



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

Σχολή Επιστήμης της Πληροφορίας και Πληροφορικής

ΤΜΗΜΑ ΑΡΧΕΙΟΝΟΜΙΑΣ, ΒΙΒΛΙΟΘΗΚΟΝΟΜΙΑΣ ΚΑΙ ΜΟΥΣΕΙΟΛΟΓΙΑΣ

Δεκέμβριος 2015

Υπηρεσίες Πληροφόρησης Βασισμένες σε Ελεγχόμενα Λεξιλόγια και στις Τεχνολο- γίες του Σημασιολογικού Ι- στού

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Κωνσταντίνος Κυπριανός

Κέρκυρα



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

Σχολή Επιστήμης της Πληροφορίας και Πληροφορικής

ΤΜΗΜΑ ΑΡΧΕΙΟΝΟΜΙΑΣ, ΒΙΒΛΙΟΘΗΚΟΝΟΜΙΑΣ ΚΑΙ ΜΟΥΣΕΙΟΛΟΓΙΑΣ
Ιωάννου Θεοτόκη 72, 49100 Κέρκυρα, τηλ. 26610-87423 fax 26610-87433

Κωνσταντίνος Κυπριανός

**Υπηρεσίες Πληροφόρησης Βασισμένες σε Ελεγ-
χόμενα Λεξιλόγια και στις Τεχνολογίες του Ση-
μασιολογικού Ιστού**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Επιβλέπων: Ιωάννης Παπαδάκης, Επίκουρος Καθηγητής

Τριμελής συμβουλευτική επιτροπή
Σαράντος Καπιδάκης, Καθηγητής
Μάριος Πούλος, Αναπληρωτής Καθηγητής
Ιωάννης Παπαδάκης, Επίκουρος Καθηγητής

Κέρκυρα

Δεκέμβριος, 2015



© 2015, Κωνσταντίνος Κυπριανός

**Αναφορά Δημιουργού-Μη Εμπορική Χρήση (Attribution – Non Commercial)
4.0**

Σύμφωνα με την αναγραφόμενη άδεια Creative Commons επιτρέπεται η αναδιανομή, η διασκευή, η τροποποίηση και η δημιουργία παράγωγων έργων με την προϋπόθεση να γίνεται αναφορά στο δημιουργό του πρωτοτύπου έργου, ενώ απαγορεύεται η χρήση αυτού του έργου με κύριο σκοπό την εμπορική του εκμετάλλευση.



Ionian University

Faculty of Information Science and Informatics

DEPARTMENT OF ARCHIVES, LIBRARY SCIENCE AND MUSEOLOGY

Ioannou Theotoki 72, 49100 Corfu, tel. 26610-87423 fax 26610-87433

Konstantinos Kyprianos

Information Services Based on Controlled Vo- cabularies and Semantic Web Technologies

DOCTORAL THESIS

Supervisor: Ioannis Papadakis, Assistant Professor

Advisory committee

Sarantos Kapidakis, Professor

Marios Poulos, Associate Professor

Ioannis Papadakis, Assistant Professor

Corfu

December, 2015



© 2015, Konstantinos Kyprianos

Attribution – Non Commercial 4.0

According to the displayed Creative Commons license, you are free to copy and reuse any of this work as long as you tell people where it is from and you are not making money from it. You do not need my permission to post my work or ideas on your website, just mention that (some or all) of it came from here, and include a clear and obvious link back to the appropriate page. You are also welcome to use my work or ideas in publications, curricula for seminars, and other media (all with appropriate attribution).

Περίληψη

Μια από τις πιο βασικές λειτουργίες που επιτελούν παραδοσιακά οι βιβλιοθήκες και τα ιδρύματα μνήμης γενικά, είναι η παροχή υπηρεσιών πληροφόρησης προς τους χρήστες. Οι υπηρεσίες αυτές έχουν ως στόχο να βοηθήσουν τους χρήστες στην ανάκτηση πληροφοριών έτσι ώστε να μπορέσουν να ικανοποιήσουν τις πληροφοριακές τους ανάγκες. Μια από τις πιο σημαντικές διαδικασίες για τη βοήθεια των χρηστών στην ανάκτηση πληροφοριών που παρείχαν ανέκαθεν οι βιβλιοθήκες είναι η θεματική επεξεργασία της πληροφορίας. Η θεματική επεξεργασία γίνεται με τη χρήση συγκεκριμένων θεματικών εργαλείων τα οποία ονομάζονται ελεγχόμενα λεξιλόγια. Μέσα στο χρόνο οι βιβλιοθηκονόμοι και γενικότερα οι επιστήμονες της πληροφόρησης έχουν δείξει μεγάλη συνέπεια στη δημιουργία ποικίλων ειδών ελεγχόμενων λεξιλογίων για την κάλυψη συγκεκριμένων αναγκών που υπάρχουν σε διαφορετικά επιστημονικά πεδία. Όμως, η πρακτική αυτή οδήγησε στη δημιουργία σημασιολογικά παρόμοιων ελεγχόμενων λεξιλογίων που είναι συντακτικά και γλωσσικά ετερογενή με αλληλοεπικαλυπτόμενα μέρη, με αποτέλεσμα η αναζήτηση και η ανάκτηση πληροφορίας ανάμεσα σε διαφορετικούς οργανισμούς να έχει γίνει μια δύσκολη και κουραστική διαδικασία. Επιπλέον, οι χρήστες πολλές φορές αγνοούν τη γλώσσα ευρετηρίασης που χρησιμοποιείται από τις βιβλιοθήκες με αποτέλεσμα να υπάρχει δυσκολία στον εντοπισμό των πηγών που χρειάζονται για την κάλυψη των πληροφοριακών τους αναγκών. Το πρόβλημα γίνεται ακόμα πιο οξύ, αφού τα ήδη υπάρχοντα βιβλιοθηκονομικά συστήματα τις περισσότερες φορές δίνουν τη δυνατότητα στο χρήστη να ανακτήσει πληροφορίες μόνο μέσω της χρήσης ενός αλφαβητικού ευρετηρίου, αφήνοντας ανεκμετάλλευτο τον τεράστιο σημασιολογικό πλούτο που μπορούν να παρέχουν τα ελεγχόμενα λεξιλόγια.

Στις μέρες μας, η τεχνολογική εξέλιξη και οι ολοένα αυξανόμενες απαιτήσεις των χρηστών για καλύτερη, γρηγορότερη και ποιοτικότερη εύρεση πληροφοριών έχει αυξήσει τους ρυθμούς ανάπτυξης νέων ψηφιακών υπηρεσιών πληροφόρησης. Η χρήση του διαδικτύου και των τεχνολογιών του έχει ωφελήσει και ενισχύσει σε μεγάλο βαθμό την παροχή τέτοιων υπηρεσιών και από τις βιβλιοθήκες. Μάλιστα, τα τελευταία χρόνια με την εμφάνιση του σημασιολογικού ιστού, οι βιβλιοθήκες έχουν στα χέρια τους ένα ισχυρό εργαλείο που θα τους βοηθήσει στη λύση προβλημάτων που τις ταλανίζουν εδώ και δεκαετίες, ιδιαίτερα στον τομέα της διαλειτουργικότητας και στην αποτελεσματικότερη αναζήτηση και ανάκτηση πληροφοριών.

Κύρια επιδίωξη λοιπόν, της παρούσας διατριβής είναι η έρευνα και η δημιουργία καινοτόμων υπηρεσιών, μεθοδολογιών και εργαλείων, με τα οποία τα παραδοσιακά

εργαλεία που χρησιμοποιούνται για τη θεματική επεξεργασία του υλικού των βιβλιοθηκών, να μπορέσουν να εξελιχθούν σε σύγχρονα εργαλεία αναζήτησης με τη βοήθεια του σημασιολογικού ιστού. Ανώτερος στόχος είναι οι νέες αυτές υπηρεσίες να βοηθούν τον τελικό χρήστη στην ευκολότερη και γρηγορότερη ανάκτηση πληροφοριών με ένα διαφανή και ξεκάθαρο τρόπο.

Λέξεις κλειδιά: Θεματική ευρετηρίαση, Ελεγχόμενα λεξιλόγια, Σημασιολογικός ιστός, Ανοιχτά συνδεδεμένα δεδομένα, Λαοταξινομίες, MARC

Abstract

One of the basic functions that traditionally libraries and memory institutions, in general, perform is the provision of information services to the users. Such services are aiming at helping users with information retrieval so that they can satisfy their information needs. One of the most important processes in helping users with information retrieval that libraries have always provided is the subject processing of information. Subject processing is achieved through the employment of subject tools, namely controlled vocabularies. Throughout the years, librarians and information scientists, in general, have demonstrated great consistency in creating various types of controlled vocabularies to cover specific needs that exist in different scientific fields. However, such a practice led to the creation of semantically similar controlled vocabularies that are syntactically and linguistically heterogeneous with overlapping parts. Consequently, information search and retrieval process among different organizations and more specifically libraries has become a difficult and tedious task. Additionally, in many cases, users ignore the indexing language that is used by libraries rendering it difficult to locate the resources that they need to fulfill their information needs. The problem gets even worse, since the existing library systems, in many cases, give the opportunity to the users to retrieve information only by the employment of alphabetical indexes, leaving the semantic richness that controlled vocabularies may provide intact.

Nowadays, the technological advancements and the users' increasing demands for better, faster and more qualitative information discovery has increased the development pace of new digital information services. The employment of the internet and its technologies has greatly benefited and enhanced the provision of such services from libraries. In fact, over the past few years, with the advent of semantic web, libraries have a new powerful tool that will help them to solve problems that have been plaguing them for decades, especially in the area of interoperability and provide a more effective search and retrieval process.

Therefore, the main objective of the specific thesis is the research and the creation of innovative services, tools and methodologies, with which the traditional tools that are employed for the subject process of libraries' resources, will evolve into modern search tools with the help of the semantic web. The ultimate goal is for such new services to help the end-user search and retrieve information in a more transparent and efficient way.

Keywords: Subject indexing, Controlled vocabularies, Semantic web, Linked open data, Folksonomies, MARC

Πρόλογος

Η παρούσα διδακτορική διατριβή για τη μελέτη των υπηρεσιών πληροφόρησης που βασίζονται στα ελεγχόμενα λεξιλόγια και στις τεχνολογίες του σημασιολογικού ιστού είχε ως έναυσμα, αφενός τις ανάγκες που διαπιστώθηκαν στο συγκεκριμένο τομέα από την καθημερινή μου επαφή με τους χρήστες της βιβλιοθήκης του Πανεπιστημίου Πειραιά, αφετέρου το πάθος μου με τις νέες τεχνολογίες και κατά πόσο αυτές θα μπορούσαν να εφαρμοστούν και στον τομέα των βιβλιοθηκών.

Πιο συγκεκριμένα, κατά τη διάρκεια της εργασίας μου παρατηρήθηκε ότι οι χρήστες αντιμετωπίζουν προβλήματα κατά τη διάρκεια της αναζήτησης και του εντοπισμού πηγών που σχετίζονται με τις πληροφοριακές τους ανάγκες. Μέσα σε αυτό το πλαίσιο, μελετήθηκαν τα παραδοσιακά εργαλεία, όπως είναι τα ελεγχόμενα λεξιλόγια και οι δυνατότητες που μπορούν να προσφέρουν σε μια βιβλιοθήκη, καθώς και οι δυνατότητες που προσφέρουν οι νέες τεχνολογίες του σημασιολογικού ιστού στις βιβλιοθήκες. Παράλληλα, μελετήθηκαν τα κενά και τα προβλήματα που προκύπτουν κατά τη χρήση και την ανάπτυξη αυτών των εργαλείων, με απώτερο στόχο να δημιουργηθούν καινοτόμες υπηρεσίες και μεθοδολογίες που θα είναι ικανές να βοηθήσουν τους χρήστες στην ευκολότερη και γρηγορότερη κάλυψη των πληροφοριακών τους αναγκών.

Σε αυτό το σημείο, θα ήθελα να ευχαριστήσω τον επιβλέποντα επίκουρο καθηγητή μου κ. Ιωάννη Παπαδάκη, για την άψογη συνεργασία που είχαμε καθ' όλη τη διάρκεια της εκπόνησης της διδακτορικής μου διατριβής και για την καθοδήγηση που μου παρείχε, βοηθώντας με να κατανοήσω πολλά θέματα της θεωρίας αλλά και στην ουσιαστική συμβολή του κατά την υλοποίηση των μεθοδολογιών και των υπηρεσιών που παρουσιάζονται σε αυτή τη διατριβή. Η βοήθειά του ήταν καθοριστική για να μπορέσω να εντρυφήσω όσο το δυνατόν περισσότερο στα θέματα που πραγματεύεται η παρούσα διδακτορική διατριβή και να κατανοήσω τις τεχνολογίες που απαιτούνται για την υλοποίηση αυτών.

Παράλληλα, θα ήθελα να ευχαριστήσω τα υπόλοιπα μέλη της τριμελούς επιτροπής, τον καθηγητή κ. Σαράντο Καπιδάκη και τον καθηγητή κ. Μάριο Πούλο για τις εύστοχες παρατηρήσεις που μου έκαναν για την ουσιαστική βελτίωση του περιεχομένου της διατριβής.

Εκτός όμως από την τριμελή επιτροπή θα ήθελα να ευχαριστήσω και όλους τους ανθρώπους από τον κοινωνικό μου περίγυρο, συγγενείς και φίλους, για την υπομονή

και τη συμπαράσταση που μου παρείχαν καθ' όλη τη διάρκεια της εκπόνησης της συγκεκριμένης διατριβής. Ολοκληρώνοντας, θα ήθελα να ευχαριστήσω και τη Βιβλιοθήκη του Πανεπιστημίου Πειραιά για την πρόσβαση που μας έδωσε στις ψηφιακές της υπηρεσίες για να μπορέσουμε να εφαρμόσουμε τις μεθοδολογίες, τις τεχνικές και τις υπηρεσίες που παρουσιάζονται στη διατριβή.

Κωνσταντίνος Κυπριανός,
Κέρκυρα, 10 Δεκεμβρίου 2015

Περιεχόμενα

Περίληψη.....	vi
Abstract.....	viii
Πρόλογος.....	x
Κατάλογος εικόνων.....	xvi
Κατάλογος πινάκων.....	xix
1. Εισαγωγή.....	1
1.1. Αντικείμενο της διατριβής.....	1
1.2. Συνεισφορά της διατριβής.....	2
1.3. Αποτύπωση του ερευνητικού χώρου που κινείται η διατριβή.....	3
1.3.1. Θεματική ευρετηρίαση στις βιβλιοθήκες.....	3
1.3.2. Σημασιολογικός ιστός.....	5
1.4. Ερευνητική μεθοδολογία.....	9
1.5. Συναφείς δημοσιεύσεις.....	10
1.6. Διάρθρωση της διατριβής.....	11
2. Θεματική ευρετηρίαση.....	15
2.1. Θεματική ευρετηρίαση στις βιβλιοθήκες.....	15
2.2. Ελεγχόμενα λεξιλόγια.....	16
2.3. Είδη ελεγχόμενων λεξιλογίων.....	17
2.3.1. Θεματικές επικεφαλίδες.....	17
2.3.2. Θησαυροί.....	21
2.3.3. Ταξινομικά σχήματα.....	26
2.4. Θεματική αναζήτηση και ανάκτηση μέσω της χρήσης των LCSH και των θησαυρών.....	28
2.5. Προτεινόμενη προσέγγιση στη θεματική ανάκτηση πληροφοριών.....	30
2.5.1. Επεκτείνοντας τη συνδετική δομή των LCSH.....	31
2.5.2. Μοντελοποίηση των σχέσεων ανάμεσα στις LCSH σε ένα οντολογικό σχήμα.....	31
2.5.3. Από τη θεωρία στην πράξη: σημασιολογική ανάκτηση πληροφοριών με τη χρήση των LCSH.....	33
2.5.4. Μελέτη περίπτωσης.....	36
2.6. Μη ελεγχόμενα λεξιλόγια.....	45
2.6.1. Λασταξινομίες.....	46
2.7. Συμπεράσματα.....	47
3. Συγχώνευση ελεγχόμενων και μη λεξιλογίων.....	51
3.1. Αιτίες που οδηγούν στη συγχώνευση λεξιλογίων.....	51

3.2. Προβλήματα που προκύπτουν κατά τη συγχώνευση ελεγχόμενων λεξιλογίων	52
3.3. Συγχώνευση θησαυρών.....	52
3.4. Συγχώνευση θεματικών επικεφαλίδων με θησαυρούς	53
3.5. Συγχώνευση ταξινομικών σχημάτων με θεματικές επικεφαλίδες	54
3.6. Συγχώνευση λαοταξινομιών με ελεγχόμενα λεξιλόγια	55
3.7. Παρουσίαση μεθοδολογίας για τη συγχώνευση θησαυρών με τις LCSH	56
3.7.1. Μεθοδολογία αντιστοίχισης όρων LCSH και θησαυρών	59
3.7.2. Μοντελοποίηση οντολογίας στο SKOS.....	62
3.7.3. Τεκμηρίωση της μεθοδολογίας	63
3.7.4. Αξιολόγηση της εφαρμογής.....	66
3.8. Συμπεράσματα	68
4. Τα Συνδεδεμένα Δεδομένα	71
4.1. Γενικές πληροφορίες	71
4.2. Τι είναι τα Συνδεδεμένα Δεδομένα.....	72
4.3. Κανόνες των Συνδεδεμένων Δεδομένων	72
4.4. Τεχνολογίες των Συνδεδεμένων Δεδομένων	74
4.4.1. URIs (καλά URIs).....	75
4.4.2. Μοντελοποίηση Συνδεδεμένων Δεδομένων	77
4.4.3. Σειριοποίηση του RDF	80
4.4.4. SKOS	84
4.4.5. Πρόσβαση στα δεδομένα	88
4.5. Συμπεράσματα	90
5. Υπηρεσία πλοήγησης θεματικών επικεφαλίδων βασισμένη στα Συνδεδεμένα Δεδομένα.....	93
5.1. Ελεγχόμενα λεξιλόγια ως Συνδεδεμένα Δεδομένα	93
5.2. Προτεινόμενη προσέγγιση στη δημιουργία υπηρεσίας θεματικής πλοήγησης βασισμένη στα Συνδεδεμένα Δεδομένα	95
5.2.1. Αρχιτεκτονική	96
5.2.2. Μοναδικά URIs στις υπηρεσίες θεματικής πλοήγησης	98
5.2.3. Συνδεδεμένα Δεδομένα βασισμένα στο SKOS.....	98
5.2.4. Διαδικασία ευθυγράμμισης	99
5.2.5. Γραφικό περιβάλλον χρήσης	100
5.2.6. Αξιολόγηση	106
5.3. Συμπεράσματα	109

6. Ευθυγράμμιση ελεγχόμενων λεξιλογίων μέσω της χρήσης υπηρεσιών σε Συνδεδεμένα Δεδομένα	111
6.1. Τεχνικές για την αντιστοίχιση ελεγχόμενων λεξιλογίων	111
6.2. Τεχνικές ταιριάσματος σε επίπεδο στοιχείων	111
6.3. Προτεινόμενη μεθοδολογία στη συγχώνευση ελεγχόμενων λεξιλογίων σε Συνδεδεμένα Δεδομένα	113
6.3.1. Εφαρμογή της μεθοδολογίας	115
6.3.2. Υλοποίηση της μεθοδολογίας	118
6.3.3. Συγκριτική αξιολόγηση	122
6.4. Συμπεράσματα	124
7. Συνδεδεμένα Δεδομένα, MARC και βιβλιοθήκες	125
7.1. MARC και βιβλιοθήκες	125
7.2. MARC και Συνδεδεμένα Δεδομένα	126
7.2.1. MARC21, UNIMARC και URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί τοπικά	128
7.2.2. MARC21, UNIMARC και URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί σε απομακρυσμένα σημεία	130
7.3. Εθνικές βιβλιοθήκες, καθιερωμένες εγγραφές, MARC και Συνδεδεμένα Δεδομένα	138
7.3.1. Βιβλιοθήκη του Κογκρέσου	139
7.3.2. Βρετανική Βιβλιοθήκη	140
7.3.3. Εθνική Βιβλιοθήκη της Γαλλίας	141
7.3.4. Εθνική Βιβλιοθήκη της Γερμανίας	142
7.3.5. Εθνική Βιβλιοθήκη της Ισπανίας	143
7.3.6. Εθνική Βιβλιοθήκη της Σουηδίας	144
7.3.7. Εθνική Βιβλιοθήκη της Ουγγαρίας	145
7.4. Δημοσίευση καθιερωμένων εγγραφών του MARC ως Συνδεδεμένα Δεδομένα	145
7.4.1. Μοντελοποίηση δεδομένων	146
7.4.2. Πρόσβαση στα δεδομένα	147
7.4.3. Διασύνδεση δεδομένων	149
7.5. Συμπεράσματα	150
8. Λασταξινομίες και βιβλιοθήκες	153
8.1. Το πρόβλημα της χρονικής διάστασης της πληροφορίας στις βιβλιοθήκες	153
8.2. Χρονική διάσταση της πληροφορίας (επικαιρότητα)	154
8.2.1. Χρονική διάσταση της πληροφορίας και βιβλιοθήκες	154
8.2.2. Web 2.0, χρονική διάσταση της πληροφορίας και βιβλιοθήκες	157

8.2.3. Twitter και βιβλιοθήκες.....	158
8.2.4. Υλοποίηση της ιδέας.....	159
8.2.5. Εφαρμογή της προτεινόμενης υπηρεσίας.....	163
8.2.6. Αξιολόγηση της υπηρεσίας.....	164
8.3. Συμπεράσματα.....	167
9. Συμπεράσματα.....	169
9.1. Ανακεφαλαίωση.....	169
9.2. Πιθανή μελλοντική εργασία.....	173
9.2.1. Εκμετάλλευση της σημασιολογικής εκφραστικότητας των LCSH.....	174
9.2.2. Μεθοδολογία για τη συγχώνευση ελεγχόμενων λεξιλογίων.....	174
9.2.3. Μεθοδολογία μετατροπής θεματικών ευρετηρίων σε μορφή των Συνδεδεμένων Δεδομένων.....	175
9.2.4. Μεθοδολογία για την αντιστοίχιση σημασιολογικά παρόμοιων ελεγχόμενων λεξιλογίων μέσω τεχνολογιών των Συνδεδεμένων Δεδομένων.....	175
9.2.5. Τα hashtags στην αντιμετώπιση του προβλήματος της χρονικής διάστασης της πληροφορίας.....	176
Βιβλιογραφία.....	177
Αρκτικόλεξα – Ακρωνύμια.....	195
Απόδοση ξένων όρων.....	197
Ευρετήριο.....	207

Κατάλογος εικόνων

Εικόνα 1.1: Δομικά στοιχεία του σημασιολογικού ιστού	6
Εικόνα 2.1: Το σχήμα της οντολογίας.....	32
Εικόνα 2.2: Πεδίο αυτόματης συμπλήρωσης	34
Εικόνα 2.3: Σχέσεις μεταξύ των θεματικών επικεφαλίδων	35
Εικόνα 2.4: Συνδέσεις μεταξύ των θεματικών επικεφαλίδων.....	35
Εικόνα 2.5: Εκτελώντας μια παραδοσιακή θεματική αναζήτηση	37
Εικόνα 2.6: Αποτελέσματα παραδοσιακής θεματικής αναζήτησης.....	38
Εικόνα 2.7: Παραδοσιακή προσέγγιση: αναζήτηση του θέματος "Instructional systems"	39
Εικόνα 2.8: Παραδοσιακή προσέγγιση: τεκμήρια που περιγράφονται με το θέμα "Instructional systems"	39
Εικόνα 2.9: Προτεινόμενη προσέγγιση: αναζήτηση του θέματος "Instructional systems"	40
Εικόνα 2.10: Προτεινόμενη προσέγγιση: τεκμήρια που περιγράφονται με το θέμα "Instructional systems - Design"	40
Εικόνα 2.11: Προτεινόμενη προσέγγιση: τεκμήρια που περιγράφονται με το θέμα "Web- based instruction - Design"	41
Εικόνα 2.12: Παραδοσιακή προσέγγιση: αναζήτηση του θέματος "Computer networks - Security measures"	41
Εικόνα 2.13: Παραδοσιακή προσέγγιση: τεκμήρια που περιγράφονται με το θέμα "Computer networks - Security measures"	42
Εικόνα 2.14: Προτεινόμενη προσέγγιση: αναζήτηση του θέματος "Computer networks - Security measures", μέρος 1 ^ο	42
Εικόνα 2.15: Προτεινόμενη προσέγγιση: αναζήτηση του θέματος "Computer networks - Security measures", μέρος 2 ^ο	43
Εικόνα 2.16: Παραδοσιακή προσέγγιση: αναζήτηση του θέματος "Accounting"	44
Εικόνα 2.17: Προτεινόμενη προσέγγιση: εξειδίκευση του θέματος "Accounting"	44
Εικόνα 3.1: Συνολική αρχιτεκτονική του συστήματος.....	58
Εικόνα 3.2: Προτεινόμενη μεθοδολογία αντιστοίχισης.....	60
Εικόνα 3.3: Κοινή δομή.....	61
Εικόνα 3.4: Απεικόνιση της οντολογίας σε SKOS.....	63
Εικόνα 3.5: Πεδίο αυτόματης συμπλήρωσης	64
Εικόνα 3.6: Σχεδιασμένο πλαίσιο του όρου "Accounting"	65
Εικόνα 3.7: Αποτελέσματα αναζήτησης για τον όρο "Accounting".....	65

Εικόνα 3.8: Ο όρος "Professional employees" σχετίζεται μέσω της σχέση NT με τους όρους: "Business people", "Engineering Personnel" και "Accountants"	65
Εικόνα 3.9: Οι όροι "Professional employees" και "Accountants" σχετίζονται μέσω της σχέσης NT	65
Εικόνα 3.10: Αποτελέσματα αναζήτησης για τον όρο "Accountants"	66
Εικόνα 4.1: Linked Data 5 star rating system	74
Εικόνα 4.2: Παράδειγμα τριπλετών	77
Εικόνα 4.3: Ιεραρχική δομή του RDFS	79
Εικόνα 4.4: Δομή της OWL 2	80
Εικόνα 4.5: Παράδειγμα σύνταξης σε RDF/XML	81
Εικόνα 4.6: Παράδειγμα σύνταξης σε RDFa	82
Εικόνα 4.7: Παράδειγμα σύνταξης σε Turtle	83
Εικόνα 4.8: Παράδειγμα σύνταξης σε N-Triples	84
Εικόνα 4.9: Παράδειγμα σύνταξης σε RDF/JSON	84
Εικόνα 5.1: Συνολική αρχιτεκτονική του συστήματος	97
Εικόνα 5.2: Η δίγλωσση λειτουργικότητα του πεδίου αυτόματης συμπλήρωσης	101
Εικόνα 5.3: Σχεδιασμένο πλαίσιο για το "Medical care" στις δύο γλώσσες, στα αγγλικά και στα ελληνικά	102
Εικόνα 5.4: Η θεματική επικεφαλίδα "Financial statements -- standards" και οι αντίστοιχες θεματικές επικεφαλίδες "standards" στις δύο γλώσσες και μόνο στα αγγλικά	103
Εικόνα 5.5: Οι θεματικές επικεφαλίδες "Advertising" και "Publicity" σχετίζονται μέσω της σχέσης "NT"	103
Εικόνα 5.6: Γραφικό περιβάλλον χρήστη με εμφάνιση και των σχετικών τεκμηρίων από την ψηφιακή βιβλιοθήκη (Διώνη)	104
Εικόνα 5.7: Η αρχική οθόνη του τελικού σημείου SPARQL της υπηρεσίας	105
Εικόνα 6.1: Προτεινόμενη μεθοδολογία	114
Εικόνα 6.2: Εφαρμογή της προτεινόμενης μεθοδολογίας	117
Εικόνα 6.3: Αποτελέσματα διαδικασίας	122
Εικόνα 6.4: Αποτελέσματα συγκριτικής ανάλυσης	123
Εικόνα 7.1: Παράδειγμα εγγραφής σε MARC	126
Εικόνα 7.2: Διασύνδεση δεδομένων βιβλιοθηκών στο σύννεφο των Συνδεδεμένων Δεδομένων	149
Εικόνα 8.1: Η συνολική αρχιτεκτονική της υπηρεσίας	160
Εικόνα 8.2: Αρχική σελίδα του OPAC της Βιβλιοθήκης και Κέντρου Πληροφόρησης του Ιονίου Πανεπιστημίου	163

Εικόνα 8.3: Αναζητήσεις που έγιναν στο σύστημα	165
Εικόνα 8.4: Αποτελέσματα του συστήματος και της προτεινόμενης υπηρεσίας.....	166
Εικόνα 8.5: Αποτελέσματα συστήματος	166
Εικόνα 8.6: Αποτελέσματα προτεινόμενης υπηρεσίας	167

Κατάλογος πινάκων

Πίνακας 4.1: Λεξιλόγιο SKOS για έννοιες και σχήματα εννοιών	86
Πίνακας 4.2: Λεξιλόγιο SKOS για λεξιλογικές ετικέτες	87
Πίνακας 4.3: Λεξιλόγιο SKOS για σημασιολογικές σχέσεις	87
Πίνακας 4.4: Λεξιλόγιο SKOS για τεκμηρίωση	88
Πίνακας 4.5: Λεξιλόγιο SKOS για σχέσεις αντιστοίχισης	88
Πίνακας 5.1: Χρήση λεξιλογίου SKOS	99
Πίνακας 6.1: Ποσοστό επιτυχίας της διαδικασίας συμφιλίωσης	120
Πίνακας 7.1: Πεδία του MARC που μπορούν να φιλοξενήσουν URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί τοπικά	128
Πίνακας 7.2: Πεδία του MARC που μπορούν να φιλοξενήσουν URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί σε απομακρυσμένα σημεία	131
Πίνακας 7.3: Πεδία UNIMARC/MARC21 και τα σημασιολογικά ισοδύναμα κατηγορήματα του SKOS	146
Πίνακας 7.4: Εθνικές βιβλιοθήκες και υπηρεσίες προς την κοινότητα των Συνδεδεμένων Δεδομένων	148

1. Εισαγωγή

1.1. Αντικείμενο της διατριβής

Οι υπηρεσίες πληροφόρησης αποτελούν παραδοσιακά ένα σημαντικό κομμάτι των υπηρεσιών που παρέχουν οι βιβλιοθήκες. Παραδοσιακά, μια από τις πιο σημαντικές βιβλιοθηκονομικές διαδικασίες για τη βοήθεια των χρηστών στην ανάκτηση πληροφοριών είναι η θεματική επεξεργασία της πληροφορίας. Η θεματική επεξεργασία γίνεται με τη χρήση συγκεκριμένων θεματικών εργαλείων τα οποία ονομάζονται ελεγχόμενα λεξιλόγια. Μέσα στο χρόνο οι βιβλιοθηκονόμοι έχουν δείξει μεγάλη συνέπεια στη δημιουργία ποικίλων ειδών ελεγχόμενων λεξιλογίων (π.χ. θεματικές επικεφαλίδες, θησαυροί, ταξινομικά σχήματα κ.ά.) για την κάλυψη συγκεκριμένων αναγκών που υπάρχουν σε διαφορετικά επιστημονικά πεδία.

Πολλές φορές οι βιβλιοθήκες χρησιμοποιούν διαφορετικά ελεγχόμενα λεξιλόγια για τη θεματική περιγραφή πηγών που υπάρχουν σε διαφορετικές υπηρεσίες, οι οποίες βρίσκονται ακόμα και μέσα στην ίδια βιβλιοθήκη. Αυτή η πρακτική είχε ως αποτέλεσμα τη δημιουργία σημασιολογικά παρόμοιων ελεγχόμενων λεξιλογίων που όμως είναι συντακτικά και γλωσσικά ετερογενή με αλληλοεπικαλυπτόμενα μέρη (Lin & Sandkuhl, 2008, σ. 341). Επομένως, σε πολλές περιπτώσεις, η αναζήτηση και η ανάκτηση πληροφορίας ανάμεσα σε διαφορετικούς οργανισμούς ή ακόμα και μέσα στους κόλπους του ίδιου οργανισμού έχει γίνει μια δύσκολη και κουραστική διαδικασία. Επιπλέον, τα ήδη υπάρχοντα βιβλιοθηκονομικά συστήματα τις περισσότερες φορές δίνουν τη δυνατότητα στο χρήστη να ανακτήσει πληροφορίες μόνο μέσω της χρήσης ενός αλφαβητικού ευρετηρίου, όπως είναι η δυνατότητα θεματικής αναζήτησης του Online Public Access Catalog – OPAC (Antelman, Lynema & Pace, 2006, σ. 128), αφήνοντας ανεκμετάλλευτο τον τεράστιο σημασιολογικό πλούτο που μπορούν να παρέχουν τα ελεγχόμενα λεξιλόγια. Παράλληλα, οι χρήστες τις περισσότερες φορές αγνοούν τη γλώσσα ευρετηρίασης που χρησιμοποιούν οι βιβλιοθήκες (δηλ. την ορολογία των ελεγχόμενων λεξιλογίων) με αποτέλεσμα να εντείνεται ακόμα περισσότερο το πρόβλημα της επιτυχημένης αναζήτησης και ανάκτησης πληροφοριών.

Τα τελευταία χρόνια, η χρήση του διαδικτύου και των τεχνολογιών του έχει αυξήσει τους ρυθμούς ανάπτυξης νέων ψηφιακών υπηρεσιών από τις βιβλιοθήκες, οι οποίες μπορούν να βελτιώσουν σημαντικά την αναζήτηση και ανάκτηση πληροφοριών. Ιδιαίτερα, με την εμφάνιση του σημασιολογικού ιστού, οι βιβλιοθήκες έχουν στα χέρια

τους ένα ισχυρό εργαλείο που θα συμβάλλει στη λύση προβλημάτων που τις ταλανίζουν εδώ και δεκαετίες, ιδιαίτερα στον τομέα της διαλειτουργικότητας. Έτσι, ο σημασιολογικός ιστός μπορεί να εξυπηρετήσει και να καλύψει την ανάγκη για ενιαία οργάνωση και διάθεση των δεδομένων, ώστε το διαδίκτυο να γίνει μια αποδοτική πλατφόρμα αυτόματης επεξεργασίας και ανταλλαγής πληροφοριών που προέρχονται από ποικίλες και ετερογενείς πηγές (Isaac et al., 2008, σ. 188). Όμως, παρά την ευρεία διαθεσιμότητα τέτοιων τεχνολογιών, φαίνεται να υπάρχει έλλειψη αντίστοιχων υπηρεσιών, οι οποίες να στοχεύουν στους τελικούς χρήστες. Οι βιβλιοθήκες δείχνουν να ενδιαφέρονται περισσότερο για το μετασχηματισμό και την έκθεση της θεματικής τους πληροφορίας σε μορφή που να είναι συμβατή με τις τεχνολογίες του σημασιολογικού ιστού, χωρίς να δίνουν αντίστοιχη βαρύτητα και προσοχή στην ευχρηστία και στη χρησιμότητα των υπηρεσιών που προκύπτουν για τους τελικούς χρήστες.

