγγελματίες του χώρου της Υγείας δεν έχουν συγκεκριμένο σταθμό εργασίας. Αντίθετα, μετακινούνται συχνά και αναζητούν πληροφορίες σε διάφορα μέρη. Η χρήση ΕΚΥ (Health Professional Cards - HPCs) εμπλουτισμένων με λογισμικό κρυπτογραφικού περιεχομένου, ικανοποιούν τις απαιτήσεις για ευελιξία κίνησης (mobility), ασφάλεια και μυστικότητα (privacy) που χαρακτηρίζουν το χώρο της Υγείας. Σε αυτήν την κατεύθυνση, το Ευρωπαϊκό έργο TRUSTworthy HEALTH Telematics (Trusthealth I-II) [Trusthealth I-II, 1999], αναγνώρισε τη σημασία της χρήσης έξυπνων καρτών (smartcards) σε κατασκευασμένα συστήματα με υψηλές απαιτήσεις ασφάλειας, οπότε και πρότεινε την ΕΚΥ, μια έξυπνη κάρτα εξειδικευμένη για το χώρο της Υγείας. Όπως αναφέρεται και στο [Blobel, 1998], “η ΕΚΥ είναι μια έξυπνη κάρτα που αποτελείται από μια ενσωματωμένη μνήμη και ένα ηλεκτρονικό κύκλωμα για εσωτερική επεξεργασία δεδομένων, αποθήκευση συνταρτήσεων ασφάλειας, κρυπτογράφηση δεδομένων και ψηφιακή υπογραφή”.

Η ευαίσθητη φύση των δεδομένων που είναι αποθηκευμένα στις ΕΚΥ, συνιστά τη διαχείρισή τους από μια ΕΤΟ. Οι ΕΤΟ μπορούν να παίξουν πολύ σημαντικό ρόλο στην ασφάλεια, εκμετάλλευση και διάδοση των εφαρμογών της τηλεϊατρικής. Οργανωτικά, μια ΕΤΟ είναι ένας ανεξάρτητος και έμπιστος φορέας, ο οποίος δεν έχει άμεσο ενδιαφέρον για το περιεχόμενο της πληροφορίας που διαχειρίζεται. Ο σκοπός μιας ΕΤΟ είναι να παρέχει το επιθυμητό επίπεδο εμπιστοσύνης, το οποίο διασφαλίζει ότι οι εφαρμοσμένες τεχνικές ασφάλειας πληρούν τις αντίστοιχες προδιαγραφές που έχουν τεθεί για το σύστημα στο οποίο αναφέρονται. Για να μπορεί μια ΕΤΟ να χρησιμοποιήσει την ΕΚΥ, η ΕΤΟ αυτή πρέπει να υποστηρίζει ένα σύνολο από υπηρεσίες υποδομής, όπως εγγραφή, ονοματολογία, υπηρεσίες καταλόγου (directory services), διαχείριση πιστοποιητικών και διαχείριση κλειδιών ασφάλειας [Blobel, 1998]. Για το σκοπό αυτό, η ΕΤΟ που παρουσιάζεται σε αυτό το κεφάλαιο αποτελείται από συστατικά τα οποία είναι αφιερωμένα σε συγκεκριμένες λειτουργίες.



Ετσι λοιπόν, η αρχή έκδοσης ονομάτων (naming authority) χρησιμοποιείται για την έκδοση μοναδικών ονομάτων στους εξουσιοδοτημένους χρήστες του συστήματος, σύμφωνα με το πρότυπο X.500 [Chadwick, 1993]. Οι χρήστες συμπληρώνουν μια φόρμα εγγραφής, αφού πρώτα πιστοποιηθούν με ένα κατάλληλο έγγραφο (π.χ. διαβατήριο).

Η αρχή εγγραφής χρηστών (registration authority) αναλαμβάνει την εγγραφή των χρηστών στο σύστημα, επιτρέποντάς τους με αυτόν τον τρόπο την έκδοση μηνυμάτων προς το ενδιάμεσο στρώμα της αρχιτεκτονικής της ιατρικής ψηφιακής βιβλιοθήκης. Τα δικαιώματα πρόσβασης κάθε μηνύματος που δέχεται ο EEC μεταβιβάζονται στα αντίστοιχα μηνύματα που αποστέλλονται από τον EEC προς τον αποθηκευτικό χώρο.

Η διακίνηση πιστοποιητικών στην ψηφιακή βιβλιοθήκη υπαγορεύει την ύπαρξη μιας αρχής έκδοσης πιστοποιητικών (certification authority), της οποίας ο ρόλος είναι να εκδίδει και να αποθηκεύει πιστοποιητικά. Η δομή των πιστοποιητικών της περιγραφόμενης ιατρικής ψηφιακής βιβλιοθήκης ακολουθεί τις προδιαγραφές του προτύπου X.509 [Housley, 1999]. Τα διάφορα τμήματα του συστήματος πρέπει να επιβεβαιώσουν τα δικαιώματα πρόσβασης των οντοτήτων που ζητούν πρόσβαση στο περιεχόμενό τους. Αυτό επιτυγχάνεται εξετάζοντας τα δικαιώματα πρόσβασης (access privileges), όπως αυτά αναγράφονται στα πιστοποιητικά των οντοτήτων, τα οποία είναι προσπελάσιμα μέσω της υπηρεσίας δημοσίου καταλόγου (public directory service). Η υπηρεσία αυτή αποτελείται από έναν κατάλογο, ο οποίος χρησιμεύει ως χώρος αποθήκευσης πληροφοριών αναγνώρισης και αυθεντικοποίησης των χρηστών [Spinellis, 1999]. Επίσης, διατηρείται και μια λίστα ληγμένων πιστοποιητικών (certification revocation list), έτσι ώστε να ελέγχονται τα ληγμένα πιστοποιητικά.

Για την προστασία των πιστοποιητικών είναι απαραίτητη η χρήση ενός πλήθους κλειδιών, τα οποία τα διαχειρίζεται η αρχή δημιουργίας κλειδιών (key generation authority). Τα κλειδιά αυτά μπορούν να δημιουργηθούν τόσο από τους κατόχους τους όσο και από την ΕΤΟ. Σε κάθε περίπτωση, για την αρχικοποίηση της διαδικασίας δημιουργίας των κλειδιών χρησιμοποιείται μια ασφαλής γεννήτρια τυχαίων αριθμών. Τα ζεύγη κλειδιών που προκύπτουν (δημόσια-ιδιωτικά) είναι μοναδικά για κάθε χρήστη. Το ιδιωτικό κλειδί βρίσκεται στην κατοχή του χρήστη, ο οποίος είναι υπεύθυνος για την ασφαλή φύλαξή του. Το δημόσιο κλειδί βρίσκεται αποθηκευμένο στην ΕΤΟ, έτσι ώστε να είναι διαθέσιμο οποιαδήποτε στιγμή και αν ζητηθεί.

#### 4.7 Συμπεράσματα και ανοιχτά θέματα

Στο κεφάλαιο αυτό περιγράφηκε η αρχιτεκτονική μιας ασφαλούς ιατρικής ψηφιακής βιβλιοθήκης, ικανής να αντιμετωπίσει τους κινδύνους ασφάλειας που υπάρχουν σε ένα

κατανεμημένο περιβάλλον, όπως είναι αυτό του χώρου της Υγείας και ειδικότερα του τομέα της ιατρικής περίθαλψης. Βασικό χαρακτηριστικό της αρχιτεκτονικής είναι ότι βασίζεται σε μια ΕΤΟ για τη δημιουργία και διανομή των αναγκαίων πληροφοριών ασφάλειας. Ιδιαίτερο βάρος δόθηκε στην έρευνα σχετικά με τον τρόπο διαχείρισης των πληροφοριών αυτών. Συγκεκριμένα, προτάθηκε η ενσωμάτωση των δεδομένων ασφάλειας σε ΕΚΥ, έτσι ώστε να μπορούν οι χρήστες-κάτοχοι των καρτών αυτών να εισέρχονται στο σύστημα από περισσότερους από ένα σταθμούς εργασίας όσο το δυνατόν πιο εύκολα και γρήγορα. Η χρήση των έξυπνων καρτών αυξάνει την ασφάλεια της ιατρικής ψηφιακής βιβλιοθήκης, ενώ ταυτόχρονα καθιστά την αλληλεπίδραση με το σύστημα φιλικότερη προς τον χρήστη.

Παράλληλα με την τεχνολογική έρευνα που καθιστά βιώσιμη την προσέγγιση αυτή, είναι αναγκαία η ανάπτυξη ενός κοινού νομικού πλαισίου για τις χώρες της Ευρωπαϊκής Κοινότητας, που θα διασφαλίζει τη συμβατότητα των ΕΚΥ σε όλες τις εμπλεκόμενες χώρες. Επίσης, πρέπει να κατοχυρωθεί μια κοινή πολιτική διαχείρισης των καρτών αυτών, των ΕΤΟ, των κρυπτογραφικών συναρτήσεων, των κλειδιών και γενικότερα των προδιαγραφών όλων των δεδομένων ασφάλειας που χρησιμοποιούνται.

## 5. Ανάκτηση πληροφοριών σε συστήματα ψηφιακών βιβλιοθηκών βασισμένων στο Web

### 5.1 Εισαγωγή

Παραδοσιακά, οι εφαρμογές ανάκτησης πληροφοριών στο Web καθώς επίσης και στο χώρο των ψηφιακών βιβλιοθηκών, αξιολογούνται με βάση την ικανότητά τους να ανακτούν αποτελέσματα αναζήτησης σχετικά (*relevant*) με τα αντίστοιχα ερωτήματα των χρηστών. Για το σκοπό αυτό χρησιμοποιούνται δυο δείκτες που μετράνε την *ακρίβεια* (*precision*) και την *ανάκληση* (*recall*) [Salton, 1997] κάθε ενός από τα μέλη του συνόλου με τα αποτελέσματα της εκάστοτε αναζήτησης. Ο πρώτος δείκτης υπολογίζει την ακρίβεια με την οποία τα ανακτούμενα έγγραφα προσεγγίζουν νοηματικά το αρχικό ερώτημα του χρήστη, ενώ ο δεύτερος δείκτης αναφέρεται στο ποσοστό των σχετικών εγγράφων επί του συνόλου των εγγράφων που ανακλήθηκαν από την εφαρμογή [Pinkerton, 1994]. Παρολαυτά, όπως θα φανεί και από τις επόμενες ενότητες, μια αξιολόγηση αυτού του είδους μπορεί να αποδειχθεί ανακριβής και εσφαλμένη, καθώς υπάρχουν περιπτώσεις όπου οι χρήστες είναι δυσαρεστημένοι με τα αποτελέσματα των αναζητήσεών τους ακόμα και αν σε αυτά περιέχονται αναφορές σε πηγές αρχικά σχετικές με το ερώτημα που αποτέλεσε την αφετηρία της διαδικασίας ανάκτησής τους.

Σε μια εκτενή λίστα αποτελεσμάτων αναζήτησης, τα αντικείμενά της ταξινομούνται συνήθως σύμφωνα με τη σχετικότητα (*relevancy*) τους ως προς το αρχικό ερώτημα του χρήστη. Πολλές φορές ωστόσο, οι χρήστες δεν είναι σε θέση να περιγράψουν ακριβώς τι γυρεύουν όταν υποβάλλουν ένα ερώτημα. Σε αυτήν την περίπτωση, εισάγουν απλά μια γενική περιγραφή της



ζητούμενης πληροφορίας στη μηχανή αναζήτησης και στη συνέχεια επιλέγουν μετά από έρευνα τις πιο κατάλληλες πηγές μέσα από ένα σύνολο πολλών δεκάδων αποτελεσμάτων. Έτσι λοιπόν, όταν η μηχανή αναζήτησης επιστρέφει μεγάλες λίστες με αποτελέσματα, δεν υπάρχει η δυνατότητα διαχείρισης των αποτελεσμάτων αυτών (εκτός βέβαια από την πλοήγηση και τη μεταπήδηση από σελίδα σε σελίδα) με τέτοιο τρόπο, ώστε οι χρήστες να έχουν τη δυνατότητα να εντοπίζουν με ευκολία τις πιο κατάλληλες πηγές. Η παροχή μιας στατικής οπτικοποίησης των συνόλων που περιέχουν τα αποτελέσματα των αναζητήσεων ταξινομημένων σύμφωνα με τη σχετικότητά τους, αποτρέπει τους χρήστες από το να εντοπίζουν χρήσιμες πηγές, ιδιαίτερα μάλιστα όταν αυτές βρίσκονται στο τέλος των συνόλων αυτών.

Στο κεφάλαιο αυτό εισάγεται η έννοια των *αλληλεπιδραστικών αποτελεσμάτων αναζήτησης*, (interactive search results) όπως αυτή βρίσκει εφαρμογή σε υπάρχοντα συστήματα ανάκτησης πληροφοριών στο Web με χαρακτηριστικό παράδειγμα τις μηχανές αναζήτησης. Η προσέγγιση που ακολουθείται βασίζεται στη χρήση μεταδεδομένων, με συνέπεια να βρίσκει ιδιαίτερη εφαρμογή στον τομέα των ψηφιακών βιβλιοθηκών, όπου τα μεταδεδομένα χρησιμοποιούνται πολύ συχνά. Η λειτουργικότητα που προκύπτει παρέχεται από ένα τμήμα λογισμικού, το οποίο βρίσκεται σε ένα ενδιάμεσο στρώμα μεταξύ της μηχανής αναζήτησης και του περιβάλλοντος διεπαφής με τους χρήστες της. Σύμφωνα με αυτό το πλάνο, τα σύνολα με τα αποτελέσματα των αναζητήσεων θεωρούνται δυναμικά έγγραφα, τα οποία μπορεί να διαχειριστεί κανείς με διάφορους τρόπους, έτσι ώστε να αποδοθεί με μεγαλύτερη σαφήνεια το περιεχόμενό τους.

Επίσης, παρουσιάζονται επιχειρήματα που υποστηρίζουν την άποψη πως η σπουδαιότητα (importance) μιας πηγής δεν εξαρτάται μόνο από τη σχετικότητά της με το αρχικό ερώτημα. Αντίθετα, περιγράφεται μια μέθοδος σύμφωνα με την οποία λαμβάνεται υπόψη και η δημοτικότητα (popularity) μιας πηγής (κατά τη διάρκεια του κύκλου ζωής της) για τον υπολογισμό της σημαντικότητάς της για τον χρήστη. Ο δείκτης της δημοτικότητας υπολογίζεται από τον αριθμό προσπελάσεων των πηγών, ο οποίος υπολογίζεται από τις επισκέψεις των χρηστών στις πηγές αυτές μέσω των αντίστοιχων συνδέσμων που βρίσκονται στα σύνολα με τα αποτελέσματα αναζήτησης.

Ένα χαρακτηριστικό γνώρισμα των εφαρμογών που βασίζονται στην έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης, εκτός από το γεγονός ότι λειτουργούν ανεξάρτητα και επικουρικά προς τη μηχανή αναζήτησης, αποτελεί και το γεγονός ότι εκμεταλλεύονται τις παροχές και δυνατότητες των σταθμών εργασίας των χρηστών χωρίς να επιβαρύνουν περαιτέρω τους εξυπηρετητές που φιλοξενούν τις μηχανές αναζήτησης. Κάτι τέτοιο είναι ειρικό μέσα από τη χρήση του προτύπου της XML, όπως θα δειχθεί και στις επόμενες ενότητες.

Οι παραπάνω σκέψεις αποτέλεσαν τη βάση για την ανάπτυξη μιας πρότυπης εφαρμογής, η οποία προσφέρει στους χρήστες μιας μηχανής αναζήτησης τη δυνατότητα να αλληλεπιδρούν με τα αποτελέσματα των αναζητήσεών τους. Η εφαρμογή αυτή χρησιμοποιεί ως δεδομένα εισόδου τη λίστα αποτελεσμάτων μιας μηχανής αναζήτησης, η οποία αναφέρεται σε ένα υπάρχον σύστημα ψηφιακής βιβλιοθήκης στο Web. Επίσης, έχουν εξαχθεί χρήσιμα συμπεράσματα αναφορικά με τη λειτουργικότητα που παρέχεται από μια τέτοιου είδους εφαρμογή, μέσα από την αξιολόγηση της πρότυπης εφαρμογής, όπως αυτή παρουσιάζεται στο τέλος του παρόντος κεφαλαίου.