1.2. Συνεισφορά της διατριβής

Η βασική επιδίωξη της διατριβής συνίσταται στην πρόταση ενός συνόλου τεχνικών και μεθοδολογιών μέσω των οποίων τα παραδοσιακά εργαλεία που χρησιμοποιούνται για τη θεματική επεξεργασία του υλικού των βιβλιοθηκών, δηλαδή τα ελεγχόμενα και μη λεξιλόγια να συνδυαστούν με τις νέες τεχνολογίες του σημασιολογικού ιστού. Απώτερος στόχος είναι η δημιουργία καινοτόμων υπηρεσιών πληροφόρησης, οι οποίες θα βοηθούν τον τελικό χρήστη στην αποτελεσματικότερη και γρηγορότερη ανάκτηση πληροφοριών από διάφορες ετερογενείς πηγές.

Ειδικότερα, η συμβολή στην επιστήμη της πληροφόρησης και της βιβλιοθηκονομίας έχει ως εξής:

- Παρουσίαση μιας μεθοδολογίας για την εκμετάλλευση της άμεσης και έμμεσης σημασιολογικής εκφραστικότητας των ελεγχόμενων λεξιλογίων.
- Ανάπτυξη μιας μεθοδολογίας για τη συγχώνευση ελεγχόμενων λεξιλογίων.
- Παρουσίαση μιας υπηρεσίας θεματικής πλοήγησης βασισμένη στις τεχνολογίες του σημασιολογικού ιστού.
- Ανάπτυξη μιας μεθοδολογίας για την αντιστοίχιση σημασιολογικά παρόμοιων ελεγχόμενων λεξιλογίων μέσω τεχνολογιών του σημασιολογικού ιστού.
- Διερεύνηση της δυνατότητας που έχουν τα παραδοσιακά εργαλεία που χρησιμοποιούνται στις βιβλιοθήκες για την καταχώρηση της θεματικής περιγραφής του υλικού να φιλοξενήσουν πληροφορία σε μορφή συμβατή με τις τεχνολογίες του σημασιολογικού ιστού.

- Τέλος, μελέτη των μη ελεγχόμενων λεξιλογίων και κατά πόσο αυτά είναι ικανά να αντιμετωπίσουν το πρόβλημα της χρονικής διάστασης της πληροφορίας.

Οι τομείς που αναφέρθηκαν παραπάνω αποτελούν τους βασικούς πυλώνες πάνω στους οποίους επικεντρώθηκε η ερευνητική προσπάθεια της παρούσας διατριβής και λειτουργούν ως οδηγοί για την περαιτέρω ανάλυση, ανάπτυξη και οργάνωση των ερευνητικών βημάτων.

1.3. Αποτύπωση του ερευνητικού χώρου που κινείται η διατριβή

1.3.1. Θεματική ευρετηρίαση στις βιβλιοθήκες

Μια από τις πιο κοινές και χρονοβόρες αρμοδιότητες των βιβλιοθηκονόμων και γενικά των επιστημόνων της πληροφόρησης κατά την εκτέλεση των καθημερινών τους καθηκόντων είναι η ανάθεση κατάλληλων τιμών στο στοιχείο μεταδεδομένων «θέμα» (*“subject”*). Τα θέματα χρησιμοποιούνται για να περιγράψουν τις πηγές ενός αποθετηρίου. Αρκετά συχνά, στα θέματα αναθέτονται τιμές που ανήκουν σε ένα ελεγχόμενο λεξιλόγιο.

Τα ελεγχόμενα λεξιλόγια περιγράφονται ως προσεχτικά προκαθορισμένες λίστες από λέξεις και φράσεις που χρησιμοποιούνται για να περιγράψουν ψηφιακές πηγές και για να οργάνωσουν τη γνώση (Williamson, 1996, σ. 160). Χρησιμοποιούνται ευρέως από τις βιβλιοθήκες και άλλα ιδρύματα μνήμης για να συλλάβουν το νόημα των συλλογών και των πηγών τους. Επομένως, τα ελεγχόμενα λεξιλόγια μπορούν να θεωρηθούν ως ο ακρογωνιαίος λίθος των εργαλείων αναζήτησης και ανάκτησης πληροφοριών που διευκολύνουν την πρόσβαση σε αυτές τις πηγές. Τα πιο διαδεδομένα ελεγχόμενα λεξιλόγια στο χώρο των βιβλιοθηκών είναι οι θεματικές επικεφαλίδες, οι θησαυροί και τα ταξινομικά σχήματα.

Ο χρόνος και η προσπάθεια που αφιερώνει το κατάλληλα εκπαιδευμένο προσωπικό για την ανάπτυξη αυτής της γνώσης, ενισχύει την ποιότητα των υπηρεσιών ανάκτησης που παρέχεται από κάθε οργανισμό ξεχωριστά. Μάλιστα, πολλοί οργανισμοί έχουν δημιουργήσει τα δικά τους ελεγχόμενα λεξιλόγια για να μπορέσουν να περιγράψουν καλύτερα τις πηγές τους και για να καλύψουν τις πληροφοριακές ανάγκες της εκάστοτε κοινότητας. Έτσι, δημιουργήθηκαν πολλά όμοια σηματολογικά ελεγχόμενα λεξιλόγια που, όμως, είναι συντακτικά και γλωσσικά ετερογενή. Αυτό είχε ως συνέπεια να μην μπορεί να επιτευχθεί η ανταλλαγή, η αναζήτηση και η ανάκτηση πληροφοριών ανά-

μεσα σε διαφορετικούς οργανισμούς και να επηρεάζεται αρνητικά η διαλειτουργικότητα. Πιο συγκεκριμένα, στα πλαίσια των ελεγχόμενων λεξιλογίων, η διαλειτουργικότητα αναφέρεται στην ικανότητα δύο ή περισσότερων λεξιλογίων και στα συστήματά τους ή στα συστατικά των συστημάτων τους να αντιστοιχήσουν τα δεδομένα τους, με απώτερο στόχο την ανταλλαγή πληροφοριών και την ενίσχυση της ανακάλυψης (Harpring, 2010, σ. 83). Η διαλειτουργικότητα ασχολείται με τις δύο αντικρουόμενες απαιτήσεις που διέπουν την ανάπτυξη και τη χρήση των ελεγχόμενων λεξιλογίων: α) την απαίτηση για τη δημιουργία εξειδικευμένων λεξιλογίων για την εξυπηρέτηση μιας συγκεκριμένης κοινότητας και β) την απαίτηση των χρηστών να εκτελούν/διεξάγουν μια μοναδική αναζήτηση για την ανάκτηση πληροφοριών από διαφορετικές πηγές.

Συμφώνα με τους Isaac et al. (2008, σ. 189), υπάρχουν δυο βασικά προβλήματα ετερογένειας ανάμεσα στα λεξιλόγια, τα οποία θα πρέπει να αντιμετωπιστούν:

- α) ετερογένεια στην αναπαράσταση: τα λεξιλόγια τις περισσότερες φορές έρχονται σε διαφορετικές μορφές (π.χ. οι θησαυροί περιέχουν όρους, ενώ τα ταξινομικά σχήματα περιέχουν κλάσεις).
- β) ετερογένεια στις έννοιες: συνήθως δύο διαφορετικά λεξιλόγια μπορεί να περιέχουν έννοιες που έχουν ίδιο ή παρόμοιο νόημα αλλά διαφορετικές ετικέτες ή ονόματα.

Σε μια πρώτη φάση για την αντιμετώπιση του προβλήματος της ετερογένειας και κατά συνέπεια του προβλήματος της διαλειτουργικότητας ανάμεσα σε δύο διαφορετικά ελεγχόμενα λεξιλόγια προτάθηκαν λύσεις, όπως η αντιστοίχιση, η ευθυγράμμιση και η συγχώνευση αυτών μέσω της χρήσης διαφόρων μεθόδων και τεχνικών (π.χ. τεχνικές αντιστοίχισης στο επίπεδο στοιχείων, τεχνικές αντιστοίχισης στο επίπεδο της δομής κλπ.) (Euzenat & Shvaiko, 2013, σ. 79-82).

Παράλληλα, η εμφάνιση του διαδικτύου και των τεχνολογιών του έχει αυξήσει τους ρυθμούς ανάπτυξης νέων ψηφιακών υπηρεσιών από τις βιβλιοθήκες, οι οποίες μπορούν να βελτιώσουν σημαντικά την αναζήτηση και ανάκτηση πληροφοριών. Όμως, η ανάπτυξη του διαδικτύου αύξησε την πολυπλοκότητα της πληροφορίας, καθώς οι τεχνολογίες του έδωσαν τη δυνατότητα στους χρήστες να δημοσιεύουν μεγάλο όγκο πληροφοριών με ελάχιστες τεχνικές προϋποθέσεις και κυρίως χωρίς κανένα έλεγχο. Μέσα σε αυτό το πλαίσιο δημιουργήθηκαν και οι λαοταξινομίες (folksonomies).

Οι λαοταξινομίες είναι μια μεθοδολογία ανάκτησης πληροφοριών βασισμένες στο διαδίκτυο και αποτελούνται από συλλογικά δημιουργημένες, ανοικτού τύπου ετικέτες,

οι οποίες κατηγοριοποιούν το περιεχόμενο του διαδικτύου. Οι λαοταξινομίες δημιουργούνται από τους ίδιους τους χρήστες που χρησιμοποιούν τις πηγές και όχι από εξειδικευμένους επιστήμονες. Για αυτό το λόγο δεν υπάρχει κανένας έλεγχος κατά τη δημιουργία τους (Noruzi, 2006, σ. 199). Σε αυτό το σημείο θα μπορούσε να τονιστεί και η θεμελιώδης διαφορά που υπάρχει ανάμεσα στα ελεγχόμενα και μη ελεγχόμενα λεξιλόγια. Τα ελεγχόμενα λεξιλόγια μπορούν να ενισχύσουν την ακρίβεια και την ανάγκη αλλά το επιτυγχάνουν μέσω της επιβολής μιας προκαθορισμένης ορολογίας με την οποία μπορεί να μην είναι εξοικειωμένη η κοινότητα των χρηστών, ενώ οι λαοταξινομίες επιτρέπουν στους χρήστες να περιγράψουν όπως αυτοί θέλουν το περιεχόμενο, αλλά μια τέτοια δυνατότητα μπορεί να κοστίσει σε όρους ακρίβειας και ανάκτησης (Tan & Clarke, 2007, σ. 123).

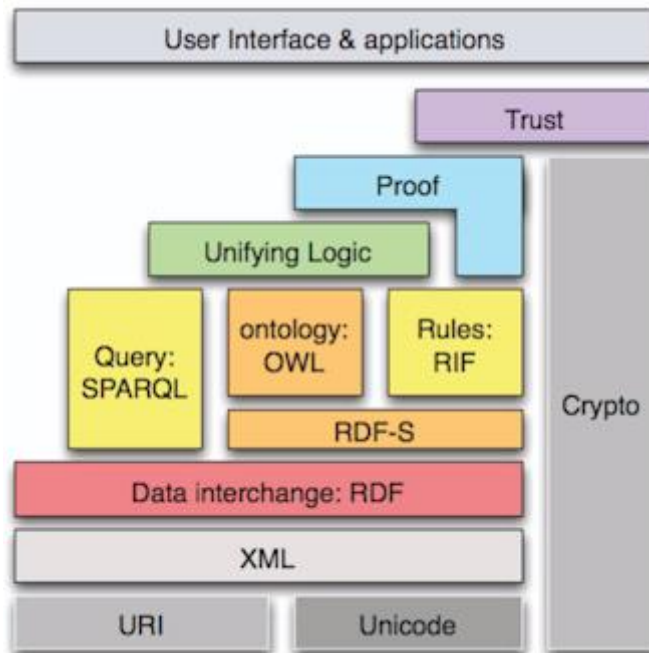
Σε αυτό το σημείο, λοιπόν, έρχονται οι τεχνολογίες του σημασιολογικού ιστού (semantic web), οι οποίες μπορούν να αντιμετωπίσουν όλα αυτά τα προβλήματα που προαναφέρθηκαν. Πιο συγκεκριμένα, ο σημασιολογικός ιστός μπορεί να εξυπηρετήσει την ανάγκη για ενιαία οργάνωση των δεδομένων, ώστε το διαδίκτυο να γίνει μια αποδοτική πλατφόρμα αυτόματης επεξεργασίας και ανταλλαγής πληροφοριών που προέρχονται από ποικίλες και ετερογενείς πηγές.

1.3.2. Σημασιολογικός ιστός

Ο σημασιολογικός ιστός (Berners-Lee, Hendler & Lassila, 2001) έχει ως στόχο να εξελίξει το σημερινό διαδίκτυο έτσι ώστε οι πληροφορίες που διακινούνται σε αυτό να είναι κατανοητές όχι μόνο από τους χρήστες του διαδικτύου αλλά και από τους υπολογιστές. Ο σημασιολογικός ιστός είναι μια επέκταση και βελτιωμένη έκδοση του σημερινού Παγκόσμιου Ιστού (World Wide Web), κυρίως όσον αφορά στην κωδικοποίηση της σημασίας της πληροφορίας. Ο ορισμός του σημασιολογικού ιστού που δόθηκε από τον Tim Berners-Lee αναφέρει ότι: "*The Semantic web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation*" (Berners-Lee, Hendler & Lassila, 2001).

Ο σημασιολογικός ιστός παρέχει ένα κοινό πλαίσιο, το οποίο επιτρέπει στα δεδομένα να διαμοιραστούν και να επαναχρησιμοποιηθούν από διάφορες εφαρμογές, οργανισμούς και κοινότητες χωρίς όρια. Το όραμα του σημασιολογικού ιστού είναι να επεκτείνει τις αρχές του διαδικτύου από τα έγγραφα στα δεδομένα. Τα δεδομένα θα πρέπει να προσεγγίζονται μέσω της χρήσης της αρχιτεκτονικής του διαδικτύου.

Μάλιστα, ο όρος σημασιολογικός ιστός τις περισσότερες φορές χρησιμοποιείται για να αναφερθούμε στις τεχνολογίες που τον υλοποιούν. Στη συνέχεια, θα δοθεί μια μικρή συνοπτική περιγραφή των βασικών δομικών στοιχείων του σημασιολογικού ιστού, δηλ. της XML, του RDF, της OWL, του SKOS, της RIF κ.ά. (Berners-Lee et al., 2006, σ. 19-22).



Εικόνα 1.1: Δομικά στοιχεία του σημασιολογικού ιστού¹

Η Extensible Markup Language - XML² είναι μια ευρέως διαδεδομένη γλώσσα περιγραφής για την ανταλλαγή δεδομένων. Τα δεδομένα αυτά με τη χρήση της μπορούν εύκολα να διαβαστούν και να επεξεργαστούν από ανθρώπους και υπολογιστές. Η XML δεν επιβάλλει κανένα σημασιολογικό περιορισμό στα δεδομένα που περιγράφει. Για την περιγραφή των δεδομένων η XML χρησιμοποιεί το Document Type Definition - DTD³. Για την παρουσίαση της πληροφορίας χρησιμοποιούνται ένα ή περισσότερα Cascading Style Sheets - CSS⁴. Τα βασικά στοιχεία ενός DTD εγγράφου είναι τα στοιχεία (elements) και τα γνωρίσματα (attributes). Λόγω του γεγονότος ότι η DTD έχει περιορισμένες δυνατότητες έχουν δημιουργηθεί τα XML Schemas. Πιο συγκεκριμένα,

¹ Berners-Lee et al., 2006, σ. 22.

² W3.org, (2015). *Extensible Markup Language (XML)*. [online] Διαθέσιμο στο: <http://www.w3.org/XML/> [Πρόσβαση: 17/07/2015].

³ W3schools.com, (2015). *DTD Tutorial*. [online] Διαθέσιμο στο: http://www.w3schools.com/xml/xml_dtd_intro.asp [Πρόσβαση: 17/07/2015].

⁴ W3.org, (2015). *Cascading Style Sheets*. [online] Διαθέσιμο στο: <http://www.w3.org/Style/CSS/> [Πρόσβαση: 17/07/2015].

τα XML Schemas (Fallside & Walmsley, 2004) είναι μια γλώσσα για την παροχή και τον περιορισμό της δομής και των στοιχείων που περιέχονται σε ένα XML έγγραφο.

Το Resource Description Framework – RDF (Klyne & Carroll, 2004) είναι ένα απλό λεξιλόγιο για την περιγραφή μεταδεδομένων στο διαδίκτυο. Προσφέρει ένα μοντέλο δεδομένων για την περιγραφή των πληροφοριών έτσι ώστε να είναι δυνατή η ανάγνωση και η κατανόησή τους από ανθρώπους και υπολογιστές. Ένα μοντέλο RDF μπορεί να αναπαρασταθεί με διάφορες συντάξεις, π.χ. RDF/XML, N3, Turtle, και RDFa. Το RDF είναι ένα βασικό πρότυπο για το σημασιολογικό ιστό. Το RDF μοντέλο δεν προσφέρει τη δυνατότητα δήλωσης περιορισμών στις σχέσεις που υπάρχουν ανάμεσα στις ιδιότητες και τις πηγές. Για αυτό το λόγο έχει δημιουργηθεί το RDF Schema – RDFS (Brickley & Guha, 2014). Η χρήση του επεκτείνει τη χρήση του RDF και είναι ένα λεξιλόγιο που μπορεί να ορίσει τις πηγές ως αντικείμενα κλάσεων. Με τη χρήση του RDF Schema δίνεται και η δυνατότητα περιγραφής απλών οντολογιών.

Οι οντολογίες αποτελούν το δομικό στοιχείο του σημασιολογικού ιστού. Με τον όρο οντολογία εννοείται η ακριβής περιγραφή των πραγμάτων και των εννοιών καθώς και των σχέσεων που υπάρχουν μεταξύ τους. Ο πιο δημοφιλής ορισμός για τον όρο οντολογία δόθηκε από τον Gruber (1993, σ. 199): "*An ontology is an explicit specification of a conceptualization*". Μια οντολογία αποτελείται από τρία είδη συστατικών: α) τις κλάσεις (classes), οι οποίες αναπαριστούν έννοιες, β) τα στιγμιότυπα (instances), τα οποία είναι αντικείμενα - μέλη των κλάσεων, και γ) τις ιδιότητες (properties), οι οποίες αναπαριστούν δυαδικές σχέσεις ανάμεσα στα στιγμιότυπα των κλάσεων.

Η Web Ontology Language – OWL 2nd ed. (Hitzler et al., 2012) αποτελεί την πιο γνωστή γλώσσα περιγραφής οντολογιών στο διαδίκτυο. Είναι πιο εκφραστική σε σχέση με το RDF και το RDF Schema. Η OWL προσφέρει καλύτερο και ευρύτερο λεξιλόγιο για την περιγραφή ιδιοτήτων και κλάσεων, καθώς και σχέσεων. Αποτελεί την πιο πρόσφατα ανεπτυγμένη γλώσσα για τη δημιουργία οντολογιών. Η OWL μπορεί να κατηγοριοποιηθεί σε τρεις υπο-γλώσσες: α) στην OWL-Lite, β) στην OWL-DL και γ) στην OWL-Full. Η OWL-Full είναι συμβατή με το RDF Schema και είναι δυνατό να ενισχύσει τη σημασία του ήδη υπάρχοντος RDF λεξιλογίου.

Η SPARQL Protocol and RDF Query Language – SPARQL (Prud' hommeaux & Seaborne, 2008) είναι μια γλώσσα ερωτημάτων που χρησιμοποιείται στο σημασιολογικό ιστό και κατ' επέκταση και στο RDF. Με τη χρήση της μπορεί ο χρήστης να εκφράζει ερωτήματα σε αποθήκες δεδομένων RDF.

Μια διαδεδομένη προσέγγιση για την εκμετάλλευση των ελεγχόμενων λεξιλογίων από εφαρμογές του σημασιολογικού ιστού είναι η αναπαράστασή τους σε κατάλληλα πρότυπα όπως είναι η γλώσσα RDF/OWL και το σύστημα Simple Knowledge Organization System – SKOS (Isaac & Summers, 2009). Το SKOS (Miles et al., 2005, σ. 3), το οποίο αναπτύσσεται από το World Wide Web Consortium – W3C⁵ είναι μια γλώσσα σήμανσης, η οποία έχει σχεδιαστεί για την αναπαράσταση θησαυρών, θεματικών επικεφαλίδων και γενικά ελεγχόμενων λεξιλογίων. Το SKOS βασίζεται στο RDF και RDFS λεξιλόγιο και ο κύριος στόχος του είναι να βοηθήσει στην εύκολη δημοσίευση ελεγχόμενων λεξιλογίων στο σημασιολογικό ιστό.

Τέλος, ένα ακόμα δομικό στοιχείο του σημασιολογικού ιστού είναι το Rule Interchange Format – RIF. Το RIF είναι μια XML γλώσσα για τη δήλωση διαδικτυακών κανόνων, τους οποίους οι υπολογιστές μπορούν να κατανοήσουν. Το RIF παρέχει διάφορες εκδόσεις που ονομάζονται διάλεκτοι. Η βασική διάλεκτος είναι η Core dialect, η οποία μπορεί να επεκταθεί στη Basic Logic Dialect – BLD και στη Production Rule Dialect – PRD (Kifer & Boley, 2013).

Ολοκληρώνοντας την αναφορά μας στο σημασιολογικό ιστό θα πρέπει να τονιστεί το γεγονός ότι για να γίνει πραγματικότητα ο σημασιολογικός ιστός είναι απαραίτητο να υπάρχουν ένας μεγάλος όγκος δεδομένων στο διαδίκτυο σε μια τυποποιημένη, προσβάσιμη και διαχειρίσιμη μορφή. Επιπλέον, οι σχέσεις ανάμεσα στα δεδομένα θα πρέπει να είναι διαθέσιμες. Αυτή η συλλογή των αλληλένδετων δεδομένων στο διαδίκτυο αναφέρεται και ως Συνδεδεμένα Δεδομένα (Linked Data)⁶. Ο όρος Συνδεδεμένα Δεδομένα βρίσκεται στην καρδιά του σημασιολογικού ιστού και χρησιμοποιείται για να περιγράψει μια διαδικασία που εκτελείται για την προβολή, το διαμοιρασμό και τη διασύνδεση δεδομένων, πληροφοριών και γνώσης στο σημασιολογικό ιστό με τη χρήση Ενιαίων Αναγνωριστικών Πόρων (Uniform Resource Identifiers – URIs⁷) και του RDF.

Έχοντας λοιπόν κατά νου τις δυνατότητες που παρέχει ο σημασιολογικός ιστός και τα προβλήματα που έρχεται να λύσει, παρατηρείται ότι, τα τελευταία χρόνια, ένας

⁵ W3.org, (2015). *World Wide Web Consortium (W3C)*. [online] Διαθέσιμο στο: <http://www.w3.org/> [Πρόσβαση: 08/07/2015].

⁶ W3.org, (2015). *Linked Data - W3C*. [online] Διαθέσιμο στο: <http://www.w3.org/standards/semanticweb/data> [Πρόσβαση: 17/07/2015].

⁷ Tools.ietf.org, (2015). *RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax*. [online] Διαθέσιμο στο: <https://tools.ietf.org/html/rfc3986> [Πρόσβαση: 07/10/2015].

ολόένα αυξανόμενος αριθμός σημασιολογικής πληροφορίας της πολιτιστικής κληρονομιάς παρέχεται ως Συνδεδεμένα Δεδομένα. Πιο συγκεκριμένα, τα ενδιαφερόμενα μέρη των ιδρυμάτων μνήμης και ιδιαίτερα οι βιβλιοθήκες συνειδητοποίησαν σχετικά σύντομα ότι η υιοθέτηση τεχνολογιών των Συνδεδεμένων Δεδομένων θα τους βοηθήσει όχι μόνο στην αντιμετώπιση των σημαντικών θεμάτων διαλειτουργικότητας που μαστιάζουν την κοινότητα εδώ και δεκαετίες, αλλά θα τους παρέχουν και την απαραίτητη υποδομή για τη δημιουργία νέων υπηρεσιών προστιθέμενης αξίας. Τέτοιες υπηρεσίες θα είναι ικανές να ενσωματώσουν σημασιολογική πληροφορία, η οποία υπάρχει στα παραδοσιακά ιδρυματικά αποθετήρια με πληροφορίες τρίτων, όπως είναι τα αρχεία εικόνων και βίντεο ή οι βάσεις γνώσεων όπως είναι η DBpedia⁸ κ.ά. Η υιοθέτηση τεχνολογιών των Συνδεδεμένων Δεδομένων, όπως είναι τα αποθετήρια τριπλετών (triplestores) των Συνδεδεμένων Δεδομένων και τα αντίστοιχα τελικά σημεία SPARQL (SPARQL endpoints) από σημαντικούς παρόχους δεδομένων των ιδρυμάτων πολιτιστικής κληρονομιάς είναι ένα σημαντικό βήμα προς την κατεύθυνση της παροχής ιδιαίτερα διαλειτουργικών ιδρυματικών αποθετηρίων. Παρά την ευρεία διαθεσιμότητα τέτοιων τεχνολογιών, φαίνεται να υπάρχει έλλειψη υπηρεσιών βασισμένων σε Συνδεδεμένα Δεδομένα, οι οποίες να στοχεύουν στους τελικούς χρήστες. Τα ιδρύματα μνήμης ενδιαφέρονται περισσότερο για το μετασχηματισμό και την έκθεση της θεματικής τους πληροφορίας σε μορφή των Συνδεδεμένων Δεδομένων, χωρίς να δίνεται ιδιαίτερη βαρύτητα και η απαραίτητη προσοχή στη δημιουργία εύχρηστων και χρήσιμων υπηρεσιών που βασίζονται σε Συνδεδεμένα Δεδομένα για τους τελικούς χρήστες.

1.4. Ερευνητική μεθοδολογία

Στο πρώτο στάδιο της προτεινόμενης μεθοδολογίας γίνεται μια εκτεταμένη βιβλιογραφική έρευνα. Η βιβλιογραφική έρευνα έχει ως στόχο την ανάλυση των θεμελιακών οντοτήτων που περιγράφονται στη διατριβή (δηλ. ελεγχόμενα λεξιλόγια και σημασιολογικός ιστός). Με τη χρήση της βιβλιογραφικής έρευνας μας δίνεται η ευκαιρία να συστηματοποιήσουμε τις υπάρχουσες πληροφορίες που υπάρχουν για το ερευνητικό αντικείμενο της διατριβής και να έχουμε μια συνολική εικόνα της κατάστασης (Παρασκευόπουλος, 1993, σ. 22).

⁸ Dbpedia.org, (2015). *DBpedia*. [online] Διαθέσιμο στο: <http://dbpedia.org/> [Πρόσβαση: 18/07/2015].

Πιο συγκεκριμένα, γίνεται μια εμπειριστατωμένη μελέτη για τα ελεγχόμενα λεξιλόγια, τις κατηγορίες ελεγχόμενων λεξιλογίων που υπάρχουν, ερευνώνται οι υφιστάμενες υπηρεσίες που βασίζονται στα ελεγχόμενα λεξιλόγια, οι πρακτικές και οι τεχνικές που υπάρχουν για τη συγχώνευση ή την αντιστοίχισή τους. Παράλληλα, μελετώνται τα μη ελεγχόμενα λεξιλόγια και η επίδραση που έχουν στους χρήστες όσον αφορά στην αναζήτηση και ανάκτηση πληροφοριών. Ταυτόχρονα, γίνεται λεπτομερής έρευνα για τις υπηρεσίες και τα εργαλεία του σημασιολογικού ιστού, ώστε να γίνει κατανοητή η χρήση και η λειτουργία τους. Μελετώνται διάφορες υπηρεσίες των βιβλιοθηκών που παρέχονται μέσω Συνδεδεμένων Δεδομένων και αξιολογούνται οι δυνατότητές τους και η χρηστικότητά τους από την πλευρά του τελικού χρήστη. Ολοκληρώνοντας τη βιβλιογραφική έρευνα, καταλήγουμε στη μελέτη των παραδοσιακών τεχνολογιών που χρησιμοποιούνται από τις βιβλιοθήκες (π.χ. MACHINE-Readable Cataloguing – MARC) και κατά πόσο μπορούν αυτές οι τεχνολογίες να συνεργαστούν και να συνδυαστούν με τις νέες τεχνολογίες των Συνδεδεμένων Δεδομένων.

Με αφετηρία τις ελλείψεις και τα προβλήματα που εντοπίζονται βάσει της βιβλιογραφικής έρευνας, στη διατριβή αναπτύσσονται διάφορα εργαλεία, μεθοδολογίες και τεχνικές όπου μπορούν να δώσουν λύση στα προβλήματα που εντοπίστηκαν. Για την τεκμηρίωση της ορθότητας των τεχνικών και των μεθοδολογιών που προτείνονται, δημιουργήθηκαν καινοτόμες υπηρεσίες, οι οποίες έχουν ως στόχο την παροχή στους τελικούς χρήστες βελτιωμένων αποτελεσμάτων σε όρους ακρίβειας και ανάκτησης πληροφοριών. Παράλληλα, οι υπηρεσίες αξιολογούνται μέσω της χρήσης διαφόρων τεχνικών αξιολόγησης (π.χ. συγκριτική αξιολόγηση, πειραματική αξιολόγηση, αξιολόγηση μέσω έρευνας χρηστών κλπ.).

1.5. Συναφείς δημοσιεύσεις

Στο πλαίσιο της διδακτορικής διατριβής δημοσιεύθηκαν οι παρακάτω εργασίες σε επιστημονικά περιοδικά, σε εθνικά και διεθνή συνέδρια:

- I. Papadakis, K. Kyprianos, M. Stefanidakis. Linked Data URIs and Libraries: The Story So Far. D-Lib Magazine, 21 (5/6), May/June 2015. Διαθέσιμο στο: <doi: 10.1045/may2015-papadakis>
- I. Papadakis, K. Kyprianos. Merging controlled vocabularies through semantic alignment based on linked data. 7th Metadata and Semantics Research Conference (MTSR 2013), Thessaloniki, Greece, 20-22 November 2013

- Ε. Μακρή, Ι. Παπαδάκης, Κ. Καρδάμης, Μ. Ασλανίδη, Κ. Κυπριανός, Μ. Στεφανιδάκης. Ακαδημαϊκές μουσικές βιβλιοθήκες: Βιβλιοθήκη Τμήματος Μουσικών Σπουδών Ιονίου Πανεπιστημίου Καθιερωμένοι όροι αναφορικά με τη μουσική γενικά και την ελληνική μουσική ειδικότερα. 22^ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών, Πάτρα, 24 - 25 Οκτωβρίου 2013
- Κ. Kyprianos, I. Papadakis, Providing LOD-based functionality in digital libraries. 6th Metadata and Semantics Research Conference (MTSR 2012), Cadiz, Spain, 28-30 November 2012
- Κ. Κυπριανός, Ι. Παπαδάκης, Καθιστώντας μια υπηρεσία θεματικής πλοήγησης στο διαδίκτυο συμβατή με τις τεχνολογίες των συνδεδεμένων δεδομένων. 21^ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών, Πειραιάς, Αθήνα, 18-19 Οκτωβρίου 2012
- Ι. Papadakis, Κ. Kyprianos, Merging Controlled Vocabularies for More Efficient Subject-based Search, International Journal of Knowledge Management, 7(3), July-September 2011, pp. 76-90
- Ι. Papadakis, Κ. Kyprianos, R. Mavropodi, Μ. Stefanidakis, Subject-based Information Retrieval within Digital Libraries Employing LCSHs, D-Lib Magazine, 15(9/10), September/October 2009. Διαθέσιμο στο: <doi: 10.1045/september2009-papadakis>

1.6. Διάρθρωση της διατριβής

Η δομή της διατριβής είναι οργανωμένη σε κεφάλαια, το καθένα εκ των οποίων πραγματεύεται συγκεκριμένες θεματικές περιοχές.

Πιο αναλυτικά, με το παρόν κεφάλαιο έγινε μια προσπάθεια να παρουσιαστεί το θεωρητικό πλαίσιο μέσα στο οποίο κινείται η συγκεκριμένη διατριβή και να προσδιοριστούν τα κίνητρα για την εκπόνησή της. Στη συνέχεια, στο κεφάλαιο 2, επιχειρείται να παρουσιαστεί αναλυτικά η κατάσταση που επικρατεί στις βιβλιοθήκες σε σχέση με τη θεματική ευρετηρίαση και αναζήτηση. Συγκεκριμένα, αναφέρονται τα διαφορετικά είδη των ελεγχόμενων και μη ελεγχόμενων λεξιλογίων που μπορούν να χρησιμοποιηθούν για τη θεματική ευρετηρίαση στις βιβλιοθήκες. Ιδιαίτερη μνεία γίνεται στις θεματικές επικεφαλίδες του Κογκρέσου, αφού αποτελούν το πιο διαδεδομένο εργαλείο θεματικής ευρετηρίασης καθώς χρησιμοποιείται από πληθώρα βιβλιοθηκών ανά τον

κόσμο. Στη συνέχεια, παρουσιάζονται άλλες τεχνικές θεματικής ευρετηρίασης, όπως είναι οι θησαυροί και τα ταξινομικά σχήματα. Ύστερα, περιγράφονται οι σύγχρονες τάσεις στη θεματική αναζήτηση και ανάκτηση πληροφοριών και παρουσιάζεται μια μεθοδολογία που εκμεταλλεύεται την άμεση και έμμεση σημασιολογική εκφραστικότητα των ελεγχόμενων λεξιλογίων. Η μεθοδολογία υλοποιείται μέσω μιας υπηρεσίας θεματικής πλοήγησης που βασίζεται στη σημασιολογική εκφραστικότητα των θεματικών επικεφαλίδων της Βιβλιοθήκης του Κογκρέσου. Η υπηρεσία αξιολογείται μέσω της συγκριτικής αξιολόγησης της παραδοσιακής θεματικής λειτουργικότητας που προσφέρεται από την ψηφιακή βιβλιοθήκη (Διώνη), η οποία είναι το ιδρυματικό αποθετήριο μεταπτυχιακών και διδακτορικών διατριβών του Πανεπιστημίου Πειραιά και της θεματικής λειτουργικότητας που προσφέρεται μέσω της προτεινόμενης υπηρεσίας. Κατόπιν, περιγράφονται οι λαοταξινομίες, ως το κατεξοχήν μη ελεγχόμενο λεξιλόγιο, που χρησιμοποιείται ιδιαίτερα στο διαδίκτυο. Το κεφάλαιο 2 ολοκληρώνεται με το σχολιασμό των παραγόντων που θα πρέπει οι εκάστοτε οργανισμοί να έχουν υπ' όψιν τους όταν καλούνται να επιλέξουν το ελεγχόμενο λεξιλόγιο ή τα ελεγχόμενα λεξιλόγια που είναι κατάλληλα για τη θεματική περιγραφή του υλικού τους.

Το κεφάλαιο 3 πραγματεύεται το πρόβλημα της ύπαρξης ποικίλων ελεγχόμενων λεξιλογίων και πως μπορούν αυτά να συγχωνευθούν ή να αντιστοιχηθούν με απώτερο στόχο την επίτευξη διαλειτουργικότητας και κατ' επέκταση τη βελτίωση της αναζήτησης και της ανάκτησης πληροφοριών. Πιο συγκεκριμένα, γίνεται εμπεριστατωμένη έρευνα και παρουσίαση των τεχνικών που υπάρχουν για τη συγχώνευση θησαυρών, τη συγχώνευση θησαυρών με θεματικές επικεφαλίδες, τη συγχώνευση ταξινομικών σχημάτων και την αντιστοίχιση των λαοταξινομιών με ελεγχόμενα λεξιλόγια. Στη συνέχεια, προτείνεται μια μεθοδολογία για τη συγχώνευση ελεγχόμενων λεξιλογίων. Η ορθότητα της προτεινόμενης μεθοδολογίας τεκμηριώνεται μέσα από την εφαρμογή της στις θεματικές επικεφαλίδες της ψηφιακής βιβλιοθήκης (Διώνη) και στο θησαυρό του International Labor Organization – ILO. Το κεφάλαιο ολοκληρώνεται με την αξιολόγηση της εφαρμογής της μεθοδολογίας.

Το κεφάλαιο 4 έρχεται να καλύψει ένα μεγάλο μέρος της θεωρίας που σχετίζεται με τα Συνδεδεμένα Δεδομένα και πως αυτά μπορούν να συμβάλλουν στην αντιμετώπιση των προβλημάτων διαλειτουργικότητας ανάμεσα σε διαφορετικές υπηρεσίες και οργανισμούς. Πιο αναλυτικά, γίνεται αναφορά για το τι είναι τα Συνδεδεμένα Δεδομένα, ποιοι είναι οι κανόνες που τα διέπουν, ποιες είναι οι τεχνολογίες που είναι απαραίτητες για τη δημοσίευση δεδομένων ως Συνδεδεμένα Δεδομένα και ποιοι είναι οι

τρόποι με τους οποίους μπορεί να αναπαρασταθεί η πληροφορία σε μορφή των Συνδεδεμένων Δεδομένων.

Στο κεφάλαιο 5 γίνεται παρουσίαση των προσπαθειών που έχουν γίνει για την παροχή θεματικής πληροφορίας διαφόρων παρόχων σε μορφή συμβατή με τα Συνδεδεμένα Δεδομένα. Παράλληλα, παρουσιάζεται η προσέγγιση που ακολουθήθηκε για τη δημιουργία μιας υπηρεσίας θεματικής πλοήγησης σε μορφή των Συνδεδεμένων Δεδομένων. Πιο συγκεκριμένα, η προτεινόμενη υπηρεσία θεματικής πλοήγησης είναι ικανή να ενσωματώνει πηγές που προέρχονται από την ψηφιακή βιβλιοθήκη (Διώνη) και από τη βάση άρθρων των New York Times – NYT. Η υπηρεσία καταφέρνει όχι μόνο να παρέχει τις θεματικές επικεφαλίδες ως Συνδεδεμένα Δεδομένα, αλλά πετυχαίνει επίσης να παρέχει στους τελικούς χρήστες επιπλέον πηγές προερχόμενες από ένα απομακρυσμένο αποθετήριο. Η αντιστοίχιση των όρων της ψηφιακής βιβλιοθήκης (Διώνη) και των NYT επιτυγχάνεται μέσω της χρήσης της μεθοδολογίας που παρουσιάστηκε στο κεφάλαιο 3 για τη συγχώνευση ελεγχόμενων λεξιλογίων. Τέλος, η συγκεκριμένη υπηρεσία αξιολογείται κατάλληλα μέσω μιας έρευνας χρηστών και της εκτέλεσης πραγματικών σεναρίων χρήσης.

Το κεφάλαιο 6 πραγματεύεται την αντιστοίχιση όρων ελεγχόμενων λεξιλογίων μέσω της χρήσης γλωσσικών συνόλων δεδομένων με απώτερο σκοπό την εύρεση περισσότερων σημείων ταύτισης ανάμεσα σε δύο λεξιλόγια. Αρχικά, παρουσιάζονται τεχνικές που ήδη έχουν δημιουργηθεί για την ευθυγράμμιση/αντιστοίχιση ελεγχόμενων λεξιλογίων. Στη συνέχεια, προτείνεται μια μεθοδολογία που στοχεύει στο συγκερασμό σημασιολογικά παρόμοιων αλλά διαφορετικών ελεγχόμενων λεξιλογίων μέσω της χρήσης σημασιολογικής ευθυγράμμισης των υποκείμενων όρων. Η σημασιολογική ευθυγράμμιση πραγματοποιείται με την εκμετάλλευση τεχνολογιών των Συνδεδεμένων Δεδομένων για την ταύτιση σημασιολογικά ισοδύναμων όρων ανάμεσα σε διαφορετικά ελεγχόμενα λεξιλόγια. Η μεθοδολογία εφαρμόζεται στην ψηφιακή βιβλιοθήκη (Διώνη) και τις θεματικές επικεφαλίδες των NYT μέσω της χρήσης των συνόλων δεδομένων της DBpedia και του WordNet. Η υλοποίησή της γίνεται μέσω της χρήσης του Google Refine⁹. Τέλος, η προτεινόμενη μεθοδολογία αξιολογείται μέσω της συγκριτικής αξιολόγησης, η οποία συγκρίνει τα αποτελέσματα της εφαρμογής της συγκεκριμένης προσέγγισης με τα αποτελέσματα της εφαρμογής μιας προσέγγισης που βασίζεται στην

⁹ Openrefine.org, (2015). *OpenRefine*. [online] Διαθέσιμο στο: <http://openrefine.org/> [Πρόσβαση: 02/10/2015].

ανακάλυψη μόνο λεξιλογικών ομοιοτήτων, η οποία αναφέρθηκε στο προηγούμενο κεφάλαιο.

Στο κεφάλαιο 7 παρουσιάζεται μια ενδελεχής έρευνα που έγινε στα πεδία καθιερωμένων εγγραφών του MARC21 και του UNIMARC με σκοπό να εντοπιστούν τα πεδία που μπορούν να αποθηκεύσουν πληροφορίες σε μορφή των Συνδεδεμένων Δεδομένων. Πιο συγκεκριμένα, η έρευνα χωρίζεται σε δυο μέρη. Στο πρώτο μέρος εξετάζονται τα πεδία του MARC που μπορούν να αποθηκεύσουν URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί τοπικά, ενώ στο δεύτερο μέρος εξετάζονται τα πεδία του MARC που μπορούν να αποθηκεύσουν URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί σε απομακρυσμένα σημεία. Στη συνέχεια, μελετώνται οι εθνικές βιβλιοθήκες ανά τον κόσμο για να εντοπιστεί εάν παρέχουν στους παραδοσιακούς τους OPAC καταλόγους πληροφορίες σε μορφή των Συνδεδεμένων Δεδομένων και ποια είναι η διαδικασία που ακολουθούν για να δημοσιεύσουν τα δεδομένα τους σε νέες υπηρεσίες βασισμένες στα Συνδεδεμένα Δεδομένα.

Το κεφάλαιο 8 ασχολείται με το πρόβλημα της χρονικής διάστασης της πληροφορίας και πως αυτό μπορεί να αντιμετωπιστεί μέσω της χρήσης όρων μη ελεγχόμενων λεξιλογίων που προέρχονται από υπηρεσίες κοινωνικής δικτύωσης (π.χ. Hashtags). Πιο συγκεκριμένα, αρχικά γίνεται αναφορά στις τεχνικές που χρησιμοποιούν παραδοσιακά οι βιβλιοθήκες για να αντιμετωπίσουν το πρόβλημα της επικαιρότητας και πώς μπορούν να παρέχουν στους χρήστες επίκαιρη και τρέχουσα πληροφόρηση. Στη συνέχεια, παρουσιάζονται οι τεχνολογίες του Web 2.0 και πιο συγκεκριμένα τα hashtags του Twitter ως ένα εργαλείο που μπορεί να βοηθήσει τις βιβλιοθήκες να μελετήσουν τις τρέχουσες ανάγκες των χρηστών τους. Έπειτα, παρουσιάζεται η ιδέα που προτάθηκε για την παροχή επίκαιρης πληροφορίας στους χρήστες μιας βιβλιοθήκης μέσω της ενσωμάτωσης των hashtags των τρεχόντων tweets στον OPAC της. Η προτεινόμενη υπηρεσία βρίσκει εφαρμογή στη Βιβλιοθήκη και Κέντρο Πληροφόρησης του Ιονίου Πανεπιστημίου και αξιολογείται αντίστοιχα μέσω της εκτέλεσης ενός σεναρίου εργασιών από χρήστες και της ανάλυσης των αρχείων καταγραφής συμβάντων.

Τέλος, στο κεφάλαιο 9 παρουσιάζονται τα συμπεράσματα της διατριβής, τα οποία αποτελούνται από μια γενική ανασκόπηση στην έρευνα που πραγματοποιήθηκε και στη συνεισφορά της διατριβής στο συγκεκριμένο ερευνητικό πεδίο. Παράλληλα, διερευνώνται ανοιχτά ζητήματα, τα οποία θα μπορούσαν να αποτελέσουν αντικείμενο μελλοντικής εργασίας και προβληματισμού.

2. Θεματική ευρετηρίαση

2.1. Θεματική ευρετηρίαση στις βιβλιοθήκες

Η θεματική ευρετηρίαση είναι η διαδικασία της περιγραφής ή της ταξινόμησης ενός τεκμηρίου ή μιας πηγής με όρους ευρετηρίασης με απώτερο στόχο να δείξει σε τι αναφέρεται η πηγή, να συνοψίσει το περιεχόμενό της και να αυξήσει την ανιχνευσιμότητά (findability) της (Singh, 2011, σ. 151). Τα ευρετήρια κατασκευάζονται ξεχωριστά σε τρία διαφορετικά επίπεδα (Singh, 2011, σ. 152): α) όροι σε μια πηγή (π.χ. ένα βιβλίο), β) αντικείμενα σε μια συλλογή (π.χ. η συλλογή μιας βιβλιοθήκης) και γ) τεκμήρια σε ένα πεδίο γνώσης (π.χ. βιβλία και άρθρα).

Ο βασικός λόγος για τη χρήση της θεματικής ευρετηρίασης είναι η αποτελεσματική και αποδοτική πρόσβαση στην πληροφορία, είτε μέσω δομημένων πηγών, όπως είναι τα βιβλία και οι βάσεις δεδομένων, είτε μέσω τυχαίων αποθετηρίων πληροφόρησης, όπως είναι η πληροφορία που εντοπίζεται μέσω των μηχανών αναζήτησης του διαδικτύου (Cleveland & Cleveland, 2013, σ. 11).

Τα ευρετήρια συνήθως χρησιμοποιούν ένα είδος μιας σηματολογικής δομής για το διαχωρισμό των ομογράφων και για τη σύνδεση όρων που σχετίζονται νοηματικά. (Cleveland & Cleveland, 2013, σ. 11). Η ευρετηρίαση είναι ένα σημαντικό εργαλείο, όχι μόνο για το βιβλιογραφικό έλεγχο αλλά και σε οποιαδήποτε άλλη περίπτωση, όπου είναι απαραίτητη η οργάνωση, η ανάκτηση και η χρήση της πληροφορίας. Γενικά, η θεματική ευρετηρίαση είναι ένας βασικός παράγοντας του τομέα των βιβλιοθηκών και γενικά της επιστήμης της πληροφόρησης. Πιο συγκεκριμένα, η θεματική ανάκτηση πληροφοριών στο πλαίσιο των βιβλιοθηκών προϋποθέτει ότι στις υποκείμενες πληροφοριακές πηγές θα πρέπει να έχει ανατεθεί ένας ικανοποιητικός αριθμός θεματικών περιγραφών επαρκούς ποιότητας. Η ποσότητα μαζί με την ποιότητα της θεματικής ευρετηρίασης μπορεί να ασκήσει τεράστια επιρροή στην ποιότητα και τη δυναμική των αποτελεσμάτων στο σημείο της πραγματικής αναζήτησης. Άλλωστε, ό,τι δεν έχει ευρετηριαστεί δεν μπορεί να ανακτηθεί (Bates, 2002, σ. 387).

Μάλιστα, οι επιστήμονες της πληροφόρησης έχουν καταβάλει στη διάρκεια του χρόνου, μεγάλες προσπάθειες για να αναπτύξουν ευρετήρια λεξιλογίων για την καλύτερη οργάνωση και περιγραφή της γνώσης (Cleveland & Cleveland, 2013, σ. 30). Οι προσπάθειες αυτές κατέληξαν στη δημιουργία διάφορων ελεγχόμενων λεξιλογίων

(controlled vocabularies), τα οποία περιλαμβάνουν όλη την ανθρώπινη γνώση και μπορούν να χρησιμοποιηθούν για την περιγραφή του θέματος των βιβλίων και γενικά όλου του υποκείμενου πληροφοριακού υλικού.

2.2. Ελεγχόμενα λεξιλόγια

Τα ελεγχόμενα λεξιλόγια είναι μια οργανωμένη διάταξη λέξεων και φράσεων που χρησιμοποιούνται για να ευρετηριαστεί το περιεχόμενο μιας πηγής ή/και να ανακτηθεί το περιεχόμενό της μέσω της περιήγησης ή της αναζήτησης. Συνήθως περιλαμβάνουν προτιμώμενους και μη όρους, εναλλακτικούς όρους και έχουν έναν προκαθορισμένο σκοπό ή περιγράφουν έναν ή παραπάνω συγκεκριμένους τομείς (Harpring, 2010, σ. 12).

Ο λόγος για τον οποίο χρησιμοποιούνται τα ελεγχόμενα λεξιλόγια είναι για να οργανώσουν την πληροφορία και να παρέχουν ορολογία κατά την καταλογογράφηση και την ανάκτηση της πληροφορίας. Μια από τις πιο σημαντικές λειτουργίες των ελεγχόμενων λεξιλογίων είναι η συγκέντρωση των διαφόρων όρων και των συνώνυμων όρων για τις έννοιες που περιγράφουν καθώς και η ταξινόμηση αυτών σε κατηγορίες ή η τοποθέτησή τους σε μια λογική σειρά (Harpring, 2010, σ. 12). Πιο αναλυτικά, σύμφωνα με τους Macgregor & McCulloch (2006, σ. 292-293) η σημασιολογική και ιεραρχική δομή που παρέχεται από τα ελεγχόμενα λεξιλόγια συμβάλλει:

- Στον έλεγχο των συνώνυμων καθορίζοντας μόνο μία μορφή του όρου. Αυτό οδηγεί στη χρήση των ίδιων όρων για να περιγραφούν όμοιες ή παρόμοιες έννοιες, συμβάλλοντας έτσι στη μείωση της πιθανότητας ίδιες πηγές να μην ανακτηθούν κατά τη διάρκεια της αναζήτησης (π.χ. *"car"*, *"automobile"*, *"motor vehicle"* κ.ά.).
- Στο διαχωρισμό των ομώνυμων, επιτρέποντας στον ευρετηριαστή να επιλύσει προβλήματα που προκύπτουν στη σύγκρουση νοημάτων που προκύπτουν με όρους, οι οποίοι γράφονται με την ίδια μορφή αλλά έχουν διαφορετικό νόημα (π.χ. η λέξη *"Java"* που αναφέρεται στη γλώσσα προγραμματισμού, ή η λέξη *"Java"* που αναφέρεται στον καφέ, ή η λέξη *"Java"* που αναφέρεται στο νησί που ανήκει στο Αρχιπέλαγος της Ινδονησίας στη νοτιοανατολική Ασία). Με τον έλεγχο των ομώνυμων, η πιθανότητα θορύβου στα αποτελέσματα των χρηστών μειώνεται.
- Στον έλεγχο των λεξιλογικών διαφορών, ελαχιστοποιώντας τα προβλήματα που μπορεί να προκύψουν από περιττά λεξιλόγια ή γραμματικές παραλλαγές των

όρων, που θα μπορούσαν δυνητικά να δημιουργήσουν περαιτέρω θόρυβο στα αποτελέσματα των χρηστών.

- Στη σύνδεση όμοιων όρων ή στη συστηματική αναφορά του ευρητηριαστή σε κοντινούς εναλλακτικούς όρους, για να διασφαλιστεί το γεγονός ότι οι όμοιες ή οι σχετικές πηγές μπορούν να συνδυαστούν. Αυτό συνήθως επιτυγχάνεται με την εμφάνιση της σχέσης γένους/είδους ανάμεσα στους όρους μέσω μιας μορφής σηματολογικής ιεραρχικής σχέσης.
- Στην εξημέρηση συντακτικών συσχετίσεων (δηλ. όχι ιεραρχικών σχέσεων), όπου αυτό είναι εφικτό (π.χ. ο όρος "*language*" σχετίζεται συντακτικά με τον όρο "*indexing*", ακόμα και αν δεν συνδέονται αυστηρά ιεραρχικά οι δύο όροι).

Ιδιαίτερα στον τομέα των βιβλιοθηκών, τα ελεγχόμενα λεξιλόγια μπορούν να περιγραφούν ως προσεχτικά επιλεγμένες λίστες από λέξεις και φράσεις που χρησιμοποιούνται για να περιγράψουν ψηφιακές πηγές και άλλα πράγματα (Williamson, 1996, σ. 160). Πιο αναλυτικά, μέσα στα χρόνια οι βιβλιοθήκες έχουν δημιουργήσει ένα σημαντικό αριθμό θεματικών ελεγχόμενων λεξιλογίων, τα οποία μπορούν να ταξινομηθούν σε τρεις μεγάλες γενικές κατηγορίες (Macgregor & McCulloch, 2006, σ. 292-293; Harper & Tillett, 2007, σ.54): λίστες θεματικών επικεφαλίδων (subject headings lists), θησαυροί (thesauri) και ταξινομίες (taxonomies) ή ταξινομικά σχήματα (classification schemes).

2.3. Είδη ελεγχόμενων λεξιλογίων

2.3.1. Θεματικές επικεφαλίδες

Για τη δημιουργία λιστών θεματικών επικεφαλίδων δεν υπάρχει κάποιο συγκεκριμένο διεθνές πρότυπο. Η καθιέρωση αυτού του είδους ελεγχόμενου λεξιλογίου στους online καταλόγους επιτυγχάνεται μέσω της κοινής χρήσης μιας τυποποιημένης λίστας θεματικών επικεφαλίδων, η οποία τις περισσότερες φορές συνοδεύεται από ένα εγχειρίδιο εφαρμογής και χρήσης. Συνήθως αυτά τα εργαλεία δημιουργούνται και διατηρούνται από εθνικές βιβλιοθήκες, οι οποίες είναι υπεύθυνες για την ευρητηρίαση όλων των εκδόσεων ενός κράτους. Έτσι, οι περισσότερες λίστες θεματικών επικεφαλίδων γίνονται πρότυπο που χρησιμοποιείται από την εκάστοτε βιβλιοθήκη μέσω της συνεργασίας (Jahns, 2012, σ. 20). Το πιο διαδεδομένο εργαλείο θεματικών επικεφαλίδων

είναι οι Θεματικές Επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου (Library of Congress Subject Headings – LCSH¹⁰).

Πιο συγκεκριμένα, οι LCSH αποτελούν ένα ελεγχόμενο λεξιλόγιο, που υποστηρίζεται από τη Βιβλιοθήκη του Κογκρέσου (Library of Congress – LoC)¹¹, για τη θεματική περιγραφή των βιβλιογραφικών εγγραφών. Οι LCSH βρίσκουν εφαρμογή σε κάθε τεκμήριο και γενικά στο υλικό που υπάρχει στη συλλογή μιας βιβλιοθήκης και βοηθούν τους χρήστες στο να έχουν πρόσβαση σε τεκμήρια που έχουν το ίδιο θέμα. Οι LCSH υποστηρίζονται ενεργά από το 1898. Λόγω της συνέπειας και της σταθερότητας που έχουν δείξει μέσα στο χρόνο, οι LCSH έχουν υιοθετηθεί από πολλές βιβλιοθήκες των ΗΠΑ, αλλά και από άλλες βιβλιοθήκες ανά τον κόσμο, πολλές φορές σε μετάφραση (Κυπριανός & Παπαδάκης, 2012, σ. 5). Είναι εμφανές, λοιπόν, ότι οι LCSH κυριάρχησαν τον τελευταίο αιώνα στον τομέα των βιβλιοθηκών ως το «*ντε φάκτο*» εργαλείο για τη θεματική περιγραφή των τεκμηρίων και γενικά των πηγών που υπάρχουν στις βιβλιοθήκες (Stone, 2000). Ακόμα και σήμερα, παρά την εμφάνιση πολλών θησαυρών και άλλων ελεγχόμενων λεξιλογίων που θεωρούνται πιο κατάλληλα για χρήση στο online περιβάλλον, ακόμα μεγάλος αριθμός OPAC και ψηφιακών βιβλιοθηκών παγκοσμίως υιοθετούν τις LCSH. Μάλιστα, παρά τη συζήτηση (Calhoun, 2006, σ. 13-15) που αφορά στην ανάλυση «*κόστους-οφέλους*» (“*cost-benefit analysis*”) με τη δημιουργία των LCSH για τη θεματική ευρετηρίαση των πληροφοριακών πηγών, όπως αναφέρεται στους Schatz et al. (1996, σ. 127) και Schwartz (2008, σ. 835), η εφαρμογή των LCSH από τους καταλογογράφους δεν έχει αλλάξει ουσιαστικά και ακόμα κάποιες νέες ψηφιακές βιβλιοθήκες χρησιμοποιούν τις LCSH.

Επιπλέον, ρίχνοντας μια πιο προσεκτική ματιά στις LCSH, ένας αριθμός από σημαντικά πλεονεκτήματα παρατηρούνται, τα οποία θα μπορούσαν να εξηγήσουν την ευρεία διάδοσή τους στα συστήματα OPAC και τις ψηφιακές βιβλιοθήκες. Έτσι, όπως αναφέρθηκε από τον Dean (2004, σ. 332), οι LCSH αποτελούν ένα ευρέως χρησιμοποιούμενο (προωθώντας με αυτόν τον τρόπο τη διαλειτουργικότητα ανάμεσα στις βιβλιοθήκες ή/και τα συστήματα ψηφιακών βιβλιοθηκών) πλούσιο λεξιλόγιο, το οποίο καλύπτει όλες τις θεματικές περιοχές και απολαμβάνει ισχυρή θεσμική υποστήριξη από

¹⁰ Authorities.loc.gov, (2015). *Library of Congress Authorities (Search for Name, Subject, Title and Name/Title)*. [online] Διαθέσιμο στο: <http://authorities.loc.gov/> [Πρόσβαση: 15/07/2015].

¹¹ The Library of Congress, (2015). *Library of Congress Home*. [online] Διαθέσιμο στο: <https://www.loc.gov/> [Πρόσβαση: 03/10/2015].

τη Βιβλιοθήκη του Κογκρέσου. Επιπλέον, οι LCSH επιβάλλουν έλεγχο συνωνύμων και ομογράφων καθώς και πολύ καλά τεκμηριωμένη ιστορία.

Η σηματολογική ποιότητα των LCSH μπορεί να παρατηρηθεί περαιτέρω όταν ληφθεί υπ' όψιν η θεωρία που στηρίζει την εφαρμογή τους στις πληροφοριακές πηγές. Ως εκ τούτου, όπως αναφέρθηκε από την Bates (2002, σ. 390-391), η κατάλληλη χρήση των LCSH μας διδάσκει ότι δεν εννοούνται ως όροι μονοσήμαντοι. Αντιθέτως, η προκύπτουσα συμβολοσειρά της θεματικής επικεφαλίδας (που αποτελείται από την κύρια επικεφαλίδα μαζί με τις υποδιαιρέσεις της) έχει ως στόχο να περιγράψει από μόνη της το σύνολο θεματικής περιγραφής της πληροφοριακής πηγής. Επομένως, η εκφραστικότητα της προκύπτουσας συμβολοσειράς της θεματικής επικεφαλίδας είναι γενικά αρκετά ευρεία, συχνά πιο ευρεία από μεμονωμένα ευρετηριασμένους εννοιολογικούς όρους. Αντί να αποδοθούν αρκετές λέξεις κλειδιά στην κάθε πληροφοριακή πηγή από εννοιολογικές γλώσσες ευρετηρίασης, μια ή δύο θεματικές επικεφαλίδες εφαρμόζονται συνήθως σε μια πληροφοριακή πηγή. Τα είδη των υποδιαιρέσεων που επιτρέπονται είναι αυστηρά ελεγχόμενα, όπως είναι και η σειρά εμφάνισής τους στη θεματική επικεφαλίδα.

Σύμφωνα με τις οδηγίες των LCSH (Library of Congress, 2007), τα θέματα χρησιμοποιούν λεκτικές συμβολοσειρές για την περιγραφή των πηγών μέσα σε ένα αποθετήριο. Επίσης, οι οδηγίες ορίζουν τις σχέσεις ανάμεσα σε αυτές τις συμβολοσειρές για να καταδείξουν τα συνώνυμα και τις συσχετίσεις μεταξύ των θεμάτων. Ένα θέμα αποτελείται από την κύρια επικεφαλίδα (main heading), η οποία αντιστοιχεί στην κυρίαρχη έννοια της πηγής. Στην περίπτωση που οι πηγές έχουν πιο ειδικές έννοιες, η κύρια επικεφαλίδα μπορεί να εξειδικευτεί περισσότερο με τη χρήση υποδιαιρέσεων (subdivisions). Αυτές οι υποδιαιρέσεις αναπαριστούν διάφορες απόψεις της κύριας επικεφαλίδας. Υπάρχουν τέσσερα είδη υποδιαιρέσεων σύμφωνα με τις οδηγίες των LCSH: θεματικές, γεωγραφικές, χρονολογικές υποδιαιρέσεις και υποδιαιρέσεις μορφής (Papadakis et al., 2009).

- Οι θεματικές υποδιαιρέσεις (topical subdivisions) αναπαριστούν τη θεματική πλευρά της κύριας επικεφαλίδας και περιορίζουν την έννοια που εκφράζεται από την κύρια επικεφαλίδα προς ένα δευτερεύον θέμα.
- Οι υποδιαιρέσεις μορφής (form subdivisions) αναπαριστούν τη βιβλιογραφική ή τη λογοτεχνική ή την καλλιτεχνική μορφή στην οποία το υλικό για ένα θέμα

είναι οργανωμένο ή παρουσιάζεται. Μια υποδιαίρεση μορφής συνήθως υποδεικνύει «*Τι είναι το έργο*» (“*what the work is*”) παρά «*Περί τίνος πρόκειται*» (“*what it is about*”).

- Οι γεωγραφικές υποδιαίρεσεις (geographic subdivisions) αναδεικνύουν την προέλευση ή τον τόπο της κύριας επικεφαλίδας. Οι γεωγραφικές υποδιαίρεσεις παρέχονται κάτω από τα θέματα, τα οποία παρουσιάζουν σημαντικές διακυμάνσεις κατά την επεξεργασία, ή σε σχέση με, διαφορετικές τοποθεσίες.
- Οι χρονικές υποδιαίρεσεις (chronological subdivisions) αναδεικνύουν τη χρονική περίοδο της κύριας επικεφαλίδας ή υποδηλώνουν την ημερομηνία έκδοσης της πληροφοριακής πηγής. Η χρονική υποδιαίρεση μπορεί να ακολουθεί απευθείας την κύρια επικεφαλίδα ή να εμφανίζεται ως υπο-υποδιαίρεση κάτω από άλλη υποδιαίρεση.

Οι οδηγίες των LCSH θέτουν ορισμένους περιορισμούς όσον αφορά στη σειρά των υποδιαίρεσεων μέσα σε μια θεματική επικεφαλίδα. Πιο αναλυτικά, μια τέτοια σειρά κανονικά συμφωνεί με το πρότυπο «θέμα-τόπος-χρόνος-μορφή»:

[κύρια επικεφαλίδα] -- [θεματική υποδιαίρεση] -- [γεωγραφική υποδιαίρεση] -- [χρονολογική υποδιαίρεση] -- [υποδιαίρεση μορφής]

Το παραπάνω πρότυπο μπορεί να τροποποιηθεί, ιδιαίτερα όσον αφορά στη θέση της γεωγραφικής υποδιαίρεσης, η οποία μπορεί να εμφανίζεται σε οποιαδήποτε θέση μετά την κύρια επικεφαλίδα.

Οι LCSH είναι πλέον κατάλληλες για προ-συνδυασμένα (pre-coordinated) συστήματα ευρετηρίασης. Στην προ-συνδυασμένη ευρετηρίαση οι όροι που εκφράζουν τα θέματα είναι συνδυασμένοι κατά την ευρετηρίαση, ώστε να αποτελούν έτοιμες θεματικές επικεφαλίδες¹² (Bodoff & Kambil, 1998, σ. 1255-1257). Με αυτό τον τρόπο οι ευρετηριασμένες συμβολοσειρές δεν χρειάζεται να συνδυαστούν κατά τη διάρκεια της αναζήτησης. Θεωρητικά, η προ-συνδυασμένη ευρετηρίαση ανακτά όλες τις σχετικές πληροφορίες και μόνο τις σχετικές πληροφορίες για ένα συγκεκριμένο ερώτημα μέσα ένα αποθετήριο. Σύμφωνα με την παραδοσιακή αναζήτηση για την ανάκτηση πληροφοριών, ο προ-συνδυασμός οδηγεί τα συστήματα στο να παρουσιάζουν 100% ακρίβεια (precision) και 100% ανάκτηση (recall). Ωστόσο, η συνολική ποιότητα ενός συστήματος ανάκτησης πληροφοριών απέχει πολύ από την τελειότητα. Αυτό μπορεί να

¹² Σκανδάλη, Ν.Α. (s.d.). *Η τυποποίηση στη βιβλιογραφική επεξεργασία*. [online] Διαθέσιμο στο: http://www.library.tee.gr/digital/m993/m993_skandali.pdf [Πρόσβαση: 07/07/2015].

οφείλεται στο γεγονός ότι οι παραπάνω μετρήσεις χρησιμοποιούνται για να μετρήσουν τη σχετικότητα μιας λίστας αποτελεσμάτων αναζήτησης συγκρινόμενη με το ερώτημα, το οποίο έχει εκφραστεί ελεύθερα από έναν ερευνητή που δεν είναι προφανώς η περίπτωση στον προ-συνδυασμό. Ως εκ τούτου, στην περίπτωση της προ-συνδυασμένης ευρετηρίασης, η ποιότητα των αποτελεσμάτων σε ένα σύστημα ανάκτησης πληροφοριών εξαρτάται σε μεγάλο βαθμό από την ικανότητα των συστημάτων να βοηθούν τους ερευνητές στον εντοπισμό του ερωτήματος, το οποίο περιγράφει όσο το δυνατόν καλύτερα τις πληροφοριακές τους ανάγκες. Εάν πετύχουν στην αναζήτησή τους, τότε η μέγιστη ακρίβεια και ανάκτηση είναι εγγυημένη από την διαδικασία ευρετηρίασης. Αυτός είναι ο λόγος που τα μοντέρνα συστήματα ανάκτησης πληροφοριών που βασίζονται στα θέματα και προσανατολίζονται στις βιβλιοθήκες (Tuominen et al., 2008) παρέχουν Γραφικό Περιβάλλον Χρήστη (Graphical User Interface – GUI), το οποίο δίνει τη δυνατότητα διαδραστικής διαμόρφωσης ερωτημάτων, όπου η ακρίβεια και η ανάκτηση είναι λιγότερο κατάλληλες μετρήσεις (Lagergren & Over, 1998, σ. 170).

2.3.2. Θησαυροί

Σχετικά πρόσφατα, σε σχέση με τις θεματικές επικεφαλίδες, οι θησαυροί έκαναν την εμφάνισή τους ως ένα εναλλακτικός τύπος ελεγχόμενων λεξιλογίων, οι οποίοι ορίζονται ως τα λεξιλόγια που είναι διατεταγμένα σε μια γνωστή σειρά και δομημένα ώστε οι διάφορες σχέσεις ανάμεσα στους όρους να εμφανίζονται ξεκάθαρα και να προσδιορίζονται από τυποποιημένες σχέσεις δεικτών (Aitchison and Clarke, 2004, σ. 6). Μάλιστα, οι θησαυροί βασίζονται πάνω σε αυστηρά πρότυπα για τη δημιουργία τους. Πιο αναλυτικά, στις αρχές της δεκαετίας του 1970, οι διεθνείς οργανισμοί United Nations Educational, Scientific and Cultural Organization – UNESCO¹³ και International Standards Organization – ISO¹⁴ δημοσίευσαν τις πρώτες οδηγίες για την κατασκευή μονόγλωσσων θησαυρών¹⁵. Στη συνέχεια, το πρότυπο του ISO 2788:1974 ενημερώθηκε με τη δεύτερη έκδοσή του το 1986¹⁶. Παράλληλα, το 1985 δημιουργήθηκε το

¹³ UNESCO, (2015). *UNESCO*. [online] Διαθέσιμο στο: <http://en.unesco.org/> [Πρόσβαση: 07/07/2015].

¹⁴ Iso.org, (2015). *ISO - International Organization for Standardization*. [online] Διαθέσιμο στο: <http://www.iso.org/iso/home.html> [Πρόσβαση: 07/07/2015].

¹⁵ Guidelines for the Establishment of Monolingual Scientific and Technical Thesauri for Information Retrieval. Paris: UNESCO, 1970; ISO 2788:1974 Guidelines for the establishment and development of monolingual thesauri. Geneva: ISO, 1974.

¹⁶ ISO 2788:1986 Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri. 2nd edition. Geneva: ISO, 1986.

συμπληρωματικό πρότυπο ISO 5964 για τη δημιουργία πολύγλωσσων θησαυρών¹⁷. Αυτά τα δύο πρότυπα υιοθετήθηκαν από πολλές χώρες ως τα βασικά εργαλεία για τη δημιουργία εθνικών προτύπων για τη δημιουργία θησαυρών (π.χ. το ANSI/NISO Z39.19-2005¹⁸ ή το BS 8723-2:2005¹⁹ κ.ά.). Στην Ελλάδα, το 1993 δημιουργήθηκε το πρότυπο 1321²⁰ από τον Ελληνικό Οργανισμό Τυποποίησης – ΕΛΟΤ²¹, το οποίο βασίζεται στο πρότυπο του ISO 2788 και παρέχει οδηγίες για τη δημιουργία μονόγλωσσων θησαυρών.

Πρόσφατα (2011), τα πρότυπα ISO 2788 και 5964 αντικαταστάθηκαν από το ISO 25964²². Το εν λόγω πρότυπο αποτελείται από δύο μέρη, τα οποία εκδόθηκαν το 2011 και το 2013. Πιο αναλυτικά, ένας θησαυρός που είναι συμβατός με το πρότυπο 25964-1²³ παραθέτει όλες τις έννοιες που διατίθενται για ευρετηρίαση σε ένα συγκεκριμένο πλαίσιο και επισημαίνει κάθε μια από αυτές μέσω ενός προτιμώμενου όρου (preferred term) καθώς και τα συνώνυμα που μπορεί να έχει αυτή η έννοια. Οι σχέσεις ανάμεσα στις έννοιες και ανάμεσα στους όρους είναι εμφανείς, κάνοντας εύκολη την πλοήγηση ανάμεσα στους όρους του πεδίου κατά τη δημιουργία ενός ερωτήματος αναζήτησης. Οι κύριοι τύποι σχέσεων ανάμεσα στις έννοιες είναι: ισοδύναμη σχέση (equivalence) π.χ. ανάμεσα στους συνώνυμους και σχεδόν συνώνυμους όρους, ιεραρχική σχέση (hierarchical) π.χ. ανάμεσα σε ευρύτερες (broader) και στενότερες (narrower) έννοιες και συνειρμική σχέση (associative) π.χ. ανάμεσα σε έννοιες που έχουν στενή σχέση μεταξύ τους αλλά δεν σχετίζονται ιεραρχικά (Will, 2012, σ. 48-50). Στους πολύγλωσσους θησαυρούς, η ισοδύναμη σχέση βρίσκει εφαρμογή και στους αντίστοιχους όρους

¹⁷ ISO 5964:1985 Documentation – Guidelines for the Establishment and Development of Multilingual Thesauri. Geneva: ISO, 1985.