Το κεφάλαιο αυτό έχει διαρθρωθεί ως εξής: Στην αμέσως επόμενη ενότητα γίνεται μια ανασκόπηση της διεθνούς βιβλιογραφίας αναφορικά με την οργάνωση των αποτελεσμάτων αναζήτησης, όπως αυτά επιστρέφονται από διάφορα συστήματα ανάκτησης πληροφοριών. Κατόπιν, ακολουθεί η παρουσίαση της έννοιας των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης στο Web, ενώ περιγράφεται και η αρχιτεκτονική που πρέπει να διέπει τις εφαρμογές που στηρίζονται στην έννοια αυτή. Στην επόμενη ενότητα αναλύεται ο ρόλος των μεταδομένων για την εύρυθμη λειτουργία των αντίστοιχων αλληλεπιδραστικών εφαρμογών στο Web, ενώ ορίζεται και η σπουδαιότητα ως ένα εναλλακτικό κριτήριο ταξινόμησης των αποτελεσμάτων μιας αναζήτησης. Στη συνέχεια, αναφέρονται επιχειρήματα υπέρ της εκτέλεσης των αλληλεπιδραστικών εφαρμογών στην πλευρά του πελάτη, έτσι ώστε να μην επιβαρύνεται το δίκτυο αλλά και οι πόροι των συστημάτων που φιλοξενούν τις εφαρμογές αυτές. Για την αξιολόγηση της χρησιμότητας (usability) της περιγραφόμενης μεθόδου, παρουσιάζεται μια πρότυπη εφαρμογή, η οποία και αξιολογείται μέσα από μια κατάλληλα διαμορφωμένη μέθοδο αξιολόγησης που στηρίζεται σε ένα ερωτηματολόγιο. Τέλος, το κεφάλαιο ολοκληρώνεται με την παράθεση των συμπερασμάτων που προκύπτουν από την έρευνα πάνω στα αλληλεπιδραστικά αποτελέσματα αναζήτησης.

## **5.2 Ανασκόπηση της Διεθνούς βιβλιογραφίας**

Η οργάνωση των αναφορών των πηγών που επιστρέφουν οι εφαρμογές ανάκτησης πληροφοριών κατά τέτοιο τρόπο ώστε να μπορούν οι χρήστες να επιλέγουν με ευκολία τις πιο κατάλληλες γι αυτούς πηγές καθώς και η παρουσίασή τους έτσι ώστε να περιγράφεται όσο το δυνατόν πιο ξεκάθαρα το περιεχόμενό τους, αποτελεί εδώ και καιρό αντικείμενο έρευνας και συζήτησης πολλών ερευνητών ανά τον κόσμο. Η ραγδαία εξάπλωση των εφαρμογών ανάκτησης πληροφοριών στο Web και κυρίως των μηχανών αναζήτησης, καθιστά ακόμα πιο επιτακτική την ανάγκη εξεύρεσης μιας βιώσιμης λύσης που θα διευκολύνει τους χρήστες στον εντοπισμό των πιο σημαντικών πηγών, όπως αυτές αναφέρονται στα αποτελέσματα των αναζητήσεων.

### **5.2.1 Ταξινομημένες λίστες με αποτελέσματα αναζήτησης σε ψηφιακές βιβλιοθήκες**

Η πιο δημοφιλής μέθοδος οργάνωσης αναφορών σε πηγές όπως αυτές επιστρέφονται από μια εφαρμογή ανάκτησης πληροφοριών, είναι η ταξινομημένη λίστα, όπου οι αναφορές στις πηγές ταξινομούνται με βάση την πιθανότητα να είναι σχετικές με το αρχικό ερώτημα του χρήστη που προκάλεσε τη δημιουργία της λίστας. Η πιο ψηλά ταξινομημένη αναφορά είναι εκείνη που σχετίζεται περισσότερο με το αρχικό ερώτημα, η επόμενη αναφορά σχετίζεται λίγο λιγότερο και με αυτόν τον τρόπο ταξινομούνται και τα υπόλοιπα μέλη της λίστας. Ο τρόπος κατάταξης είναι απλός και αποδοτικός, ενώ οι χρήστες ακολουθούν τη σειρά κατάταξης καθώς εξετάζουν τα αποτελέσματα της αναζήτησης.

Οι μέθοδοι αξιολόγησης αυτής της προσέγγισης έχουν δείξει πως η ταξινομημένη λίστα συμπεριφέρεται αξιόπιστα και αποδοτικά σε πολλές περιπτώσεις [Harman, 1998, Harman, 1999]. Το βασικό μειονέκτημα της ταξινομημένης λίστας είναι το γεγονός ότι μόλις καθοριστεί η μέθοδος ταξινόμησης, η σειρά παράθεσης των αναφορών δεν μπορεί να αλλάξει. Έτσι, μια μικρή ασάφεια στη σύνταξη του ερωτήματος από τους χρήστες μπορεί να έχει ως αποτέλεσμα την ταξινόμηση χρησίμων αναφορών χαμηλά στη λίστα, καθιστώντας τις αντίστοιχες πηγές πολύ δύσκολο να εντοπιστούν. Ο μόνος τρόπος αναδιοργάνωσης των αποτελεσμάτων είναι να ξαναδιατυπωθεί το αρχικό ερώτημα και να ξαναχρησιμοποιηθεί η εφαρμογή ανάκτησης πληροφοριών.

Η διαδικασία της ανάδρασης σχετικότητας (relevance feedback [Rocchio, 1971]) αποτελεί μια επέκταση της ταξινομημένης λίστας. Σύμφωνα με αυτή τη μέθοδο, οι χρήστες έχουν τη δυνατότητα να σημειώνουν τις σχετικές πηγές κατά τη διάρκεια εξέτασης των αποτελεσμάτων και να προσαρμόζουν το αρχικό τους ερώτημα έτσι ώστε να αντιστοιχεί στις σημειωμένες πηγές. Κατόπιν, οι χρήστες απευθύνουν το προσαρμοσμένο ερώτημα στην εφαρμογή ανάκτησης πληροφοριών, έτσι ώστε να παραλάβουν τα αποτελέσματα της αναζήτησης ταξινομημένα διαφορετικά. Σύμφωνα με αναλύσεις που έχουν γίνει [Salton, 1990], η προσέγγιση αυτή βελτιώνει κατά πολύ την απόδοση των αντίστοιχων συστημάτων.

### **5.2.2 Ομαδοποίηση αποτελεσμάτων αναζήτησης σε ψηφιακές βιβλιοθήκες**

Μια από τις πιο σημαντικές προσπάθειες στο χώρο της οργάνωσης μεγάλων συλλογών από έγγραφα, έγινε από τον van Rijsbergen, ο οποίος παρατήρησε πως “*στενά συσχετιζόμενα έγγραφα τείνουν να συσχετίζονται με τα ίδια ερωτήματα*” [van Rijsbergen, 1979]. Ο στόχος λοιπόν



είναι να ομαδοποιούνται τα έγγραφα σε ομάδες (clusters) όπου η συσχέτιση είναι ισχυρή μεταξύ των μελών της ίδιας ομάδας και ισχνή μεταξύ μελών διαφορετικών ομάδων. Ο βαθμός συσχέτισης μεταξύ των εγγράφων μπορεί να υπολογιστεί με δυο τρόπους. Σύμφωνα με τον πρώτο, εντοπίζονται οι κοινές βιβλιογραφικές αναφορές μεταξύ των εγγράφων. Κατόπιν, τα έγγραφα που μοιράζονται τις ίδιες βιβλιογραφικές αναφορές, εντάσσονται στην ίδια ομάδα. Ο δεύτερος τρόπος υπαγορεύει πως κάθε έγγραφο συγκρίνεται με ένα σύνολο από λέξεις ευρετηρίου και ανάλογα με το ποσοστό συνάφειάς του με το σύνολο αυτό, εντάσσεται στην κατάλληλη ομάδα. Το ευρετήριο που χρησιμοποιείται για τις συγκρίσεις μπορεί να κατασκευαστεί είτε από ειδικευμένο προσωπικό ή αυτόματα [Leouski, 1996]. Αφού υπολογιστεί ο βαθμός συσχέτισης των εγγράφων, ακολουθείται κάποια μέθοδος ένταξης των εγγράφων σε ομάδες. Οι μέθοδοι αυτές μπορούν να χωριστούν σε δυο κατηγορίες: στις μη ιεραρχικές και τις ιεραρχικές. Τα συστήματα που στηρίζονται σε μη ιεραρχικές μεθόδους διαιρούν τη συλλογή με τα έγγραφα σε ένα πλήθος από ομάδες, οι οποίες περιέχουν έγγραφα σχετικά το ένα με το άλλο, όπου οι διαφορετικές ομάδες δεν σχετίζονται [Salton, 1971, Cutting, 1992]. Οι περισσότερες μη ιεραρχικές μέθοδοι έχουν χαμηλή πολυπλοκότητα που κυμαίνεται από  $O(N)$  σε  $O(N \log N)$ , όπου  $N$  είναι το πλήθος των εγγράφων του αποθηκευτικού χώρου. Από την άλλη μεριά, η δομή της συλλογής των εγγράφων εξαρτάται καθολικά από το  $K$ -κεντρικό διάγραμμα κάθε ομάδας εγγράφων ( $K$  centroid [Liu, 1998]) και συχνά δεν ανταποκρίνεται στην πραγματική διάρθρωση της υποκείμενης συλλογής [Leouski, 1996]. Τα συστήματα που χρησιμοποιούν ιεραρχικές μεθόδους, καταλήγουν να δημιουργούν δενδροειδείς δομές, όπου έγγραφα με μεγάλο βαθμό συσχέτισης βρίσκονται σε ομάδες που είναι μέλη μεγαλύτερων ομάδων, που με τη σειρά τους περιέχουν λιγότερο συναφή έγγραφα [Gennari, 1989]. Όπως και τα μη ιεραρχικά συστήματα, έτσι και τα ιεραρχικά αποδείχθηκαν αρκετά αποδοτικά. Παρολαυτά, έχουν μεγαλύτερο υπολογιστικό κόστος, ενώ επίσης είναι δύσκολο να καταλάβει κανείς το περιεχόμενο μιας ομάδας από την περιγραφή της [Leouski, 1996], καθώς αποτελείται από αδόμητα συγκεντρωμένους όρους.

Βασισμένοι στις παραπάνω παρατηρήσεις, ορισμένοι ερευνητές του χώρου της ανάκτησης πληροφοριών προσπάθησαν να οργανώσουν τα έγγραφα όπως αυτά επιστρέφονται από τις μηχανές αναζήτησης σε ομάδες εγγράφων παραπλήσιου περιεχομένου [Hearst, 1996] κατά τέτοιο τρόπο ώστε να διευκολυνθούν οι χρήστες στην επιλογή του πιο χρήσιμου για αυτούς εγγράφου. Δοκιμασμένες τεχνικές ομαδοποίησης εγγράφων χρησιμοποιούν μεγάλο αριθμό διαστάσεων για την αναπαράσταση των ομάδων αυτών αλλά και των εγγράφων που περιέχουν. Κάθε διάσταση αντιστοιχεί σε έναν άξονα, ο οποίος διαφοροποιεί τις ομάδες μεταξύ τους. Το κόστος που προκύπτει από την παροχή τόσο μεγάλης ακρίβειας, έγκνεται στην αντίστοιχα

μεγάλη πολυπλοκότητα των διαδικασιών αναζήτησης και ταξινόμησης, η οποία αυξάνει τους χρόνους απόκρισης των συστημάτων που στηρίζονται στις μεθόδους αυτές. Αντίστοιχα, όταν χρησιμοποιείται μικρός αριθμός διαστάσεων, η απεικόνιση των σχέσεων μεταξύ των εγγράφων είναι λιγότερο ακριβής, ενώ αντίθετα, η πρόσβαση σε αυτά επιτυγχάνεται πιο γρήγορα [Chalmers, 1992].

Μια μέθοδος που βασίζεται στις παραπάνω σκέψεις για την οργάνωση των αποτελεσμάτων μιας μηχανής αναζήτησης για μια βάση δεδομένων που αποτελείται από έγγραφα είναι και η μέθοδος Scatter/Gather [Hearst, 1996], η οποία αναπτύχθηκε στο κέντρο ερευνών Xerox Palo Alto. Η μέθοδος αυτή λαμβάνει ακόμα υπόψη της το γεγονός ότι *“τα ίδια σύνολα εγγράφων συμπεριφέρονται διαφορετικά (αναφορικά με την ταξινόμηση του) όταν διαφοροποιηθεί το νοηματικό πλαίσιο από το οποίο δημιουργήθηκε”*. Σε γενικές γραμμές, η λογική της μεθόδου αυτής είναι η ομαδοποίηση των αποτελεσμάτων των αναζητήσεων σε *k* ομάδες. Κατόπιν, δίνεται η ευκαιρία στους χρήστες να επιλέξουν κάποιες από τις ομάδες αυτές και να επανακαταθέσουν το σύνολο των περιεχομένων εγγράφων σε άλλες *k* ομάδες.

### 5.2.3 Οπτικοποίηση των αποτελεσμάτων

Εξίσου σημαντικό ρόλο στον εντοπισμό των κατάλληλων πηγών από τις εφαρμογές ανάκτησης πληροφοριών αποτελεί και ο τρόπος οπτικοποίησης των αποτελεσμάτων που επιστρέφουν. Τα τελευταία χρόνια, έχουν προταθεί διάφορες προσεγγίσεις πάνω σε αυτό το θέμα. Γενικά, τα περισσότερα συστήματα απαρτίζονται από πλοηγητικά συστήματα σχεδιασμένα κατά τέτοιο τρόπο ώστε να αποκαλύπτουν τις σχέσεις που υπάρχουν μεταξύ των διαφορετικών πηγών που απαρτίζουν το σύνολο των αποτελεσμάτων. Η μορφοποίηση της παρουσίασης διαφέρει από σύστημα σε σύστημα. Αρχικά συχνά, η οργάνωση των ανακτούμενων πηγών παρουσιάζεται με τη βοήθεια γραφικών. Τα έγγραφα, οι παράγραφοι και οι έννοιες παρουσιάζονται συνήθως ως σημεία ή αντικείμενα στο χώρο με τη σχετική τους θέση και απόσταση να υποδηλώνει το βαθμό συσχέτισής τους [Allan, 1997].

#### 5.2.3.1 Διδιάστατες απεικονίσεις

Το σύστημα Vibe [Dubin, 1995] χρησιμοποιεί μια μέθοδο διδιάστατης απεικόνισης, σύμφωνα με την οποία ο βαθμός συσχέτισης των περιεχομένων πηγών καταδεικνύεται από διαστάσεις που επιλέγονται από τους χρήστες. Οι πηγές τοποθετούνται στο κέντρο ενός κύκλου. Οι χρήστες έχουν τη δυνατότητα να εντοπίσουν οποιοδήποτε αριθμό όρων εντός των ορίων του

κύκλου. Επίσης, οι χρήστες μπορούν να μετατοπίσουν τη θέση των όρων και να μεταβάλλουν τη βαρύτητά τους, έτσι ώστε να κατανοήσουν καλύτερα τις σχέσεις μεταξύ των πηγών.

Ενα ακόμα σύστημα διδιάστατης οπτικοποίησης των αποτελεσμάτων όπως αυτά επιστρέφονται από μια εφαρμογή ανάκτησης πληροφοριών, προτείνεται από τον Shneiderman et al. στο [Schneiderman, 1995]. Συγκεκριμένα, στην εργασία αυτή, παρουσιάζεται μια μεθοδολογία σύμφωνα με την οποία τα αποτελέσματα μιας αναζήτησης αναπαρίστανται κατά τέτοιο τρόπο ώστε να συνδυάζονται τα προτερήματα της ιεραρχικής πλοήγησης (hierarchical browsing) με τα προτερήματα της διδιάστατης απεικόνισης. Στους χρήστες παρουσιάζεται μια περιήληψη από έγχρωμες κουκίδες ή γραφήματα-στίβες (bar-charts) τακτοποιημένα σε ένα πλέγμα και οργανωμένα ανά κατηγορίες. Παρέχεται επίσης και δυνατότητα μεγέθυνσης σε επιλεγμένες κατηγορίες καθώς και μετάβασης σε διαφορετική κατηγορία. Το βασικότερο πλεονέκτημα αυτής της μεθόδου είναι η παράθεση πολλών πηγών ταυτόχρονα στην οθόνη των χρηστών.

### 5.2.3.2 Τριδιάστατες απεικονίσεις

Η εξέλιξη των υπολογιστών και ιδιαίτερα η ραγδαία βελτίωσή τους στην υποστήριξη τριδιάστατων γραφικών, ώθησε πολλούς ερευνητές να ασχοληθούν με την αναπαράσταση των σχέσεων μεταξύ εγγράφων διαφόρων μορφών στον τριδιάστατο χώρο. Το σύστημα Lyberworld [Hemphie, 1994] είναι μια εξέλιξη του συστήματος Vibe, το οποίο αναφέρθηκε νωρίτερα. Σύμφωνα με το σύστημα αυτό, οι πηγές που αποτελούν τα αποτελέσματα των αναζητήσεων αναπαρίστανται σύμφωνα με το τριδιάστατο μοντέλο. Σε αυτή την περίπτωση, οι χρήστες καλούνται να επιλέξουν όρους, οι οποίοι τοποθετούνται στην επιφάνεια μιας σφαίρας. Η επιπλέον διάσταση επιτρέπει στους χρήστες να εντοπίζουν τις διαφορές μεταξύ των πηγών πιο έγκαιρα και με μεγαλύτερη σιγουριά.