¹⁸ ANSI/NISO Z39.19-2005 – Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Baltimore, MD: NISO, 2005.

¹⁹ BS 8723-2:2005 Structured Vocabularies for Information Retrieval. Guide. Thesauri. London: The British Standards Institution, 2005.

²⁰ Ελληνικό πρότυπο 1321: Τεκμηρίωση - Κατευθυντήριες Οδηγίες για τη Συγκρότηση και Ανάπτυξη Μονόγλωσσων Θησαυρών. Αθήνα: ΕΛΟΤ, 1993.

²¹ Elot.gr, (2015). *ΕΛΟΤ - Αρχική Σελίδα*. [online] Διαθέσιμο στο: <http://www.elot.gr/> [Πρόσβαση: 07/07/2015].

²² Niso.org, (2015). *ISO 25964 Thesaurus Schemas - National Information Standards Organization*. [online] Διαθέσιμο στο: <http://www.niso.org/schemas/iso25964/> [Πρόσβαση: 07/07/2015].

²³ Iso.org, (2015). *ISO 25964-1:2011 - Information and Documentation -- Thesauri and Interoperability with Other Vocabularies -- Part 1: Thesauri for Information Retrieval*. [online] Διαθέσιμο στο: http://www.iso.org/iso/catalogue_detail.htm?csnumber=53657 [Πρόσβαση: 14/07/2015].

που προέρχονται από διαφορετικές φυσικές γλώσσες. Το 2013 εκδόθηκε και το δεύτερο μέρος του προτύπου ISO 25964-2²⁴, το οποίο βρίσκει εφαρμογή σε θησαυρούς και άλλα είδη λεξιλογίων ή συστήματα οργάνωσης γνώσης (Knowledge Organization Systems – KOS) που χρησιμοποιούνται για την ανάκτηση πληροφοριών. Πιο συγκεκριμένα, περιγράφει, συγκρίνει και αντιπαραβάλλει τα στοιχεία και τα χαρακτηριστικά των λεξιλογίων που εμπλέκονται όταν απαιτείται διαλειτουργικότητα. Δίνει οδηγίες για την καθιέρωση και τη διατήρηση αντιστοιχίσεων ανάμεσα σε πολλαπλούς θησαυρούς ή ανάμεσα σε θησαυρούς και άλλα είδη ελεγχόμενων λεξιλογίων.

Οι περιγραφείς των θησαυρών είναι μικρότεροι σε μήκος και λιγότερο περίπλοκοι σε σχέση με τους όρους των LCSH. Ως εκ τούτου, είναι κατάλληλοι για συστήματα υστερο-συνδυσασμένης (post-coordinated) ευρετηρίασης. Η υστερο-συνδυσασμένη ευρετηρίαση είναι το αντίθετο της προ-συνδυσασμένης ευρετηρίασης. Οι όροι μιας ένωσης συμβολοσειράς θεμάτων μπορεί αποτελεσματικά να αναδιαταχθεί ώστε να ταιριάζει σε οποιοδήποτε ερώτημα (Bodoff & Kambil, 1998, σ. 1257-1258). Στην περίπτωση της υστερο-συνδυσασμένης ευρετηρίασης, το ερώτημα εκφράζεται απευθείας από τον ερευνητή μέσω της χρήσης ενός είδους λογικής Boolean και επομένως είναι πιο κοντά στις πληροφοριακές του ανάγκες. Από μια άλλη οπτική γωνία, αυτή η ελευθερία στην έκφραση ενός ερωτήματος συχνά καταλήγει σε μια συμβολοσειρά που περιέχει διαφορετικούς όρους που υποφέρουν από πολυσημία ή/και συνωνυμία (Miller, 1995, σ. 40). Για παράδειγμα, ένα ερώτημα που περιέχει τον όρο "*Jaguar*" θα επέστρεφε αποτελέσματα "*Jaguar*" και για το ζώο και για το αμάξι, επηρεάζοντας έτσι την ακρίβεια. Παρομοίως, ένα ερώτημα που περιέχει τον όρο "*Contemporary art*" δεν θα επέστρεφε τις πηγές που είναι ευρετηριασμένες με τον περιγραφέα θησαυρών "*Modern art*", επηρεάζοντας έτσι την ανάκτηση.

Κατά καιρούς έχουν δημιουργηθεί διάφορα είδη θησαυρών (μονόγλωσσων – πολύγλωσσων) είτε γενικοί, που αναφέρονται σε ποικίλα πεδία γνώσης, είτε ειδικοί που αναφέρονται σε συγκεκριμένα επιστημονικά πεδία (π.χ. Art & Architecture Thesaurus

²⁴ Iso.org, (2015). *ISO 25964-2:2013 - Information and Documentation -- Thesauri and Interoperability with Other Vocabularies -- Part 2: Interoperability with Other Vocabularies*. [online] Διαθέσιμο στο: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53658 [Πρόσβαση: 14/07/2015].

– AAT²⁵, UNBIS Thesaurus²⁶, UNESCO Thesaurus²⁷, AGROVOC Multilingual agricultural thesaurus²⁸, EuroVoc πολύγλωσσος θησαυρός της Ευρωπαϊκής Ένωσης²⁹ κ.ά.). Στη συνέχεια θα παρουσιαστούν τα πλεονεκτήματα και τα μειονεκτήματα κατά τη χρήση των LCSH ή των θησαυρών.

2.3.2.1. LCSH vs Θησαυροί

Η γλώσσα που χρησιμοποιείται από τις LCSH είναι μακράν η πιο διαδεδομένη γλώσσα θεματικής ευρετηρίασης παγκοσμίως (Anderson & Hofmann, 2006, σ. 9). Οι LCSH έχουν τις καταβολές τους πίσω στον Cutter (1904) και είναι το προϊόν ενός ευρύτατου συνόλου κανόνων. Έκτοτε, αυτοί οι κανόνες έχουν επεκταθεί και αναπτυχθεί από τη Βιβλιοθήκη του Κογκρέσου και άλλες εθνικές βιβλιοθήκες. Οι LCSH *«δεν είναι ένας τυπικός κώδικας ή πρότυπο κατά τα νομικά πλαίσια και έννοιες· δεν προκύπτουν από μια επίσημη συμφωνία με μια αντιπροσωπευτική ομάδα χρηστών· ούτε η ευρεία αποδοχή τους βασίζεται πάνω σε κάποια εξαιρετική ποιότητα που διαθέτουν. Μάλλον, είναι το προϊόν της πρακτικής και των καθημερινών ενεργειών που εκτελούνται σε μια βιβλιοθήκη κατά τη λειτουργία της»* (Williamson, 1996, σ. 160).

Η υπεροχή των LCSH οφείλεται στους ακόλουθους παράγοντες (ALCTS/CCS/SAC/Subcommittee, 1999):

- Είναι ένα πλούσιο λεξιλόγιο που καλύπτει όλες τις θεματικές περιοχές, ουσιαστικά το μεγαλύτερο γενικό λεξιλόγιο ευρετηρίασης στην αγγλική γλώσσα.
- Υπάρχει έλεγχος για συνώνυμα και ομόγραφα.
- Περιέχει πλούσιες σχέσεις (οι παραπομπές υποδεικνύουν συσχετίσεις) ανάμεσα στους όρους. Είναι ένα προ-συνδυσασμένο σύστημα που εξασφαλίζει ακρίβεια κατά την ανάκτηση.

²⁵ Getty.edu, (2015). *Art & Architecture Thesaurus (Getty Research Institute)*. [online] Διαθέσιμο στο: <http://www.getty.edu/research/tools/vocabularies/aat/index.html> [Πρόσβαση: 08/07/2015].

²⁶ Lib-thesaurus.un.org, (2015). *UNBIS Thesaurus*. [online] Διαθέσιμο στο: <http://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf> [Πρόσβαση: 08/07/2015].

²⁷ Databases.unesco.org, (2015). *UNESCO THESAURUS*. [online] Διαθέσιμο στο: <http://databases.unesco.org/thesaurus/> [Πρόσβαση: 07/07/2015].

²⁸ Aims.fao.org, (1980). *AGROVOC Multilingual agricultural thesaurus | Agricultural Information Management Standards (AIMS)*. [online] Διαθέσιμο στο: <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus> [Πρόσβαση: 08/07/2015].

²⁹ Eurovoc.europa.eu, (2015). *Eurovoc, the EU's multilingual thesaurus*. [online] Διαθέσιμο στο: <http://eurovoc.europa.eu/> [Πρόσβαση: 08/07/2015].

- Εξυπηρετούν την περιήγηση σε θέματα πολύπλευρα και με πολλές έννοιες, και
- Το γεγονός ότι έχουν μεταφραστεί και έχουν υιοθετηθεί ως μοντέλο για την ανάπτυξη συστημάτων θεματικών επικεφαλίδων από πολλές χώρες ανά τον κόσμο, καθιστά τις LCSH ως το «*ντε φάκτο*» παγκόσμιο ελεγχόμενο λεξιλόγιο.

Ωστόσο, ενώ το λεξιλόγιο, ή η σηματολογία, των LCSH έχει τόσα να προσφέρει στην περιγραφή και τη διαχείριση των αντίστοιχων πηγών, ο τρόπος με τον οποίο εφαρμόζονται επί του παρόντος στα μοντέρνα συστήματα ανάκτησης πληροφοριών έχει ορισμένους περιορισμούς:

- Οι LCSH είναι ακριβές για να συντηρηθούν· επειδή έχουν περίπλοκη σύνταξη και κανόνες εφαρμογής, η ανάθεση των LCSH στις πηγές σύμφωνα με τις τρέχουσες πολιτικές της Βιβλιοθήκης του Κογκρέσου απαιτεί κατάλληλα εκπαιδευμένο προσωπικό.
- Οι LCSH, στην παρούσα μορφή και εφαρμογή τους, δεν είναι συμβατές στη σύνταξη με άλλα ελεγχόμενα λεξιλόγια, και
- Οι LCSH δεν είναι επιδεκτικές σε μηχανές αναζήτησης έξω από το περιβάλλον του OPAC.

Η μετατόπιση της ανταλλαγής πληροφοριών στο διαδίκτυο και η συνακόλουθη απαίτηση να συμφωνεί κανείς με τις κυρίαρχες τεχνολογίες του φέρνει στην επιφάνεια την ανάγκη για ένα πιο κατάλληλο είδος ελεγχόμενου λεξιλογίου από τις LCSH. Μέσα σε αυτό το πλαίσιο, οι θησαυροί εμφανίστηκαν ως τα θεματικά ελεγχόμενα λεξιλόγια, που είναι ιδιαίτερα συμβατά με το διαδίκτυο. Αυτό οφείλεται στο γεγονός ότι τα πρότυπα, οι οδηγίες και οι κανόνες των θησαυρών είναι πιο συστηματικοί και ακριβείς σε σχέση με τους κανόνες των LCSH. Η σύνταξη των θησαυρών συνήθως συνοδεύεται από την παράλληλη ανάπτυξη οδηγιών για την κατασκευή τους. Με αυτό τον τρόπο, η ανάπτυξη ενός θησαυρού είναι σταδιακή και αθροιστική και τις περισσότερες φορές βασίζεται σε επίσημους κανόνες (Williamson, 1996, σ. 168). Επιπλέον, οι περιγραφείς των θησαυρών καθορίζονται από τη συγκεκριμένη κοινότητα που στην πραγματικότητα τους χρησιμοποιεί, ενώ οι LCSH καθορίζονται από βιβλιοθηκονόμους/επιστήμονες της πληροφόρησης. Επομένως, όχι μόνο ανταποκρίνονται γρηγορότερα οι θησαυροί σε τυχόν αλλαγές στην ορολογία (Chan, 1995, σ. 31), αλλά επίσης, αποτελούνται από πιο ειδικά θέματα σε σχέση με τις LCSH, παρέχοντας με αυτό τον τρόπο περισσότερη ακρίβεια στην ανάκτηση πληροφοριών.

Ολοκληρώνοντας, φαίνεται ότι οι LCSH εφαρμόζονται πάνω στις αρχές της οικονομίας κατά την είσοδο στο σύστημα και του πλεονασμού κατά την αναζήτηση, ενώ οι θησαυροί δίνουν έμφαση στον πλεονασμό κατά την είσοδο στο σύστημα και στην οικονομία κατά την αναζήτηση. Τελευταίο αλλά εξίσου σημαντικό είναι το γεγονός ότι, σε αντίθεση με τις LCSH, οι θησαυροί μπορούν να παρέχουν μεταφρασμένους όρους ως επίσημες καταχωρίσεις τους (πολύγλωσσοι θησαυροί). Αυτό προωθεί τη διαλειτουργικότητα ανάμεσα στα συστήματα ανάκτησης πληροφοριών που απευθύνονται σε πολύγλωσσες κοινότητες.

2.3.3. Ταξινομικά σχήματα

Τα ταξινομικά σχήματα είναι και αυτά ελεγχόμενα λεξιλόγια αλλά όχι με την έννοια των θησαυρών και των θεματικών επικεφαλίδων. Πιο συγκεκριμένα, τα περισσότερα ταξινομικά σχήματα βασίζονται κυρίως σε μια αυστηρά ιεραρχική ταξινόμηση για ένα προκαθορισμένο επιστημονικό πεδίο. Συνήθως αποτελούνται από ένα ελεγχόμενο λεξιλόγιο όρων (μόνο προτιμώμενων όρων), το οποίο είναι δομημένο ιεραρχικά. Κάθε όρος, δηλαδή, σε ένα ταξινομικό σχήμα έχει σχέση της μορφής στενότερος/ευρύτερος όρος με τους υπόλοιπους (Harrington, 2010, σ. 22).

Ένα ταξινομικό σχήμα μπορεί να διαφέρει από ένα θησαυρό σε σχέση με το γεγονός ότι το πρώτο έχει μικρότερες ιεραρχίες και λιγότερο περίπλοκη δομή. Γενικά, τα ταξινομικά σχήματα μπορούν να προσφέρουν συστηματική επισκόπηση της γνώσης με τη χρήση λογικά βελτιωμένων δυνατοτήτων θεματικής περιήγησης, καθιστώντας εύκολη τη μετακίνηση ανάμεσα σε ευρύτερες και στενότερες έννοιες, αφού οι σχέσεις ανάμεσα σε αυτές ορίζονται ρητά. Επίσης, τα ταξινομικά σχήματα σε διεθνές επίπεδο μπορούν να προσφέρουν πολύγλωσση πρόσβαση και να βελτιώσουν τη διαλειτουργικότητα με άλλες υπηρεσίες (Jahns, 2012, σ. 23). Τα πιο διαδεδομένα ταξινομικά σχήματα είναι: α) η Δεκαδική Ταξινόμηση Dewey (Dewey Decimal Classification – DCC³⁰), β) η Παγκόσμια Δεκαδική Ταξινόμηση (Universal Decimal Classification – UDC³¹) και γ) η ταξινόμηση της Βιβλιοθήκης του Κογκρέσου (Library of Congress Classification – LCC³²).

³⁰ Oclc.org, (2015). *Dewey Services*. [online] Διαθέσιμο στο <http://www.oclc.org/dewey/> [Πρόσβαση: 08/07/2015].

³¹ Udcc.org, (2015). *UDC Consortium*. [online] Διαθέσιμο στο: <http://www.udcc.org/pub.htm> [Πρόσβαση: 08/07/2015].

³² Loc.gov, (2015). *Library of Congress Classification*. [online] Διαθέσιμο στο: <http://www.loc.gov/catdir/cpsolcc.html> [Πρόσβαση: 08/07/2015].

Πιο συγκεκριμένα, το DDC επινοήθηκε από τον Melvil Dewey το 1873 και δημοσιεύτηκε για πρώτη φορά το 1876 (σήμερα, Δεκέμβριος 2015, βρίσκεται στην 23^η έκδοσή του που δημοσιεύτηκε το 2011). Αποτελεί ένα εργαλείο για την οργάνωση όλης της ανθρώπινης γνώσης και ανανεώνεται σε τακτά χρονικά διαστήματα για να συμβαδίζει με την εξέλιξή της. Το σύστημα επεκτείνεται περαιτέρω με την κατασκευή αριθμών, με τις μεταφράσεις του και τις σχέσεις του με κατηγοριοποιημένο περιεχόμενο και τέλος με τις αντιστοιχήσεις του σε άλλα θεματικά σχήματα (Jahns, 2012, σ. 24). Είναι δομημένο αυστηρά ιεραρχικά με πλούσιο δίκτυο σχέσεων ανάμεσα στα θέματα του και έχει μια σημειογραφία βασισμένη στους διεθνώς αναγνωρισμένους αραβικούς αριθμούς (Vizine-Goetz, 2002, σ. 11-12).

Το UDC έχει και αυτό τις καταβολές του στα τέλη του 19^{ου} αιώνα και ξεκίνησε από την απόπειρα δύο Βέλγων δικηγόρων να ξεκινήσουν ένα φιλόδοξο έργο για τη δημιουργία μιας περιεκτικής συστηματικής λίστας όλης της γνώσης από τη στιγμή που ανακαλύφθηκε η εκτύπωση. Στο έργο αυτό χρειάζονταν μια κατάλληλη βιβλιογραφική ταξινόμηση. Έτσι, σε συμφωνία με τον Dewey επέκτειναν το σχήμα του και πρόσθεσαν έναν αριθμό συνθετικών διατάξεων και βοηθητικούς πίνακες. Η πρώτη ολοκληρωμένη μορφή του έργου εκδόθηκε το 1907. Πιο συγκεκριμένα, το UDC είναι ένα συνθετικό και ιεραρχικό ταξινομικό σχήμα. Αποτελείται από συστηματικά τακτοποιημένους κύριους πίνακες και βοηθητικούς πίνακες για τις έννοιες που μπορεί να καλύπτουν όλη την ανθρώπινη γνώση σε όλα τα επιστημονικά πεδία. Η σημειογραφία του αποτελείται από αραβικούς αριθμούς και αρκετά σύμβολα (McIlwaine, 2007).

Τέλος, το LCC είναι το σύστημα που εφευρέθηκε από τον Thomas Jefferson για να οργανώσει την προσωπική του βιβλιοθήκη και αγοράστηκε από τη Βιβλιοθήκη του Κογκρέσου. Η ανάπτυξή του ξεκίνησε στα τέλη του 19^{ου} αιώνα για την οργάνωση της συλλογής της. Το σύστημα χωρίζει όλη την ανθρώπινη γνώση σε 21 κύριες κατηγορίες, οι οποίες προσδιορίζονται από ένα γράμμα του αλφαβήτου. Οι περισσότερες από αυτές τις κατηγορίες χωρίζονται περαιτέρω σε υποκατηγορίες, που προσδιορίζονται από συνδυασμούς δύο γραμμάτων ή σε κάποιες περιπτώσεις και τριών γραμμάτων. Κάθε υποκατηγορία περιλαμβάνει χαλαρή ιεραρχική διάταξη των θεμάτων που σχετίζονται με την υποκατηγορία, πηγαίνοντας από το πιο γενικό στο πιο ειδικό. Σε κάθε θέμα εκχωρείται ένας μοναδικός αριθμός ή μια έκταση αριθμών (Jahns, 2012, σ. 26).

Στη συνέχεια, θα παρουσιαστεί ο λόγος για τον οποίο η θεματική αναζήτηση και ανάκτηση πληροφοριών φαίνεται να μην είναι ο πιο δημοφιλής τρόπος αναζήτησης σε

μια ψηφιακή βιβλιοθήκη, παρ' όλο το χρόνο που καταναλώνει εξειδικευμένο προσωπικό για τη θεματική ευρετηρίαση των πηγών στις βιβλιοθήκες και κατ' επέκταση στις ψηφιακές βιβλιοθήκες (Paradakis et al., 2009).

2.4. Θεματική αναζήτηση και ανάκτηση μέσω της χρήσης των LCSH και των θησαυρών

Ο σημασιολογικός πλούτος που ενθυλακώνεται στις LCSH μαζί με την ευρεία διάδοσή τους θα μπορούσε να οδηγήσει στο συμπέρασμα ότι οι τελικοί χρήστες προτιμούν να κάνουν αναζητήσεις βάσει θέματος στις ψηφιακές βιβλιοθήκες. Αυτό όμως απέχει πολύ από την πραγματικότητα. Η θεματική αναζήτηση φαίνεται να μην είναι δημοφιλής τρόπος για να απευθύνει κανείς ερωτήματα σε μια ψηφιακή βιβλιοθήκη. Αυτό θα μπορούσε να οφείλεται σε πολλούς λόγους, όπως θα δούμε στη συνέχεια.

Σύμφωνα με την κοινή πρακτική, στη θεματική ανάκτηση πληροφορίας στις ψηφιακές βιβλιοθήκες, οι ερευνητές επιλέγουν λέξεις και φράσεις που είναι σημασιολογικά πιο κοντά στις πληροφοριακές τους ανάγκες και αντίστοιχα τις συνδυάζουν με τη χρήση λογικών τελεστών Boolean³³, ελπίζοντας ότι με αυτό τον τρόπο θα έχουν κάποια ταύτιση μέσα στο υποκείμενο ευρετήριο. Ωστόσο, ένα τέτοιο σενάριο μπορεί να στεφθεί με επιτυχία μόνο στην περίπτωση που η ορολογία των ερευνητών ταιριάζει με την ορολογία που χρησιμοποιείται στις υποκείμενες θεματικές επικεφαλίδες (Bennett, 2006). Επιπλέον, ακόμα και στην περίπτωση που ένα μέρος από μια λέξη-κλειδί ταιριάζει με μια συμβολοσειρά των θεματικών επικεφαλίδων, το σύστημα ανάκτησης πληροφοριών δεν θα εκμεταλλευτεί την έμφυτη ικανότητα των LCSH να μεταβούν από τη μια θεματική επικεφαλίδα στην άλλη, μέσω των ορισμένων συνδετικών σχέσεων. Πολύ συχνά, η μόνη επιλογή για να πλοηγηθεί κανείς στις θεματικές επικεφαλίδες είναι μέσω ενός αλφαβητικού ευρετηρίου πριν από την έναρξη της διαδικασίας ανάκτησης πληροφοριών. Ωστόσο, αυτό το είδος αναζήτησης ελαττώνει τη σημασιολογική εκφραστικότητα των LCSH.

³³ Οι λογικοί τελεστές Boolean χρησιμοποιούνται για να συνδέσουν και να ορίσουν τη σχέση ανάμεσα στους όρους αναζήτησης. Όταν γίνεται αναζήτηση σε ηλεκτρονικές βάσεις δεδομένων, οι τελεστές Boolean μπορούν να χρησιμοποιηθούν είτε για να περιοριστούν είτε για να διευρυνθούν τα αποτελέσματα αναζήτησης. Υπάρχουν 3 βασικοί λογικοί τελεστές Boolean ΚΑΙ (AND), Η (OR) και ΟΧΙ (NOT). (Μετάφραση από: Library.albany.edu, (2015). Boolean Operators. [online] Διαθέσιμο στο: <http://library.albany.edu/subject/tutorials/education/boolean.html> [Πρόσβαση: 08/07/2015]).

Τα γλωσσικά προβλήματα μπορούν επιπλέον να επηρεάσουν την ποιότητα της διαδικασίας ανάκτησης πληροφοριών. Πιο συγκεκριμένα, τα ερωτήματα αναζήτησης εκφραζόμενα ως συνεχόμενοι όροι είναι πιθανό να πάσχουν από πολυσημία ή/και συνωνυμία στους όρους που αποτελούν το ερώτημα αναζήτησης. Η πολυσημία συμβαίνει όταν μια λέξη έχει παραπάνω από μια έννοια και η συνωνυμία συμβαίνει όταν δύο ή περισσότερες λέξεις μοιράζονται το ίδιο νόημα (Miller, 1995, σ. 40). Αυτά τα προβλήματα πηγάζουν από το συνολικό σχεδιασμό του συστήματος και όχι το λεξιλόγιο που χρησιμοποιείται - οι LCSH συνήθως αποτελούνται από πολλές λέξεις που περιγράφουν ένα συγκεκριμένο σημασιολογικό πλαίσιο.

Επομένως, όταν πρόκειται να αξιολογηθεί η αποτελεσματικότητα της χρήσης των LCSH σε υπηρεσίες θεματικής ανάκτησης πληροφοριών, τα αποτελέσματα είναι μάλλον αποθαρρυντικά (Calhoun, 2006, σ. 24-25). Θα μπορούσε κάποιος να ισχυριστεί ότι αφού οι LCSH σχεδιάστηκαν για ένα διαφορετικό περιβάλλον, δεν δουλεύουν καλά με τη λειτουργικότητα αναζήτησης από τη χρήση Boolean τελεστών, η οποία παρέχεται από την πλειονότητα των σύγχρονων συστημάτων ανάκτησης πληροφοριών μέσω λέξεων κλειδιών στους OPAC ή/και στις ψηφιακές βιβλιοθήκες.

Για να αντιμετωπιστούν μερικά από τα προαναφερθέντα ζητήματα που αφορούν τη θεματική ανάκτηση πληροφορίας, χρησιμοποιούνται οι θησαυροί (Asghar, Revie & Ghowdhury, 2002, σ. 113; Gordon & Domeshek, 1998, σ. 127). Κατά τη διάρκεια της ανάκτησης και της αναζήτησης, οι θησαυροί είναι χρήσιμοι στη γεφύρωση του χάσματος που υπάρχει ανάμεσα στα μεταδεδομένα που παρέχονται από τους ευρετηριαστές και τις έννοιες που παρουσιάζονται από τους ερευνητές (Bates, 1998). Επιπλέον, οι θησαυροί μπορούν να χρησιμοποιηθούν για να προωθήσουν την καθοδηγούμενη ή/και την πολύπλευρη πλοήγηση αξιοποιώντας τη σημασιολογική δομή ανάμεσα στους όρους (Bechhofer & Goble, 2001, σ. 26). Μια τέτοια πλοήγηση οδηγεί τελικά στην κατασκευή ερωτήματος, το οποίο απευθύνεται στην υποκείμενη μηχανή αναζήτησης.

Παρά τα εγγενή μειονεκτήματα τους, όπως αναφέρθηκαν από τους Bechhofer & Goble (2001, σ. 29), οι θησαυροί έχουν αποδειχτεί αρκετά αποτελεσματικοί για την ανακάλυψη πληροφοριών σε έντυπα εργαλεία. Ωστόσο, εξ' αιτίας του γεγονότος ότι η σημασιολογική τους δομή είναι λιγότερο ανεπτυγμένη για τους χρήστες στα διαδικτυακά περιβάλλοντα, η επίδρασή τους στη θεματική ανάκτηση πληροφορίας είναι κάτω του αναμενόμενου. Ο Schwartz (2008, σ. 832-836) παρέχει μια εκτενή και με σχόλια επισκόπηση της πρόσφατης έρευνας στη θεματική ανάκτηση πληροφορίας. Α-

κολουθώντας αυτή την έρευνα, θα μπορούσε να ειπωθεί ότι τα συστήματα που βασίζονται στις οντολογίες φαίνεται να είναι ικανά να αντιμετωπίζουν πολλά από τα προβλήματα που αφορούν στη μοντελοποίηση της θεματικής δομής των θησαυρών, συμβάλλοντας έτσι στην ανάπτυξη πιο αποτελεσματικών συστημάτων ανάκτησης θεματικής πληροφορίας.

Μέσα σε αυτό το πλαίσιο, ένα σύστημα σχετικό με τους θησαυρούς που θα μπορούσε να εφαρμοστεί για να εξυπηρετήσει και τις LCSH είναι το ONKI-SKOS (Tuominen et al., 2008). Ο εξυπηρετητής ONKI-SKOS παρέχει έναν ομοιόμορφο τρόπο πρόσβασης θησαυρών που αναπαρίστανται ως λεξιλόγια του SKOS. Όσον αφορά στο περιβάλλον διεπαφής χρήστη, ο εξυπηρετητής ONKI-SKOS παρέχει μια λειτουργία αυτόματης συμπλήρωσης, η οποία λειτουργεί ως ένα σημείο εισαγωγής για ολόκληρη τη διαδικασία της ανάκτησης πληροφοριών βασισμένη σε θησαυρούς. Κατόπιν, παρέχεται στους χρήστες ένα περιβάλλον διεπαφής με υπερσυνδέσμους για να διατρέξουν το θησαυρό, ο οποίος είναι κωδικοποιημένος σε μορφή OWL. Το ONKI-SKOS μπορεί να χρησιμοποιηθεί για την περιήγηση, την αναζήτηση και την οπτικοποίηση οποιουδήποτε λεξιλογίου που συμμορφώνεται με τις προδιαγραφές του SKOS. Οι θησαυροί στο ONKI-SKOS μπορούν να αναπαριστούν βασικές σχέσεις SKOS ανάμεσα σε έννοιες, όπως είναι οι ιεραρχικές σχέσεις και τα συνώνυμα.

Στη συνέχεια, θα παρουσιαστεί η προσέγγιση που προτάθηκε από τους Paradakis et al. (2009) για τη θεματική αναζήτηση και ανάκτηση πληροφοριών στις ψηφιακές βιβλιοθήκες.

2.5. Προτεινόμενη προσέγγιση στη θεματική ανάκτηση πληροφοριών

Η προτεινόμενη προσέγγιση προσπαθεί να εκμεταλλευτεί την εγγενή σημασιολογική εκφραστικότητα που υπάρχει στις LCSH με σκοπό να παρέχει πιο αποτελεσματική θεματική ανάκτηση πληροφοριών στις ψηφιακές βιβλιοθήκες.

Το πρώτο βήμα προς αυτή την κατεύθυνση είναι να επεκταθεί η συνδετική δομή (δηλ. ευρύτεροι, στενότεροι, σχετικοί όροι) που διατρέχει τις LCSH, με τη χρήση σχέσεων βασισμένες στις υποδιαιρέσεις. Όπως θα παρουσιαστεί αργότερα σε αυτή την ενότητα, η παροχή μεγαλύτερης συνδεσιμότητας ανάμεσα στις LCSH καταλήγει στη δημιουργία αποτελεσματικότερων υπηρεσιών ανάκτησης θεματικής πληροφορίας στις ψηφιακές βιβλιοθήκες. Έπειτα, η προκύπτουσα εκτεταμένη συνδετική δομή αντιστοιχεί σε ένα κατάλληλα σχεδιασμένο οντολογικό σχήμα. Αυτό το σχήμα μπορεί να

εφαρμοστεί σε οποιαδήποτε ψηφιακή βιβλιοθήκη που περιέχει θεματικούς περιγραφείς σύμφωνα με τις οδηγίες των LCSH.

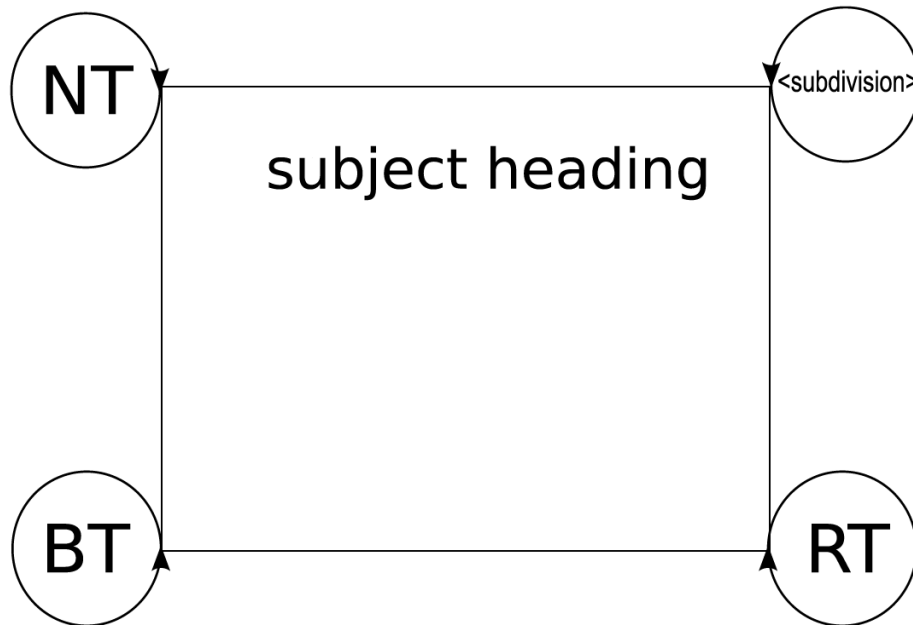
Τέλος, προκειμένου να αξιοποιηθεί ο σημασιολογικός πλούτος της υποκείμενης οντολογίας, παρουσιάζεται ένα κατάλληλο γραφικό περιβάλλον διεπαφής χρήστη. Το περιβάλλον διεπαφής δίνει τη δυνατότητα στους χρήστες να εξοικειωθούν με την ορολογία που υπάρχει στις LCSH και τους καθοδηγεί να επιλέξουν το θέμα που ταιριάζει καλύτερα στις πληροφοριακές τους ανάγκες. Η καθοδηγούμενη πλοήγηση βασίζεται στις σχέσεις ανάμεσα στις θεματικές επικεφαλίδες που προέρχονται από την εκτεταμένη συνδετική δομή που αναφέρθηκε προηγουμένως.

2.5.1. Επεκτείνοντας τη συνδετική δομή των LCSH

Η συνδετική δομή των LCSH αποτελείται από τρεις σχέσεις, δηλαδή ευρύτερες, στενότερες (δηλ. ιεραρχικές σχέσεις) και σχετικές ("*related*") (δηλ. συνειρμικές σχέσεις) (Marshall, 2003). Εκτός από τη συνδετική δομή ("*syndetic structure*"), οι LCSH σχετίζονται εμμέσως μέσω των (πιθανών) κοινών υποδιαιρέσεων. Οι θεματικές επικεφαλίδες που σχετίζονται μεταξύ τους μέσω των υποδιαιρέσεων, ορίζουν την εκτεταμένη συνδετική δομή ("*extended syndetic structure*"). Όπως έχει ήδη αναφερθεί (βλέπε ενότητα 2.3.1), σύμφωνα με τις οδηγίες των LCSH υπάρχουν τέσσερις τύποι υποδιαιρέσεων (δηλ. θεματικές, γεωγραφικές, χρονικές και μορφής). Ένας λοιπόν από τους στόχους της προτεινόμενης προσέγγισης είναι να εκμεταλλευτεί αυτούς τους τέσσερις τύπους υποδιαιρέσεων για να επεκτείνει τη συνδετική δομή των LCSH. Οι σχέσεις που βασίζονται στις υποδιαιρέσεις δεν χρησιμοποιούνται από τα σύγχρονα συστήματα ανάκτησης θεματικής πληροφορίας στις βιβλιοθήκες, μειώνοντας έτσι την πιθανότητα να παρέχονται δυνατότητες αναζήτησης βασισμένες στις κοινές υποδιαιρέσεις των θεματικών επικεφαλίδων.

2.5.2. Μοντελοποίηση των σχέσεων ανάμεσα στις LCSH σε ένα οντολογικό σχήμα

Για να εκφραστεί επίσημα η εκτεταμένη συνδετική δομή των θεματικών επικεφαλίδων, θα μπορούσε να χρησιμοποιηθεί η γλώσσα OWL. Πιο συγκεκριμένα, η έκδοση OWL-DL βασίζεται στην Περιγραφική Λογική (Description Logics – DL) (Baader, 2007), η οποία έχει ελκυστική και καλά κατανοητή υπολογιστική εκφραστικότητα. Η OWL θεωρείται μια από τις θεμελιώδεις τεχνολογίες που υποστηρίζουν το σημασιολογικό ιστό. Η εικόνα 2.1 απεικονίζει το σχήμα της προτεινόμενης οντολογίας:



Εικόνα 2.1: Το σχήμα της οντολογίας

Το παραπάνω σχήμα δημιουργεί οντολογίες, οι οποίες περιέχουν μια ιεραρχία κλάσεων που μοντελοποιούν τις θεματικές επικεφαλίδες ως κλάσεις. Σύμφωνα με τη συνδεδετική δομή, μια θεματική επικεφαλίδα μπορεί να σχετίζεται με μία άλλη μέσω της χρήσης των σχέσεων: ευρύτερος όρος (Broader Term – BT), στενότερος όρος (Narrower Term – NT) ή σχετικός όρος (Related Term – RT). Πιο αναλυτικά, οι σχέσεις «ευρύτερος» και «στενότερος» όρος, όχι πάντα αλλά αρκετά συχνά (Bechhofer & Goble, 2001), αναφέρονται στην ιεραρχική δομή των θεματικών επικεφαλίδων, ενώ η σχέση «σχετικός όρος» φέρνει κοντά θεματικές επικεφαλίδες που μοιράζονται μια παρόμοια έννοια. Ως εκ τούτου, η ιεραρχική δομή των θεματικών επικεφαλίδων αντιστοιχείται στην οντολογική ιεραρχία κλάσεων. Το σύνολο σχέσεων <subdivision> αναφέρεται στις σχέσεις που μοντελοποιούν την εκτεταμένη συνδεδετική δομή. Πιο συγκεκριμένα, οι υποδιαίρεσεις μοντελοποιούνται ως συμμετρικές ιδιότητες αντικειμένου ανάμεσα στις κλάσεις, οι οποίες αντιστοιχούν σε θεματικές επικεφαλίδες που σχετίζονται μεταξύ τους μέσω των υποδιαίρεσεών τους. Για παράδειγμα, οι θεματικές επικεφαλίδες:

1. Automobile industry and trade -- Law and legislation -- Greece
2. Automobiles -- Law and legislation -- Greece

Μοιράζονται τις υποδιαίρεσεις “*Law and Legislation*” και “*Greece*”. Επομένως, η οντολογία συνδέει τις αντίστοιχες κλάσεις με δύο επιπλέον ιδιότητες αντικειμένου, δηλαδή “*Law and Legislation*” και “*Greece*”.

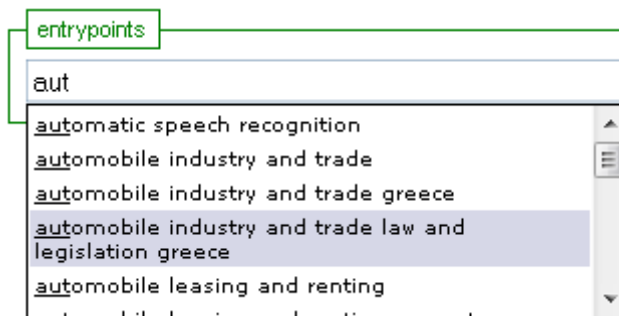
2.5.3. Από τη θεωρία στην πράξη: σημασιολογική ανάκτηση πληροφοριών με τη χρήση των LCSH

Όταν σειριοποιηθεί, αποθηκευτεί και ευρετηριαστεί με αποτελεσματικό τρόπο, η προτεινόμενη οντολογία μπορεί να χρησιμεύσει ως βάση για την ανάπτυξη αποδοτικών και σημασιολογικά πλούσιων υπηρεσιών ανάκτησης πληροφοριών στις ψηφιακές βιβλιοθήκες. Αυτές οι υπηρεσίες θα μπορούσαν να αντικαταστήσουν τα παραδοσιακά ευρετήρια λέξεων κλειδιών στα συστήματα υποστήριξης που χρησιμοποιούνται από τις περισσότερες υπηρεσίες ανάκτησης πληροφοριών στις σύγχρονες ψηφιακές βιβλιοθήκες.

Για να είναι αποτελεσματική η οντολογία, είναι απαραίτητη η ύπαρξη ενός γραφικού περιβάλλοντος ικανό να προβάλλει την προαναφερθείσα σημασιολογική εκφραστικότητα. Τα σύγχρονα γραφικά περιβάλλοντα στις ψηφιακές βιβλιοθήκες για τη θεματική ανάκτηση πληροφορίας πολλές φορές αποτυγχάνουν στο να καλύψουν μια τέτοια απαίτηση. Συνήθως παρέχουν ένα πλαίσιο αναζήτησης, το οποίο εφαρμόζει ένα πρότυπο αντιστοίχισης των λέξεων κλειδιών που παρέχονται από τους χρήστες με ένα ευρετήριο θεματικών συμβολοσειρών. Στην περίπτωση των θεματικών συμβολοσειρών των LCSH, όλα τα οφέλη από τον έλεγχο και την τυποποίηση αυτών των συμβολοσειρών χάνονται και οι ερευνητές αφήνονται σε στατιστικούς αλγόριθμους για να μπορέσουν να ικανοποιήσουν τις πληροφοριακές τους ανάγκες. Επομένως, υπάρχει η ανάγκη για ένα περιβάλλον διεπαφής χρήστη ικανό να εμφανίζει στον τελικό χρήστη την εκφραστικότητα των συμβολοσειρών των LCSH.

Όπως αναφέρθηκε νωρίτερα σε αυτή την ενότητα, μια θεμελιώδης απαίτηση για ένα τέτοιο περιβάλλον διεπαφής χρήστη είναι η ικανότητα να γεφυρώνει το χάσμα ανάμεσα στις πληροφοριακές ανάγκες των χρηστών και της ορολογίας που χρησιμοποιείται για την ανάπτυξη των υποκείμενων θεματικών συμβολοσειρών των LCSH. Επιπλέον, η παρεχόμενη λειτουργικότητα θα πρέπει να βρει τη χρυσή τομή στο γραφικό περιβάλλον, διατηρώντας παράλληλα τα βασικά χαρακτηριστικά ευχρηστίας (Krug, 2000, σ. 95), που αποτελούν το σύνολο του συστήματος ανάκτησης πληροφοριών ελκυστικό για τους χρήστες του. Άλλωστε, όπως αναφέρθηκε από την Bates (2002), η τελική μέτρηση της απόδοσης ενός συστήματος θα είναι πάντα το πως οι χρήστες στην πραγματικότητα χρησιμοποιούν το σύστημα και όχι το πως κάποιος περιμένει να το χρησιμοποιήσουν.

Έχοντας τα παραπάνω κατά νου, παρουσιάζεται ένα περιβάλλον διεπαφής, το οποίο είναι ικανό να καθιστά την εκφραστικότητα της υποκείμενης οντολογίας άμεσα διαθέσιμη στους χρήστες, ενώ παράλληλα διατηρεί την απλότητα και την ευχρηστία. Το προτεινόμενο περιβάλλον διεπαφής χτίζεται πάνω σε προηγούμενη δουλειά (Paradakis, Stefanidakis & Tzali, 2007) σχετική με την οπτικοποίηση οντολογίας, παρέχοντας μια επέκταση της συνδεσιμότητας των θεματικών επικεφαλίδων όπως αυτή εκφράζεται μέσω των υποδιαίρεσεων. Αποτελείται από ένα πεδίο αυτόματης συμπλήρωσης (autosuggest), όπου οι χρήστες καλούνται να πληκτρολογήσουν τα πρώτα γράμματα των λέξεων που περιγράφουν καλύτερα τις πληροφοριακές τους ανάγκες. Το γραφικό περιβάλλον επιστρέφει μια λίστα με τις θεματικές επικεφαλίδες που περιέχουν τη συμβολοσειρά που παρέχεται από τον χρήστη. Κατόπιν, ο χρήστης μπορεί να επιλέξει τη θεματική επικεφαλίδα που ταιριάζει καλύτερα στις πληροφοριακές του ανάγκες (βλέπε εικόνα 2.2).

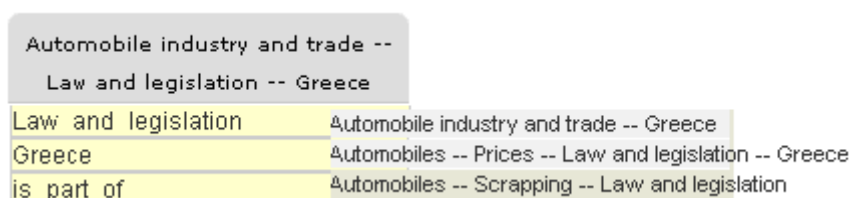


Εικόνα 2.2: Πεδίο αυτόματης συμπλήρωσης

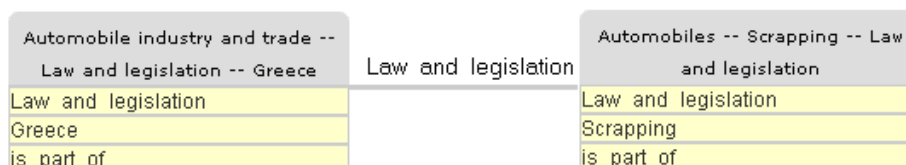
Με αυτό τον τρόπο, οι χρήστες εκτελούν μια αρχική αναζήτηση ταιριάσματος μοτίβου (pattern-matching search) στις θεματικές επικεφαλίδες που βασίζονται στις LCSH, οι οποίες χρησιμοποιούνται για να περιγράψουν το θέμα των υποκείμενων πληροφοριακών πηγών. Ωστόσο, η οντολογία θα μπορούσε να περιέχει επίσης θεματικές επικεφαλίδες που είναι σημασιολογικά κοντά στις πληροφοριακές ανάγκες των χρηστών χωρίς να περιέχουν τη συγκεκριμένη συμβολοσειρά. Όπως αναφέρθηκε νωρίτερα σε αυτή την ενότητα, μια τέτοια προσέγγιση στις πληροφοριακές ανάγκες των χρηστών δεν μπορεί να υλοποιηθεί από τα τρέχοντα συστήματα.

Έτσι, για να βοηθηθούν οι χρήστες στην εύρεση των θεματικών επικεφαλίδων, η προτεινόμενη προσέγγιση παρέχει μια διαδικασία περιήγησης, η οποία εκμεταλλεύεται την εκτεταμένη συνδετική δομή της υποκείμενης οντολογίας, όπως ορίστηκε στην ενότητα 2.5.1.

Πιο συγκεκριμένα, κατά την επιλογή μιας θεματικής επικεφαλίδας από το πλαίσιο αυτόματης συμπλήρωσης, σχεδιάζεται ένα πλαίσιο που αναπαριστά τη θεματική επικεφαλίδα. Το πλαίσιο περιέχει τον τίτλο της επικεφαλίδας ακολουθούμενη από τις πιθανές σχέσεις που μπορεί να έχει με άλλες θεματικές επικεφαλίδες στην οντολογία (βλέπε εικόνα 2.3). Πατώντας πάνω σε μια σχέση, εμφανίζεται στο χρήστη μια λίστα με θεματικές επικεφαλίδες, οι οποίες σχετίζονται με τη θεματική επικεφαλίδα που σχεδιάστηκε προηγουμένως μέσω της επιλεγμένης σχέσης (βλέπε εικόνα 2.3). Ύστερα, πατώντας πάνω σε μια από τις θεματικές επικεφαλίδες, ένα άλλο πλαίσιο σχεδιάζεται δίπλα στο πρώτο. Τα δύο πλαίσια συνδέονται με μια γραμμή στην οποία αναγράφεται η περιγραφή της επιλεγμένης σχέσης (βλέπε εικόνα 2.4). Αυτή η διαδικασία μπορεί να επαναληφθεί μέχρις ότου ο χρήστης εντοπίσει μια θεματική επικεφαλίδα που ικανοποιεί τις πληροφοριακές του ανάγκες.



Εικόνα 2.3: Σχέσεις μεταξύ των θεματικών επικεφαλίδων



Εικόνα 2.4: Συνδέσεις μεταξύ των θεματικών επικεφαλίδων

Στο τέλος, η προκύπτουσα θεματική επικεφαλίδα μπορεί να απευθυνθεί ως έχει στην υποκείμενη μηχανή αναζήτησης με σκοπό να ανακτηθούν πληροφοριακές πηγές που έχουν χαρακτηριστεί με τη συγκεκριμένη θεματική επικεφαλίδα. Αυτό το τελικό βήμα εμπλέκει ταίριασμα μοτίβου με το παραδοσιακό θεματικό ευρετήριο που περιέχει τις LCSH. Ωστόσο, το ταίριασμα μοτίβου δεν εφαρμόζεται ανάμεσα στις συμβολοσειρές που παρέχονται από τους χρήστες και τις συμβολοσειρές που περιέχονται στο σύστημα, όπως ακριβώς συμβαίνει με τις σημερινές υπηρεσίες ανάκτησης πληροφοριών. Σύμφωνα με την προτεινόμενη προσέγγιση, το ταίριασμα μοτίβου εφαρμόζεται ανάμεσα στις συμβολοσειρές που προτείνονται από το σύστημα (και αντίστοιχα επιλέγονται από το χρήστη) και στις συμβολοσειρές που περιέχονται στο σύστημα. Με αυτό τον τρόπο, οι τελικοί χρήστες εξοικειώνονται με την ειδική ορολογία που χρησιμοποιείται για να περιγραφούν τα θέματα των υποκείμενων πηγών και αντίστοιχα να απευθύνουν στοχευμένα ερωτήματα στο σύστημα ανάκτησης πληροφοριών.

Ολοκληρώνοντας, η προτεινόμενη προσέγγιση παρέχει μια υπηρεσία θεματικής ανάκτησης πληροφοριών που χτίζεται πάνω σε θέματα που συμμορφώνονται με τις οδηγίες των LCSH. Αποτελείται από ένα επίπεδο παρουσίασης (presentation layer) και ένα επίπεδο πρόσβασης στα δεδομένα (data access layer), τα οποία είναι ενσωματωμένα σε ένα ενοποιημένο σύστημα ανάκτησης πληροφοριών.

Ένα τέτοιο σύστημα θα μπορούσε να χρησιμοποιηθεί από διάφορες ψηφιακές βιβλιοθήκες βασισμένες στις LCSH που επιθυμούν να παρέχουν θεματική ανάκτηση πληροφοριών στους χρήστες τους. Με μικρές τροποποιήσεις στο σχεδιασμό της υποκείμενης οντολογίας, η προτεινόμενη μεθοδολογία θα μπορούσε να εφαρμοστεί ουσιαστικά σε οποιοδήποτε είδος θησαυρού που χρησιμοποιείται για να μοντελοποιήσει τα θέματα των πληροφοριακών πηγών μιας ψηφιακής βιβλιοθήκης, παρέχοντας έτσι ένα γενικό εργαλείο για τη διαδραστική, θεματική ανάκτηση πληροφοριών με επίκεντρο το χρήστη.

2.5.4. Μελέτη περίπτωσης

Η προτεινόμενη προσέγγιση εφαρμόστηκε στην ψηφιακή βιβλιοθήκη μεταπτυχιακών και διδακτορικών διατριβών του Πανεπιστημίου Πειραιά (Διώνη)³⁴. Πιο συγκεκριμένα, αναπτύχθηκε μια υπηρεσία θεματικής ανάκτησης πληροφοριών και ενσωματώθηκε στην αντίστοιχη ψηφιακή βιβλιοθήκη που είναι βασισμένη στο DSpace³⁵.

Σύμφωνα με την προτεινόμενη προσέγγιση, η συνδετική δομή εκφράζεται μέσω του ορισμού τριών σχέσεων (δηλ. "*contains*", "*is_part_of*" και "*seeAlso*"). Για να εκφραστεί η εκτεταμένη συνδετική δομή, ορίστηκε ένας αριθμός σχέσεων που αντιστοιχούν στις υποδιαιρέσεις κάθε μιας από τις LCSH της ψηφιακής βιβλιοθήκης (Διώνη). Ως εκ τούτου, τα θέματα που μοιράζονται μια κοινή υποδιαίρεση συνδέονται μεταξύ τους με μία αντίστοιχα ορισμένη σχέση.

Στην επόμενη ενότητα, μια συγκριτική αξιολόγηση της προτεινόμενης προσέγγισης θα παρουσιαστεί σε σχέση με την παραδοσιακή διαδικασία θεματικής αναζήτησης που παρέχεται από την ψηφιακή βιβλιοθήκη (Διώνη).

³⁴ Dione.lib.unipi.gr, (2015). *Διώνη: Αρχική*. [online] Διαθέσιμο στο: <http://dione.lib.unipi.gr/> [Πρόσβαση: 04/10/2015].

³⁵ Dspace.org, (2015). *DSpace / DSpace is a Turnkey Institutional Repository Application*. [online] Διαθέσιμο στο: <http://www.dspace.org> [Πρόσβαση: 10/07/2015].

2.5.4.1. Συγκριτική αξιολόγηση

Σύμφωνα με την τρέχουσα πρακτική, οι χρήστες αλληλεπιδρούν με το σύστημα μέσω μιας διεπαφής σε μορφή φόρμας, όπου καλούνται να επιλέξουν το θέμα που έχουν στο μυαλό τους μέσα από ένα αλφαβητικό ευρετήριο ή αντίστοιχα να πληκτρολογήσουν τα πρώτα γράμματα του θέματος σε ένα πλαίσιο αναζήτησης. Ύστερα, το σύστημα προσπαθεί να ταιριάξει το κείμενο που εισάγει ο χρήστης με θεματικές συμβολοσειρές του συστήματος που έχουν το ίδιο πρόθεμα. Στην περίπτωση ενός επιτυχημένου ταιριάσματος, το αλφαβητικό ευρετήριο εμφανίζει το θέμα που ταίριαξε ακριβώς με το κείμενο του χρήστη μαζί με τα υπόλοιπα που είναι κοντά βάσει της συμβολοσειράς τους (βλέπε εικόνα 2.5). Τέλος, οι χρήστες μπορούν να επιλέξουν ένα θέμα του ευρετηρίου προκειμένου να ανακτήσουν τα τεκμήρια στα οποία έχει εκχωρηθεί το συγκεκριμένο θέμα (βλέπε εικόνα 2.6). Η μόνη πιθανότητα για να περιηγηθεί κανείς στα σχετικά θέματα είναι να επιλέξει ένα τεκμήριο και κατόπιν να επιλέξει κάποιο άλλο θέμα από αυτά που έχουν εκχωρηθεί στο συγκεκριμένο τεκμήριο.

Browse by Subject

Jump to: 0-9
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω
or enter first few letters:

Showing subjects 1-21 of 2551.

[Next page](#)

Έλαιο
Έλαιο-Απόβλητα
ABC (συστήμα κωδικοποίησης)
Academic achievement
Accelerated life testing
Acceptance sampling
Accountants-Professional ethics
Accounting-data processing
Accounting-Moral and ethical aspects
Accounting-standards
Accrual basis accounting
Activity-based costing
Adjustable rate mortgages
Adult education
Advertising
Advertising laws-European Economic Community countries
Aeronautics, Commercial-Freight-Safety measures
Agricultural biotechnology
Air conditioning-Equipment and supplies

Εικόνα 2.5: Εκτελώντας μια παραδοσιακή θεματική αναζήτηση

Items for Subject "Advertising"

[Return to Browse by Subject](#)

[Sort by Title](#) **Sorting by Date**

Showing 1 items.

Issue Date	Title	Author(s)
19 Ιούν 2007	Ολοκληρωμένη δοκιμή επικοινωνίας : η περίπτωση της εταιρείας BMW	Ξανθοπούλου, Λουίζα Κ.

Εικόνα 2.6: Αποτελέσματα παραδοσιακής θεματικής αναζήτησης

Έτσι, η περιήγηση στα θέματα μέσω των τεκμηρίων είναι μια μάλλον κουραστική διαδικασία, η οποία δεν εκμεταλλεύεται την (εκτεταμένη) συνδυαστική δομή των υποκείμενων θεμάτων των LCSH.

Για να αξιολογηθεί η αποτελεσματικότητα της προτεινόμενης προσέγγισης, παρουσιάζονται και αντίστοιχα αξιολογούνται τρία σενάρια αναζήτησης βάσει θέματος. Κάθε σενάριο εφαρμόζεται στο παραδοσιακό σύστημα καθώς και στην προτεινόμενη υπηρεσία θεματικής ανάκτησης πληροφοριών.

Σενάριο 1: ο χρήστης πρέπει να βρει τεκμήρια για το θέμα "*Instructional systems*".

Σύμφωνα με την παραδοσιακή προσέγγιση, ο χρήστης εντοπίζει το θέμα "*Instructional systems - Design*" (βλέπε εικόνα 2.7).

Jump to: 0-9
[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)
[Α](#) [Β](#) [Γ](#) [Δ](#) [Ε](#) [Ζ](#) [Η](#) [Θ](#) [Ι](#) [Κ](#) [Λ](#) [Μ](#) [Ν](#) [Ξ](#) [Ο](#) [Π](#) [Ρ](#) [Σ](#) [Τ](#) [Υ](#) [Φ](#) [Χ](#) [Ψ](#) [Ω](#)

or enter first few letters:

Showing subjects 487-507 of 2551.

Information warfare
Input-output analysis
Insider trading in securities-Law and legislation-Greece
Instructional systems-Design
Insurance, Automobile
Insurance, Automobile-Cyprus
Insurance, Disaster-Attica (Greece)
Insurance-Finance
Integrated software
Intelligent agents (Computer Software)
Interbank market
Intercultural communication
Interest rate
Interest rates
Interest rates-European Union countries
Interest rates-Mathematical models
Integrated logistics support
International business enterprises-Personnel management
International business enterprises-Personnel management-Case studies
International cooperation
International finance

Εικόνα 2.7: Παραδοσιακή προσέγγιση: αναζήτηση του θέματος "Instructional systems"

Πατώντας πάνω στο θέμα, υπάρχει μόνο ένα τεκμήριο διαθέσιμο (βλέπε εικόνα 2.8).

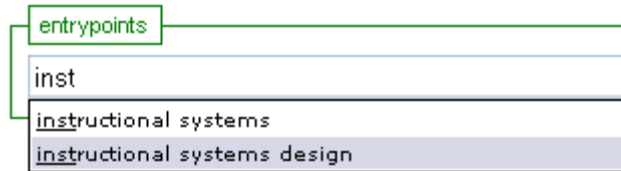
[Sort by Title](#) **Sorting by Date**

Showing 1 items.

Issue Date	Title	Author(s)
20-Φεβ-2007	Μεθοδολογίες σχεδίασης εκπαιδευτικών δραστηριοτήτων σε περιβάλλοντα ηλεκτρονικής μάθησης	Κοκκινός, Αντώνιος

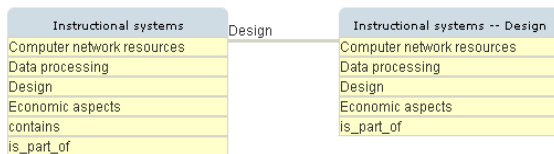
Εικόνα 2.8: Παραδοσιακή προσέγγιση: τεκμήρια που περιγράφονται με το θέμα "Instructional systems"

Από την άλλη μεριά, σύμφωνα με την προτεινόμενη προσέγγιση, ο χρήστης πληκτρολογεί τα πρώτα γράμματα του θέματος που έχει στο μυαλό του στο πεδίο αυτόματης συμπλήρωσης (βλέπε εικόνα 2.9).



Εικόνα 2.9: Προτεινόμενη προσέγγιση: αναζήτηση του θέματος "Instructional systems"

Υστερα, ο χρήστης επιλέγει το θέμα "*Instructional systems*" και πατώντας πάνω στη σχέση "*Design*" του παρουσιάζεται σε ένα πλαίσιο το θέμα "*Instructional systems - Design*" μαζί με το αντίστοιχο αποτέλεσμα, το οποίο είναι προφανώς ίδιο με την παραδοσιακή προσέγγιση (βλέπε εικόνα 2.10).



Subject	"Instructional systems - Design"	and
Subject		and
Subject		

Search

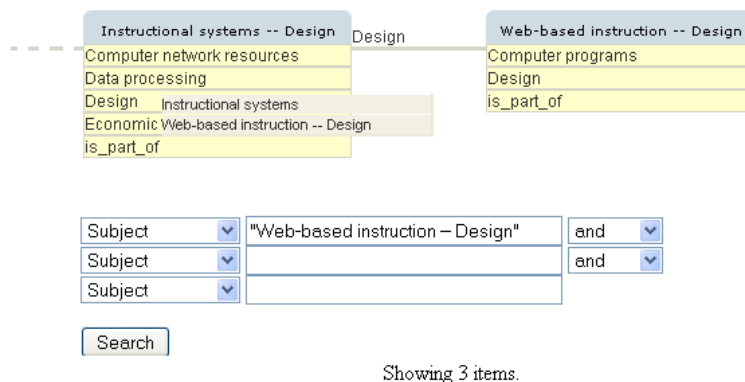
Showing 1 items.

Issue Date	Title	Author(s)
20-Φεβ-2007	Μεθοδολογίες σχεδίασης εκπαιδευτικών δραστηριοτήτων σε περιβάλλοντα ηλεκτρονικής μάθησης	Κοκκινός, Αντώνιος

Εικόνα 2.10: Προτεινόμενη προσέγγιση: τεκμήρια που περιγράφονται με το θέμα "Instructional systems - Design"

Μέχρι τώρα, οι δύο προσεγγίσεις εμφανίζουν την ίδια περίπου συμπεριφορά και επομένως ο χρήστης παίρνει τα ίδια αποτελέσματα. Ωστόσο, η προτεινόμενη προσέγγιση παρέχει περισσότερες αλληλεπιδράσεις, όπως θα παρουσιαστεί παρακάτω.

Πατώντας, λοιπόν, πάνω σε μια υποδιαίρεση (που αναπαρίσταται ως κίτρινη ετικέτα κάτω από τον τίτλο του πλαισίου), παρουσιάζονται στους χρήστες τα θέματα που μοιράζονται την ίδια υποδιαίρεση με το τρέχον θέμα (δηλ. "*Instructional systems - Design*"). Εάν ο χρήστης βρει ένα κατάλληλο θέμα, μπορεί να επιλέξει το συγκεκριμένο θέμα και να παρατηρήσει τα αντίστοιχα τεκμήρια. Στο περιγραφόμενο σενάριο, πατώντας πάνω στο θέμα "*Design*", παρουσιάζεται στο χρήστη το θέμα "*Web-based instruction - Design*", το οποίο ανακτά τρία τεκμήρια (βλέπε εικόνα 2.11).

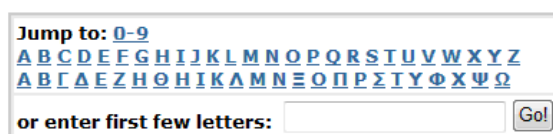


Issue Date	Title	Author(s)
20-Φεβ-2007	Μαθησιακές τεχνολογίες σε περιβάλλοντα υποστήριξης δικτυακών μαθησιακών κοινοτήτων	Καστραντάς, Κωνσταντίνος
16-Ιουλ-2008	Ανάλυση ικανοποίησης χρηστών πανεπιστημιακού portal	Τζικόπουλος, Αργύριος
29-Ιουλ-2008	Πρακτικές σχεδίασης συστημάτων ηλεκτρονικής μάθησης	Γεωργιάδη, Σταυρούλα Δ.

Εικόνα 2.11: Προτεινόμενη προσέγγιση: τεκμήρια που περιγράφονται με το θέμα "Web-based instruction - Design"

Σενάριο 2: ο χρήστης πρέπει να βρει τεκμήρια για "Security measures in Computer networks".

Σύμφωνα με την παραδοσιακή προσέγγιση, παρουσιάζεται στους χρήστες το θέμα "Computer networks - Security measures" (βλέπε εικόνα 2.12).



Showing subjects 162-182 of 2551.

Computer graphics
Computer interfaces
Computer network protocols
Computer networks
Computer networks-Access control
Computer networks-Management
Computer networks-Security measures
Computer programs
Computer security
Computer vision-Mathematics
Computer-aided software engineering
Computer-Assisted instruction
Computer-assisted instruction
Computer-assisted instruction-Computer programs
Computer-assisted instruction.
Computers and civilization
Computers and literacy

Εικόνα 2.12: Παραδοσιακή προσέγγιση: αναζήτηση του θέματος "Computer networks - Security measures"

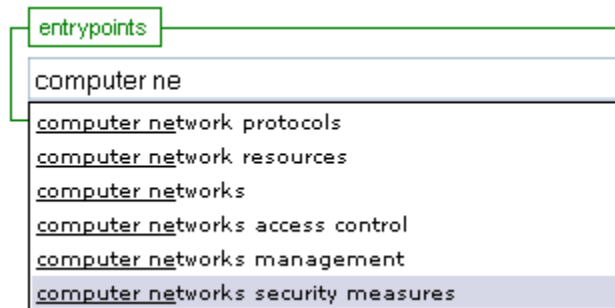
Πατώντας πάνω στο θέμα, υπάρχουν τρία τεκμήρια διαθέσιμα (βλέπε εικόνα 2.13).

Showing 3 items.

Issue Date	Title	Author(s)
21-Αύγ-2007	Αναλυτής αρχείων καταγραφής για τσίχη ηλεκτρονικής προστασίας στο λειτουργικό σύστημα Linux	Λαγός, Μιχαήλ
13-Ιούν-2008	Συστήματα ανίχνευσης παρεισφρήσεων με χρήση εξελικτικών νευρωνικών δικτύων	Μιχαηλίδης, Εμμανουήλ
26-Ιούν-2008	Μελέτη και ανάλυση των συστημάτων ανίχνευσης εισβολών - intrusion detection systems	Οικονόμου, Εοσθήμαος

Εικόνα 2.13: Παραδοσιακή προσέγγιση: τεκμήρια που περιγράφονται με το θέμα “Computer networks - Security measures”

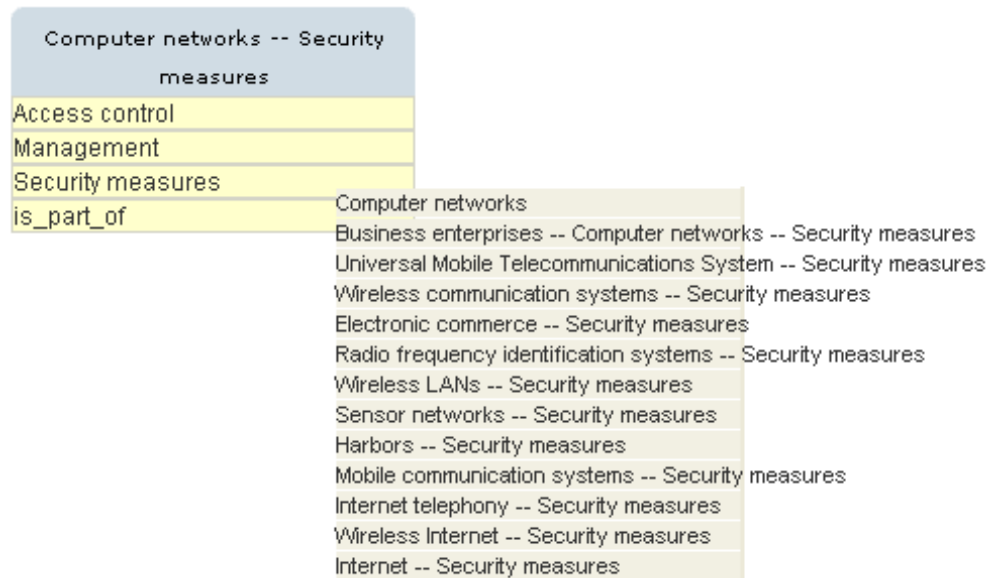
Από την άλλη μεριά, σύμφωνα με την προτεινόμενη προσέγγιση, ο χρήστης πληκτρολογεί τα πρώτα γράμματα του θέματος που έχει στο μυαλό του στο πεδίο αυτόματης συμπλήρωσης (βλέπε εικόνα 2.14).



Εικόνα 2.14: Προτεινόμενη προσέγγιση: αναζήτηση του θέματος “Computer networks - Security measures”, μέρος 1^ο

Ύστερα, ο χρήστης επιλέγει το θέμα “Computer networks - Security measures” και του παρουσιάζεται ένα πλαίσιο με το αντίστοιχο θέμα μαζί με τα τρία τεκμήρια, τα οποία είναι ίδια με την παραδοσιακή προσέγγιση.

Περαιτέρω αλληλεπίδραση με το σύστημα παρέχεται πατώντας πάνω στην υποδιαίρεση “Security measures” (βλέπε εικόνα 2.15).



Εικόνα 2.15: Προτεινόμενη προσέγγιση: αναζήτηση του θέματος “Computer networks - Security measures”, μέρος 2^ο

Έτσι, εάν ο χρήστης βρει ένα κατάλληλο θέμα, μπορεί να επιλέξει το συγκεκριμένο θέμα και να παρατηρήσει τα αντίστοιχα τεκμήρια.

Σενάριο 3: ο χρήστης πρέπει να βρει τεκμήρια για το γενικό θέμα “*Accounting*”.

Σύμφωνα με την παραδοσιακή προσέγγιση, στο χρήστη παρουσιάζονται τρία θέματα σχετικά με το θέμα “*Accounting*”, δηλαδή: “*Accounting - Data processing*”, “*Accounting - Moral and ethical aspects*” και “*Accounting - Standards*” (βλέπε εικόνα 2.16).

Academic achievement
Accelerated life testing
Acceptance sampling
Accountants-Professional ethics
Accounting-data processing
Accounting-Moral and ethical aspects
Accounting-standards
Accrual basis accounting
Activity-based costing
Adjustable rate mortgages
Adult education
Advertising
Advertising laws-European Economic Community countries
Aeronautics, Commercial-Freight-Safety measures
Agricultural biotechnology
Air conditioning-Equipment and supplies
Air quality management
Airplanes-Noise-Environmental aspects-Greece
Ajax (Web site development technology)
Allocation
Aluminum-Anodic oxidation

Εικόνα 2.16: Παραδοσιακή προσέγγιση: αναζήτηση του θέματος "Accounting"

Πατώντας πάνω στο κάθε θέμα, ο χρήστης θα καταλήξει σε ένα σύνολο τριών σχετικών τεκμηρίων.