Το σύστημα Bead [Chalmers, 1992] χρησιμοποιεί έναν αλγόριθμο κατασκευής γραφήματος, ο οποίος ονομάζεται spring embedding για την τοποθέτηση πολυδιάστατων αντικειμένων σε περιβάλλον λιγότερων διαστάσεων. Σύμφωνα με τη μεθοδολογία αυτού του συστήματος, οι επιστρεφόμενες αναφορές απεικονίζονται στο τριδιάστατο χώρο λαμβάνοντας υπόψη την ομοιότητα μεταξύ των πηγών. Δεν αποτελεί ερευνητικό στόχο του συστήματος ο διαχωρισμός μεταξύ σχετικών και μη πηγών. Ενα ακόμα αδύνατο σημείο του συστήματος Bead αποτελεί το γεγονός ότι έχει σχεδιαστεί κατά τέτοιο τρόπο ώστε να χειρίζεται αναφορές πολύ μικρού μεγέθους. Στην ουσία, η βάση δεδομένων του Bead αποτελείται από βιβλιογραφικές εγγραφές αναπαριστώμενες από λέξεις-κλειδιά, οι οποίες έχουν καταχωρηθεί από ειδικευμένο προσωπικό. Μια παρόμοια προσέγγιση για την οπτικοποίηση αποτελεσμάτων υιοθετήθηκε από



τους Leuski & Allan στο [Leuski, 1998], οι οποίοι επέκτειναν τη μεθοδολογία του συστήματος Bead, έτσι ώστε να μπορεί να εφαρμοστεί σε ολόκληρωμένα έγγραφα πλήρους μεγέθους.

#### **5.2.4 Συστήματα επεξεργασίας αποτελεσμάτων αναζήτησης στο Web**

Οι λύσεις που έχουν παρατεθεί μέχρι τώρα, αν και είναι αρκετά δημοφιλείς στο χώρο των παραδοσιακών ψηφιακών βιβλιοθηκών, εντούτοις δε μπορούν να εφαρμοστούν με την ίδια ευκολία στον ευρύτερο χώρο του Web και ειδικότερα στις ψηφιακές βιβλιοθήκες που βασίζονται στον Web browser για να φιλοξενήσουν το περιβάλλον διεπαφής με τους χρήστες. Το γεγονός αυτό οφείλεται στο ότι: α) τα περισσότερα από αυτά τα συστήματα απαιτούν πολύπλοκο λογισμικό επεξεργασίας γραφικών στην πλευρά του χρήστη, β) είναι σχεδιασμένα για να εξυπηρετούν μικρό όγκο δεδομένων και παρουσιάζουν προβλήματα όταν επεκτείνονται σε ολόκληρο το Web, γ) απαιτούν εκτεταμένη χρήση των πηγών των υπολογιστών στους οποίους εκτελούνται και δ) συνδυάζουν όλα τα παραπάνω [Berenci, 1999]. Οι ιδιαίτερες απαιτήσεις της ανάκτησης πληροφοριών στο Web όπως επίσης και η ανάγκη να απεμπλακεί η διαδικασία της επεξεργασίας των αποτελεσμάτων μιας αναζήτησης από το μηχανισμό εντοπισμού και ανάκτησης των πηγών που συνθέτουν τα αποτελέσματα αυτά, αποτελούν μείζοντα θέματα της ανάκτησης πληροφοριών στα οποία δυστυχώς δεν έχουν αναφερθεί επαρκώς οι ερευνητές του χώρου αυτού [Theng, 1999].

Σε αυτήν την κατεύθυνση, οι Zamir και Etzioni [Zamir, 1998] πρότειναν μια μέθοδο ανακατάταξης και σημειώσης (tagging) των πιο σχετικών από τις πηγές που επιστρέφουν οι μηχανές αναζήτησης, βασισμένοι σε έναν αλγόριθμο ομαδοποίησης. Ουσιαστικά, η συνεισφορά της εργασία αυτής έγκειται στη συνδυασμένη χρήση αλγορίθμων ταξινόμησης λιστών με μεθόδους ομαδοποίησης πηγών, όπως αυτές εμφανίζονται στα αποτελέσματα μιας αναζήτησης στο Web. Οι χρήστες έχουν τη δυνατότητα να σημειώσουν τις πιο σχετικές πηγές και να εφαρμόσουν την παρεχόμενη μέθοδο, έτσι ώστε να εντοπίσουν πηγές που δεν βρίσκονται στις πρώτες θέσεις της αρχικής λίστας με τα αποτελέσματα της αναζήτησης.

Παραμφερούς φιλοσοφίας είναι και το σύστημα VIEWER [Berenci, 1999], το οποίο παρέχει ένα ενδιάμεσο επίπεδο μεταξύ της φάσης σχηματισμού και επεξεργασίας του ερωτήματος και της φάσης αναπαράστασης των επιστρεφόμενων αποτελεσμάτων. Σύμφωνα με τη μεθοδολογία που ακολουθείται στο σύστημα αυτό, οι χρήστες έχουν τη δυνατότητα να αλληλεπιδράσουν δυναμικά με τα αποτελέσματα αυτά, εφαρμόζοντας κατάλληλους χειρισμούς στις διάφορες "όψεις", όπου η όψη ορίζεται ως το υποσύνολο των ανακτούμενων πηγών που περιέχουν ένα συγκεκριμένο υποσύνολο των όρων που απαρτίζουν το αρχικό ερώτημα. Για τη

δημιουργία μιας νέας όψης των αποτελεσμάτων, οι χρήστες πρέπει να μετατρέψουν το αρχικό τους ερώτημα και να καλέσουν εκ νέου τη μηχανή αναζήτησης.

### 5.2.5 Μηχανές αναζήτησης στο Web και παρουσίαση αποτελεσμάτων

Εκτός από τις προαναφερθείσες αναφορές στη βιβλιογραφία της ανάκτησης πληροφοριών, ένας σημαντικός αριθμός εμπορικών εφαρμογών ανάκτησης πληροφοριών ασχολείται ενεργά με τα προβλήματα που προκύπτουν στη διαχείριση και παρουσίαση των αποτελεσμάτων αναζήτησης [The Best Search Engines, 2002]. Παρολαυτά, το μεγαλύτερο μέρος της έρευνας στον τομέα των εμπορικών μηχανών αναζήτησης δε συνοδεύεται δυστυχώς από αντίστοιχη δημοσιοποίηση των τεχνικών προδιαγραφών πάνω στις οποίες βασίζονται οι μηχανές αυτές. Ως εκ τούτου, η τεχνολογία των εμπορικών μηχανών αναζήτησης στο Web παραμένει σε μεγάλο ποσοστό κρυφή από το ευρύ κοινό [Page, 1998a].

Εξάιρεση στον κανόνα αποτελεί η μηχανή αναζήτησης Google [Page, 1998a], για την οποία έχουν δημοσιευτεί οι τεχνικές της προδιαγραφές στο Web. Η διαδικτυακή αυτή εφαρμογή έχει αναπτυχθεί από τον Πανεπιστήμιο Stanford. Σύμφωνα με τη λογική λειτουργίας του Google, τα αποτελέσματα κάθε αναζήτησης παρουσιάζονται σε μορφή λίστας, η οποία είναι ταξινομημένη με βάση τις ακόλουθες αρχές: α) Η κατάταξη μιας πηγής είναι υψηλότερη όσο πιο πολλοί είναι οι σύνδεσμοι που δείχνουν προς αυτή την πηγή. β) Σύνδεσμοι που προέρχονται από πηγές οι οποίες είναι ταξινομημένες σε υψηλές θέσεις, έχουν μεγαλύτερο βάρος από συνδέσμους που προέρχονται από πηγές χαμηλότερης ταξινόμησης. Ειδικότερα, σε κάθε πηγή που αναφέρεται στη λίστα με τα αποτελέσματα αντιστοιχεί ένα βάρος, το οποίο υπολογίζεται σε ένα μεγάλο ποσοστό σύμφωνα με τον αλγόριθμο Pagerank [Page, 1998b]. Ο αλγόριθμος αυτός υπολογίζει ουσιαστικά το βάρος (δηλ. πόσο σημαντική είναι) κάθε πηγής σύμφωνα με τις πληροφορίες που λαμβάνονται από τη δομή συνδέσμων (link structure) στο Web. Στην ουσία, η δομή αυτή είναι ένα γράφημα, το οποίο αποτελείται από κόμβους (nodes) και συνδέσμους (links). Κάθε πηγή φιλοξενεί έναν αριθμό συνδέσμων που δείχνουν σε άλλες πηγές (εμπροσθοσύνδεσμοι - forward links), ενώ ταυτόχρονα αποτελεί την κατάληξη συνδέσμων που δείχνουν σε αυτή, ξεκινώντας από άλλες πηγές (οπισθοσύνδεσμοι - backlinks). Σύμφωνα με τον αλγόριθμο Pagerank, μια πηγή έχει μεγάλο βάρος αν το άθροισμα των βαρών των οπισθοσυνδέσμων της είναι μεγάλο. Ο ορισμός αυτός καλύπτει τις εξής περιπτώσεις: α) μια πηγή έχει πολλούς οπισθοσυνδέσμους, β) μια πηγή έχει λίγους οπισθοσυνδέσμους από πηγές με μεγάλα βάρη. Η καθιέρωση της μηχανής αναζήτησης Google ως μια από τις πιο δημοφιλείς, αξιόπιστες και αποδοτικές εφαρμογές ανάκτησης πληροφοριών στο Web, καταδεικνύει εκτός

των άλλων, την επιτυχία του εναλλακτικού τρόπου ταξινόμησης των αποτελεσμάτων που εισάγει ο αλγόριθμος Pagerank.

### **5.2.6 Αποτελέσματα αναζήτησης και ψηφιακές βιβλιοθήκες βασισμένες στο Web**

Στο χώρο των ψηφιακών βιβλιοθηκών βασισμένων στο Web, οι εφαρμογές ResearchIndex [Lawrence, 1999] και Open Journal Project [Hitchcock, 2000] έχουν εισάγει τη δική τους μέθοδο υπολογισμού της σημασίας των πηγών, έτσι ώστε να προσφέρουν πιο αποδοτική ταξινόμηση για τους χρήστες. Η ψηφιακή βιβλιοθήκη ResearchIndex, η οποία περιέχει υλικό που ανήκει στην ακαδημαϊκή βιβλιογραφία (scientific literature), υπολογίζει τη σπουδαιότητα των εγγράφων της βασισμένη στο σύστημα Autonomous Citation Indexing (ACI). Το σύστημα αυτό διατηρεί ένα ευρετήριο, το οποίο καταλογογραφεί αυτόματα τις αναφορές που βρίσκονται στα έγγραφα συνδέοντας τις αναφορές αυτές με τα αντίστοιχα έγγραφα. Έτσι, τα έγγραφα που συνθέτουν τη λίστα με τα αποτελέσματα κάθε αναζήτησης ταξινομούνται σύμφωνα με τις αναφορές που έχουν. Επίσης, η ψηφιακή βιβλιοθήκη ResearchIndex λαμβάνει υπόψη το χρονικό διάστημα που έχει παρέλθει από τη στιγμή που δημοσιεύτηκε το κάθε έγγραφο [Lawrence, 1999]. Παρόμοια προσέγγιση (αν και όχι τόσο εφαρμόσιμη όσο αυτή που εισάγει η ψηφιακή βιβλιοθήκη ResearchIndex, καθώς βασίζεται σε συγκεκριμένη μορφοποίηση εγγράφων) ακολουθείται από το Open Journal Project. Σύμφωνα με το σύστημα αυτό, ορίζεται μια υπηρεσία συνδέσμων, η οποία συσχετίζει κάθε εργασία με τις υπόλοιπες μέσω των αναφορών που υπάρχουν σε αυτές. Οι σύνδεσμοι παρέχονται από μια βάση δεδομένων συνδέσμων. Όπως και στο σύστημα ResearchIndex, δεν παρέχεται αλληλεπίδραση με τους χρήστες.

Η ψηφιακή βιβλιοθήκη τεχνικών αναφορών Πληροφορικής NCSTRL [NCSTRL, 2002], βασίζεται στο Web και παρέχει τη δυνατότητα στους χρήστες να επιλέξουν τον τρόπο με τον οποίο θα καταταχθούν τα αποτελέσματα της εκάστοτε αναζήτησης. Συγκεκριμένα, μόλις ολοκληρώσουν τη σύνταξη του ερωτήματος, οι χρήστες μπορούν να επιλέξουν να ομαδοποιήσουν και να ταξινομήσουν τα αναμενόμενα αποτελέσματα σύμφωνα με την σχετικότητα τους ή την ημερομηνία εισαγωγής στην ψηφιακή βιβλιοθήκη. Διευκρινίζεται πως οι χρήστες δεν έχουν τη δυνατότητα να αλλάξουν με δυναμικό τρόπο την ομαδοποίηση ή/και κατάταξη των αποτελεσμάτων της αναζήτησης χωρίς να απευθύνουν ξανά το ερώτημά τους προς την ψηφιακή βιβλιοθήκη.



### 5.3 Αλληλεπιδραστικά αποτελέσματα αναζήτησης

#### 5.3.1 Αρχιτεκτονική

Η έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης αναφέρεται σε μια μεθοδολογία σύμφωνα με την οποία η λειτουργικότητα που έχει να κάνει με την ανάκτηση πληροφοριών εξακολουθεί να παρέχεται στους χρήστες ακόμα και μετά την κατασκευή του συνόλου με τα αποτελέσματα της αναζήτησης. Αυτού του είδους η λειτουργικότητα απαιτεί αφενός μεν ενεργή συμμετοχή από τους χρήστες, αφετέρου δε χώρο εκτέλεσης (execution space) στην πλευρά του πελάτη μέσω του αντίστοιχου Web browser. Σύμφωνα με τη μεθοδολογία που παρουσιάζεται σε αυτό το κεφάλαιο, οι σελίδες με τα αποτελέσματα αναζήτησης θεωρούνται δυναμικά, ομοιογενή έγγραφα που περιέχουν αναφορές σε πηγές εμπλουτισμένες από μεταδομένα (δηλ. δεδομένα που περιγράφουν τα κυρίως δεδομένα [Rusch-Feja, 1998]), τα οποία συνοδεύουν τα περιεχόμενα των εγγράφων αυτών. Οι πληροφορίες αυτές παρέχονται από τη μηχανή αναζήτησης, η οποία, με τη σειρά της, συλλέγει τα μεταδομένα από τον αποθηκευτικό χώρο της ψηφιακής βιβλιοθήκης. Στην περίπτωση των μηχανών αναζήτησης στο Web, το παρεχόμενο σύνολο μεταδομένων είναι αρκετά περιορισμένο.

Στα πλαίσια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης, οι χρήστες έχουν τη δυνατότητα να μετασχηματίσουν την αρχική σελίδα με τα αποτελέσματα της αναζήτησης εκμεταλλευόμενοι την πληροφορία για το περιεχόμενό τους, που παρέχουν τα αντίστοιχα μεταδομένα που τα συνοδεύουν.

Ετσι, οι αποφάσεις για τη χρησιμότητα ή όχι των περιεχόμενων πηγών λαμβάνονται πιο γρήγορα και με μεγαλύτερη ακρίβεια. Μεγάλες λίστες αποτελεσμάτων αναζήτησης μετατρέπονται από τους χρήστες σε λίστες με λίγα αποτελέσματα, έτσι ώστε να μπορούν να γίνουν πιο εύκολα κατανοητές όσον αφορά το περιεχόμενό τους. Επίσης, οι δισοληψίες (transactions) ολοκληρώνονται σε μικρότερο χρονικό διάστημα και έχουν μεγαλύτερη ακρίβεια, γεγονός που επιφέρει οικονομία αφενός στη χρήση του εύρους του δικτύου και αφετέρου στην εκμετάλλευση των πόρων του εξυπηρετητή. Η παρεχόμενη λειτουργικότητα αναφέρεται στη δυνατότητα που έχουν οι χρήστες να ομαδοποιούν, φιλτράρουν ή/και ταξινομούν τα αποτελέσματα της αναζήτησης ανάλογα με τα μεταδομένα που συνοδεύουν τις πηγές που απαρτίζουν τα αποτελέσματα αυτά.

Η ανάπτυξη εφαρμογών ικανών να παρέχουν την περιγραφόμενη λειτουργικότητα, προϋποθέτει την υποστήριξή τους από κατάλληλες αρχιτεκτονικές. Ετσι, όπως φαίνεται και στο σχ. 1, τα δεδομένα εξόδου μιας μηχανής αναζήτησης κατευθύνονται σε μια ενδιάμεση εφαρμογή, η οποία εκτελείται στην πλευρά του πελάτη. Οι χρήστες έχουν τη δυνατότητα να αλληλεπιδρούν με την εφαρμογή αυτή και να μεταβάλλουν την κατάταξη ή/και τον πληθυσμό των αντικειμένων που απαρτίζουν τα αποτελέσματα αναζήτησης. Το εύρος των επιλογών των



Σχήμα 1: Αρχιτεκτονική αλληλεπιδραστικών αποτελεσμάτων αναζήτησης

χρηστών αναφορικά με την παρεχόμενη λειτουργικότητα, καθορίζεται από την ποικιλία των μεταδεδομένων που παρέχονται από τη μηχανή αναζήτησης. Ένα πλούσιο σύνολο μεταδεδομένων εγγυάται στους χρήστες ένα εξίσου πλούσιο σύνολο από επιλογές αλληλεπίδρασης με τα αποτελέσματα.

### 5.3.2 Μεταδεδομένα

Το πιο διαδεδομένο κριτήριο ταξινόμησης αποτελεσμάτων όπως αυτά παρέχονται από μια μηχανή αναζήτησης, είναι ο βαθμός σχετικότητάς τους με το αρχικό ερώτημα που προκάλεσε τη δημιουργία τους. Τα τελευταία χρόνια, σε μια προσπάθεια βελτίωσης της μεθόδου ταξινόμησης των αποτελεσμάτων έτσι ώστε να τοποθετούνται οι πιο χρήσιμες για τους χρήστες πηγές στις πιο υψηλές θέσεις ταξινόμησης, αναζητήθηκαν αναλλακτικές μέθοδοι παράθεσης των αποτελεσμάτων αυτών [The Best Search Engines, 2002].

Στο κεφάλαιο αυτό παρουσιάζεται μια μέθοδος κατάταξης των αποτελεσμάτων αναζήτησης λαμβάνοντας υπόψη τη δημοτικότητά (popularity) τους. Η δημοτικότητα των πηγών που αντιστοιχούν στα αποτελέσματα υπολογίζεται με βάση την παρατήρηση ότι "πηγές που δέχονται πολλές επισκέψεις θεωρούνται πιο δημοφιλείς από άλλες που δε δέχονται επισκέψεις τόσο συχνά". Η δημοτικότητα μιας πηγής υπολογίζεται σε συνάρτηση με τον κύκλο ζωής της πηγής αυτής. Ετσι λοιπόν, η "ηλικία" μιας πηγής αποτελεί ένα μεταδεδομένο που επηρεάζει τη δημοτικότητά

της. Σε αντίθεση με τη σχετικότητα, πληροφορίες όπως “δημοτικότητα” και “ηλικία” είναι ανεξάρτητες από το αρχικό ερώτημα του χρήστη και πρέπει να παρέχονται από τη μηχανή αναζήτησης. Η *σπουδαιότητα* (importance) ορίζεται ως ο συνδυασμός της δημοτικότητας και της σχετικότητας μιας πηγής. Το ποσοστό συμμετοχής των δυο αυτών συμπληρωματικών μεταδεδομένων (δημοτικότητα-σχετικότητα) στη σπουδαιότητα μιας πηγής καθορίζεται από τους χρήστες. Για παράδειγμα, ένα ιδεατό σενάριο υπογορεύει ότι τα αποτελέσματα μιας αναζήτησης ταξινομούνται με βάση τη σπουδαιότητά τους, στην οποία η σχετικότητα συμμετέχει με ποσοστό 90%, ενώ η δημοτικότητα με 10%.

Εκτός από τα παραπάνω μεταδεδομένα, τα οποία χαρακτηρίζουν όλες τις πηγές στο Web, υπάρχουν επίσης αρκετά μεταδεδομένα, τα οποία αναφέρονται σε συγκεκριμένες συλλογές. Για παράδειγμα, σε μια ψηφιακή βιβλιοθήκη τεχνικών αναφορών (π.χ. NCSTRL [NCSTRL, 2002]), συναντώνται πληροφορίες όπως “τίτλος”, “συγγραφέας” κλπ. Επίσης, οι ψηφιακές βιβλιοθήκες συγκεκριμένης μορφοποίησης (π.χ. ήχου, εικόνας, video) στο Web περιέχουν μεταδεδομένα που χαρακτηρίζουν τις μορφοποιήσεις αυτές. Συνεπώς, είναι ευθύνη της εκάστοτε εφαρμογής διαχείρισης των αποτελεσμάτων να εκμεταλλευτεί τις πληροφορίες που της παρέχονται (μεταδεδομένα πηγών), έτσι ώστε να παρέχει εκτεταμένη λειτουργικότητα στους χρήστες της.

### 5.3.3 Χώρος εκτέλεσης της εφαρμογής

Σύμφωνα με τη μεθοδολογία που περιγράφεται σε αυτό το κεφάλαιο, οι εφαρμογές που υλοποιούν την έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης, πρέπει να εκμεταλλεύονται τους διαθέσιμους πόρους στο σταθμό εργασίας στην πλευρά του πελάτη. Η λειτουργικότητα που παρέχουν οι εφαρμογές ανάκτησης πληροφοριών (π.χ. μηχανές αναζήτησης, κατάλογοι στο Web) προσελκύουν το μεγαλύτερο ενδιαφέρον των χρηστών του Web, άρα και των ψηφιακών βιβλιοθηκών που βασίζονται σε αυτό για την αλληλεπίδρασή τους με τους χρήστες. Η ολοκληρωτική εξάρτηση από τις δυνατότητες των εξυπηρετητών σε υπολογιστική ισχύ, μνήμη και δυνατότητες δικτύου για την εκπλήρωση των αυξημένων αναγκών των χρηστών, κλιμακώνει καθημερινά το κόστος λειτουργίας αυτού του είδους των εφαρμογών. Για την αποφυγή αυτού του επιπλέον κόστους, προτείνεται η ανάπτυξη εφαρμογών που διαχειρίζονται τα αποτελέσματα των αναζητήσεων να χρησιμοποιούν τις δυνατότητες των σταθμών εργασίας στην πλευρά του πελάτη.

Στο κοντινό παρελθόν, ήταν δύσκολο, αν όχι αδύνατο, να βρεθεί η κατάλληλη τεχνολογία για να ικανοποιεί τις απαιτήσεις για ένα ομοιογενή/ διαφανή τρόπο διαχείρισης των αποτελεσμάτων αναζήτησης στο Web ανεξάρτητα από το λειτουργικό σύστημα και τον Web



browser που είναι εγκατεστημένα στους διάφορους σταθμούς εργασίας στην πλευρά του πελάτη. Ευτυχώς, η καθιέρωση της XML ως ένα πρότυπο στο Web για την ανταλλαγή και παρουσίαση πληροφοριών, καθώς και η ευρεία αποδοχή της από όλους σχεδόν τους Web browsers, καθιστά το πρότυπο αυτό ως μια τεχνολογία ικανή να υποστηρίξει το προτεινόμενο μοντέλο. Ακόμα, η ιδιότητα της XML να διαχωρίζει το περιεχόμενο από τον τρόπο παρουσίασης του στους χρήστες, αλλά και να μετατρέπει έγγραφα από μια μορφή σε μια άλλη, αποτελούν δυο πολύ σημαντικά χαρακτηριστικά της XML που βρίσκουν εφαρμογή σε αλληλεπιδραστικά μοντέλα στο Web όπως αυτό που παρουσιάζεται στο παρόν κεφάλαιο.

### 5.3.4 Πρότυπη Υλοποίηση

#### 5.3.4.1 Η Εφαρμογή

Στα πλαίσια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης, παρουσιάζεται μια εφαρμογή διαχείρισης αποτελεσμάτων αναζήτησης, η οποία επέκτεινε τη λειτουργικότητα μιας υπάρχουσας ψηφιακής βιβλιοθήκης (University of Piraeus Lecture Notes digital library – UPLN [UPLN, 2002]). Η πρότυπη αυτή εφαρμογή λειτουργεί ως σύστημα Recommender [Oard, 1998] με την έννοια ότι παρέχει βοήθεια στους χρήστες για τον εντοπισμό χρήσιμων πηγών. Η UPLN είναι μια ψηφιακή βιβλιοθήκη βασισμένη στο Web, η οποία παρέχει τυπικές μεθόδους αναζήτησης του περιεχομένου της στους χρήστες. Ο χώρος αποθήκευσης περιέχει διδακτικό υλικό από εννέα τμήματα εντός του Πανεπιστημίου Πειραιά. Στην ουσία, η UPLN φιλοξενεί τις σημειώσεις του διδακτικού προσωπικού του Πανεπιστημίου οργανωμένες κατά Τμήμα και διαθέσιμες σε διάφορες μορφοποιήσεις (pdf, doc, tif, jpg κλπ). Η μηχανή αναζήτησης της ψηφιακής αυτής βιβλιοθήκης συντηρεί ένα ευρετήριο με βιβλιογραφικά μεταδεδομένα όπως “συγγραφείς”, “ημερομηνία παράδοσης”, “τίτλος”, τα οποία βρίσκονται σε ξεχωριστά αρχεία για κάθε ένα σύνολο σημειώσεων (βλ. σχ. 2):

```
BIB-VERSION:: CS-TR-v2.1
ID:: unipi.csd//csd-001
TITLE:: "Database Systems"
ENTRY:: February 12, 1999
AUTHOR:: George Chondrokoukis
ABSTRACT:: ="Homework for Relational Schemas..."
END:: unipi.csd//csd-001
COUNTER:: 624
```

Σχήμα 2: Ένα αρχείο βιβλιογραφικών μεταδεδομένων

Τα δεδομένα αυτά σχετίζονται άμεσα με τα μεταδεδομένα που χρειάζονται για την αλληλεπιδραστική διαχείριση των αποτελεσμάτων αναζήτησης. Έτσι λοιπόν, η πρότυπη εφαρμογή που περιγράφεται σε αυτό το κεφάλαιο χρησιμοποιεί μεταδεδομένα όπως “τίτλος”, “συγγραφέας”, “ημερομηνία εισαγωγής”, “περίληψη”, “μετρητής (Counter)”, όπου το μεταδεδομένο “μετρητής” παρακολουθεί τον αριθμό των προσβάσεων στο αντίστοιχο έγγραφο-πηγή από τους χρήστες της ψηφιακής βιβλιοθήκης.

Απαραίτητη προϋπόθεση για την προσθήκη της περιγραφόμενης λειτουργικότητας στην ήδη υπάρχουσα ψηφιακή βιβλιοθήκη, είναι η τροποποίηση της μηχανής αναζήτησης κατά τέτοιο τρόπο, ώστε τα να παρέχει τα δεδομένα εξόδου της όχι πλέον σε μορφή HTML, αλλά σε μορφή XML, η οποία περιέχει τα κατάλληλα μεταδεδομένα για κάθε ένα αποτέλεσμα αναζήτησης (βλ. σχ. 3):

```
<?xml version="1.0" encoding="ISO-8859-7"?>
<?xml-stylesheet type="text/xsl" href="hits.xsl"
xmlns:xsl="http://www.w3.org/TR/WD-xsl"?>
<list>
<resource date="November 13, 1999" counter="450"...
url="http://thalis.gr/Rep/U.I/prod-001/index.html"
relevancy="7.654" popularity="15.3"
abstract="Database Systems..." id="prod-001"
importance="1">
<author>George Hondrokoukis</author>
</resource>
<resource date="November 13, 1999" counter="353"...
url="http://thalis.cs.unipi.gr/a.html"
relevancy="7.654" popularity="12"
abstract="Homework for Rel. Schemas..." id="csd-003"
importance="0.7">
<author>George Vassilakopoulos</author>
</resource>
</list>
```

Σχήμα 3: Αποτελέσματα αναζήτησης σε μορφή XML

Κάθε αντικείμενο του συνόλου αποτελεσμάτων αντιστοιχεί σε ένα συστατικό (element) "resource" της παραπάνω XML δομής, το οποίο περιέχει μεταδεδομένα όπως "URL", "abstract", τα οποία αντιστοιχούν σε πληροφορίες που ανακτώνται από τον αποθηκευτικό χώρο της ψηφιακής βιβλιοθήκης και παρέχονται από τη μηχανή αναζήτησης. Το συστατικό "popularity" αντιστοιχεί στον ρυθμό πρόσβασης του εκάστοτε εγγράφου, ο οποίος υπολογίζεται σε προσβάσεις (hits) ανά μήνα. Το συστατικό "σπουδαιότητα" υπολογίζεται από τις εξισώσεις 1-3 στο σχ. 4, όπου  $imp_i$  είναι ο συντελεστής σπουδαιότητας για την πηγή  $i$ ,  $rel_i$  είναι η σχετικότητα της πηγής  $i$  ( $0 \leq i \leq n$ , όπου  $n$  είναι ο συνολικός αριθμός των αποτελεσμάτων αναζήτησης) όπως λαμβάνονται από τη μηχανή αναζήτησης,  $pop_i$  είναι η δημοτικότητα μιας πηγής υπολογισμένη σε προσβάσεις ανά μήνα ( $counter/30$ ),  $k$  είναι η στάθμιση της σχετικότητας στη σπουδαιότητα και  $1-k$  είναι η στάθμιση της δημοτικότητας στη σπουδαιότητα. Οι τιμές της σταθεράς  $k$  κυμαίνονται από 0 μέχρι 1 ( $0 \leq k \leq 1$ ). Ο συντελεστής  $k$  αποφασίζει ποιά από τα δυο συστατικά (σχετικότητα- δημοτικότητα) θα επηρεάσει περισσότερο τη σπουδαιότητα της πηγής. Τιμές κοντά στο "0" ευνοούν τη δημοτικότητα, ενώ τιμές κοντά στο "1" ευνοούν τη σχετικότητα. Στην ιδανική περίπτωση, ο χρήστης πρέπει να έχει τη δυνατότητα να ρυθμίσει την τιμή της σταθεράς  $k$  σύμφωνα με τις προσωπικές του απαιτήσεις. Παρολαυτά, για λόγους απλοποίησης, στην περιγραφόμενη εφαρμογή η σταθερά  $k$  έχει την τιμή 0.8. Τέλος, η σπουδαιότητα  $imp_i$  κανονικοποιείται στο διάστημα τιμών [0,1].



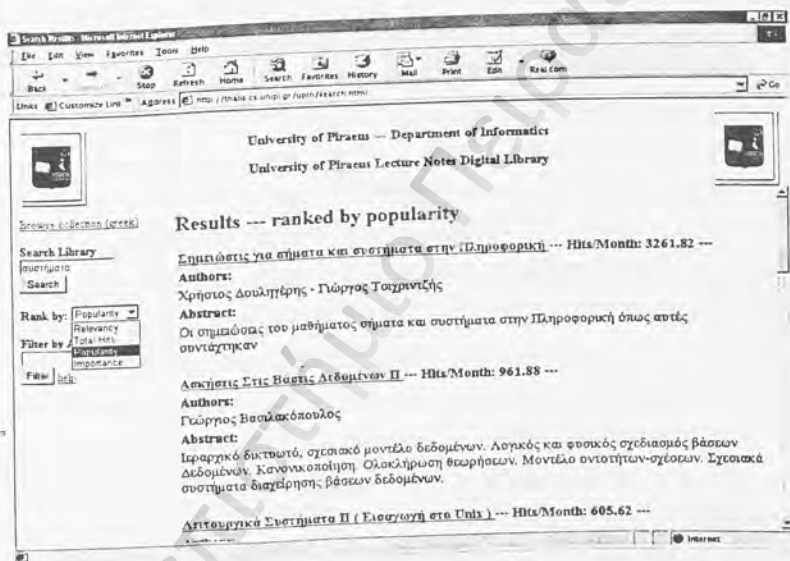
$$\overline{imp}_i = rel_i * k + pop_i * (1 - k), \quad (1)$$

$$imp_{max} = \max(\overline{imp}_i), \quad (2)$$

$$imp_i = \overline{imp}_i / imp_{max} \quad (3)$$

Σχήμα 4: Εξισώσεις υπολογισμού της σπουδαιότητας

Η δομή XML που προκύπτει εισάγεται στην εφαρμογή, η οποία μορφοποιεί τα δεδομένα σε HTML σύμφωνα με τις οδηγίες που υπάρχουν σε ένα αντίστοιχο XSL αρχείο. Αρχικά, τα αποτελέσματα ταξινομούνται με βάση τη σχετικότητα τους με το αρχικό ερώτημα. Όπως φαίνεται στο σχ. 5, η αλληλεπίδραση με τους χρήστες παρέχεται από τις επιλογές του



Σχήμα 5: Αποτελέσματα αναζήτησης ταξινομημένα ανά δημοτικότητα

αριστερού πλαισίου του περιβάλλοντος διεπαφής της εφαρμογής.

Οι χρήστες έχουν τη δυνατότητα να αναταξινομήσουν τα αντικείμενα του συνόλου με τα αποτελέσματα μιας αναζήτησης σύμφωνα με α) τη δημοτικότητά τους β) το συνολικό αριθμό προσβάσεων από τους χρήστες και γ) τη σπουδαιότητά τους, όπως έχει ήδη οριστεί. Επίσης, οι χρήστες μπορούν να επιλέξουν να φιλτράρουν τα αποτελέσματα με βάση το όνομα του συγγραφέα. Έτσι, στο σχ. 5 τα αποτελέσματα αναζήτησης είναι ταξινομημένα με βάση τη δημοτικότητά τους. Στην περίπτωση όπου ζητούνται τα έγγραφα ενός και μόνο συγγραφέα, οι

χρήστες μπορούν να σημειώσουν το όνομά του στο πεδίο "Filter by Author" και να εκτελέσουν τη λειτουργία "filter".