Από την άλλη μεριά, σύμφωνα με την προτεινόμενη προσέγγιση, επιλέγοντας το θέμα "*Accounting*", ο χρήστης μπορεί να ανακαλύψει διάφορα θέματα που εξειδικεύουν (σύμφωνα με την αντίστοιχη συνδετική δομή των LCSH) τον αρχικό όρο "*Accounting*", πατώντας πάνω στη σχέση "*contains*" (βλέπε εικόνα 2.17).

Accounting	
Data processing	
Moral and ethical aspects	
Psychological aspects	
Standards	
contains	
is_part_of	Accounting -- Data processing
	Accounting -- Standards
	Accounting -- Moral and ethical aspects
	Auditing
	Cost accounting
	Cost analysis
	Disclosure in accounting
	Financial statements
	Managerial accounting
	Valuation

Εικόνα 2.17: Προτεινόμενη προσέγγιση: εξειδίκευση του θέματος "Accounting"

Έτσι, εάν ο χρήστης βρει ένα κατάλληλο θέμα, μπορεί να επιλέξει το συγκεκριμένο θέμα και να παρατηρήσει τα αντίστοιχα τεκμήρια.

Στη συνέχεια, παρουσιάζονται τα μη ελεγχόμενα λεξιλόγια, που είναι πολύ δημοφιλή ανάμεσα στους χρήστες του διαδικτύου και μπορούν να χρησιμοποιηθούν επίσης στη θεματική ευρετηρίαση.

2.6. Μη ελεγχόμενα λεξιλόγια

Στον αντίποδα των ελεγχόμενων λεξιλογίων έρχονται τα μη ελεγχόμενα λεξιλόγια (uncontrolled vocabularies), όπου οι χρήστες μπορούν να κάνουν αναζητήσεις με οποιοδήποτε λέξεις κλειδιά τους έρχονται στο μυαλό σχετικές με τις πληροφοριακές τους ανάγκες χωρίς να συμβουλευονται κάποιο από τα προαναφερθέντα ελεγχόμενα λεξιλόγια. Επομένως, τα μη ελεγχόμενα λεξιλόγια βασίζονται στη φυσική γλώσσα.

Τα συστήματα που δεν χρησιμοποιούν ελεγχόμενα λεξιλόγια μπορούν να χαρακτηριστούν: α) βάσει της ευρετηρίασης που παρέχουν μέσω λέξεων κλειδιών (δηλ. σημαντικές λέξεις που εξάγονται από τον τίτλο και τις περιλήψεις των πηγών) και β) βάσει του είδους της αναζήτησης που επιτρέπουν χωρίς τη δημιουργία οποιουδήποτε ευρετηρίου (δηλ. αναζήτηση ελεύθερου κειμένου, αναζήτηση ολόκληρου κειμένου κλπ.) (Svenonius, 2003, σ. 822).

Η υιοθέτηση μη ελεγχόμενων λεξιλογίων δίνει την ελευθερία στους χρήστες να αποθηκεύσουν οποιαδήποτε πληροφορία μπορεί να τους φανεί χρήσιμη στο μέλλον και να την οργανώσουν με όποιο τρόπο επιθυμούν. Παράλληλα, επιτρέπει στους χρήστες να μοιράζονται και να ανταλλάσσουν πληροφορίες εύκολα και ελεύθερα. Επιπλέον, είναι πιο ευέλικτα στη χρήση, πιο οικονομικά και λιγότερο χρονοβόρα σε σχέση με τα ελεγχόμενα λεξιλόγια (Sinclair & Cardew-Hall, 2008, σ. 16).

Από την άλλη μεριά, οι πολέμιοι των μη ελεγχόμενων λεξιλογίων υποστηρίζουν ότι τα μη ελεγχόμενα λεξιλόγια μπορούν να δημιουργήσουν πλεονασμό όρων, αφού επιτρέπουν τη χρήση ενικού και πληθυντικού, καθώς και ασυνέπεια στην ορθογραφία, στη σύνταξη των λέξεων, στη χρήση κεφαλαίων - μικρών κ.ά. Επίσης, δεν είναι αξιόπιστα αφού οι χρήστες μπορούν να αποθηκεύσουν οποιαδήποτε πληροφορία, η οποία μπορεί να μην έχει κανένα νόημα ή να είναι ανακριβής (Spiteri, 2007).

Η διαμάχη ανάμεσα στη χρήση ελεγχόμενων και μη λεξιλογίων δεν είναι καινούρια. Ήδη από τις αρχές του 20^{ου} αιώνα υπήρχαν πολλές συζητήσεις για το ποιες από τις δύο πρακτικές είναι καλύτερες. Μάλιστα, σύμφωνα με τη Svenonius (2003, σ. 836),

αυτή η διαμάχη μπορεί να χωριστεί σε 3 περιόδους. Η πρώτη διαμάχη ξεκίνησε στις αρχές του 20^{ου} αιώνα, όπου η ευρετηρίαση βάσει του τίτλου προτάθηκε από τον Samson Low και τον βοηθό του. Αυτοί πρότειναν ότι η πρόσβαση στα βιβλία με λέξεις από τον τίτλο θα βοηθούσε στο πρόβλημα της ακρίβειας που προκαλούνταν από τον ταξινομημένο κατάλογο. Στον αντίποδα αυτών, ήταν ο Cutter όπου υποστήριζε ότι ο τίτλος μπορεί να μην εκφράζει επακριβώς το θέμα του βιβλίου και με αυτόν τον τρόπο θα διαχωρίζονταν βιβλία που έχουν το ίδιο θέμα αλλά διαφορετικές λέξεις στον τίτλο. Η δεύτερη διαμάχη ξεκίνησε στα μέσα περίπου του 20^{ου} αιώνα, όταν έκαναν την εμφάνισή τους οι υπολογιστές, όπου έδιναν τη δυνατότητα ευρετηρίασης φυσικής γλώσσας. Τέλος, η τρίτη διαμάχη φτάνει περίπου ως τις μέρες μας, όπου η τεχνολογία των υπολογιστών εξελίχθηκε και δημιουργήθηκε το διαδίκτυο ως ένας νέος τρόπος επικοινωνίας και ανταλλαγής γνώσης και πληροφοριών. Μάλιστα, ιδιαίτερα στις μέρες μας, με την εμφάνιση των τεχνολογιών του Web 2.0, η χρήση μη ελεγχόμενων λεξιλογίων έχει εξαπλωθεί σε μεγάλο βαθμό και νέες έννοιες δημιουργήθηκαν, όπως το social tagging και οι λαοταξινομίες.

2.6.1. Λαοταξινομίες

Οι λέξεις κλειδιά – ετικέτες (tags) – που δημιουργούνται από τους χρήστες έχουν προταθεί ως ένας εύκολος τρόπος για την ενίσχυση της περιγραφής των online πληροφοριακών πηγών, βελτιώνοντας έτσι την πρόσβαση σε αυτές μέσω ευρύτερης ευρετηρίασης. Η έννοια social tagging αναφέρεται στην πρακτική της δημόσιας επισήμανσης ή κατηγοριοποίησης των πληροφοριακών πηγών σε ένα κοινόχρηστο online περιβάλλον. Η προκύπτουσα συνάθροιση ετικετών σχηματίζει μια λαοταξινομία: συνδυασμός των λέξεων "*folks*" (λαοί) και "*taxonomy*" (ταξινομία), όπου χρησιμοποιείται για την αναφορά σε μια άτυπη, οργανική συνάθροιση της σχετικής ορολογίας (Trant, 2008, σ. 1).

Υπάρχουν δύο είδη λαοταξινομιών: α) ευρείες λαοταξινομίες, όπου χαρακτηρίζονται από το εύρος του αριθμού των ετικετών που έχουν καθιερωθεί για να παραπέμπουν σε μια πηγή και από το γεγονός ότι οι χρήστες μπορούν να ανακατανεύμουν μια ετικέτα που έχει ήδη καταχωρηθεί σε μια πηγή και β) στενές λαοταξινομίες, όπου χαρακτηρίζονται από τον περιορισμό στον αριθμό των ετικετών που θα εκχωρηθούν σε μια πηγή. Αυτός ο περιορισμός ξεκινά από τον περιορισμό στον αριθμό των χρηστών που έχουν δικαίωμα να εκχωρήσουν μια ετικέτα και στον περιορισμό των ετικετών που μπορεί μια πηγή να δεχτεί (Auray, 2007, σ. 73).

2.7. Συμπεράσματα

Τα ελεγχόμενα και μη ελεγχόμενα λεξιλόγια είναι συστήματα για την οργάνωση των δεδομένων και των πληροφοριών και βοηθούν στην αποθήκευση και ανάκτηση αυτών. Τα ελεγχόμενα λεξιλόγια είναι κεντριοποιημένα συστήματα, στα οποία η πληροφορία συγκεντρώνεται και τοποθετείται σε μια ιεραρχική δομή. Για παράδειγμα, τα ταξινομικά σχήματα ακολουθούν μια ιεραρχική δομή στην οποία τοποθετούνται μια έννοια ή ένα σύνολο σχετικών εννοιών μαζί, ταξινομημένες από την ευρύτερη έννοια στις πιο ειδικές. Ο θησαυρός είναι ένα άλλο παράδειγμα ελεγχόμενου λεξιλογίου, όπου ο κύριος λόγος ύπαρξής του είναι να προσδιορίσει οποιοδήποτε όρο είναι διφορούμενος για τον χρήστη. Ο θησαυρός χωρίζει τις έννοιες σε δύο κύριες κατηγορίες: ευρύτερους όρους και στενότερους όρους. Το όφελος από τη χρήση αυτών των συστημάτων έγκειται στο γεγονός ότι μειώνουν το χρόνο αναζήτησης, αφού παρέχουν στους χρήστες ειδικούς όρους με την κατάλληλη ορθογραφία και διασφαλίζουν τη σταθερότητα των όρων και τη μοναδική σημασία τους. Από την άλλη όμως, υπάρχουν κάποια προβλήματα όσων αφορά στη χρήση αυτών των λεξιλογίων. Πιο συγκεκριμένα, τα ελεγχόμενα λεξιλόγια είναι ακριβά και χρονοβόρα, ιδιαίτερα εάν χρειάζεται να εκπαιδευτεί κανείς στη χρήση τους. Επιπλέον, τα συστήματα αυτά θα πρέπει να ενημερώνονται ανά τακτά χρονικά διαστήματα για να είναι σύγχρονα και να ελέγχονται συχνά για την ύπαρξη ανθρώπινων λαθών κατά τη δημιουργία τους.

Από την άλλη μεριά, τα μη ελεγχόμενα λεξιλόγια δεν είναι κεντριοποιημένα, όπως τα ελεγχόμενα λεξιλόγια και είναι πιο δημοφιλή στις μέρες μας, ιδιαίτερα στο διαδίκτυο. Οι λασοταξινομίες ή το social tagging επιτρέπουν στους χρήστες να βάζουν ετικέτες και να αποθηκεύουν οποιαδήποτε πληροφορία βρίσκουν χρήσιμη κάτω από οποιοδήποτε όρο που μπορεί να είναι σχετικός και να μοιράζονται αυτή την πληροφορία με χρήστες που έχουν τα ίδια ενδιαφέροντα. Δίνουν την ελευθερία και τη δυνατότητα στους χρήστες να αποθηκεύσουν και να ονομάσουν την πληροφορία όπως αυτοί νομίζουν καλύτερα, το οποίο μπορεί να είναι και μειονέκτημα παράλληλα. Οι χρήστες συνήθως κάνουν ορθογραφικά λάθη ή μπορεί οι όροι που χρησιμοποιούν να μην είναι ορθοί για να περιγράψουν μια πληροφορία. Αυτό θα μπορούσε να έχει ως αποτέλεσμα την παραπλάνηση των υπόλοιπων χρηστών. Οι οργανισμοί θα πρέπει να εστιάσουν καλά στα πλεονεκτήματα και τα μειονεκτήματα και των δύο ειδών λεξιλογίων, να λάβουν υπ' όψιν τους τα κόστη και τα οφέλη που μπορούν να επιτευχθούν όταν υιοθετείται το ένα ή το άλλο είδος, σύμφωνα με τις ανάγκες τους, την κουλτούρα τους και του όγκου της πληροφορίας που διαχειρίζονται.

Ταυτόχρονα, η θεματική ευρετηρίαση και κατ' επέκταση η θεματική ανάκτηση πληροφοριών στο πλαίσιο των βιβλιοθηκών απαιτεί οι υποκείμενες πληροφοριακές πηγές να έχουν έναν ικανοποιητικό αριθμό θεματικών περιγραφών επαρκούς ποιότητας. Η ποσότητα μαζί με την ποιότητα της θεματικής ευρετηρίασης μπορεί να ασκήσει τεράστια επιρροή στην ποιότητα και τη δυναμική των αποτελεσμάτων στο σημείο της πραγματικής αναζήτησης. Παράλληλα, η δημιουργία υπηρεσιών που εκμεταλλεύονται πλήρως τη σημασιολογία των ελεγχόμενων λεξιλογίων μπορεί να ενισχύσει σημαντικά την αναζήτηση και την ανάκτηση πληροφοριών. Μέσα σε αυτό το πλαίσιο, παρουσιάστηκε μια προσπάθεια για την εκμετάλλευση της άμεσης και έμμεσης σημασιολογικής εκφραστικότητας των LCSH, προς όφελος της δημιουργίας πιο αποδοτικών υπηρεσιών θεματικής ανάκτησης πληροφοριών στις βιβλιοθήκες. Με μικρές τροποποιήσεις στο σχεδιασμό της υποκείμενης οντολογίας, η προτεινόμενη μεθοδολογία θα μπορούσε να εφαρμοστεί ουσιαστικά σε οποιοδήποτε είδος θησαυρού που χρησιμοποιείται για τη μοντελοποίηση των θεμάτων των πληροφοριακών πηγών στις βιβλιοθήκες, παρέχοντας με αυτό τον τρόπο ένα γενικό εργαλείο για τη διαδραστική, θεματική ανάκτηση πληροφοριών με επίκεντρο τον χρήστη.

Για την επιβεβαίωση της σύλληψης της ιδέας της προτεινόμενης προσέγγισης, ένα πρωτότυπο σύστημα αναπτύχθηκε ικανό να παρέχει στους τελικούς χρήστες θεματική λειτουργικότητα κατά την ανάκτηση πληροφοριών μιας ψηφιακής βιβλιοθήκης που χρησιμοποιεί τις LCSH. Βασιζόμενοι στις πιο σύγχρονες τάσεις της ανάπτυξης του διαδικτύου, το σύστημα προωθεί συγκυριακή ή περιστασιακή ανακάλυψη «*ενδιαφέρουσας*» πληροφορίας συνδεδεμένη με τη διαδικασία αλληλεπιδραστικής πλοήγησης. Κατά τη διάρκεια της διαδικασίας, οι χρήστες υποστηρίζουν τη γνωστική τους μάθηση, αφού μπορούν να ανακαλύψουν ποια θέματα αντιστοιχούν στις πληροφοριακές τους ανάγκες.

Παράλληλα, όπως έγινε αντιληπτό από τις προηγούμενες ενότητες, η ύπαρξη μεγάλου όγκου συστημάτων (ελεγχόμενων και μη ελεγχόμενων λεξιλογίων) που έχουν δημιουργηθεί για την εξυπηρέτηση διαφορετικών κοινοτήτων καθώς και το γεγονός ότι αυτά τα λεξιλόγια χρησιμοποιούνται από διάφορους οργανισμούς για την κάλυψη συγκεκριμένων αναγκών έχει οδηγήσει αναπόφευκτα στη συγχώνευση

(merging³⁶), την αντιστοίχιση (mapping³⁷) και την ενσωμάτωση (integrating³⁸) αυτών των λεξιλογίων για την αποδοτικότερη και πληρέστερη ανάκτηση πληροφοριών ανάμεσα σε διαφορετικούς οργανισμούς ή ακόμα και στους κόλπους του ίδιου οργανισμού. Στο επόμενο κεφάλαιο θα παρουσιαστούν οι διάφορες προσπάθειες που έχουν γίνει προς αυτή την κατεύθυνση.

³⁶ “*Merging*” είναι η διαδικασία με την οποία τα αλληλεπικαλυπτόμενα μέρη των ελεγχόμενων λεξιλογίων συνδυάζονται, απομακρύνονται τα διπλότυπα, αλλά αυξάνονται οι μη προτιμώμενοι όροι (Hedden, H. (2010). Combining Controlled Vocabularies).

³⁷ “*Mapping*” είναι η διαδικασία με την οποία τα αλληλεπικαλυπτόμενα μέρη των ελεγχόμενων λεξιλογίων παραμένουν διαχωρισμένα, το ένα χρησιμοποιείται στη θέση του άλλου σε μια συγκεκριμένη εφαρμογή (π.χ. ευρετηρίαση έναντι ανάκτησης ελεγχόμενων λεξιλογίων) (Hedden, H. (2010). Combining Controlled Vocabularies).

³⁸ “*Integrating*” είναι η διαδικασία με την οποία τα μη αλληλεπικαλυπτόμενα μέρη των ελεγχόμενων λεξιλογίων συνδυάζονται και συμπληρώνουν το ένα το άλλο για τη δημιουργία ενός μεγαλύτερου ελεγχόμενου λεξιλογίου (Hedden, H. (2010). Combining Controlled Vocabularies).

3. Συγχώνευση ελεγχόμενων και μη λεξιλογίων

3.1. Αιτίες που οδηγούν στη συγχώνευση λεξιλογίων

Το γεγονός ότι ένα μόνο ελεγχόμενο λεξιλόγιο δεν είναι ικανό, τις περισσότερες φορές, να καλύψει όλη την ορολογία που υπάρχει σε ένα συγκεκριμένο τομέα έχει οδηγήσει στην ανάγκη για τη συγχώνευση ή και ακόμα το συνδυασμό ενός ή περισσότερων λεξιλογίων. Πιο συγκεκριμένα, οι βιβλιοθήκες χρησιμοποιούν διάφορα ελεγχόμενα λεξιλόγια για να ευρετηριάσουν διαφορετικές συλλογές που υπάρχουν είτε μέσα στην ίδια βιβλιοθήκη είτε σε διαφορετικές βιβλιοθήκες. Τα ελεγχόμενα λεξιλόγια επιλέγονται βάσει της γλώσσας, τη θεματική εστίαση, το βαθμό εμπάθουσας των θεμάτων, τη διαθεσιμότητα κ.ά. Για παράδειγμα, μια ακαδημαϊκή βιβλιοθήκη μπορεί να χρησιμοποιεί διαφορετικά ευρετήρια για να περιγράψει τα τεκμήρια στην κύρια συλλογή της, στις ειδικές συλλογές της, στα ψηφιακά της αποθετήρια κλπ. Αυτό έχει ως αποτέλεσμα να μην υπάρχει διαλειτουργικότητα ακόμα και ανάμεσα σε όρους και έννοιες που χρησιμοποιούνται στους κόλπους μιας βιβλιοθήκης³⁹. Με αυτό τον τρόπο οι χρήστες δεν μπορούν να βρουν όλη την πληροφορία που τους παρέχει η βιβλιοθήκη τους και έτσι δημιουργούνται προβλήματα κατά την αναζήτηση και ανάκτηση πληροφοριών. Επομένως, οι βιβλιοθήκες θα πρέπει να βρουν τρόπους για να συνδυάσουν αυτά τα ελεγχόμενα λεξιλόγια, βελτιώνοντας έτσι όχι μόνο τη διαλειτουργικότητα ανάμεσα στις υπηρεσίες μιας συγκεκριμένης βιβλιοθήκης, αλλά και την επικοινωνία ανάμεσα σε διαφορετικές βιβλιοθήκες και άλλους οργανισμούς.

Η διαλειτουργικότητα εν γένει ανάμεσα σε δύο ή περισσότερα ελεγχόμενα λεξιλόγια έχει απασχολήσει τους ερευνητές περίπου από τη δεκαετία του 1960 (Harpring, 2010, σ. 83). Μάλιστα εκείνη την περίοδο δυο διαφορετικοί όροι επινοήθηκαν, ο πρώτος για τα λεξιλόγια και ο δεύτερος για τα συστήματα: α) "*convertibility*", όπου ορίζεται ως η ικανότητα να μεταφερθεί κανείς από το ένα λεξιλόγιο ευρετηρίασης στο άλλο και β) "*compatibility*", όπου ορίζεται ως η ικανότητα ενός πληροφοριακού συστήματος να αποδεχτεί την πρωτότυπη ευρετηρίαση των δεδομένων από ένα άλλο πληροφοριακό σύστημα για οποιαδήποτε θεματική περιοχή που καλύπτεται και από τα

³⁹ W3.org, (2015). *Use Case Vocabulary Merging - Library Linked Data*. [online] Διαθέσιμο στο: https://www.w3.org/2005/Incubator/llid/wiki/Use_Case_Vocabulary_Merging [Πρόσβαση: 20/07/2015].

δύο συστήματα (Svenonius, 2003, σ. 834). Σύμφωνα με τον Doerr (2001), η συγχώνευση λεξιλογίων μπορεί να οριστεί ως «η διαδικασία αναγνώρισης όρων, εννοιών και ιεραρχικών σχέσεων που είναι περίπου ισοδύναμοι».

3.2. Προβλήματα που προκύπτουν κατά τη συγχώνευση ελεγχόμενων λεξιλογίων

Οι Lancaster & Smith (1983, σ. 74-77) δημοσίευσαν μια επισκόπηση των θεμάτων που προκύπτουν κατά την ενσωμάτωση διαφορετικών ειδών λεξιλογίων. Περιέγραψαν διάφορους παράγοντες που επηρεάζουν πόσο επιτυχημένα ένα λεξιλόγιο μπορεί να συσχετιστεί με ένα άλλο. Ένας παράγοντας είναι η έκταση της επικάλυψης της έννοιας που περιγράφεται από κάθε όρο του ελεγχόμενου λεξιλογίου. Ένας άλλος παράγοντας είναι το επίπεδο της εξειδίκευσης των όρων μέσα στο ίδιο λεξιλόγιο (δηλ. ένα ελεγχόμενο λεξιλόγιο μπορεί να έχει πιο εξειδικευμένους όρους σε ένα τομέα). Επιπλέον, τα διαφορετικά επίπεδα του προ- και υστερο- συνδυασμένου συντονισμού που υπάρχει ανάμεσα σε διαφορετικά λεξιλόγια και ο τρόπος με τον οποίο τοποθετούν τους κοινούς όρους μέσα στην υποκείμενη δομή τους μπορεί να επηρεάσει την ενοποίηση. Στην ίδια γραμμή σκέψεων, οι Doerr (2001), Olson & Strawn (1997) ανακάλυψαν διαφορές στη χρήση των λέξεων (π.χ. κοινών έναντι επιστημονικών ονομάτων) και διαφορές στο νόημα που προκύπτει από τη διαφορά στην ταξινόμηση των όρων.

Στη συνέχεια, θα παρουσιαστούν κάποια παραδείγματα εργασιών που έχουν γίνει για τη συγχώνευση θησαυρών.

3.3. Συγχώνευση θησαυρών

Έχοντας τα παραπάνω κατά νου, πολλές προσπάθειες πρόσφατα έχουν γίνει που στοχεύουν στην ενσωμάτωση διαφορετικών θησαυρών (Francesconi et al., 2008; Lauser et al., 2008; Liang et al., 2006; Tordai et al., 2007). Αυτές οι προσπάθειες βασίζονται πάνω σε διάφορες μεθοδολογίες ταιριάσματος προτύπων με σκοπό να αναπτύξουν οντολογίες βασισμένες στο SKOS⁴⁰. Οι οντολογίες χρησιμοποιούνται για να αντιστοιχίσουν και να ευθυγραμμίσουν περιγραφείς από διαφορετικούς θησαυρούς στο διαδίκτυο.

⁴⁰ W3.org, (2015). *SKOS Simple Knowledge Organization System - home page*. [online] Διαθέσιμο στο: <http://www.w3.org/2004/02/skos/> [Πρόσβαση: 06/07/2015].

Πολλές ερευνητικές προσπάθειες έχουν σαν στόχο ειδικά την ενσωμάτωση των LCSH με τους θησαυρούς, τα οποία θεωρούνται πιο συνεκτικά είδη ελεγχόμενων λεξιλογίων (Chaplan, 1995; Giunchiglia et al., 2009; Olson & Strawn, 1997; Vizine-Goetz et al., 2005; Whitehead, 1990). Η επόμενη ενότητα παρουσιάζει κάποιες ενδεικτικές προσπάθειες που αφορούν στην αντιστοίχιση των LCSH με θησαυρούς.

3.4. Συγχώνευση θεματικών επικεφαλίδων με θησαυρούς

Στο άρθρο των Olson & Strawn (1997), γίνεται μια προσπάθεια να αντιστοιχηθούν οι περιγραφείς του θησαυρού ERIC⁴¹ στις LCSH. Πιο συγκεκριμένα, τα λεξιλόγια υπό αντιστοίχιση πρώτα μετατρέπονται στο πρότυπο καθιερώσεων του MARC21 (MARC21 Authority Format⁴²). Τα πολλαπλά στιγμιότυπα των ευρύτερων όρων, στενότερων όρων και σχετικών όρων που αποθηκεύονται σε ένα πεδίο του ERIC κωδικοποιούνται ως ξεχωριστά πεδία στο πρότυπο του MARC21. Οι σχέσεις σχετικών όρων καταλήγουν στην παραγωγή διαφόρων πεδίων των αντίστοιχων εγγραφών στο MARC21. Το πεδίο με την ετικέτα χρησιμοποίησε για (Used For – UF) μετατρέπεται παρόμοια σε πολλαπλά πεδία στο MARC21. Οι όροι που βρίσκονται μέσα σε παρενθέσεις στο θησαυρό ERIC υποδεικνύουν έναν παλαιότερα έγκυρο όρο, ο οποίος έχει αντικατασταθεί από κάποιο άλλο όρο. Με την κωδικοποίηση των λεξιλογίων πηγή (source) και στόχος (target) σε ένα κοινό πρότυπο (δηλ. MARC21 Authorities Format), η αναπαράσταση παρόμοιας πληροφορίας ομογενοποιείται και η ικανότητα της συσχέτισης των λεξιλογίων βελτιώνεται.

Μια προσπάθεια για να αντιστοιχηθεί ο AAT στις LCSH παρουσιάζεται από τον Whitehead (1990). Όπου δεν υπάρχει ακριβές ταίριασμα όρων ανάμεσα στον AAT και τις LCSH, ο πιο κοντινός όρος των LCSH καταγράφεται. Η συνδυαστική δομή (δηλ. BT, NT και RT) των LCSH δεν αντιστοιχείται στον AAT, αφού ο AAT παρέχει μια εξαντλητική ταξινόμηση γένους - είδους για τα πεδία που καλύπτει.

Μια παρόμοια προσέγγιση για την αντιστοίχιση του Laborline Thesaurus στις LCSH παρουσιάζεται από τον Chaplan (1995). Αρχικά, προσδιορίζονται οι ακριβείς αντιστοι-

⁴¹ Eric.ed.gov, (2015). *ERIC - Education Resources Information Center*. [online] Διαθέσιμο στο: <http://eric.ed.gov/> [Πρόσβαση: 06/07/2015].

⁴² Loc.gov, (2015). *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/specifications/> [Πρόσβαση: 06/07/2015].

χίες ανάμεσα στα δύο λεξιλόγια. Αυτές ακολουθούνται από τις παραπομπές ή τα εναλλακτικά θέματα και τέλος, υποδεικνύονται οι ευρύτερες, στενότερες και σχετικές σχέσεις για κάθε αντιστοίχιση.

Μια πιο πρόσφατη προσέγγιση παρουσιάζεται στο (Giunchiglia et al., 2009). Πιο συγκεκριμένα, οι συγγραφείς προσπαθούν να αντιστοιχίσουν τις LCSH στον National Agricultural Library Thesaurus – NALT⁴³ χρησιμοποιώντας τη τεχνολογία KOS. Η προσέγγισή τους βασίζεται σε έννοιες· χρησιμοποιούν προηγμένες γλωσσικές τεχνικές για να εκφράσουν αυτόματα όρους από ένα KOS ως προτασιακούς τύπους Περιγραφικής Λογικής και κατόπιν χρησιμοποιούν αυτούς τους τύπους για να υπολογίσουν την ελάχιστη αντιστοίχιση ανάμεσα στους κόμβους του KOS. Αντί να αξιολογηθεί η αντιστοίχιση των όρων σε σχέση με την ακρίβεια και την ανάκτηση, ο κύριος στόχος της εργασίας είναι να μάθουν από τα λάθη και να καταλάβουν πως είναι δυνατόν να βελτιώσουν σταδιακά τη διαδικασία της αντιστοίχισης.

Στη συνέχεια, θα παρουσιαστούν οι προσπάθειες που έχουν γίνει για την αντιστοίχιση ταξινομικών σχημάτων με άλλα ελεγχόμενα λεξιλόγια.

3.5. Συγχώνευση ταξινομικών σχημάτων με θεματικές επικεφαλίδες

Το 1994, το προσωπικό του Online Computer Library Center, Inc. – OCLC⁴⁴ ξεκίνησε την απόπειρα σύνδεσης των LCSH με το ταξινομικό σχήμα DDC. Τα ζεύγη DDC/LCSH δημιουργήθηκαν από τις εγγραφές του OCLC WorldCat⁴⁵, οι οποίες περιείχαν αριθμούς DDC και LCSH ταυτόχρονα (Vizine-Goetz et al., 2004).

Επίσης, η Βιβλιοθήκη του Κογκρέσου έχει δημιουργήσει το Classification Plus⁴⁶, το οποίο είναι ένα ηλεκτρονικό εργαλείο που συνδυάζει το ταξινομικό σχήμα LCC με τις LCSH.

⁴³ Agclass.nal.usda.gov, (2015). *Agricultural Thesaurus and Glossary Home Page*. [online] Διαθέσιμο στο: <http://agclass.nal.usda.gov/> [Πρόσβαση: 06/07/2015].

⁴⁴ Oclc.org, (2015). *OCLC: Worldwide, Member-driven Library Cooperative*. [online] Διαθέσιμο στο: <https://www.oclc.org/en-europe/home.html> [Πρόσβαση: 06/07/2015].

⁴⁵ Oclc.org, (2015). *WorldCat*. [online] Διαθέσιμο στο: <https://www.oclc.org/worldcat.en.html> [Πρόσβαση: 06/07/2015].

⁴⁶ Loc.gov, (2015). *Classification Plus (Ιούνιος 24 – Ιούλιος 8, 1996) - Library of Congress Information Bulletin*. [online] Διαθέσιμο στο: <http://www.loc.gov/loc/lcib/9612/cplus.html> [Πρόσβαση: 06/07/2015].

Παράλληλα, πολλές προσπάθειες έχουν γίνει για την ταύτιση του UDC με άλλα εργαλεία θεματικών επικεφαλίδων όπως είναι οι General Finnish Subject Headings, Subject Headings of the National Library of Madrid κ.ά. (San Segundo, 2009, σ. 289).

3.6. Συγχώνευση λαοταξινομιών με ελεγχόμενα λεξιλόγια

Όσον αφορά στη συγχώνευση ελεγχόμενων λεξιλογίων με λαοταξινομίες, δεν υπάρχει κάποια έρευνα που να αντιμετωπίζει αυτό το θέμα. Αντίθετα, οι περισσότερες έρευνες εστιάζουν στο κατά πόσο οι λαοταξινομίες μοιάζουν με τα ελεγχόμενα λεξιλόγια και αν θα μπορούσε έτσι να ενισχυθεί η αναζήτηση και η ανάκτηση πληροφοριών.

Πιο συγκεκριμένα, οι Yi & Chan (2009) προσπάθησαν να αντιστοιχίσουν τις ετικέτες που δημιουργούνται στο Delicious⁴⁷ λεκτικά με τις LCSH. Οι LCSH μεταμορφώθηκαν σε ένα δέντρο και έγινε μια λεπτομερής εξέταση για να ελεγχθούν τα χαρακτηριστικά των ετικετών, η αλληλοεπικάλυψή τους με τις LCSH και η κατανομή των ετικετών μέσα στο δέντρο των LCSH. Τα αποτελέσματα της έρευνας αποκάλυψαν ότι περίπου τα 2/3 των ετικετών είχαν πιθανά ταιριάσματα στις LCSH, επομένως οι ετικέτες θα μπορούσαν να βοηθήσουν στην αναζήτηση και ανάκτηση πληροφοριών.

Στο ίδιο μήκος κύματος είναι και η έρευνα που διεξήχθη από τους Thomas, Caudle & Schmitz (2009), όπου κατέληξαν στο συμπέρασμα ότι οι ετικέτες μπορούν να ενισχύσουν τις θεματικές επικεφαλίδες και αντίστοιχα να παρέχουν επιπρόσθετη πρόσβαση στις πηγές.

Επίσης, ο White (2013) διεξήγαγε μια έρευνα με την οποία προσπάθησε να ανακαλύψει τις σχέσεις που υπάρχουν ανάμεσα στις λέξεις κλειδιά ελεύθερου κειμένου με τους όρους διαφόρων ελεγχόμενων λεξιλογίων που χρησιμοποιούνται στις επιστήμες. Κατέληξε στο συμπέρασμα ότι οι LCSH έχουν τις περισσότερες αντιστοιχίες σε σχέση με οποιοδήποτε άλλο ελεγχόμενο λεξιλόγιο και θα πρέπει τα ελεγχόμενα και μη λεξιλόγια να χρησιμοποιούνται συμπληρωματικά.

Τέλος, σε αντίθετο τελείως συμπέρασμα κατέληξε η έρευνα που πραγματοποιήθηκε από τους Lee & Schleyer (2010). Οι συγγραφείς συνέκριναν τις ετικέτες και τους ευ-

⁴⁷ Delicious.com, (2015). *Delicious*. [online] Διαθέσιμο στο: <https://delicious.com/> [Πρόσβαση: 20/07/2015].

ρητηριασμένους όρους που περιγράφουν ένα συγκεκριμένο αριθμό εργασιών που υπάρχουν στο CiteULike⁴⁸ και στη MEDLINE⁴⁹ αντίστοιχα και αξιολόγησαν τις διαφορές τους. Τα αποτελέσματα τους δείχνουν ότι οι ετικέτες και οι ιατρικές θεματικές επικεφαλίδες (Medical Subject Headings – MeSH⁵⁰) έχουν πολύ μικρή αλληλοεπικάλυψη και υπάρχει μεγάλη ετερογένεια στο πως συλλαμβάνουν οι δύο υπηρεσίες τις έννοιες των υποκείμενων πηγών.

Στη συνέχεια, θα παρουσιαστεί αναλυτικά μια μεθοδολογία που προτάθηκε από τους Paradakis & Kyrgianos (2011). Η συγκεκριμένη μεθοδολογία αναφέρεται στη συγχώνευση ενός θησαυρού με τις LCSH, αλλά θα μπορούσε επίσης να εφαρμοστεί σε άλλες συλλογές πέρα από τον τομέα των βιβλιοθηκών, με την προϋπόθεση ότι αυτές οι συλλογές χρησιμοποιούν ελεγχόμενα λεξιλόγια που διαθέτουν κάποιο είδος συνδετικής δομής.

3.7. Παρουσίαση μεθοδολογίας για τη συγχώνευση θησαυρών με τις LCSH

Όπως έχει ήδη γίνει σαφές, η καθιέρωση των LCSH ως το πιο ευρέως χρησιμοποιούμενο ελεγχόμενο λεξιλόγιο για τη θεματική ευρετηρίαση στον τομέα των βιβλιοθηκών ανά τον κόσμο (Williamson, 1996, σ. 159) είναι πέρα από κάθε αμφιβολία. Ωστόσο, φαίνεται ότι οι LCSH δυσκολεύονται στο να προσαρμοστούν στις συγκεκριμένες απαιτήσεις που επιβάλλονται από το διαδίκτυο. Αυτή η ασυμβατότητα προέρχεται από το γεγονός ότι οι LCSH αναδύθηκαν σε μια εποχή που η πληροφορία δεν ήταν ψηφιακή. Επομένως, οι LCSH δεν λαμβάνουν υπ' όψιν τους πολλές πτυχές της ψηφιακής πληροφορίας και του τρόπου με τον οποίο θα πρέπει η εν λόγω πληροφορία να διαδίδεται μέσω του διαδικτύου.

Σε μια προσπάθεια να βρεθεί λύση σε αυτό το θέμα, πολλές φορές οι όροι των LCSH αντικαθίστανται από περιγραφείς των θησαυρών. Το γεγονός ότι οι θησαυροί εξελίχθηκαν σχετικά πρόσφατα (Markham & Avery, 1998, σ. 350) και κυρίως μέσα στα πλαίσια του διαδικτύου, κατέληξε στην ανάγκη ενός εργαλείου θεματικής ευρετηρίασης ικανό να προσαρμοστεί επιτυχώς στο ψηφιακό περιβάλλον και στα αντίστοιχα

⁴⁸ Gallagher, F. (2015). *CiteULike: Everyone's Library*. [online] Διαθέσιμο στο: <http://www.citeulike.org/> [Πρόσβαση: 06/07/2015].

⁴⁹ Ncbi.nlm.nih.gov, (2015). *PubMed - NCBI*. [online] Διαθέσιμο στο: <http://www.ncbi.nlm.nih.gov/pubmed/> [Πρόσβαση: 06/07/2015].

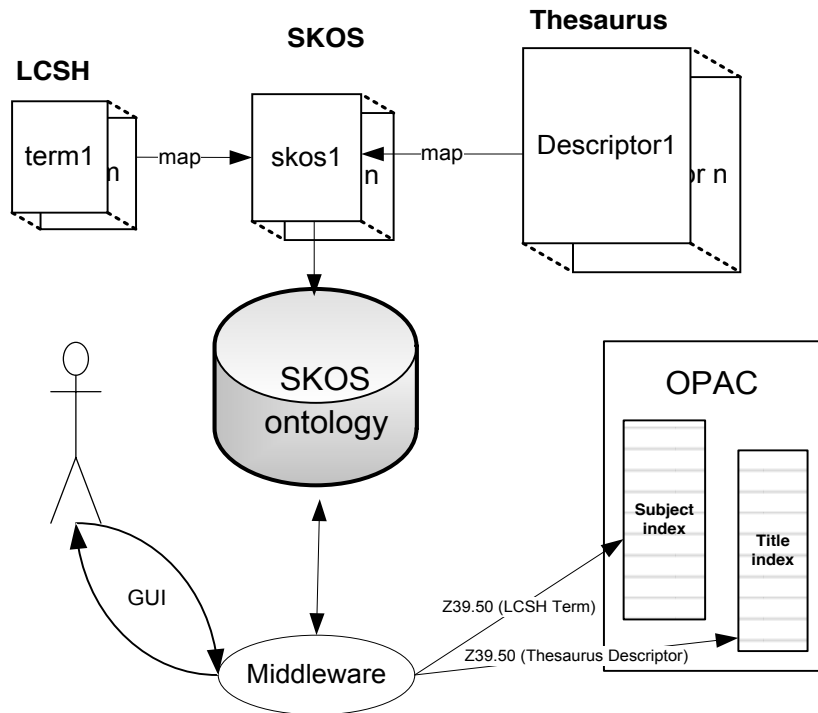
⁵⁰ Nlm.nih.gov, (2015). *Medical Subject Headings*. [online] Διαθέσιμο στο: <https://www.nlm.nih.gov/mesh/> [Πρόσβαση: 06/07/2015].

συστήματα ανάκτησης πληροφοριών. Παρ' όλα τα εγγενή πλεονεκτήματα των θησαυρών σε σχέση με τις LCSH, όσον αφορά στην ομαλή τους υιοθέτηση από τα σύγχρονα συστήματα ανάκτησης πληροφοριών, η αντικατάσταση των LCSH με τους θησαυρούς δεν είναι ακόμα διαθέσιμη. Πράγματι, το γεγονός ότι η πλειονότητα των υπάρχοντων θεματικών ευρετηρίων που βασίζεται στις LCSH δεν μπορεί να αγνοηθεί. Η αντικατάσταση των όρων των LCSH με περιγραφείς των θησαυρών στο πλαίσιο μιας διαδικασίας επανευρετηρίασης είναι πολύ ακριβή όσον αφορά σε χρόνο και χρήματα.

Κατά τη διάρκεια των τελευταίων ετών, διάφορες προσπάθειες έγιναν στο πλαίσιο εντοπισμού τρόπων προσαρμογής των ελεγχόμενων λεξιλογίων στο διαδίκτυο. Προς αυτή την κατεύθυνση, οι οντολογίες αναδύθηκαν ως μια επίσημη αναπαράσταση της γνώσης μέσω ενός συνόλου εννοιών σε ένα τομέα καθώς και οι σχέσεις ανάμεσα σε αυτές τις έννοιες. Ωστόσο, οι διάφορες γλώσσες που χρησιμοποιούνται για να εκφράσουν τις οντολογίες σε μορφή κατανοητή για τις μηχανές (π.χ. DAML+OIL⁵¹, OWL⁵²) δεν είναι πολύ δημοφιλείς ανάμεσα στα μέλη της βιβλιοθηκονομικής κοινότητας εξ' αιτίας της εγγενούς πολυπλοκότητάς τους. Ως εκ τούτου, οι βιβλιοθηκονόμοι υπέδειξαν το SKOS ως ένα απλό μοντέλο οντολογίας που είναι σε θέση να ανταποκριθεί στις συγκεκριμένες απαιτήσεις. Το SKOS είναι μια οικογένεια από επίσημες γλώσσες οντολογιών που έχει σχεδιαστεί για την αναπαράσταση διαφόρων ειδών ελεγχόμενων λεξιλογίων που έχουν μοντελοποιηθεί ως σχήματα εννοιών (Miles et al., 2005, σ. 3), συμπεριλαμβανομένων των θησαυρών, των σχημάτων ταξινόμησης, των λιστών θεματικών επικεφαλίδων, των ταξινομιών, των ορολογιών (terminologies), των γλωσσάριων (glossaries) και άλλων ειδών (αναλυτική περιγραφή του SKOS γίνεται στην ενότητα 4.4.4).

⁵¹ Daml.org, (2015). *The DARPA Agent Markup Language Homepage*. [online] Διαθέσιμο στο: <http://www.daml.org/> [Πρόσβαση: 07/07/2015].

⁵² W3.org, (2015). *OWL Web Ontology Language Reference*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/owl-ref/> [Πρόσβαση: 07/07/2015].



Εικόνα 3.1: Συνολική αρχιτεκτονική του συστήματος

Όπως φαίνεται στην εικόνα 3.1, παρουσιάζεται ένα σύστημα ανάκτησης πληροφοριών βασισμένο στα θέματα, το οποίο αποτελείται από όρους των LCSH ή/και περιγραφείς θησαυρών. Αυτά τα θέματα συνιστούν μια οντολογία, η οποία βασίζεται στις προδιαγραφές του SKOS. Σύμφωνα με την προτεινόμενη προσέγγιση, οι περιγραφείς του θησαυρού συσχετίζονται με τους υπάρχοντες όρους των LCSH βασιζόμενοι σε μια μεθοδολογία αντιστοίχισης που θα αναλυθεί αργότερα.

Η προκύπτουσα οντολογία αποθηκεύεται σε ένα κατάλληλο σύστημα βάσης δεδομένων. Οι τελικοί χρήστες αλληλεπιδρούν με την υποκείμενη οντολογία μέσω ενός αντίστοιχα σχεδιασμένου γραφικού περιβάλλοντος διεπαφής χρήστη. Εάν η αλληλεπίδραση καταλήξει σε ένα όρο των LCSH, τότε το ενδιάμεσο λογισμικό (middleware) αποστέλλει τον όρο στο θεματικό ευρετήριο του OPAC, μέσω της χρήσης του πρωτοκόλλου Z39.50⁵³. Με αυτή τη διαδικασία οδηγούμαστε σε μια λίστα αποτελεσμάτων που περιέχει πηγές που έχουν ευρετηριαστεί με αυτό τον όρο από εκπαιδευμένους βιβλιοθηκονόμους. Εάν η αλληλεπίδραση καταλήξει σε ένα περιγραφέα θησαυρού, τότε το σύστημα απαντά όπως περιγράφεται παρακάτω:

⁵³ Niso.org, (2015). *ANSI/NISO Z39.50-2003, Information Retrieval: Application Service Definition & Protocol Specification - National Information Standards Organization*. [online] Διαθέσιμο στο: http://www.niso.org/apps/group_public/document.php?document_id=6522 [Πρόσβαση: 07/07/2015].

1. Ο περιγραφέας συγκρίνεται με τις καταχωρήσεις του υποκείμενου ευρετηρίου τίτλων (title index) (βλέπε εικόνα 3.1) και μια αντίστοιχη λίστα αποτελεσμάτων αναζήτησης αποστέλλεται στον ερευνητή. Η επιλογή να συγκρίνουμε το ερώτημα με τον ευρετήριο τίτλων αποφασίστηκε λόγω του γεγονότος ότι ο τίτλος μιας πηγής θεωρείται ως μια ένδειξη του θέματός της (Svenonius, 1986, σ. 332).
2. Εάν ο περιγραφέας συσχετίζεται με έναν όρο από τις LCSH σύμφωνα με τη μεθοδολογία αντιστοίχισης (βλέπε ενότητα 3.7.1), τότε το γραφικό περιβάλλον του συστήματος παρέχει τη δυνατότητα στον ερευνητή να πλοηγηθεί σε έναν όρο των LCSH και επομένως να απευθύνει ένα ερώτημα στο υποκείμενο ευρετήριο θεμάτων (subject index).

Εν κατακλείδι, καταλήγοντας σε έναν όρο των LCSH ο χρήστης οδηγείται σε μια λίστα που περιέχει πηγές για αυτούς τους όρους, όπως αποφασίστηκε από κατάλληλα εκπαιδευμένο προσωπικό. Αυτό δεν συμβαίνει όταν καταλήγουμε σε έναν περιγραφέα θησαυρών. Στο σενάριο αυτό, η αντίστοιχη λίστα αποτελεσμάτων αναζήτησης περιέχει πηγές, οι οποίες μπορούν πιθανώς να ταιριάξουν τη συμβολοσειρά του περιγραφέα στον τίτλο τους. Επομένως, η σχετικότητα μπορεί να είναι πιθανώς πιο μεγάλη όταν το ερώτημα αναζήτησης αποτελείται από έναν όρο των LCSH.

Στην επόμενη ενότητα, απεικονίζεται η ενοποίηση δύο διαφορετικών δομών στο SKOS.

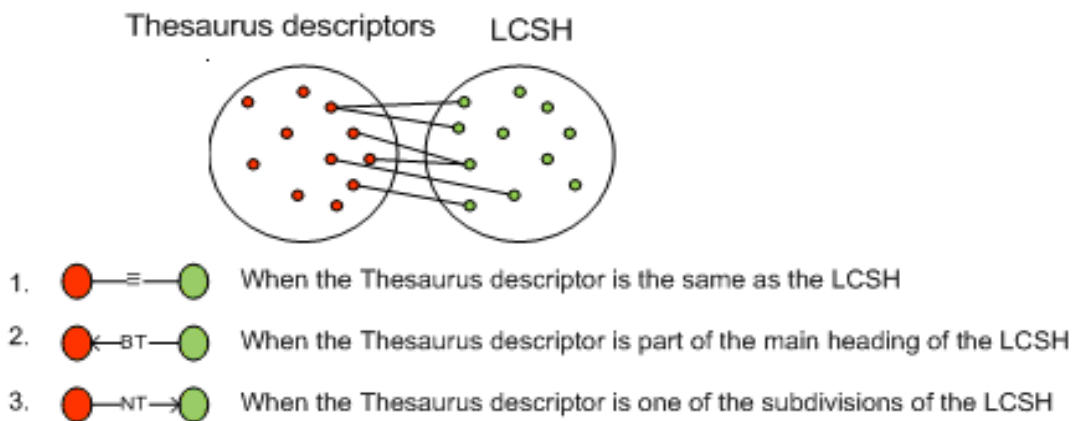
3.7.1. Μεθοδολογία αντιστοίχισης όρων LCSH και θησαυρών

Η προτεινόμενη μεθοδολογία βασίζεται στην αξίωση ότι η υποκείμενη συλλογή είναι ήδη ευρετηριασμένη με όρους των LCSH. Για να ενισχυθεί η συλλογή που είναι ευρετηριασμένη με όρους των LCSH με περιγραφείς των θησαυρών, η υποκείμενη συλλογή δεν επανευρετηριάζεται. Κάθε περιγραφέας του θησαυρού συγκρίνεται με όλους τους όρους των LCSH σύμφωνα με τον αλγόριθμο ταύτισης που είναι παρόμοιος με αυτόν που παρουσιάστηκε από τον Charlan (1995). Αυτή η επιλογή έγινε λόγω του γεγονότος ότι ο αλγόριθμος που σκιαγραφείται από τον Charlan αντιμετωπίζει και τα τρία επίπεδα συμβατότητας που αφορούν στην αντιστοίχιση LCSH και θησαυρών, σύμφωνα με τον Chamis (1991), δηλαδή:

1. Συνέπεια στις ορθογραφικές παραλλαγές, μορφές ενικού και πληθυντικού, χρόνοι ρημάτων και άλλες γραμματικές παραλλαγές.

2. Ισοδύναμοι και συνώνυμοι όροι, παραπομπές προς τον προτιμώμενο όρο, ακρωνύμια και αντώνυμα, ομόγραφα (homographs) και αλληγορίες (metaphors), πολλαπλές λέξεις και προ-συνδυασμένοι όροι.
3. Σημασιολογικές και γενικές σχέσεις, δηλ. ευρύτεροι, στενότεροι και σχετικοί όροι.

Όπως φαίνεται στην εικόνα 3.2, υπάρχουν τρία είδη πιθανών επιτυχημένων ταιριασμάτων ανάμεσα στους όρους των LCSH και τους περιγραφείς των θησαυρών, δηλαδή Ισοδυναμία (\equiv), Ευρύτερος όρος (BT) και Στενότερος όρος (NT).



Εικόνα 3.2: Προτεινόμενη μεθοδολογία αντιστοίχισης

Τα παραπάνω είδη επιτυχών ταιριασμάτων επεξηγούνται λεπτομερώς μέσω της χρήσης παραδειγμάτων.

1. Η ισοδυναμία (equivalence) σημαίνει ότι ένας περιγραφέας θησαυρού (TH) είναι όμοιος με έναν όρο των LCSH. Ως εκ τούτου, η αντίστοιχη συμβολοσειρά τοποθετείται στον ίδιο κόμβο στην ιεραρχική δομή (π.χ. Accounting (LCSH) \equiv Accounting (TH)).

Υπάρχουν πολλές περιπτώσεις όπου το ισοδύναμο ταιρίασμα καθιερώνεται ακόμα κι αν οι συμβολοσειρές των δύο λεξιλογίων δεν είναι ακριβώς όμοιες.

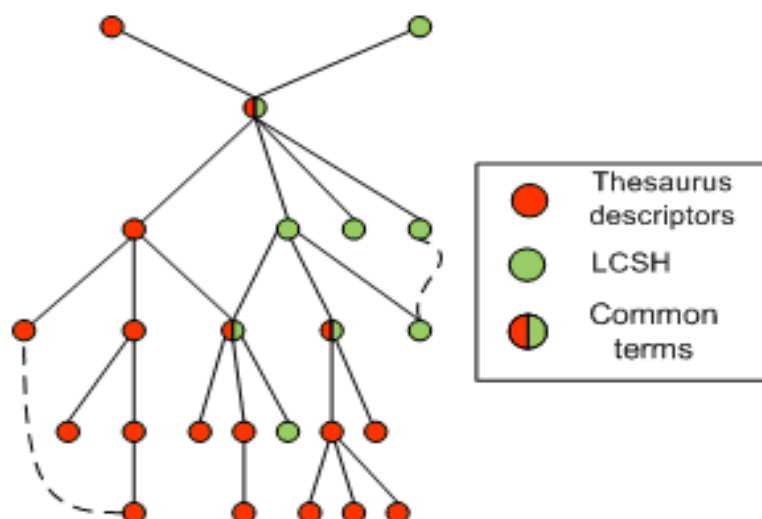
- Σχεδόν ακριβές ταιρίασμα: εάν ο περιγραφέας του θησαυρού (π.χ. Research management) είναι περίπου όμοιος με έναν όρο των LCSH (π.χ. Research (Management)), αυτές οι δύο συμβολοσειρές είναι ισοδύναμες.
- Ταίριασμα ενικού/πληθυντικού: εάν δύο συμβολοσειρές είναι στον ενικό και στον πληθυντικό αντίστοιχα, αυτές οι συμβολοσειρές θεωρούνται ισοδύναμες (π.χ. Automobiles (LCSH), Automobile (TH)).

- Παραλλαγές των καταλήξεων (suffix variation): εάν δύο συμβολοσειρές είναι διαφορετικές εξ' αιτίας μερικών παραπάνω χαρακτήρων (π.χ. Business plan (LCSH), Business planning (TH)), αυτές οι συμβολοσειρές θεωρούνται ισοδύναμες.
 - Οι συμβολοσειρές που έχουν διαφορετική σειρά στις λέξεις θεωρούνται ισοδύναμες, (π.χ. Finance, Public (LCSH), Public finance (TH)).
2. Όταν ένας περιγραφέας του θησαυρού αντιστοιχείται με την κύρια επικεφαλίδα ενός καθιερωμένου όρου των LCSH και όχι με τα υπόλοιπα στοιχεία της επικεφαλίδας, τότε μια σχέση BT καθιερώνεται ανάμεσα στον περιγραφέα και τον αντίστοιχο όρο (π.χ. Distribution (TH) **BT** Distribution (Probability theory) (LCSH)).

Αυτές οι δύο συμβολοσειρές συνδέονται με τη σχέση BT, επειδή ο όρος μέσα στην παρένθεση κάνει τη συμβολοσειρά πιο ειδική.

3. Όταν ο περιγραφέας του θησαυρού αντιστοιχείται με μια από τις υποδιαίρεσεις ενός όρου των LCSH, τότε μια σχέση NT καθιερώνεται ανάμεσα στον όρο των LCSH και τον αντίστοιχο περιγραφέα του θησαυρού (π.χ. Environmental management -- Measurement (LCSH) **NT** Measurement (TH)).

Αυτές οι δύο συμβολοσειρές συνδέονται με τη σχέση NT, επειδή ο περιγραφέας του θησαυρού χρησιμοποιείται ως υποδιαίρεση στον όρο των LCSH, έχοντας με αυτόν τον τρόπο ένα πιο αφηρημένο νόημα σε σχέση με τον όρο των LCSH.



Εικόνα 3.3: Κοινή δομή

Η προτεινόμενη μεθοδολογία καταλήγει σε μια ιεραρχική δομή, η οποία περιέχει μαζί τους όρους των LCSH και τους περιγραφείς των θησαυρών (βλέπε την εικόνα

3.3). Οι συμβολοσειρές που υπάρχουν και ως όροι των LCSH και ως περιγραφείς του θησαυρού αναπαρίστανται ως ένας κόμβος και ουσιαστικά αποτελούν το σημείο διασύνδεσης ανάμεσα στις δύο δομές.

3.7.2. Μοντελοποίηση οντολογίας στο SKOS

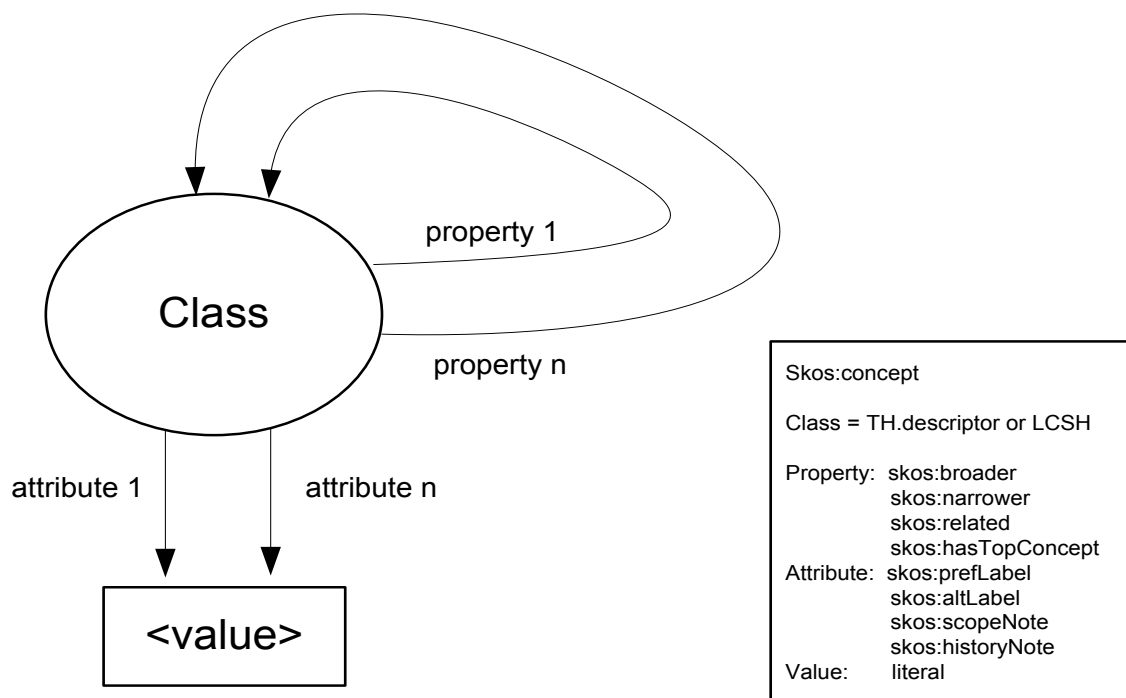
Για να αποθηκευτεί η προαναφερθείσα δομή σε μια μορφή κατανοητή από τις μηχανές, προτείνεται μια οντολογία σε μορφή SKOS (αναλυτική παρουσίαση του λεξιλογίου SKOS γίνεται στην ενότητα 4.4.4). Υπάρχουν πολλές μελέτες που χρησιμοποιούν το SKOS για να ενοποιήσουν τις LCSH με τους θησαυρούς (π.χ. Harper, 2006; Miles et al., 2005). Κατά παρόμοιο τρόπο, το SKOS λειτουργεί εδώ ως το κοινό σχήμα πάνω στο οποίο εκφράζονται οι LCSH και οι θησαυροί. Πιο συγκεκριμένα, κάθε κόμβος της δομής που παρουσιάζεται στην εικόνα 3.3, αναπαρίσταται στο SKOS ως `<skos:concept>`. Το `<skos:concept>` έχει τις ακόλουθες ιδιότητες:

Descriptor		SKOS		LCSH
Preferred	←	<code><skos:prefLabel></code>	→	USE
Non Preferred	←	<code><skos:altLabel></code>	→	UF
Scope note	←	<code><skos:scopeNote></code>	→	Scope note
HN	←	<code><skos:historyNote></code>	→	---

Επιπλέον, τα στοιχεία του `<skos:concept>` συσχετίζονται μεταξύ τους σύμφωνα με τις ακόλουθες σχέσεις:

Descriptor		SKOS		LCSH
BT	←	<code><skos:broader></code>	→	BT
NT	←	<code><skos:narrower></code>	→	NT
RT	←	<code><skos:related></code>	→	RT
Top term	←	<code><skos:hasTopConcept></code>	→	---

Το `<skos:conceptScheme>` χρησιμοποιείται για να καταγράψει την αρχική προέλευση μιας συμβολοσειράς (δηλ. όρος των LCSH ή περιγραφέας θησαυρού). Τέλος, η ιδιότητα `'xml:lang'` χρησιμοποιείται για να καταγράψει τη γλώσσα της συμβολοσειράς (βλέπε εικόνα 3.4).



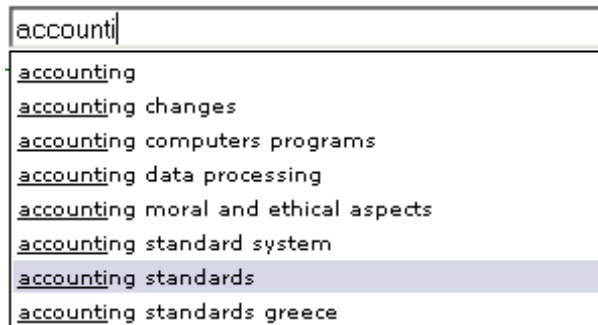
Εικόνα 3.4: Απεικόνιση της οντολογίας σε SKOS

Όπως αναφέρθηκε προηγουμένως, η οντολογία SKOS σειριοποιείται σε ένα σύστημα βάσης δεδομένων (βλέπε ενότητα 3.7.1), το οποίο είναι ικανό να ανταποκριθεί αποτελεσματικά σε ερωτήματα που απευθύνονται από τους τελικούς χρήστες. Μια πρωτότυπη εφαρμογή, που περιέχει την κοινή δομή των όρων των LCSH και των περιγραφών του θησαυρού θα παρουσιαστεί στην επόμενη ενότητα.

3.7.3. Τεκμηρίωση της μεθοδολογίας

Για να τονιστεί η δυνατότητα εφαρμογής της προτεινόμενης προσέγγισης στα σύγχρονα συστήματα ανάκτησης πληροφοριών, δημιουργήθηκε ένα πρωτότυπο σύστημα ανάκτησης πληροφοριών βασισμένο στα θέματα, το οποίο λειτουργεί ως επέκταση της λειτουργικότητας της ψηφιακής βιβλιοθήκης (Διώνη). Το πρωτότυπο σύστημα βασίζεται στην κοινή δομή που προέρχεται από την ενοποίηση των ήδη ευρετηριασμένων LCSH και του θησαυρού του ILO⁵⁴. Η κοινή δομή παράγεται με τη χρήση της προαναφερθείσας μεθοδολογίας ταύτισης και αντίστοιχα αποθηκεύεται σε μια οντολογία βασισμένη στις προδιαγραφές του SKOS.

⁵⁴ Ilo.org, (2015). *ILO Thesaurus*. [online] Διαθέσιμο στο: <http://www.ilo.org/thesaurus/> [Πρόσβαση: 07/07/2015].

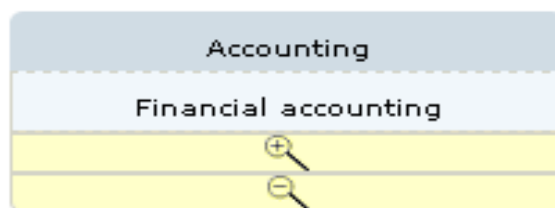


Εικόνα 3.5: Πεδίο αυτόματης συμπλήρωσης

Το προτεινόμενο σύστημα δίνει στους τελικούς χρήστες τη δυνατότητα να περιηγηθούν στην κοινή δομή και ταυτόχρονα να λάβουν σχετικές πηγές από τον υποκείμενο ΟΡΑΚ. Πιο συγκεκριμένα, οι χρήστες αρχικά καλούνται να εκφράσουν τις πληροφοριακές τους ανάγκες σε ένα πεδίο αυτόματης συμπλήρωσης (βλέπε εικόνα 3.5), το οποίο προτείνει θέματα που προέρχονται από την κοινή δομή. Ένα τέτοιο εργαλείο λειτουργεί ως σημείο εισαγωγής (entrypoint) στο γενικότερο σύστημα, το οποίο είναι ικανό να βοηθά τους χρήστες να αντιστοιχήσουν τις πληροφοριακές τους ανάγκες σε ένα υπάρχον θέμα. Το πεδίο αυτόματης συμπλήρωσης γεφυρώνει το χάσμα ανάμεσα στις πληροφοριακές ανάγκες των χρηστών και τα θέματα, τα οποία είναι αποθηκευμένα στο υποκείμενο ευρετήριο.

Κατά την επιλογή μιας πρότασης από το πεδίο αυτόματης συμπλήρωσης, δύο γεγονότα ενεργοποιούνται:

1. Ένα πλαίσιο (βλέπε εικόνα 3.6) που αναπαριστά το θέμα σχεδιάζεται κάτω από το πεδίο αυτόματης συμπλήρωσης. Το πλαίσιο αποτελείται από την ετικέτα του θέματος μαζί με τις ετικέτες των προτιμώμενων και μη προτιμώμενων όρων (preferred/non-preferred terms) (δηλ. σκούρο και ανοιχτό σκιασμένο φόντο αντίστοιχα) και τις σχέσεις NT, BT (δηλ. 🔍 και 🔍 αντίστοιχα), RT (δηλ. "see/so"). Όπως αναφέρθηκε προηγουμένως, το θέμα μπορεί να ανήκει στο λεξιλόγιο των LCSH, στο θησαυρό ILO ή και στα δύο (βλέπε εικόνα 3.3). Εάν το θέμα είναι ένας όρος των LCSH, τότε το ερώτημα απευθύνεται στο ευρετήριο "subject" του ΟΡΑΚ. Εάν το θέμα είναι ένας περιγραφέας του θησαυρού ILO, τότε το ερώτημα απευθύνεται στο ευρετήριο "title" του υποκείμενου ΟΡΑΚ. Τέλος, εάν ο όρος είναι κοινός και στα δύο λεξιλόγια, τότε το ερώτημα απευθύνεται στο ευρετήριο "subject".
2. Ένα ερώτημα που περιέχει το επιλεγμένο θέμα αποστέλλεται στον υποκείμενο ΟΡΑΚ (βλέπε εικόνα 3.7).



Εικόνα 3.6: Σχεδιασμένο πλαίσιο του όρου "Accounting"

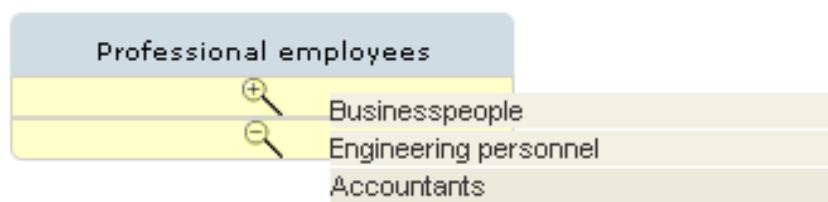
Subject	▼	"Accounting"	or	▼
Subject	▼		or	▼
Subject	▼			
Search				

Results 1-10 of 11

Date	Title
3-Oct-2008	Accounting: ethics and morality
1-Dec-2003	Budget of the cost-based activities
10-Mar-2006	International financial reporting standards
26-May-2009	The information content of the profit announcement by accruals

Εικόνα 3.7: Αποτελέσματα αναζήτησης για τον όρο "Accounting"

Οι τελικοί χρήστες μπορούν να πατήσουν πάνω σε ένα αποτέλεσμα αναζήτησης ώστε να πλοηγηθούν στη συγκεκριμένη πηγή, ή να επαναδιατυπώσουν το ερώτημά τους. Εναλλακτικά, μπορεί να επιθυμούν να συνεχίσουν την περιήγηση στην υποκείμενη συλλογή. Σε αυτή την περίπτωση, το προτεινόμενο σύστημα παρέχει τη δυνατότητα πλοήγησης στα θέματα που σχετίζονται με τον αρχικά επιλεγμένο όρο σύμφωνα με τη συνδετική δομή της κοινής δομής (δηλ. ευρύτεροι, στενότεροι και σχετικοί όροι). Αυτό είναι εφικτό πατώντας πάνω στην ετικέτα της επιθυμητής συσχέτισης (βλέπε εικόνα 3.8).



Εικόνα 3.8: Ο όρος "Professional employees" σχετίζεται μέσω της σχέση NT με τους όρους: "Business people", "Engineering Personnel" και "Accountants"



Εικόνα 3.9: Οι όροι "Professional employees" και "Accountants" σχετίζονται μέσω της σχέσης NT

Subject	"Accountants"	or
Subject		or
Subject		

Search

Results 1-10 of 11

Date	Title
3-Oct -2008	Accounting: Ethics and morality
1-Dec -2003	Cost-based activities budget
10-Aug -2006	International financial reporting standards

Εικόνα 3.10: Αποτελέσματα αναζήτησης για τον όρο "Accountants"

Κατά την επιλογή ενός θέματος, ένα άλλο πλαίσιο σχεδιάζεται δίπλα στο πρώτο. Τα δύο πλαίσια συνδέονται με μια γραμμή που περιέχει και μια ετικέτα (βλέπε εικόνα 3.9). Ταυτόχρονα, ένα νέο ερώτημα απευθύνεται στον OPAC (βλέπε εικόνα 3.10). Ξανά, ανάλογα με την προέλευση του θέματος, το ερώτημα απευθύνεται στο ευρετήριο "subject" ή στο ευρετήριο "title". Αυτή η διαδικασία μπορεί να επαναληφθεί μέχρις ότου ο τελικός χρήστης εντοπίσει μια πηγή που ικανοποιεί τις πληροφοριακές του ανάγκες.

Το προτεινόμενο γραφικό περιβάλλον είναι έτσι κατασκευασμένο ώστε να υποστηρίζει τη γνωστική ικανότητα των τελικών χρηστών. Αυτό οφείλεται στο γεγονός ότι οι χρήστες μπορούν να δουν εναλλακτικούς τρόπους για να εκφράσουν τις πληροφοριακές τους ανάγκες μέσω της ταυτόχρονης παρουσίασης και ομαδοποίησης των προτιμώμενων και μη όρων. Επιπλέον, η οπτικοποίηση της συνδετικής δομής βοηθά τους χρήστες στην εκμάθηση των σχέσεων μιας έννοιας με μια άλλη έννοια, η οποία είναι ευρύτερη, στενότερη ή σχετική.

3.7.4. Αξιολόγηση της εφαρμογής

Όπως συζητήθηκε νωρίτερα, η προτεινόμενη προσέγγιση ενισχύει ένα υπάρχον σύστημα ανάκτησης πληροφοριών βασισμένο σε θέματα μιας βιβλιοθήκης με περιγραφείς ενός θησαυρού που προέρχονται από μία εξωτερική πηγή (π.χ. ο θησαυρός ILO). Από την πλευρά του τελικού χρήστη, δεν υπάρχει καμιά προφανής διαφορά ανάμεσα στην εμφάνιση και την αίσθηση του αρχικού συστήματος ανάκτησης πληροφοριών και της εκτεταμένης έκδοσής του, που προκύπτει από τα ευρήματα της προαναφερθείσας μεθοδολογίας. Επομένως, για να αξιολογήσουμε την αποτελεσματικότητα της προτεινόμενης προσέγγισης, αποφασίστηκε να συγκρίνουμε την αρχική έκδοση του συστήματος με την προκύπτουσα έκδοση της κεντρικής συνιστώσας του συστήματος ανάκτησης πληροφοριών βασισμένο στα θέματα, δηλαδή την κοινή δομή.