#### 5.3.4.2 Θέματα υλοποίησης

Η εφαρμογή που περιγράφεται σε αυτό το κεφάλαιο εκμεταλλεύεται τις πηγές (resources) στο σταθμό εργασίας του πελάτη, μέσω της χρήσης κατάλληλων τεχνολογιών όπως των γλωσσών JavaScript και XSLT. Βασικός σχεδιαστικός στόχος της εφαρμογής αποτελεί η υποστήριξη της παρεχόμενης λειτουργικότητας μέσα από τη χρήση εντολών XSLT που εκτελούνται στον Web browser, καθώς οι περισσότερες και πιο γνωστές εταιρίες που κατασκευάζουν Web browsers υποστηρίζουν το πρότυπο αυτό. Παρ'όλα αυτά, σύντομα διαπιστώθηκε ότι δεν ήταν δυνατό να κατασκευαστεί μια εφαρμογή, η οποία να είναι συμβατή με όλους τους Web browsers που υποστηρίζουν την XSLT. Το γεγονός αυτό οφείλεται στο ότι κάθε Web browser υποστηρίζει και από ένα διαφορετικό υποσύνολο μιας διαφορετικής έκδοσης των επίσημων προδιαγραφών της XSLT.

Ουσιαστικά, δεν ήταν δυνατό να εντοπιστεί έστω και ένας Web browser που να υποστηρίζει την XSLT σε τέτοιο βαθμό, ώστε να μπορεί να εκτελεστεί η πρότυπη εφαρμογή χωρίς προβλήματα. Τελικά, η εφαρμογή υλοποιήθηκε με τη βοήθεια εντολών της Javascript και εντολών του λειτουργικού συστήματος Microsoft Windows. Αυτό είχε ως αποτέλεσμα να περιοριστεί η συμβατότητα της μόνο στον Internet Explorer έκδοσης 5.0 και νεότερης. Η έλλειψη κατάλληλης υποστήριξης της γλώσσας XSLT οδήγησε στην επιβολή ορισμένων περιορισμών αναφορικά με την αλληλεπίδραση με τους χρήστες. Έτσι, λόγω της ανεπαρκούς υποστήριξης στην αναγνώριση προτύπων σε συμβολοσειρές (pattern matching), δεν είναι δυνατός ο εντοπισμός συμβολοσειρών, οι οποίες αποτελούν τμήμα μιας μεγαλύτερης συμβολοσειράς. Αυτό έχει σαν αποτέλεσμα να απαιτείται η εισαγωγή ολόκληρου του ονοματεπώνυμου του συγγραφέα στο πεδίο "Filter by Author" προκειμένου να φιλτράρουν τα υπόλοιπα έγγραφα. Παρ'όλα αυτά, η τάση στο Web υποδεικνύει πως είναι απλά θέμα χρόνου μέχρι να υπάρξει πλήρης υποστήριξη των πρωτοκόλλων XML/XSLT από τους Web browsers.

Σύμφωνα με την αρχιτεκτονική υλοποίησης της εφαρμογής, η δομή XML που λαμβάνεται από τη μηχανή αναζήτησης, υπόκειται σε ένα πλήθος από μετασχηματισμούς XSL, οι οποίοι έχουν ως αποτέλεσμα την αναδιάταξη και ανασύσταση του συνόλου με τα αποτελέσματα. Οι μετασχηματισμοί αυτοί ενεργοποιούνται από τους χρήστες μέσα από το αριστερό πλαίσιο της αρχικής οθόνης της εφαρμογής, στο οποίο υπάρχουν και τα αντίστοιχα κριτήρια. Η μηχανή αναζήτησης της ψηφιακής βιβλιοθήκης UPLN ενεργοποιείται μόνο όταν οι χρήστες της απευθύνουν κάποιο νέο ερώτημα. Το περιγραφόμενο σύστημα εκτελείται σε έναν

εξυπηρετητή Compaq Proliant, 300 MHz στον οποίο είναι εγκατεστημένο το λειτουργικό σύστημα Linux (Redhat). Για τις αλλαγές στον κώδικα της αρχικής μηχανής αναζήτησης (Perlflect search 3.10 [Perlflect, 2001]) της ψηφιακής βιβλιοθήκης UPLN έχει επιλεγεί η γλώσσα υλοποίησης Perl.

### 5.3.4.3 Επέκταση στο Web

Η προτεινόμενη αρχιτεκτονική μπορεί να εφαρμοστεί με μικρές τροποποιήσεις σε ένα μεγάλο αριθμό ψηφιακών βιβλιοθηκών, καθώς τέτοιου είδους συστήματα χρησιμοποιούν συνήθως μεγάλα σύνολα από μεταδεδομένα. Αν και οι πηγές στο Web σπανίως συνοδεύονται από μεταδεδομένα, εντούτοις οι μηχανές αναζήτησης στο Web θα μπορούσαν να επωφεληθούν από την περιγραφόμενη λειτουργικότητα. Για παράδειγμα, εκτός από την κατάταξη με βάση τη σπουδαιότητα και σχετικότητα, τα αντικείμενα ενός συνόλου από αποτελέσματα αναζήτησης θα μπορούσαν να φιλτραριστούν/ ομαδοποιηθούν με βάση τη δικτυακή τους διεύθυνση ή/και γλώσσα. Ο μοναδικός περιορισμός αναφέρεται στο γεγονός ότι το σύνολο των αποτελεσμάτων πρέπει να είναι κωδικοποιημένο σύμφωνα με το πρότυπο της XML. Επίσης, η ιεραρχική δομή των XML δεδομένων αλλά και η αλματώδης ανάπτυξη της τεχνολογίας του υλικού των υπολογιστών (οι υπολογιστές σήμερα συνδυάζουν ιδιαίτερα υψηλές αποδόσεις με ταυτόχρονα λογικές τιμές αγοράς) διασφαλίζουν τη γρήγορη και αποδοτική διαχείριση ιδιαίτερως μεγάλων και ετερογενών συνόλων με αποτελέσματα αναζήτησης.

## 5.3.5 Αξιολόγηση

### 5.3.5.1 Η Μέθοδος

Σύμφωνα με τον Nielsen [Nielsen, 1993], οι μέθοδοι για αξιολόγηση μπορούν να καταταχθούν σε δυο κατηγορίες: σε μεθόδους δοκιμής χρηστικότητας (usability testing), όπου οι χρήστες συμμετέχουν στη διαδικασία αξιολόγησης και σε μεθόδους επίβλεψης χρηστικότητας (usability inspection), όπου δεν υπάρχει συμμετοχή των χρηστών. Η χρηστικότητα έχει να κάνει με το επίπεδο κατανόησης και χρήσης της παρεχόμενης λειτουργικότητας από τους χρήστες.

Η αξιολόγηση της πρότυπης εφαρμογής έγινε με τη μέθοδο του ερωτηματολογίου. Συγκεκριμένα, ένα κατάλληλα διαμορφωμένο ερωτηματολόγιο διανεμήθηκε σε ένα σύνολο από χρήστες, οι οποίοι είχαν προηγούμενη εμπειρία με το σύστημα ανάκτησης πληροφοριών της ψηφιακής βιβλιοθήκης UPLN. Κατόπιν, ακολούθησε μια δεκάλεπτη συζήτηση με κάθε ένα χρήστη ξεχωριστά, στην οποία εκτιμήθηκε η γενικότερή του εντύπωση για την εφαρμογή. Το



πλεονέκτημα της μεθόδου του ερωτηματολογίου είναι αφενός το χαμηλό της κόστος και αφετέρου το γεγονός ότι μπορούν εύκολα να εμπλακούν σε αυτή πολλοί χρήστες. Από την άλλη μεριά, τα αποτελέσματα αξιολόγησης που βασίζονται σε παρόμοιες μεθόδους είναι συχνά αμφιβόλου ποιότητας καθώς οι απαντήσεις των χρηστών βασίζονται συνήθως σε ότι θεωρούν πως κάνουν και όχι σε ό,τι πραγματικά κάνουν [Nielsen, 1993]. Το ερωτηματολόγιο που χρησιμοποιήθηκε στην αξιολόγηση της πρότυπης εφαρμογής είναι βασισμένο σε ένα ερωτηματολόγιο, το οποίο είναι κατάλληλα προσαρμοσμένο στο χώρο των ψηφιακών βιβλιοθηκών [Theng, 1999] και θεωρείται αρκετά αξιόπιστο από τους εμπνευστές του. Ερωτήσεις που αξιολογούν τόσο τη μορφή όσο και το περιεχόμενο της εφαρμογής έχουν αποφευχθεί, καθώς έχει αποδειχθεί πως μπλοκάρουν τους χρήστες [Lif, 1998]. Τέλος, πρέπει να ληφθεί υπόψη το γεγονός πως κύριος στόχος της μεθόδου αποτελεί η αξιολόγηση της έννοιας των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης μέσα από τη χρήση της πρότυπης εφαρμογής και όχι η ίδια η εφαρμογή.

#### **5.3.5.2 Δομή της μεθόδου αξιολόγησης**

Όπως έχει ήδη αναφερθεί, στόχος της μεθόδου είναι η αξιολόγηση της χρηστικότητας των εφαρμογών που υιοθετούν την έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης στο Web, όπως έχει οριστεί σε αυτό το κεφάλαιο. Συγκεκριμένα, σύμφωνα με τη μεθοδολογία του πειράματος που ακολουθήθηκε, οι συμμετέχοντες κλήθηκαν να θέσουν ένα σύνολο από τυχαία καθώς και προκαθορισμένα ερωτήματα στη μηχανή αναζήτησης της ψηφιακής βιβλιοθήκης UPLN, μέσα από την πρότυπη εφαρμογή. Τα προκαθορισμένα ερωτήματα έχουν σχεδιαστεί κατά τέτοιο τρόπο, ώστε να επιστρέφουν μακροσκελή αποτελέσματα. Κατόπιν, οι συμμετέχοντες είχαν τη δυνατότητα να αλληλεπιδράσουν με τα αποτελέσματα, μέσω της παρεχόμενης λειτουργικότητας της πρότυπης εφαρμογής. Στη συνέχεια, οι συμμετέχοντες κλήθηκαν να συμπληρώσουν ένα ερωτηματολόγιο, το οποίο περιείχε ερωτήσεις αξιολόγησης της εφαρμογής. Τέλος, η διαδικασία αξιολόγησης ολοκληρώθηκε με μια δεκάλεπτη συζήτηση με κάθε έναν από τους συμμετέχοντες, από τη οποία καταγράφηκε η γενικότερη εντύπωση που τους άφησε η εφαρμογή.

#### **5.3.5.3 Το περιβάλλον εκτέλεσης (Testbed)**

Η ευρεία χρήση μεταδομένων στις ψηφιακές βιβλιοθήκες και ειδικότερα στη UPLN, καθιστά το σύστημα αυτό ως το πλέον κατάλληλο περιβάλλον για να φιλοξενήσει την πρότυπη εφαρμογή, καθώς και την αξιολόγησή της. Η μηχανή αναζήτησης της UPLN λειτουργεί από το

1998 και έχει αποδειχτεί μια πολύ διαδεδομένη και αποδοτική εφαρμογή για τους φοιτητές του Πανεπιστημίου Πειραιά. Ένας ακόμα λόγος για τον οποίο επιλέχθηκε το συγκεκριμένο σύστημα, αποτελεί και η εξοικείωση των φοιτητών-συμμετεχόντων με το περιεχόμενό του και τις λειτουργίες αναζήτησης που παρέχει. Η εξοικείωση αυτή είχε ως αποτέλεσμα την ανάπτυξη ενός εύκολα κατανοητού ερωτηματολογίου, που διευκόλυε τη διαδικασία της αξιολόγησης.

#### 5.3.5.4 Οι συμμετέχοντες

Στη διαδικασία της αξιολόγησης συμμετείχαν εννέα άτομα. Τέσσερα από αυτά ήταν υποψήφιοι διδάκτορες του Τμήματος Πληροφορικής του Πανεπιστημίου Πειραιά ενώ τα υπόλοιπα πέντε ήταν προπτυχιακοί φοιτητές διαφόρων Τμημάτων του Πανεπιστημίου. Όλοι οι συμμετέχοντες στη διαδικασία της αξιολόγησης είχαν προηγούμενη εμπειρία χρήσης της συγκεκριμένης ψηφιακής βιβλιοθήκης. Η αξιολόγηση έγινε σε ένα εργαστήριο του Τμήματος Πληροφορικής, εφοδιασμένο με 10 σταθμούς εργασίας, οι οποίοι είχαν ταχύτητα επεξεργαστή στα 350 MHz, λειτουργικό σύστημα Windows NT 4.0 και Internet Explorer 5.5, ο οποίος υποστηρίζει το πρότυπο της XML 1.0.

#### 5.3.5.5 Το ερωτηματολόγιο

Το ερωτηματολόγιο περιείχε ερωτήσεις ομαδοποιημένες σε έξι νοηματικές κατηγορίες, οι οποίες αποτελούνταν από τέσσερις ερωτήσεις το πολύ η κάθε μια. Συγκεκριμένα, οι κατηγορίες ήταν οι ακόλουθες:

1. Γενικότερες αντιδράσεις προς το σύστημα
2. Περιβάλλον διεπαφής
3. Μάθηση (Learning)
4. Δυνατότητες του συστήματος και έλεγχος χρήστη (user control)
5. Πλοήγηση
6. Περάτωση στόχων (completing tasks)

Οι παραπάνω κατηγορίες είναι εμπνευσμένες από το σχήμα κατηγοριοποίησης του Yin Leng Theng's et al. [Theng, 1998], το οποίο με τη σειρά του είναι βασισμένο πάνω στην κατηγοριοποίηση του Lindgaard για ελαττώματα ευχρηστίας σε αλληλεπιδραστικά συστήματα [Lindgaard, 1994]. Συγκεκριμένα, ερωτήσεις της πρώτης κατηγορίας αξιολογούν τη γενικότερη εντύπωση του χρήστη για το σύστημα, αναφορικά με την εμφάνιση και τη λειτουργικότητά του. Η δεύτερη κατηγορία περιέχει ερωτήσεις σχετικές με την οργάνωση και την εμφάνιση στην οθόνη του υπολογιστή της παρεχόμενης πληροφορίας. Η κατηγορία "Μάθηση" έχει να κάνει με

την ευκολία χρήσης του συστήματος, η κατηγορία “Δυνατότητες του συστήματος και έλεγχος χρήστη” ελέγχει την ταχύτητα χρήσης της παρεχόμενης λειτουργικότητας από τους χρήστες και την αξιοπιστία του συστήματος και οι ερωτήσεις της κατηγορίας “Πλοήγηση” αποσκοπούν στο να ανακαλύψουν αν οι χρήστες “χάνονται” στο περιβάλλον διεπαφής του συστήματος. Τέλος, η κατηγορία “Περάτωση στόχων” αξιολογεί το βαθμό χρησιμότητας των παρεχόμενων υπηρεσιών, αναφορικά με τη δυνατότητα που παρέχει το σύστημα στους χρήστες να εκπληρώνουν τους πληροφοριακούς τους στόχους.

Όλες οι ερωτήσεις που αναφέρονται στο ερωτηματολόγιο είναι ερωτήσεις κλειστού τύπου (closed questions), καθώς εκτιμάται πως είναι πιο εύκολο να αναλυθούν από ερωτήσεις ανοικτού τύπου (open questions) [Lindgaard, 1994]. Σύμφωνα με τη ακολουθούμενη μέθοδο αξιολόγησης, οι απαντήσεις των συμμετεχόντων δόθηκαν μέσα από μια επταβάθμια κλίμακα (βλ. σχ. 6), εκτός από τις απαντήσεις που αντιστοιχούν στην ερώτηση 4.3 (βλ. Παράρτημα), η οποία απαιτούσε μια ταξινόμηση προτάσεων. Τιμές κοντά στη μονάδα αντιστοιχούν στις πιο αρνητικές απαντήσεις, ενώ τιμές κοντά στο επτά αντιστοιχούν στις πιο θετικές απαντήσεις. Τέλος, στην τιμή τέσσερα αντιστοιχούν ουδέτερες απαντήσεις.

πολύ	αρκετά	ελαφρά	ουδέτερη	ελαφρά	αρκετά	πολύ
1	2	3	4	5	6	7

Σχήμα 6: Επταβάθμια κλίμακα απαντήσεων

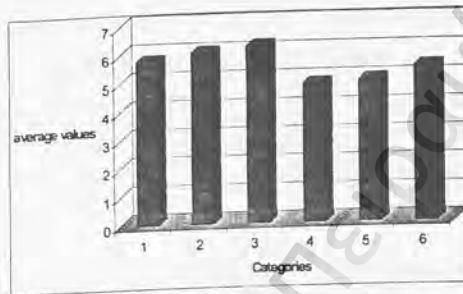
#### 5.3.5.6 Ανάλυση αποτελεσμάτων

Συνολικά, τα αποτελέσματα της μεθόδου αξιολόγησης ήταν αναμενόμενα. Με μια πρώτη ματιά στο σχ. 7, διαπιστώνεται πως οι συμμετέχοντες στο πείραμα έμειναν σε γενικές γραμμές ευχαριστημένοι με την εφαρμογή. Το σχ. 7 παρουσιάζει το μέσο όρο τιμών που αντιστοιχούν στις απαντήσεις των συμμετεχόντων στις ερωτήσεις των έξι διαφορετικών κατηγοριών.



Οι συμμετέχοντες έδειξαν ιδιαίτερα ευχαριστημένοι με την ευκολία χρήσης της εφαρμογής (κατηγορία 3). Αν και τα προηγούμενα χρόνια είχαν συνηθίσει να χειρίζονται ένα απλούστερο περιβάλλον διεπαφής της UPLN, εντούτοις προσαρμόστηκαν γρήγορα στις νέες απαιτήσεις, καθώς η εκπαίδευση πάνω στο καινούργιο περιβάλλον δεν κράτησε πάνω από δέκα λεπτά. Επίσης, προκύπτει το συμπέρασμα ότι οι συμμετέχοντες εκτίμησαν τα πλεονεκτήματα που προκύπτουν από την παρεχόμενη λειτουργικότητα.

Στην ερώτηση “αξιολογήστε τη σπουδαιότητα ως κριτήριο ταξινόμησης των



Σχήμα 7: Μέσοι όροι τιμών για τις απαντήσεις στις έξι κατηγορίες ερωτήσεων

αποτελεσμάτων συγκριτικά με τις παραδοσιακές στατικές τεχνικές εξαντλητικής πλοήγησης για τον εντοπισμό χρήσιμων εγγράφων”, τα αποτελέσματα της μεθόδου καταδεικνύουν πως οι συμμετέχοντες έκαναν εκτεταμένη χρήση όλων των αλληλεπιδραστικών δυνατοτήτων της εφαρμογής. Το ίδιο συμπέρασμα προκύπτει και από τις απαντήσεις της ερώτησης 4.3 (βλ. Παράρτημα), όπου, σύμφωνα με τους συμμετέχοντες, το κριτήριο ταξινόμησης “σχετικότητα” δεν ήταν τόσο ικανοποιητικό όσο τα υπόλοιπα κριτήρια που παρείχε η πρότυπη εφαρμογή. Συγκεκριμένα, η δυνατότητα “ταξινόμηση ανά δημοτικότητα” αποτέλεσε την πιο δημοφιλή επιλογή των συμμετεχόντων, ακολουθούμενη από την επιλογή “ταξινόμηση ανά σπουδαιότητα”. Οι επιλογές “ταξινόμηση ανά σχετικότητα” και “φιλτράρισμα” συγκέντρωσαν τις αμέσως λιγότερες ψήφους, ενώ τελευταία κατατάχθηκε η επιλογή “ταξινόμηση ανά συνολικό αριθμό επισκέψεων”.

Ενα ακόμα πολύ σημαντικό συμπέρασμα που προκύπτει, είναι ότι η πρότυπη εφαρμογή προέτρεφε τους χρήστες να εγκαταλείψουν την παρατήρηση των αποτελεσμάτων και να αρχίσουν να αλληλεπιδρούν μαζί τους, προκειμένου να καλύψουν τις πληροφοριακές τους ανάγκες. Με αυτόν τον τρόπο, μειώνεται ο χρόνος επεξεργασίας των αποτελεσμάτων, ενώ ταυτόχρονα αυξάνεται η απόδοση της αντίστοιχης μηχανής αναζήτησης. Ενα ακόμα σημαντικό χαρακτηριστικό γνώρισμα της πρότυπης εφαρμογής αποτελεί το γεγονός ότι η πλεονάζουσα

λειτουργικότητα δεν επιτυγχάνεται σε βάρος του χρόνου ενασχόλησης με το σύστημα. Η γενικότερη εντύπωση των συμμετεχόντων ήταν πως χωρίς τις επιλογές στο αριστερό πλαίσιο της οθόνης, η διαδικασία εντοπισμού χρήσιμων εγγράφων θα διαρκούσε περισσότερο χρόνο. Ο σχετικά χαμηλός μέσος όρος τιμών αναφορικά με τις ερωτήσεις της τέταρτης κατηγορίας (δηλ. δυνατότητες του συστήματος και έλεγχος χρήστη) οφείλεται στον περιορισμό σύμφωνα με τον οποίο το φίλτράρισμα ανά συγγραφέα επιτυγχάνεται μόνο αν εισαχθεί η ακριβής συμβολοσειρά που αντιστοιχεί στο ονοματεπώνυμο του εκάστοτε συγγραφέα (οι XSLT συναρτήσεις για αναγνώριση συμβολοσειρών δεν υποστηρίζονται επαρκώς στην έκδοση του Internet Explorer που χρησιμοποιήθηκε). Στους χρήστες δεν άρεσε το γεγονός ότι ήταν αναγκασμένοι να εισάγουν ολόκληρο το ονοματεπώνυμο του συγγραφέα στο αντίστοιχο πεδίο της οθόνης. Ομως, όπως έχει ήδη αναφερθεί, τέτοιου είδους ασυμβατότητες πιθανότατα να εξαλειφθούν στις επερχόμενες εκδόσεις των Web browsers.

#### **5.4 Συμπεράσματα και ανοιχτά θέματα**

Στο κεφάλαιο αυτό αναλύθηκε η έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης στο Web, η οποία παρέχει έναν εναλλακτικό τρόπο εντοπισμού χρήσιμων πηγών σε σύνολα με αποτελέσματα αναζήτησης. Η μεθοδολογία που παρουσιάστηκε βρίσκει ιδιαίτερη εφαρμογή σε μεγάλα σύνολα από αποτελέσματα, όπου οι χρήστες είναι αναγκασμένοι να ξοδεύουν πολύτιμο χρόνο, πλοηγούμενοι σε πολλές σελίδες που αποτελούνται από σύντομες περιγραφές σχετικών ή/και μη σχετικών πηγών.

Η αλληλεπίδραση με τα αποτελέσματα αναζήτησης παρέχεται από κατάλληλα τμήματα λογισμικού, τα οποία καθιστούν ικανούς τους χρήστες να ταξινομήσουν τα αποτελέσματα αυτά σύμφωνα με διάφορα κριτήρια. Επίσης, υπάρχει η δυνατότητα μείωσης του πλήθους των αποτελεσμάτων, εφαρμόζοντας διάφορα φίλτρα, τα οποία βασίζονται στα μεταδεδομένα που συνοδεύουν τις πηγές και παρέχονται από τη μηχανή αναζήτησης.

Επίσης, παρουσιάστηκε μια μέθοδος ταξινόμησης των αποτελεσμάτων σύμφωνα με το σταθμισμένο συνδυασμό της σχετικότητας και της δημοτικότητάς τους (δηλ. σπουδαιότητα).

Για λόγους βελτίωσης της απόδοσης αλλά και για εξοικονόμηση χρήσης του εύρους ζώνης του δικτύου, οι εφαρμογές που βασίζονται στην έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης στο Web, πρέπει να εκτελούνται στην πλευρά του πελάτη. Επίσης, η αποτελεσματικότητα των εφαρμογών αυτών εξαρτάται άμεσα από τη δυνατότητα του Web browser να υποστηρίζει το πρότυπο της XML και των συγγενών προτύπων (π.χ. XSLT).

Για την αξιολόγηση των παραπάνω προτάσεων, αναπτύχθηκε μια πρότυπη εφαρμογή αλληλεπιδραστικής διαχείρισης των αποτελεσμάτων αναζήτησης. Σύμφωνα με τα αποτελέσματα αυτής της αξιολόγησης, η παρεχόμενη λειτουργικότητα φαίνεται ότι βοηθά τους χρήστες να ανακτούν τις πιο χρήσιμες γι' αυτούς πηγές μέσα από μεγάλα σύνολα με αποτελέσματα αναζήτησης στο Web σε μικρότερο χρονικό διάστημα αλλά και με μεγαλύτερη ευκολία. Παρολαυτά, η τρέχουσα υποστήριξη του προτύπου της XML και των συναφών προτύπων από τους Web browsers είναι μάλλον ανεπαρκής για τη δημιουργία εφαρμογών που θα υλοποιούν την έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης. Ευτυχώς, υπάρχουν ισχυρές ενδείξεις στο χώρο του Web που υπαγορεύουν ότι οι επόμενες εκδόσεις των πιο δημοφιλών Web browsers θα ξεπεράσουν τέτοιου είδους ασυμβατότητες.



## 6. Συμπεράσματα της Διατριβής

### 6.1 Συμπεράσματα και ανοιχτά θέματα

Ο σημαντικός ρόλος των ψηφιακών βιβλιοθηκών στην ταχύτατα αναπτυσσόμενη κοινωνία της πληροφορίας, καθιστά τις ψηφιακές βιβλιοθήκες ως ένα από τα πιο εξελιξίμα ερευνητικά πεδία του ευρύτερου χώρου της Πληροφορικής. Έτσι λοιπόν, τα ερευνητικά επιτεύγματα στον τομέα των ψηφιακών βιβλιοθηκών προσελκύουν το ενδιαφέρον μεγάλου μέρους της ευρύτερης επιστημονικής κοινότητας.

Κύριο χαρακτηριστικό γνώρισμα των ψηφιακών βιβλιοθηκών αποτελεί το γεγονός ότι συγκεντρώνουν τεχνικές και μεθόδους από μια ευρεία γκάμα παραδοσιακών επιστημονικών περιοχών, συνδυάζοντας και προσαρμόζοντας τις μεθόδους αυτές σε ολοκληρωμένα συστήματα, ικανά να διαχειριστούν με συνέπεια και αποδοτικότητα μεγάλο όγκο ψηφιακών δεδομένων διαφόρων μορφοποιήσεων.

Στα πλαίσια αυτά, η παρούσα Διδακτορική Διατριβή διαπραγματεύτηκε ζητήματα που ανήκουν στους κλάδους α) του σχεδιασμού και της αρχιτεκτονικής εφαρμογών βασισμένων στο Web, β) της ασφάλειας στα ιατρικά δεδομένα και γ) της ανώρισης πληροφοριών, όπως αυτοί εφαρμόζονται σε μοντέρνα συστήματα ψηφιακών βιβλιοθηκών.

### **6.1.1 Σχεδιασμός και αρχιτεκτονική ψηφιακών βιβλιοθηκών**

Κατά τη διάρκεια της παρούσας Διατριβής, παρουσιάστηκε ένα πλαίσιο (framework) στήριξης αρχιτεκτονικών ψηφιακών βιβλιοθηκών βασισμένων στο Web και στο πρότυπο της XML. Βασικό κίνητρο της προσπάθειας αυτής ήταν να αναδείξει τις προοπτικές εκμετάλλευσης των δυνατοτήτων που παρέχει το νέο αυτό πρότυπο στο σχεδιασμό ολοκληρωμένων συστημάτων ψηφιακών βιβλιοθηκών βασισμένων στο Web. Το μοντέλο που παρουσιάστηκε, διακρίνεται για την απλότητα και ευκολία με την οποία υποστηρίζει βασικές απαιτήσεις των ψηφιακών βιβλιοθηκών όπως ανανεωσιμότητα, επεκτασιμότητα και υποστήριξη ονοματολογίας (namespaces). Επίσης, η χρήση της XML διευκολύνει τη διαχείριση των μεταδεδομένων με απλότητα και συνέπεια, ενώ ταυτόχρονα παρέχει την κατάλληλη υποδομή για την ανταλλαγή μηνυμάτων μεταξύ των συστατικών των προκλυπόμενων συστημάτων.

Λαμβάνοντας υπόψη τα συμπεράσματα αυτά, περιγράφηκε η DocML, μια ψηφιακή βιβλιοθήκη βασισμένη στο Web και στην XML για τη διαχείριση Πανεπιστημιακού υλικού. Η παρουσίαση της DocML είχε ως στόχο τον εντοπισμό και κατόπιν την αντιμετώπιση των διαφόρων θεμάτων που προκύπτουν από την υλοποίηση του αρχικού πλαισίου. Έτσι, το βασικό συμπέρασμα που ανέδειξε η υλοποίηση του συστήματος αυτού είναι η έλλειψη κατάλληλων εργαλείων για την αποδοτική διαχείριση των δεδομένων που βασίζονται στο πρότυπο της XML. Παρολαυτά, η ραγδαία εξέλιξη της XML υποδεικνύει ότι είναι θέμα χρόνου η ανάπτυξη ειδικευμένων εφαρμογών διαχείρισης XML δεδομένων, ικανών να υποστηρίξουν με συνέπεια και αποδοτικότητα τις λειτουργίες μιας ψηφιακής βιβλιοθήκης βασισμένης στην XML, όπως είναι η DocML.

Σημείο διαφοροποίησης της DocML από το πλαίσιο στήριξης αρχιτεκτονικών ψηφιακών βιβλιοθηκών αποτελεί το γεγονός ότι χρησιμοποιείται ένα σύστημα ανάκτησης πληροφοριών, το οποίο βασίζεται στο συγκεντρωτικό μοντέλο. Επίσης, αν και οι δυο προσεγγίσεις διαπραγματεύονται κατανεμημένες συλλογές Πανεπιστημιακού υλικού, εντούτοις ο αποθηκευτικός χώρος της DocML αποτελείται από κόμβους που αντιστοιχούν σε Πανεπιστήμια, σε αντίθεση με τον αποθηκευτικό χώρο του αρχικού πλαισίου, ο οποίος αποτελείται από κόμβους που αντιστοιχούν στα επί μέρους Τμήματα ενός Πανεπιστημίου.

Το πιο αξιόπιστο κριτήριο για την αξιολόγηση της λειτουργικότητας που παρέχει μια εφαρμογή της εμβέλειας μιας ψηφιακής βιβλιοθήκης, αναφέρεται στην απήχηση που έχει μεταξύ των χρηστών της. Οι δυο προσεγγίσεις που παρουσιάστηκαν στην παρούσα Διατριβή πρέπει να εξελιχθούν σε ολοκληρωμένες εφαρμογές. Με αυτόν τον τρόπο, θα

μπορέσουν να εξαχθούν πιο ασφαλή συμπεράσματα αναφορικά με τα πλεονεκτήματα αλλά και τα ενδεχόμενα μειονεκτήματα που παρουσιάζουν συγκριτικά με αντίστοιχα συστήματα διαχείρισης Πανεπιστημιακού υλικού. Επίσης, πρέπει να γίνει περαιτέρω ανάλυση και έρευνα στην κατεύθυνση των απαιτήσεων σε ασφάλεια των διαφόρων εγγράφων που διακινούνται στα δυο προτεινόμενα μοντέλα και γενικότερα στα καταναμημένα συστήματα ψηφιακών βιβλιοθηκών Πανεπιστημιακού υλικού. Στα πλαίσια αυτής της έρευνας πρέπει να δοθεί ιδιαίτερη βαρύτητα στην προστασία των πνευματικών δικαιωμάτων των δημιουργών των εγγράφων αυτών, καθώς τα Πανεπιστημιακά έγγραφα δεν αποτελούν απρόσωπα ψηφιακά δεδομένα, αλλά πολύτιμες συνεισφορές στην Πανεπιστημιακή κοινότητα.

### 6.1.2 Μουσικές ψηφιακές βιβλιοθήκες διαχείρισης αρχείων mp3

Η παρούσα Διατριβή ασχολήθηκε με την αρχιτεκτονική μιας κατανεμημένης μουσικής ψηφιακής βιβλιοθήκης βασισμένη στο Web, ικανή να διαχειριστεί μουσικά αρχεία τύπου mp3. Η περιγραφόμενη αρχιτεκτονική βασίζεται σε ένα κατανεμημένο μοντέλο τριών στρωμάτων, για την αποθήκευση και ανάκτηση του περιεχομένου της. Επίσης, η λειτουργικότητα που παρέχεται στους χρήστες στηρίζεται στις διαδεδομένες τεχνολογίες του Web εξυπηρετητή και του Web browser.

Η ψηφιακή αυτή βιβλιοθήκη έρχεται να καλύψει το κενό που υπάρχει στο Web αναφορικά με την ύπαρξη κατάλληλων εφαρμογών για την οργάνωση, διαχείριση και διάδοση αρχείων τύπου mp3 με έναν απλό και αποδοτικό τρόπο. Ταυτόχρονα, παρουσιάστηκε ένα τεχνολογικό πλαίσιο, το οποίο ευνοεί το σεβασμό των πνευματικών δικαιωμάτων των τραγουδιών που διαχειρίζεται, μέσω της τεχνικής αναμετάδοσης καναλιών ηχητικών δεδομένων συνεχούς ροής. Επίσης, η ανεξαρτησία πλατφόρμας καθώς και η εκμετάλλευση των δυνατοτήτων παγιωμένων εφαρμογών όπως ο Web browser που χαρακτηρίζουν την προτεινόμενη μουσική ψηφιακή βιβλιοθήκη, της προσδίδουν το χαρακτηριστικό της μεγάλης διαθεσιμότητας και επεκτασιμότητας για τους χρήστες του Web.

Η επιμονή στη χρήση ευρύτατα διαδεδομένων προτύπων όπως αυτό της XML για την υποστήριξη και περιγραφή των λειτουργιών που παρέχονται από τη μουσική ψηφιακή βιβλιοθήκη σε συνδυασμό με την απουσία συγκεκριμένων προτάσεων αναφορικά με τον τρόπο υλοποίησης των λειτουργιών αυτών, καθιστά την περιγραφόμενη αρχιτεκτονική ανοικτή, με συνέπεια η μουσική ψηφιακή βιβλιοθήκη να είναι επίκαιρη αλλά και συμβατή με κάθε είδους τεχνολογία που εμφανίζεται κατά καιρούς στο χώρο του Διαδικτύου. Για



παράδειγμα, η ανταλλαγή μηνυμάτων μεταξύ των επί μέρους λογισμικών τμημάτων της εφαρμογής θα μπορούσε να υλοποιηθεί σύμφωνα με τις κατευθύνσεις του πρωτοκόλλου Simple Object Access Protocol - SOAP, χωρίς όμως αυτή η πρόταση να αποτελεί δέσμευση.

Λαμβάνοντας υπόψη τις θεαματικές εξελίξεις στον τομέα των δικτύων, οι οποίες ενδεχομένως να εξασφαλίσουν στο μέλλον το επιθυμητό εύρος ζώνης δικτύου σε πολυμεσικές εφαρμογές στο Διαδίκτυο, θα μπορούσε να σχεδιαστεί ένα σύστημα αναμετάδοσης πολλαπλών καναλιών ήχου συνεχούς ροής, κάθε ένα διαφορετικό για κάθε χρήστη της εφαρμογής. Έτσι λοιπόν, ανάλογα με τις προσωπικές προτιμήσεις των χρηστών, οι οποίες θα καταγράφονται και θα φυλάσσονται στα αντίστοιχα προφίλ τους, θα κατασκευάζονται τα αντίστοιχα κανάλια ήχου, τα οποία και θα αναμεταδίδονται σε συγκεκριμένους σταθμούς εργασίας. Όπως είναι κατανοητό, αυτού του είδους η λειτουργικότητα καταναλώνει μεγάλο ποσοστό των διαθέσιμων πηγών του εκάστοτε υπολογιστικού συστήματος, ενώ ταυτόχρονα απαιτεί μεγάλο εύρος ζώνης δικτύου. Αν και οι σημερινές προδιαγραφές των περισσότερων δικτύων που χρησιμοποιούνται στο Διαδίκτυο δεν επαρκούν για να υποστηριχθεί η παραπάνω λειτουργικότητα, εντούτοις οι εξελίξεις στην έρευνα για τα δίκτυα προμηνύουν αλλαγή της κατάστασης αυτής στο άμεσο μέλλον.

### **6.1.3 Ιατρικές ψηφιακές βιβλιοθήκες**

Στην παρούσα Διατριβή περιγράφηκε η αρχιτεκτονική μιας ασφαλούς ιατρικής ψηφιακής βιβλιοθήκης, ικανής να αντιμετωπίσει τις απειλές ασφάλειας που συναντώνται στα κατανεμημένα πληροφοριακά συστήματα, και ειδικότερα στα πληροφοριακά συστήματα του χώρου της Υγείας. Η αρχιτεκτονική αυτή αντιμετώπισε τους κινδύνους ασφάλειας μιας κατανεμημένης ιατρικής ψηφιακής βιβλιοθήκης παρέχοντας τις υπηρεσίες: α) εμπιστευτικότητα, β) αυθεντικοποίηση οντοτήτων, γ) έλεγχος πρόσβασης βασισμένος σε ρόλους χρηστών, δ) ακεραιότητα δεδομένων, ε) καταλογισμός ευθύνης και στ) διαθεσιμότητα. Βασικό χαρακτηριστικό της αρχιτεκτονικής πάνω στην οποία στηρίχθηκε η ιατρική ψηφιακή βιβλιοθήκη αποτελεί το γεγονός ότι βασίστηκε στην έννοια των Εμπιστων Τρίτων Οντοτήτων - ETO για τη δημιουργία και διανομή των αναγκαίων πληροφοριών ασφάλειας. Οι πληροφορίες αυτές μπορούν να ενσωματωθούν σε Επαγγελματικές Κάρτες Υγείας - EKY (έξυπνες κάρτες) διευκολύνοντας με αυτόν τον τρόπο τη γρήγορη και αποδοτική εξυπηρέτηση των χρηστών.

Το μοντέλο της ιατρικής ψηφιακής βιβλιοθήκης που περιγράφηκε δεν υπόκειται σε γεωγραφικούς περιορισμούς. Τεχνικά, θα μπορούσε να φέρει σε επαφή ένα σύνολο από ιατρικούς οργανισμούς, κατανεμημένους γεωγραφικά σε διαφορετικούς δήμους, πόλεις, και χώρες. Η συστέταση διαφορετικών ιατρικών οργανισμών κάτω από μια κοινή ψηφιακή βιβλιοθήκη διευκολύνει σε μεγάλο βαθμό την ανταλλαγή πληροφοριών και γενικότερα τη συνεργασία μεταξύ των επαγγελματιών του χώρου της Υγείας, γεγονός που έχει ως άμεση συνέπεια την αναβάθμιση της ιατρικής περίθαλψης για τους πολίτες. Δυστυχώς όμως, η τεχνική υποδομή από μόνη της δεν επαρκεί για τη δημιουργία και συντήρηση ενός ασφαλούς κατανεμημένου ιατρικού συστήματος. Κάθε ιατρικός οργανισμός διέπεται από συγκεκριμένους κανόνες λειτουργίας και πολιτικές ασφάλειας. Στην περίπτωση των οργανισμών διαφορετικών κρατών, υπάρχουν σημαντικές διαφορές ακόμα και σε νομικό-πολιτειακό επίπεδο. Παραμένει λοιπόν ανοικτό θέμα η κατοχύρωση ενός κοινού πλαισίου, το οποίο να καθορίζει τις πολιτικές ασφάλειας για την ομαλή όσο και ασφαλή λειτουργία εφαρμογών που διαχειρίζονται ευαίσθητα ιατρικά δεδομένα στο χώρο της Υγείας.

#### 6.1.4 Ανάκτηση πληροφοριών

Στην παρούσα Διατριβή, αναλύθηκε η έννοια των αλγήλεπιδραστηκών αποτελεσμάτων αναζήτησης στο Web, η οποία παρέχει έναν εναλλακτικό τρόπο εντοπισμού χρήσιμων πηγών σε εκτενή σύνολα αποτελεσμάτων που προκύπτουν από διαδικασίες αναζήτησης στο Web. Η αλγήλεπίδραση, έγκειται στη δυνατότητα των χρηστών να διαχειρίζονται με δυναμικό τρόπο τα αποτελέσματα αυτά. Η δυνατότητα διαχείρισης των αποτελεσμάτων στηρίζεται στα μεταδεδομένα που συνοδεύουν τις πηγές, οι οποίες αναφέρονται στα αποτελέσματα αναζήτησης. Στο πλαίσιο αυτό, παρουσιάστηκε μια μέθοδος ταξινόμησης των αποτελεσμάτων σύμφωνα με τη σπουδαιότητά τους, η οποία ορίστηκε ως ο σταθμισμένος συνδυασμός της σχετικότητας (relevancy) και της δημοτικότητάς (popularity) τους. Τέλος, για λόγους βελτίωσης της απόδοσης των εφαρμογών που στηρίζονται στην παραπάνω έννοια, αλλά και για λόγους εξοικονόμησης του εύρους ζώνης του δικτύου, υποστηρίχθηκε ότι οι εφαρμογές αυτές πρέπει να εκτελούνται στην πλευρά του πελάτη (client layer).

Η υλοποίηση εφαρμογών που υποστηρίζουν την παραπάνω λειτουργικότητα, βασίζεται σε μεγάλο βαθμό στη χρήση του προτύπου της XML, αλλά και στη θεώρηση του Web browser ως το λειτουργικό σύστημα πάνω στο οποίο εκτελούνται οι εφαρμογές αυτές. Είναι συνεπώς αναγκαίο να βελτιωθεί και να επεκταθεί η υποστήριξη της XML και των

διαφόρων συναφών προτύπων (π.χ. XSLT, DOM) από τους Web browsers, με τέτοιο τρόπο, ώστε οι εφαρμογές που θα προκύψουν να μπορούν να εκτελεστούν από οποιοδήποτε σταθμό εργασίας. Αρα λοιπόν, η χρήση ευρέως αποδεκτών προτύπων, αποτελεί βασική προϋπόθεση για τη διάδοση εφαρμογών βασισμένων στο Web, όπως αυτές που στηρίζονται στην έννοια των αλληλεπιδραστικών αποτελεσμάτων αναζήτησης.

Πανεπιστήμιο Πειραιώς



## 7. Βιβλιογραφικές αναφορές

- [Allan, 1997] Allan J, *Building hypertext using information retrieval*, Information Processing and Management Vol.33, No. 2, pp. 145–159, 1997
- [Aloisio, 1999] Aloisio G, Millilo G, Williams R.D, *An XML architecture for high performance web-based analysis of remote-sensing archives*, Future Generation Computer Systems, Vol. 16, pp. 91-100, 1999
- [Baldonado, 1997] Baldonado M, Chang C.K, Gravano L, *Metadata for Digital Libraries: Architecture and Design Rationale*, Computer Science Department, Stanford University, 1997
- [Barber, 1998] Barber B, Louwse K, Davey J, *White Paper on Healthcare Information Security*, NHS executive's information management center, 1998
- [Berenci, 1999] Berenci E, Carpineto C, Giannini V, Mizzaro S, *Effectiveness of keyword-based display and selection of retrieval results for interactive searches*, Proceedings of the third European Conference in Digital Libraries (ECDL '99), LNCS, Springer Verlag, pp. 106-125, 1999
- [Blobel, 1998] Blobel B, Pharow P, *The need and practice of user authentication and TTP services in distributed Health Information Systems*, Proceedings of the Toward an Electronic Patient Record Conference 1998 (IEPR '98), Newton, pp. 34-40, 1998
- [Brando, 1995] Brando T. J., *Interoperability and the CORBA specification*, MITRE document, 1995

[Chadwick, 1993] Chadwick D, *Understanding X.500 - The Directory*, ISBN: 0-412-43020-7, Chapman & Hall, 1993

[Chalmers, 1992] Chalmers M, Chitson P, *Bead: Explorations in information visualization*, Proceedings of SIGIR Conference on Research and Development in Information Retrieval (SIGIR '92), Copenhagen, pp. 330-337, 1992

[Choi, 1997] Choi R, Na J, Lee K, Kim E, Han W, *Design of a security platform for CORBA based Application*, Proceedings of the International Conference on Information and Communications Security Conference (ICICS '97), pp. 98-108, 1997

[Cole, 2000] Cole W.T, Mischo H.W, Ferrer R, Habing G.T, *Using XML, XSLT and CSS in a Digital Library*, Proceedings of 63<sup>rd</sup> Conference of the American Society for Information Science (ASIS) 2000, pp. 430-439, 2000

[Council of Europe, 1997] Council of Europe, *Recommendation R(97)5, On the protection of medical data*, Council of Europe, February 12, 1997

[Crews, 1997] Crews K.D, *Copyright and digital libraries: The U.S. perspective and international implications*, Digital Libraries, No. 10, pp. 35-56, July 1997

[Cutting, 1992] Cutting D. R, Pederson J. O, Karger D. R, Tukey J. W, *Scatter/gather: A cluster-based approach to browsing large document collections*, Proceedings of the Fifteenth Annual International ACM - SIGIR Conference on Research and Development in Information Retrieval, pp. 318-329, 1992

[Dewire, 1998] Dewire D. T, *Thin clients: Delivering Information over the Web*, McGraw-Hill, ISBN 0-07-016738-9, 1998

[Dubin, 1995] Dubin D, *Document analysis for visualization*, Proceedings of SIGIR Conference on Research and Development in Information Retrieval (SIGIR '95), pp. 199-204, 1995

- [Dunn, 1999] Dunn J.V, Mayer C.A, *VARIATIONS: A Digital Library system at Indiana University*, Proceedings of the fourth ACM Conference on Digital Libraries (DL '99), pp. 12-19, 1999
- [Dushay, 1999] Dushay N, French J. C, Lagoze C, *Using Query Mediators for Distributed Searching in Federated Digital Libraries*, Proceedings of the fourth ACM Conference on Digital Libraries (DL '99), pp. 171-178, 1999
- [Fleischhauer, 1996] Fleischhauer C, *Digital Historical Collections: Types, Elements, and Construction*, Library of Congress technical white paper, available at: <http://memory.loc.gov/ammem/elements.html>, 1996
- [Fonss - Jorgensen, 1997] Fonss - Jorgensen E, *JUKEBOX: Final report*, LIB-JUKEBOX/4-1049. Edited report no. 2. Aarhus, Denmark: State and University Library, available at: <http://www.sb.aau.dk/Jukebox/finalrep.html>, April, 1997
- [Fox, 1997] Fox E, Eaton J, McMillan, *Networked Digital Library of Thesis and Dissertations*, D-lib Magazine, Sep. 1997, available at: [www.dlib.org](http://www.dlib.org), 1997
- [Frenot, 1999] Frenot S, Laforest F, *Medical Record Management Systems: criticism and new perspectives*, Methods of Information in Medicine, Vol. 38, pp. 89-95, 1999
- [Gennari, 1989] Gennari J. H, Langley P, Fisher D, *Models of Incremental Concept Formation*, Artificial Intelligence, No 40, pp. 11-61, 1989
- [Gnutella, 2000] *The Gnutella file sharing system*, available at: [www.gnutella.com](http://www.gnutella.com), 2000
- [Golden, 1998] Golden R. F, *The CPR, document imaging and the INTERNET*, Proceedings of the Healthcare Information and Management Systems Society Conference (HIMMS '98), pp. 46-55, 1998



[Grammatikou, 2000] Grammatikou M, Stamatiopoulos F, Maglaris B, *Distributed Information System Architecture for Primary Health Care*, Proceedings of the Medical Infobahn in Europe Conference (MIE 2000), IOS Press, pp. 978-982, 2000

[Gupta, 1999] Gupta A, Baru C, *An extensible information model for shared scientific data collections*, Future Generation Computer Systems, Vol. 16, pp. 9-20, 1999

[Hamalka, 1998] Hamalka J, Safran C, *CareWeb, a Web-based Medical Record for an Integrated Healthcare Delivery system*, Proceedings of the MedInfo Conference 1998 (Medinfo '98), IOS Press, pp. 36-39, 1998

[Harman, 1998a] Harman D, Voorhees E, *Overview of the Fifth Text REtrieval Conference (TREC-5)*, Proceedings of the 5<sup>th</sup> Text REtrieval Conference (TREC-5), pp. 1-28, 1998

[Harman, 1998b] Harman D, Voorhees E, *Overview of the Sixth Text REtrieval Conference (TREC-6)*, Proceedings of the 6<sup>th</sup> Text REtrieval Conference (TREC-6), pp. 1-24, 1999

[Hearst, 1996] Hearst M.A, Pedersen J.O, *Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results*, Proceedings of SIGIR Conference on Research and Development in Information Retrieval (SIGIR '96), Zurich, August, 1996, pp. 76-84, 1996

[Hemmje, 1994] Hemmje M, Kunkel C, Willer A, *Lyberworld - A visualization user interface supporting full text retrieval*, Proceedings of SIGIR Conference on Research and Development in Information Retrieval (SIGIR '94), 1994, pp. 249-259, 1994

[Hitchcock, 2000] Hitchcock S, Carr L, Jiao Z, Bergmark D, Hall W, Lagoze C, Harnad S, *Developing Services for Open Eprint Archives: Globalization, Integration and the Impact of Links*, Proceedings of the 5<sup>th</sup> ACM Conference on Digital Libraries, pp. 143-151, 2000

[Housley, 1999] Housley R, Ford W, Polk W, Solo D, *Internet X.509 Public Key Infrastructure Certificate and CRL Profile*, RFC 2459, 1999

- [Houstis, 1997] Houstis C, Nikolaou C, Marazakis M, Patrikalakis M, Sairamesh J, Thoma A, *THETIS: Design of a data management and data visualization system for coastal zone management of the Mediterranean sea*, D-lib Magazine, available at: <http://www.dlib.org/dlib/november97/thetis/11thetis.html>, 1997
- [Iannella, 1997] Iannella R, Waugh A, *Metadata: Enabling the Internet*, available at: [www.dstc.edu.au/RDU/publications/cause97](http://www.dstc.edu.au/RDU/publications/cause97), 1997
- [Iona, 1999] Iona Tec, *OrbixWeb*, available at: [www.iona.com/Products/Orbix/OrbixWeb/](http://www.iona.com/Products/Orbix/OrbixWeb/), 1999
- [ISHTAR, 1997] ISHTAR Project, *Deliverable D09, Current Security Issues Faced by Healthcare Establishments*, Magdeburg, UHM, 1997
- [Kohonen, 1995] Kohonen T, *Self-Organizing Maps*, Springer Verlag, Berlin, Germany, 1995
- [Labreze, 1998] Labreze L, Lagouarde P, Dakin C, Renaud-Salis J. L, *A web interface for multimedia electronic patient record: consensual validation of the Aquitaine Health Information prototypes*, Medical Informatics, Vol. 23, No. 1, pp. 75-84, 1998
- [Laforest, 2000] Laforest F, Flory A, *The electronic medical record: Using documents for information capture*, Proceedings of the Medical Infobahn in Europe Conference 2000 (MIE 2000), IOS Press, pp. 617-621, 2000
- [Lagoze, 1995] Lagoze C, *Dienst - An Architecture for Distributed Document Libraries*, Communications of the ACM, Vol. 38, No. 4, April 1995
- [Lawrence, 1999] Lawrence S, Lee Giles C, Bollacker K, *Digital Libraries and Autonomous Citation Indexing*, IEEE Computer, Vol. 32, No. 6, pp. 67-71, 1999

[Lawton, 1999] Lawton G, *Vendors battle over mobile-os market*, IEEE Computer, Vol. 32, No. 2, pp. 13-15, 1999

[Leouski, 1996] Leouski A, Croft B, *An Evaluation of Techniques for Clustering Search Results*, CIIR technical report, 1996

[Leuski, 1998] Leuski A, Allan J, *Interactive cluster visualization for information retrieval*, Proceedings of the second European Conference in Digital Libraries (ECDL '98), LNCS, Springer Verlag, pp. 535-554, 1998

[Liechti, 1998] Liechti O, Sifer M, Ichikawa T, *Structured graph format: XML metadata for describing Web site structure*, Computer Networks and ISDN systems, Vol. 30, pp. 11-21, 1998

[Lif, 1998] Lif M, *Adding usability: A case study*, Uppsala University Technical Report No. 87, Sweden, 1998

[Lindgaard, 1994] Lindgaard G, *Usability testing and system evaluation: A guide for designing useful computer systems*, Chapman & Hall, 1994

[Liu, 1998] Liu Y-H, Dantzig P, Sachs M, Corey J, Hinnebusch M, Sullivan T, Damashek M, Cohen J, *Visualizing Document Classification: A Search Aid for the D.L.*, Proceedings of the second European Conference in Digital Libraries (ECDL '98), LNCS, Springer Verlag, pp. 555-567, 1998

[Lodder, 2000] Lodder H, Wolf L, Vernee J, *Sharing medication data using the InterCare architecture*, Proceedings of the Medical Infobahn in Europe Conference (MIE 2000), IOS Press, pp. 647-655, 2000

[Lu, 1998] Lu H, Feng L, *Integrating database and World Wide Web technologies*, World Wide Web, pp. 73-86, 1998



- [Lyon, 1997] Lyon E, Maslin J, *Audio and video on-demand for the performing arts: Project PATRON*, International Journal of Electronic Library Research, Vol. 1, No 2, pp. 119-131, 1997
- [Makris, 1998] Makris L, Kopsacheilis E.V, Srintzis M.G, *Hippocrates: An integrated platform for telemedicine applications*, Medical Informatics, Vol. 23, No. 4, pp. 265-276, 1998
- [Marschal, 1999] Marschal B, *XML by Example*, Que publications, ISBN: 0789722429, pp. 32-33, 1999
- [McGrath, 1999] McGrath R, *Integrating Scientific Datasets and Digital Libraries*, Center for Excellence in Space Data and Information Science, April 1999
- [McNab, 1996] McNab R.J, Smith L.A, Witten I.H, Henderson C.L, Cunningham S.J, *Toward the digital music library: tune retrieval from acoustic input*, Proceedings of the first ACM Conference on Digital Libraries (DL '96), pp. 11-18, 1996
- [Melucci, 2000] Melucci M, Orio N, *SMILE: a System for Content-based Musical Information Retrieval Environments*, Proceedings of RIAO Conference (RIAO 2000), Paris, France, April 2000
- [Minton, 1997] Minton G, *HOP specification: A closer look*, Unix Review, January 1997
- [Napster, 2000] *The Napster Music Community*, available at: [www.napster.com](http://www.napster.com), 2000
- [NCSTRL, 2002] NCSTRL: *The Networked Computer Science Technical Reference Library*, available at: <http://www.ncstrl.org>, 2002
- [Nielsen, 1993] Nielsen J, *Usability Engineering*, Academic Press Inc, 1993

[Nürnberg, 1995] Nürnberg P, Furuta R, Leggett J, Marshall C, Shipman F.M, *Digital Libraries: Issues and Architectures*, Proceedings of Advances in Digital Libraries Conference, 1995

[Oard, 1998] Oard D.W, Kim J, *Implicit feedback for Recommender Systems*, Proceedings of the AAAI Workshop on Recommender Systems, Madison, pp. 81-83, 1998

[OMG, 1997a] OMG, *CORBA Security Services*, Specification from the OMG, 1997

[OMG, 1997b] OMG, *CORBA Transaction Services*, Specification from the OMG, 1997

[OMG, 1998] OMG, *Common Object Request Broker Architecture (CORBA)*, Specification from the OMG, December 1998

[Page, 1998a] Page L, Brin S, *The Anatomy of a Large-Scale Hypertextual Search Engine*, Proceedings of the World Wide Web Conference (WWW '98) Brisbane, Elsevier Science BV, pp. 14-18, 1998

[Page, 1998b] Page L, Brin S, Motwani R, Winograd T, *The pagerank citation ranking: Bringing order to the web*, Technical report, Computer Science Department, Stanford University, 1998

[Papadakis, 2000a] Papadakis I, Chrissikopoulos V, Polemi D, *A Secure Web-based Medical Digital Library Architecture Based on TTPs*, Proceedings of the 16<sup>th</sup> International Conference of Medical Infobahn in Europe - XVI MIE 2000, pp. 610-616, IOS Press, 2000

[Papadakis, 2000b] Papadakis I, Chrissikopoulos V, *A Digital Library Framework based on XML*, Proceedings of the 3rd International Conference of Asian Digital Library - ICADL '00, pp. 81-88, Seoul, 2000

- [Papadakis, 2001a] Papadakis I, Douligeris C, *Design and Architecture of a Digital Music Library on the Web*, New Review on Hypermedia and Multimedia -NRHM, Vol. 7, pp. 115-126, Taylor-Graham, 2001
- [Papadakis, 2001b] Papadakis I, Chrissikopoulos V, Polemi D, *Secure Medical Digital Libraries*, International Journal of Medical Informatics, Vol. 64, No. 2-3, pp. 417-428, Elsevier Science, 2001
- [Papadakis, 2001c] Papadakis J, Douligeris C, *Design and Architecture of a Digital Music Library*, Proceedings of the 3<sup>rd</sup> International Conference on Information Integration and Web-based Applications & Services - IIWAS 2001, pp. 445-447, Austria, 2001
- [Papadakis, 2002a] Papadakis I, Andreou I, Chrissikopoulos V, *Interactive Search Results*, Proceedings of 6th European Conference on Digital Libraries - ECDL 2002, pp. 448-462, Elsevier Science, Rome, 2002
- [Papadakis, 2002b] Papadakis I, Karakoidas V, Chrissikopoulos V, *DocML: A Digital Library of University Data*, Library Hi-tech, Vol. 20, No.3, pp. 378 - 387, Emerald Publications, 2002
- [Pardi, 1999] Pardi W. J., *XML in action*, Microsoft Press, ISBN: 0-7356-0562-9, 1999
- [Payette, 1998] Payette S, Lagoze C, *Flexible and Extensible Digital Object and Repository Architecture (FEDORA)*, Proceedings of European Conference on Digital Libraries (ECDL 98), LNCS, Springer Verlag, pp. 41-59, 1998
- [Perfect, 2001] Perfect: *The Perfect search engine*, available at: <http://www.perfectsearch.com>, 2001



[Pinkerton, 1994] Pinkerton B, *Finding What People Want: Experiences with the WebCrawler*, Proceedings of the World Wide Web 1994 Conference (WWW '94) Geneva, Elsevier Science BV, pp. 17-20, available at: <http://info.webcrawler.com/bp/WWW94.html>, 1994

[Polemi, 1997] Polemi D. (ed), *EUROMED-ETS, Deliverable No.3, Pilot and Validation of Security Measures in EUROMED*, INFOSEC Project 20820, CEC, DGXIII B-6, September, 1997

[Polemi, 1998] Polemi D, Marsh A, *Secure telemedicine applications*, Proceedings of the High-Performance Computing and Networking Conference (HPCN '98), pp. 1015-1017, 1998

[Powel, 1998] Powel J, Fox E, *Multilingual Federated Searching Across Heterogeneous Collections*, D-lib Magazine, Sep. 1998, available at: [www.dlib.org](http://www.dlib.org), 1998

[Prior, 1998] Prior F.W, Glicksman R, De Greef B, Wong S, *CPR Architecture based on Distributed Objects and Web Technologies*, Proceedings of the Healthcare Information and Management Systems Society Conference (HIMMS '98), pp. 30-38, 1998

[Rauber, 1999] Rauber A, *Management of Distributed Information Repositories*, Department of Software Technology, Vienna University of Technology, 1999

[Revet, 1997] Revet B, *DICOM cook book*, PHILIPS Medical Systems, 1997

[Rijsbergen, 1979] van Rijsbergen C.J, *Information Retrieval*, Butterworths, London, 1979

[Rocchio, 1971] Rocchio J. J, *Relevance feedback in information retrieval*, pp. 313-323, Prentice Hall Inc, 1971

- [Roland, 2000] Roland P, *XMLAMIR: Extensible Markup Language for Music Information Retrieval*, Proceedings of International Symposium on Music Information Retrieval (Music IR 2000), available at: [http://ciir.cs.umass.edu/music2000/papers/roland\\_paper.pdf](http://ciir.cs.umass.edu/music2000/papers/roland_paper.pdf), 2000
- [Rothenberg, 1999] Rothenberg J, *Avoiding Technological Quickstand: Finding a Viable Technical Foundation for Digital Preservation*, Rep. to Council on Library and Information Resources, available at: <http://www.clir.org/pubs/reports/Rothenberg/pub77.pdf>, 1999
- [Rusch-Feja, 1998] Rusch-Feja D, *Metadata: Standards for retrieving WWW documents*, Library and Information Services in Astronomy III, pp. 157-165, 1998
- [Salton, 1971] Salton G, *The SMART Retrieval System*, Englewood Cliffs, NJ: Prentice Hall, 1971
- [Salton, 1989] Salton G, *Automatic text processing: The transformation, analysis and retrieval of information by computer*, Addison Wesley, 1989
- [Salton, 1990] Salton G, Buckley C, *Improving retrieval performance by relevance feedback*, Journal of the American Society for Information Science, No. 41, pp. 288-297, 1990
- [Schäuble, 1998] Schäuble P, Smeaton A.F, *An International Research Agenda for Digital Libraries*, Summary Report of the Series of Joint NSF-EU Working Groups on Future Directions for Digital Libraries Research, 1998
- [Schneiderman, 2000] Schneiderman B, Feldman D, Rose A, Grau X.F, *Visualizing Digital Library Search Results with Categorical and Hierarchical Axes*, Proceedings of Digital Libraries 2000, San Antonio TX, ACM, pp. 57-65, 2000

[Spinellis, 1999] Spinellis D, Gritzalis S, Iliadis J, Gritzalis D, Katsikas S, *Trusted Third Party Services for deploying secure telemedical applications over the WWW*, Computers & Security, Vol. 18, No. 7, pp. 627-639, 1999

[The BEST Search Engines, 2002] *The BEST Search Engines*, available at: <http://lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html>, 2002

[Theng, 1999] Theng Y.L, Duncker E, Mohd-Nasir, N, Buchanan G, Thimbleby H, *Design Guidelines and User-Centered Digital Libraries*, Proceedings of the 3<sup>rd</sup> European Conference in Digital Libraries (ECDL '99), LNCS, Springer Verlag, pp. 167-183, 1999

[Trusthealth, 1999] *Trusthealth I-II*, available at: <http://www.ehto.be/projects/trusthealth/>, June, 1999

[UPLN, 2002] UPLN: *University of Piraeus Lecture Notes digital library*, available at: <http://thalis.cs.unipi.gr/upln/search.html>, 2002

[Varvitsionis, 1998] Varvitsionis A, Polemi D, Marsh A, *EUROMED-Java: Trusted Third Party Services for securing medical java applets*, Proceedings of the European Symposium On Research in Computer Security Conference (ESORICS '98), LNCS, Springer Verlag, Vol. 1485, pp. 209-220, 1998

[Verderio, 2000] Verderio V, Cooper P, *A core middleware service in H.I.S: the experience of IC\_PIDRM*, Proceedings of the Medical Infobahn in Europe Conference (MIE 2000), IOS Press, pp. 964-968, 2000

[Vinoski, 1996] Vinoski S, *CORBA: Integrating Diverse Applications Within Distributed Heterogeneous Environments*, IONA Technologies, 1996

[W3C, 1998] The W3C group: *XML Specifications*, available at: <http://www.w3.org/XML/#9802xml10>, 1998



[W3C, 1999a] The W3C group: *XML namespaces Version 1.0*, available at: <http://www.w3.org/XML/#9901names>, 1999

[W3C, 1999b] The W3C group: *XML Path Language (XPath) Version 1.0*, available at: <http://www.w3.org/TR/xpath>, 1999

[W3C, 1999c] The W3C group: *XSL Transformations (XSLT) Version 1.0*, available at: <http://www.w3.org/TR/xslt>, 1999

[Ward, 1998] Ward N, Wood A, *Emerging Technologies for Networked Information Discovery: Beyond Z39.50*, Proceedings of the Victorian Association for Library Automation 9<sup>th</sup> Biennial Conference (VALA 1998), Melbourne, Australia available at: [http://www.dstc.edu.au/RDU/publications/e\\_vala98/](http://www.dstc.edu.au/RDU/publications/e_vala98/), 1998

[Wells, 1997] Wells D, Kurien A, *Searching and Indexing*, available at: <http://www.objs.com/survey/crawl.htm>, 1997

[Wilson, 1997] Wilson W, Beznosov K, *CORBAmed Security White Paper*, OMG, November 1997

[Wreder, 1998] Wreder K, Beznosov K, Bramblett A, Butler E, D'Empaire A, Hernandez E, Navarro E, *Architecting a Computerized Patient Record with Distributed Objects*, Proceedings of the Healthcare Information and Management Systems Society Conference, (HIMMS '98), pp. 149-158, 1998

[Zamir, 1998] Zamir O, Etzioni O, *Web document clustering: a feasibility demonstration*, Proceedings of SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98), Melbourne, pp. 46-54, 1998

[Παπαδάκης, 1999] Παπαδάκης Ι, Χρυσικόπουλος Β, Πολέμη Δ, *Ασφάλεια στις Ιατρικές Ψηφιακές Βιβλιοθήκες*, Πρακτικά 2ου Πανελληνίου Συνεδρίου Βιοϊατρικής Τεχνολογίας, σελ. 153-157, Νοέμβριος 1999

## ΠΑΡΑΡΤΗΜΑ ΚΕΦΑΛΑΙΟΥ 5

### ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ (στην Αγγλική γλώσσα)

- 1.1 How helpful did you find the options "filter" and "rank by" in locating useful resources?
- 1.2 Do you believe that the provided functionality at the left frame of the UI is a good alternative to exhaustive browsing of search results?
- 1.3 How often did you change the default option of ranking search results?
- 2.1 Is the functionality provided by the system easy to access?
- 2.2 Do you feel comfortable with the overall screen layout?
- 3.1 How long did it take you to get used to the provided User Interface of the system?
- 3.2 How often did you use the provided help?
- 4.1 How long did it take for your actions at the left frame of the Use Interface to be applied at the main frame?
- 4.2 Did you feel confident with the filter and the rank by criteria of the system?

- 4.3 Please order the ranking and filtering possibilities provided at the left frame in an order that describes the usefulness of each choice regarding the top items of the corresponding search result lists (i.e. which choice ranks search results more efficiently?).
- A. Relevancy
  - B. Importance
  - C. Total hits
  - D. Popularity
  - E. Filter by
- 5.1 How easy was to move from one view of the search results to another (e.g. from the filtered view to the rank by importance view)?
- 6.1 Did you feel that in some cases, exhaustive browsing could be a faster method for locating useful search results?
- 6.2 Provided that the queries returned relatively long search results lists, did you locate any useful resources in other pages except the first one?
- 6.3 In cases where the search engine didn't return any useful results, do you think that the functionality provided at the left frame was helpful in deciding quicker that you had to rephrase the original query in order to get useful results?
- 6.4 Do you feel that ranking by relevancy should be the only ranking criterion in a search results list?