Η αρχική εκδοχή της κοινής δομής αποτελείται από θέματα που συνδέονται μεταξύ τους, όπως αυτή υποδεικνύεται από τις οδηγίες των LCSH. Σύμφωνα με το παρεχόμενο γραφικό περιβάλλον, οι τελικοί χρήστες μπορούν να πλοηγηθούν στην αρχική δομή μέσω των σχέσεων BT, NT και RT που υπάρχουν ανάμεσα στα 2856 θέματα. Η χρήση του αλγόριθμου ταύτισης κατέληξε στον προσδιορισμό 476 κοινών συμβολοσειρών ανάμεσα στους όρους των LCSH και στους περιγραφείς του θησαυρού ILO. Κάθε μια από αυτές τις συμβολοσειρές συσχετίζεται με περιγραφείς του θησαυρού ILO μέσω ενός συνόλου 2343 σχέσεων BT, NT και RT. Πιο συγκεκριμένα, το βάθος κάθε κλαδιού της αρχικής ιεραρχικής δομής που περιέχει μια κοινή συμβολοσειρά επεκτείνεται κατά δύο ακόμα επίπεδα. Αυτά τα δύο επίπεδα περιέχουν 1436 συμβολοσειρές που απαντώνται μόνο στο θησαυρό ILO και 907 συμβολοσειρές που απαντώνται και στο θησαυρό ILO και στην αρχική δομή των LCSH.

Επομένως, είναι εμφανές ότι η αρχική δομή βελτιώνεται και σε όρους ποσότητας και σε όρους ποιότητας. Το πρώτο δικαιολογείται από το γεγονός ότι 1436 νέα θέματα υιοθετήθηκαν από το σύστημα χωρίς να χρειαστεί να επανευρετηριαστεί όλη η συλλογή, ενώ το δεύτερο δικαιολογείται από το γεγονός ότι οι τελικοί χρήστες μπορούν να πλοηγηθούν σε μια σημασιολογικά πλουσιότερη δομή, η οποία περιέχει 2343 σχέσεις ανάμεσα στα θέματα.

Η προτεινόμενη προσέγγιση εφαρμόζεται σε ένα σύστημα ανάκτησης πληροφοριών που επιτρέπει τη διαδραστική θεματική περιήγηση της συλλογής μιας βιβλιοθήκης (μια λεπτομερής ανάλυση του γραφικού περιβάλλοντος παρουσιάζεται στην ενότητα 2.5.3). Το σύστημα ανάκτησης πληροφοριών χρησιμοποιείται για να δώσει την ευκαιρία στους τελικούς χρήστες να βρουν σχετικές πηγές με τη διαφανή πλοήγηση μιας θεματικής δομής, η οποία περιέχει όρους των LCSH και περιγραφείς θησαυρών. Στην περίπτωση που ο τελικός χρήστης έρχεται αντιμέτωπος με έναν όρο των LCSH, το αντίστοιχο θέμα απευθύνεται στο ευρετήριο "*subject*" του εξυπηρετητή. Από την άλλη μεριά, όταν ο τελικός χρήστης έρχεται αντιμέτωπος με έναν περιγραφέα θησαυρού, το αντίστοιχο θέμα απευθύνεται στο ευρετήριο "*title*" του εξυπηρετητή.

Ο εμπλουτισμός της υπάρχουσας γενικής δομής των LCSH με περιγραφείς που προέρχονται από ένα εξειδικευμένο ελεγχόμενο λεξιλόγιο (δηλ. ILO thesaurus) βελτιώνει τη σημασιολογική ποιότητα του συστήματος ανάκτησης πληροφοριών. Οι τελικοί χρήστες έχουν περισσότερες πιθανότητες να εντοπίσουν κάτι χρήσιμο στη συλλογή μέσω της χρήσης της προτεινόμενης προσέγγισης. Επιπλέον, οι περιγραφείς του θησαυρού εκτίθενται στους τελικούς χρήστες χωρίς να έχουν ανατεθεί αποκλειστικά

σε συγκεκριμένες πηγές. Με αυτό τον τρόπο, εξοικονομούνται χρόνος και χρήμα. Ωστόσο, σε όρους ποιότητας των αποτελεσμάτων αναζήτησης, αυτή η δυνατότητα είναι παράλληλα και μειονέκτημα. Ένα ερώτημα που περιέχει περιγραφείς θησαυρού απευθύνεται σε ένα ευρετήριο που δεν έχει δημιουργηθεί αποκλειστικά για αυτό το σκοπό.

3.8. Συμπεράσματα

Σε αυτό το κεφάλαιο, παρουσιάστηκε μια μεθοδολογία που είναι ικανή να ενσωματώνει το ευρέως χρησιμοποιούμενο ελεγχόμενο λεξιλόγιο των LCSH με ένα θησαυρό. Η ενσωμάτωση καταλήγει στη δημιουργία μιας κοινής δομής, η οποία εκφράζεται ως μια οντολογία που περιέχει μαζί τους όρους των LCSH και τους περιγραφείς του θησαυρού ακολουθώντας τις προδιαγραφές του SKOS. Η οντολογία θα μπορούσε να χρησιμοποιηθεί από διάφορα συστήματα ανάκτησης πληροφοριών βασισμένα σε θέματα, χωρίς να χρειάζεται να ξαναχτιστεί το υπάρχον θεματικό ευρετήριο. Η προτεινόμενη προσέγγιση θα μπορούσε επίσης να εφαρμοστεί σε άλλες συλλογές πέρα από τον τομέα των βιβλιοθηκών, με την προϋπόθεση ότι αυτές οι συλλογές χρησιμοποιούν ελεγχόμενα λεξιλόγια που διαθέτουν κάποιο είδος συνδετικής δομής.

Ως απόδειξη της ορθότητας της ιδέας, μια πρωτότυπη εφαρμογή κατασκευάστηκε που εφαρμόζει τη λειτουργικότητα που πηγάζει από την προτεινόμενη μεθοδολογία. Πιο συγκεκριμένα, η υπάρχουσα οντολογία σε LCSH που είναι το υπόβαθρο του διαδραστικού συστήματος ανάκτησης πληροφοριών βασισμένο σε θέματα, που βρίσκεται σε μια εγκατάσταση του DSpace στη ψηφιακή βιβλιοθήκη (Διώνη), εμπλουτίστηκε και αντικαταστάθηκε με μια αντίστοιχα σχεδιασμένη οντολογία σε SKOS που πηγάζει από την ενσωμάτωση των υπάρχοντων όρων των LCSH με περιγραφείς από το θησαυρό ILO. Η οντολογία σε SKOS είναι μεγαλύτερη σε μέγεθος και σημασιολογικά πλουσιότερη και ως εκ τούτου παρέχει τη δυνατότητα για μια εμπειρία πιο ολοκληρωμένης θεματικής πλοήγησης στους τελικούς χρήστες. Από την άλλη μεριά, οι περιγραφείς του θησαυρού απευθύνονται σε ένα ευρετήριο (δηλ. "*title*") που αποτελείται από ένα σύνολο λέξεων που εμμέσως αντιστοιχεί στο θέμα της υποκείμενης πηγής.

Τέλος, ολοκληρώνοντας αυτό εδώ το κεφάλαιο θα πρέπει να τονιστεί το γεγονός ότι υπάρχει μεγάλη ποικιλία ελεγχόμενων λεξιλογίων, τα οποία δημιουργήθηκαν για την κάλυψη των ειδικών αναγκών συγκεκριμένων επιστημονικών τομέων, δυσκολεύοντας έτσι την επικοινωνία ανάμεσα σε διαφορετικούς τομείς και οργανισμούς. Μέσα σε αυτό το πλαίσιο, πολλές προσπάθειες έγιναν μέσα στο χρόνο που είχαν ως στόχο

τη συγχώνευση και την αντιστοίχιση ελεγχόμενων λεξιλογίων για την αποτελεσματικότερη αναζήτηση και ανάκτηση πληροφοριών ανάμεσα σε διαφορετικούς οργανισμούς ή και ανάμεσα στις διάφορες υπηρεσίες του ίδιου οργανισμού. Μια τέτοια προσπάθεια ήρθε να πραγματοποιήσει και η προτεινόμενη μεθοδολογία, χρησιμοποιώντας όμως πιο εξελιγμένες τεχνολογίες, όπως είναι οι οντολογίες σε μορφή SKOS.

Μάλιστα, τα τελευταία χρόνια με την έκρηξη του διαδικτύου και τα οφέλη που προσφέρει για τη γρήγορη και άμεση διάδοση της πληροφορίας, η ανάγκη για διαλειτουργικότητα ανάμεσα στους διαφορετικούς τομείς της γνώσης και τους οργανισμούς γενικά, γίνεται ολοένα και πιο επιτακτική. Αυτό το σκοπό έρχονται να καλύψουν οι τεχνολογίες των Συνδεδεμένων Δεδομένων, οι οποίες έχουν αρχίσει να παίζουν ενεργό ρόλο και στον τομέα των βιβλιοθηκών και κατ' επέκταση στη διάθεση και τη χρήση των ελεγχόμενων λεξιλογίων.

Πιο συγκεκριμένα, στις μέρες μας, ένας μεγάλος αριθμός σημασιολογικής πληροφορίας των βιβλιοθηκών παρέχεται ως πληροφορία βασισμένη στα Συνδεδεμένα Δεδομένα. Πολλά αποθετήρια γνώσης και βιβλιοθήκες συνειδητοποίησαν από νωρίς τα οφέλη από την υιοθέτηση τεχνολογιών βασισμένων στα συνδεδεμένα δεδομένα. Έτσι, οι οργανισμοί αυτοί θα επωφεληθούν όχι μόνο από την εξεύρεση λύσεων όσον αφορά στο πρόβλημα της διαλειτουργικότητας που ταλανίζει την κοινότητα των βιβλιοθηκών εδώ και πολλές δεκαετίες (Paercke et al., 1998, σ. 34), αλλά και από τη δημιουργία της απαραίτητης υποδομής για την εισαγωγή νέων υπηρεσιών προστιθέμενης αξίας. Αυτές οι υπηρεσίες θα δίνουν τη δυνατότητα στις βιβλιοθήκες να ενσωματώνουν τη σημασιολογική πληροφορία που υπάρχει στις παραδοσιακές τους υπηρεσίες, όπως είναι ο OPAC, με πληροφορίες που προέρχονται από τρίτους, όπως είναι τα αρχεία βίντεο και εικόνων, ή οι βάσεις γνώσης (π.χ. DBpedia⁵⁵, Freebase⁵⁶ κλπ.).

Στο επόμενο κεφάλαιο, θα παρουσιαστούν διεξοδικά οι τεχνολογίες των Συνδεδεμένων Δεδομένων και πως μπορούν να συνεισφέρουν στον τομέα των βιβλιοθηκών.

⁵⁵ Dbpedia.org, (2015). *DBpedia*. [online] Διαθέσιμο στο: <http://dbpedia.org/> [Πρόσβαση: 08/07/2015].

⁵⁶ Freebase.com, (2015). *Freebase*. [online] Διαθέσιμο στο: <http://www.freebase.com/> [Πρόσβαση: 08/07/2015].

4. Τα Συνδεδεμένα Δεδομένα

4.1. Γενικές πληροφορίες

Όπως αναφέρθηκε νωρίτερα, ένα πολύ σημαντικό είδος σημασιολογικής πληροφορίας που υπάρχει στα ιδρύματα μνήμης και πιο συγκεκριμένα στις βιβλιοθήκες είναι η θεματική πληροφορία των ευρετηριασμένων πηγών. Οι βιβλιοθήκες έχουν εκδηλώσει αξιοσημείωτη σταθερότητα και συνέπεια στη δημιουργία και αντίστοιχα στην κατανάλωση θεματικής πληροφορίας όλα αυτά τα χρόνια (Paercke et al., 1998, σ. 39) μέσα στα αποθετήριά τους σύμφωνα με ευρέως αποδεκτά πρότυπα και πρακτικές. Η υιοθέτηση τεχνολογιών των Συνδεδεμένων Δεδομένων, όπως είναι τα αποθετήρια τριπλών των Συνδεδεμένων Δεδομένων και τα αντίστοιχα τελικά σημεία SPARQL από σημαντικούς παρόχους δεδομένων των ιδρυμάτων πολιτιστικής κληρονομιάς είναι ένα σημαντικό βήμα προς την κατεύθυνση της παροχής ιδιαίτερα διαλειτουργικών ιδρυματικών αποθετηρίων.

Μάλιστα, ο τομέας των βιβλιοθηκών μπορεί να επωφεληθεί ποικιλοτρόπως από την χρήση των Συνδεδεμένων Δεδομένων⁵⁷. Με τα Συνδεδεμένα Δεδομένα, δίνεται η δυνατότητα στις βιβλιοθήκες να μοιραστούν και να διαθέσουν τα δεδομένα τους με έναν τρόπο που παλαιότερα δεν ήταν δυνατόν να γίνει, καθώς τα δεδομένα αυτά βασίζονταν στο πρότυπο MARC. Τα Συνδεδεμένα Δεδομένα μπορούν να συμβάλλουν επίσης, στην απρόσκοπτη πλοήγηση ανάμεσα σε ετερογενείς βιβλιοθηκονομικές βάσεις δεδομένων και άλλους παρόχους εξωτερικών πληροφοριών, όπως είναι άλλες βιβλιοθήκες και μηχανές αναζήτησης. Επίσης, με τη χρήση των Συνδεδεμένων Δεδομένων οι χρήστες μπορούν να ανακαλύψουν υλικό και πληροφορίες, τις οποίες δεν μπορούσαν καν να φανταστούν ότι υπήρχαν. Παράλληλα, μειώνεται το κόστος αποθήκευσης δεδομένων, αφού τα δεδομένα δεν χρειάζονται να αποθηκεύονται ξανά και ξανά, αλλά να διαμοιράζονται αμοιβαία ανάμεσα στους διάφορους οργανισμούς, επεκτείνοντας με αυτό τον τρόπο και την πρόσβαση στις πληροφορίες. Ακόμα, η μεταφορά των δεδομένων των βιβλιοθηκών σε μια κοινή τεχνολογία θα συμβάλλει στη διαλειτουργικότητα όχι μόνο ανάμεσα στις βιβλιοθήκες αλλά και ανάμεσα σε διαφορετικούς οργανισμούς και παρόχους πληροφοριών (Hynönen, 2012, σ. 17-18; Alemu et al., 2012, σ.557-561; Gonzales, 2014, σ. 14-15).

⁵⁷ W3.org, (2015). *Benefits - Library Linked Data*. [online] Διαθέσιμο στο: <http://www.w3.org/2005/Incubator/lld/wiki/Benefits> [Πρόσβαση: 07/07/2015].

4.2. Τι είναι τα Συνδεδεμένα Δεδομένα

Ο όρος «*Συνδεδεμένα Δεδομένα*» αναφέρεται σε ένα σύνολο αρχών και τεχνικών για τη δημοσίευση δομημένων δεδομένων στο διαδίκτυο έτσι ώστε αυτά να είναι πιο εύκολα προσβάσιμα και πιο χρήσιμα. Τεχνικά, τα Συνδεδεμένα Δεδομένα αναφέρονται στα δεδομένα που δημοσιεύονται στο διαδίκτυο με τέτοιο τρόπο που να είναι κατανοητά από τις μηχανές (machine-readable), ενώ το νόημά τους ορίζεται ρητά και συνδέονται με άλλα εξωτερικά σύνολα δεδομένων (Bizer, Heath & Berners-Lee, 2009, σ. 2).

Μάλιστα, ο Berners-Lee (2006), ο οποίος θεωρείται και ο ιδρυτής του κινήματος των Συνδεδεμένων Δεδομένων, δημοσίευσε τέσσερις κανόνες, οι οποίοι θα πρέπει να τηρούνται όταν κάποιος θέλει να συμμετέχει στο κίνημα των Συνδεδεμένων Δεδομένων. Στη συνέχεια, θα παρουσιαστούν αυτοί οι κανόνες διεξοδικά.

4.3. Κανόνες των Συνδεδεμένων Δεδομένων

1. *Χρησιμοποίησε URIs ως ονόματα για πράγματα ("Use URIs as names for things")*

Ένα URI μπορεί να αναπαριστά οποιαδήποτε οντότητα: έναν άνθρωπο, ένα αντικείμενο, μια ιδέα κλπ. Το URI δεν είναι οντότητα το ίδιο αλλά μια αναφορά σε αυτή. Η αναφορά είναι μονοσήμαντη, δηλαδή ένα URI δηλώνει πάντα μια συγκεκριμένη οντότητα και μόνο αυτή.

2. *Χρησιμοποίησε HTTP URIs έτσι ώστε οι άνθρωποι να μπορούν να ψάξουν αυτά τα ονόματα ("Use HTTP URIs so that people can look up those names")*

Ένα HTTP URI είναι μια διεύθυνση στο διαδίκτυο που μπορεί κάποιος να προσπελάσει για να ανακτήσει πληροφορίες για την οντότητα που αναγνωρίζεται μέσω του συγκεκριμένου URI (Archer et al., 2012, σ. 4). Με τη χρήση βασικών διαδικτυακών προτύπων (δηλ. URIs) για τον προσδιορισμό της οντότητας στη μορφή των Συνδεδεμένων Δεδομένων, καθίστανται τα αντίστοιχα δεδομένα εύκολα προσβάσιμα όχι μόνο από τους υπολογιστές αλλά και από τους ανθρώπους. Επιπλέον, η εγγενής ιδιότητα ενός URI να αναφέρεται στον οργανισμό που είναι υπεύθυνος για την αντίστοιχη οντότητα διευκολύνει τη συνολική διαχείριση των URIs.

3. *Όταν κάποιος ψάχνει πληροφορία γύρω από ένα URI, να παρέχεται χρήσιμη πληροφορία με τη χρήση των προτύπων RDF (Klyne & Carroll, 2004), SPARQL (Prud'*

hommeaux & Seaborne, 2008) (*"When someone looks up a URI, provide useful information, using the standards RDF, SPARQL"*)

Στο χώρο των βιβλιοθηκών, η πιο απλή μορφή πρόσβασης σε ένα URI παρέχεται μέσω μιας ιστοσελίδας με την πληροφορία από τον κατάλογο για την αντίστοιχη οντότητα. Εκτός από αυτό, αυτή η πληροφορία θα πρέπει επίσης να παρέχεται σύμφωνα με τα πρότυπα των Συνδεδεμένων Δεδομένων (δηλ. RDF, SPARQL). Με αυτό τον τρόπο, η παρεχόμενη πληροφορία μπορεί να χρησιμοποιηθεί και από άλλες εφαρμογές υπολογιστών.

4. Συμπεριέλαβε συνδέσμους σε άλλα URIs, έτσι ώστε και άλλα πράγματα να μπορούν να ανακαλυφθούν ("Include links to other URIs, so that more things may be discovered")

Η πραγματική δύναμη των Συνδεδεμένων Δεδομένων βρίσκεται στην ικανότητά τους να φέρνουν κοντά οντότητες από διαφορετικά συστήματα. Μέσα σε αυτό το πλαίσιο, μια βιβλιοθήκη που επιθυμεί να γίνει πάροχος Συνδεδεμένων Δεδομένων θα πρέπει να συμπεριλαμβάνει URIs που αναφέρονται σε απομακρυσμένες οντότητες.

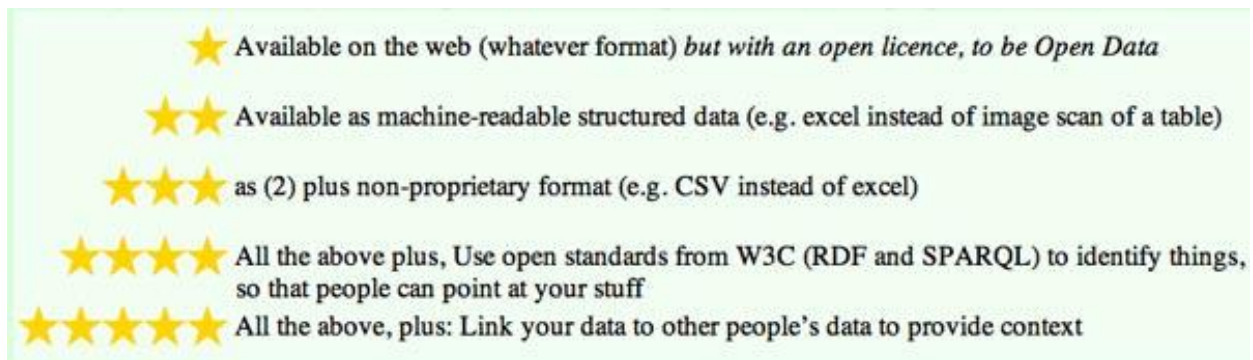
Η συμφωνία με τους παραπάνω σχεδιαστικούς κανόνες μπορεί να επιτευχθεί μέσω της χρήσης συγκεκριμένων τεχνολογιών και εργαλείων (Hannemann & Kett, 2010, σ. 3-5). Αυτές οι τεχνολογίες θα πρέπει να στηρίζονται πάνω στις παρακάτω ιδέες: α) τα URIs παρέχουν τα μέσα για τον προσδιορισμό των υποκείμενων δεδομένων, β) το RDF παρέχει ένα εννοιολογικό σχήμα για τη μοντελοποίηση των δεδομένων, γ) για τη σειριοποίηση και την αποθήκευση των δεδομένων σε μηχαναγνώσιμη μορφή χρησιμοποιούνται τα αποθετήρια τριπλετών, και τέλος, δ) τα δεδομένα γίνονται διαθέσιμα στα ενδιαφερόμενα μέρη μέσω της χρήσης ερωτημάτων που εκφράζονται σε μία αντίστοιχη γλώσσα που έχει δημιουργηθεί για αυτό το σκοπό, δηλ. SPARQL (Isaac et al., 2011).

Μάλιστα, το 2010, ο Berners-Lee πρόσθεσε μια ακόμα ενότητα στο άρθρο του για τα θέματα σχεδιασμού των Συνδεδεμένων Δεδομένων⁵⁸ σχετική με μια κλίμακα βαθμολόγησης πέντε αστεριών (5 star rating)⁵⁹. Ο στόχος αυτής της προσθήκης δεν είναι

⁵⁸ Βλέπε Berners-Lee, 2006.

⁵⁹ 5stardata.info, (2015). *5-star Open Data*. [online] Διαθέσιμο στο: <http://5stardata.info/en/> [Πρόσβαση: 02/10/2015].

να εκπαιδεύσει τους μηχανικούς στο πώς να δημιουργήσουν «καλά» Συνδεδεμένα Δεδομένα, αλλά στοχεύει στα άτομα που παίρνουν αποφάσεις για τους οργανισμούς και τα ενδιαφερόμενα μέρη. Ενθαρρύνει τους κυβερνητικούς οργανισμούς που κατέχουν δεδομένα να τα δημοσιεύσουν βασιζόμενοι στις αρχές των Συνδεδεμένων Δεδομένων ρωτώντας τους «*Είναι τα Ανοιχτά Συνδεδεμένα Δεδομένα σας 5 αστέρων;*» (Krzysztof et al., 2014, σ. 1; Hyvönen, 2012, σ. 21).



Εικόνα 4.1: Linked Data 5 star rating system⁶⁰

Πιο συγκεκριμένα, όπως φαίνεται στην εικόνα 4.1, το πρώτο αστέρι δίνεται όταν τα δεδομένα είναι απλά δημοσιευμένα στο διαδίκτυο. Όλα τα υπόλοιπα αστέρια δίνονται όταν τα δεδομένα είναι πιο εύκολο να εντοπιστούν, να χρησιμοποιηθούν και να κατανοηθούν. Επομένως, το δεύτερο αστέρι δίνεται όταν τα δεδομένα δίνονται σε μια μορφή δομημένη και μηχαναγνώσιμη. Το τρίτο αστέρι δίνεται όταν τα δεδομένα είναι σε μορφή που συμφωνεί με τα διεθνή πρότυπα ή έχουν ελεύθερες άδειες και δεν περιορίζονται σε ένα κατασκευαστή. Το τέταρτο αστέρι δίνεται όταν χρησιμοποιούνται ανοιχτά πρότυπα του W3C, όπως είναι το RDF, για τον προσδιορισμό των πηγών. Τέλος, το πέμπτο αστέρι δίνεται όταν τα τοπικά δεδομένα συνδέονται με άλλα σύνολα δεδομένων.

4.4. Τεχνολογίες των Συνδεδεμένων Δεδομένων

Τα Συνδεδεμένα Δεδομένα βασίζονται πάνω σε δύο τεχνολογίες που είναι θεμελιώδεις για το διαδίκτυο: τα URIs (Berners-Lee, 1996) και το Πρωτόκολλο Μεταφοράς Υπερκειμένου (HyperText Transfer Protocol – HTTP)⁶¹. Ενώ οι Ενιαίοι Εντοπιστές Πό-

⁶⁰ Εικόνα από: W3.org, (2015). *Linked Data - Design Issues*. [online] Διαθέσιμο στο: <http://www.w3.org/DesignIssues/LinkedData.html> [Πρόσβαση: 16/07/2015].

⁶¹ W3.org, (2015). *HTTP - Hypertext Transfer Protocol Overview*. [online] Διαθέσιμο στο: <http://www.w3.org/Protocols/> [Πρόσβαση: 08/07/2015].

ρων (Uniform Resource Locators – URLs (van Kesteren & Sam, 2014)) έχουν κατοχυρωθεί ως διευθύνσεις για έγγραφα και άλλες οντότητες που μπορούν να εντοπιστούν στο διαδίκτυο, τα URIs παρέχουν ένα πιο γενικό μέσο για τον προσδιορισμό οποιασδήποτε οντότητας που υπάρχει στον κόσμο (Bizer, Heath & Berners-Lee, 2009, σ. 3).

Όπου οι οντότητες προσδιορίζονται από URIs που χρησιμοποιούν το σχήμα "*http://*", αυτές οι οντότητες μπορούν να αναζητηθούν απλά με την ανάλυση της αναφοράς (dereference) του URI μέσω της χρήσης του πρωτοκόλλου HTTP. Τα URIs και το HTTP συμπληρώνονται από μια ακόμα τεχνολογία που είναι πολύ σημαντική για το δίκτυο των δεδομένων (Web of Data – WoD), το μοντέλο RDF (Klyne & Carroll, 2004). Ενώ η Γλώσσα Σήμανσης Υπερκειμένου (HyperText Markup Language – HTML, (Hickson et al., 2014)) παρέχει ένα μέσο για τη δόμηση και τη σύνδεση των κειμένων στο διαδίκτυο, το μοντέλο RDF παρέχει ένα περιγραφικό μοντέλο δεδομένων με το οποίο μπορούν να δομηθούν και να συνδεθούν δεδομένα που περιγράφουν τον κόσμο.

Με τη χρήση των URIs για τον προσδιορισμό των πηγών, του πρωτοκόλλου HTTP και του RDF ως μοντέλου για την περιγραφή των δεδομένων, τα Συνδεδεμένα Δεδομένα βασίζονται άμεσα στη γενική αρχιτεκτονική του διαδικτύου (Jacobs & Walsh, 2004). Επομένως, το δίκτυο των δεδομένων μπορεί να θεωρηθεί ως ένα πρόσθετο στρώμα που είναι στενά συνυφασμένο με το κλασικό διαδίκτυο και έχει πολλές κοινές ιδιότητες: α) το δίκτυο των δεδομένων είναι γενικό και μπορεί να περιέχει οποιοδήποτε είδος δεδομένων, β) οποιοσδήποτε μπορεί να δημοσιεύσει στο δίκτυο των δεδομένων, γ) αυτοί που δημοσιεύουν δεδομένα δεν περιορίζονται στην επιλογή του λεξιλογίου με το οποίο θα δημοσιευθούν τα δεδομένα και δ) οι οντότητες συνδέονται με συνδέσμους RDF, δημιουργώντας έτσι ένα παγκόσμιο γράφο δεδομένων, ο οποίος γεφυρώνει τις πηγές δεδομένων και κάνει δυνατή την ανακάλυψη νέων πηγών δεδομένων (Bizer, Heath & Berners-Lee, 2009, σ. 4).

Στη συνέχεια, θα παρουσιαστούν αναλυτικά οι τεχνολογίες των Συνδεδεμένων Δεδομένων.

4.4.1. URIs (καλά URIs)

Για να δημοσιεύσει κανείς δεδομένα στο διαδίκτυο, θα πρέπει τα δεδομένα αρχικά να προσδιοριστούν. Αυτά τα δεδομένα μπορεί να αποτελούνται από ιδιότητες και σχέσεις που τα περιγράφουν και να περιλαμβάνουν πόρους του διαδικτύου και οντότητες του πραγματικού κόσμου ή/και αφηρημένες έννοιες. Όπως ήδη αναφέρθηκε, αφού τα Συνδεδεμένα Δεδομένα βασίζονται στις αρχιτεκτονικές του διαδικτύου, τα δεδομένα

σε μορφή των Συνδεδεμένων Δεδομένων προσδιορίζονται από HTTP URIs (Heath & Bizer, 2011).

Οποιοδήποτε HTTP URI θα πρέπει να είναι "*dereferenceable*", δηλαδή θα πρέπει οι HTTP πελάτες να μπορούν να αναζητήσουν το URI με τη χρήση του πρωτοκόλλου HTTP και να ανακτούν την περιγραφή της πηγής που προσδιορίζεται με αυτό το URI. Η περιγραφή των πηγών ενσωματώνεται στη μορφή των εγγράφων του διαδικτύου. Η περιγραφή που προορίζεται να διαβαστεί από ανθρώπους συνήθως αναπαρίσταται ως HTML, ενώ η περιγραφή που προορίζεται να προσπελαστεί από μηχανές αναπαρίσταται ως RDF (Heath & Bizer, 2011).

Οι πάροχοι δεδομένων μπορούν να επιλέξουν ανάμεσα σε δύο πρότυπα χρήσης του HTTP URI για τον προσδιορισμό των οντοτήτων: α) 303 URIs⁶² και β) hash URIs⁶³. Και τα δύο πρότυπα εγγυώνται ότι μπορούν να ξεχωρίσουν τα URIs που προσδιορίζουν οντότητες του πραγματικού κόσμου και τα URIs που προσδιορίζουν έγγραφα του διαδικτύου που περιγράφουν οντότητες του πραγματικού κόσμου (Sauermann & Cyganiak, 2008).

Τα καλά (cool) URIs δεν παρέχουν μόνο περιγραφές για ανθρώπους και μηχανές, αλλά είναι σχεδιασμένα με απλότητα, σταθερότητα και ευκολία στη διαχείριση. Η απλότητα έχει να κάνει με τη χρήση μικρών και μνημονικών URIs που δεν θα σπάσουν εύκολα όταν αποστέλλονται μέσω e-mail και γενικά είναι πιο εύκολα στην απομνημόνευση. Η σταθερότητα έχει να κάνει με τη διατήρηση ενός URI στην αρχική του μορφή

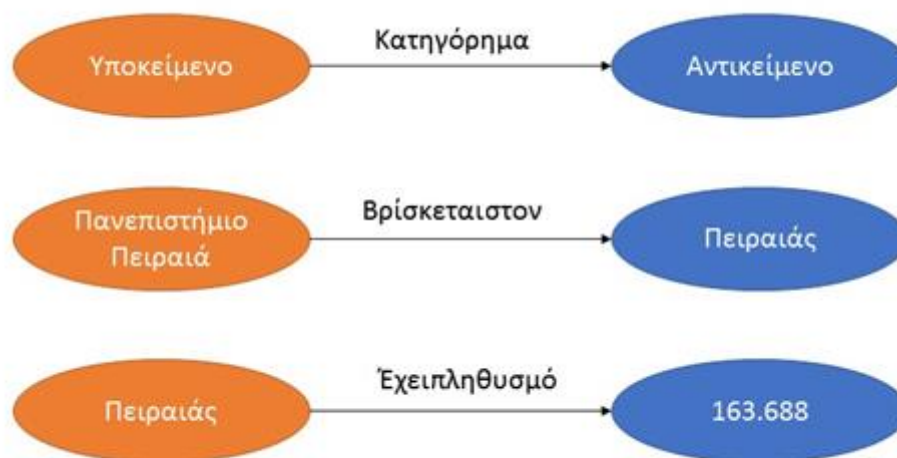
⁶² "Real-world objects, like houses or people, cannot be transmitted over the wire using the HTTP protocol. Thus, it is also not possible to directly dereference URIs that identify real-world objects. Therefore, in the 303 URIs strategy, instead of sending the object itself over the network, the server responds to the client with the HTTP response code 303 See Other and the URI of a Web document which describes the real-world object. This is called a 303 redirect. In a second step, the client dereferences this new URI and gets a Web document describing the real-world object" (Heath & Bizer, 2011).

⁶³ "A widespread criticism of the 303 URI strategy is that it requires two HTTP requests to retrieve a single description of a real-world object. One option for avoiding these two requests is provided by the hash URI strategy. The hash URI strategy builds on the characteristic that URIs may contain a special part that is separated from the base part of the URI by a hash symbol (#). This special part is called the fragment identifier. When a client wants to retrieve a hash URI, the HTTP protocol requires the fragment part to be stripped off before requesting the URI from the server. This means a URI that includes a hash cannot be retrieved directly and therefore does not necessarily identify a Web document. This enables such URIs to be used to identify real-world objects and abstract concepts, without creating ambiguity" (Heath & Bizer, 2011).

από την ημέρα που δημιουργήθηκε και για όσο το δυνατόν μεγαλύτερο χρονικό διάστημα. Τέλος, η ευκολία στη διαχείριση σχετίζεται με την έκδοση των URIs με τρόπο που να μπορούμε να τα διαχειριστούμε εύκολα (Sauermann & Cyganiak, 2008).

4.4.2. Μοντελοποίηση Συνδεδεμένων Δεδομένων

Τα URIs και το HTTP χρησιμοποιούνται στα Συνδεδεμένα Δεδομένα με την τεχνολογία του προτύπου RDF. Το RDF είναι το πρότυπο για την κωδικοποίηση μεταδεδομένων και γενικά της γνώσης στο σημασιολογικό ιστό. Παρέχει ένα γενικό μοντέλο δεδομένων με το οποίο δομούνται και συνδέονται τα δεδομένα που περιγράφουν αντικείμενα στον κόσμο. Η κωδικοποίηση των δεδομένων, σύμφωνα με το RDF, γίνεται με τη μορφή τριπλετών (triples): «*υποκείμενο (subject), κατηγορημα (predicate), αντικείμενο (object)*». Στη γενική τους μορφή, και τα τρία συστατικά της τριπλέτας αναπαρίστανται ως URIs. Το κατηγορημα καθορίζει το πώς συνδέονται μεταξύ τους το υποκείμενο και το αντικείμενο. Πολλές φορές, το αντικείμενο μπορεί να είναι απλό κείμενο (Bizer, Heath & Berners-Lee, 2009, σ. 3). Μια ομάδα τριπλετών αποτελεί έναν RDF γράφο.

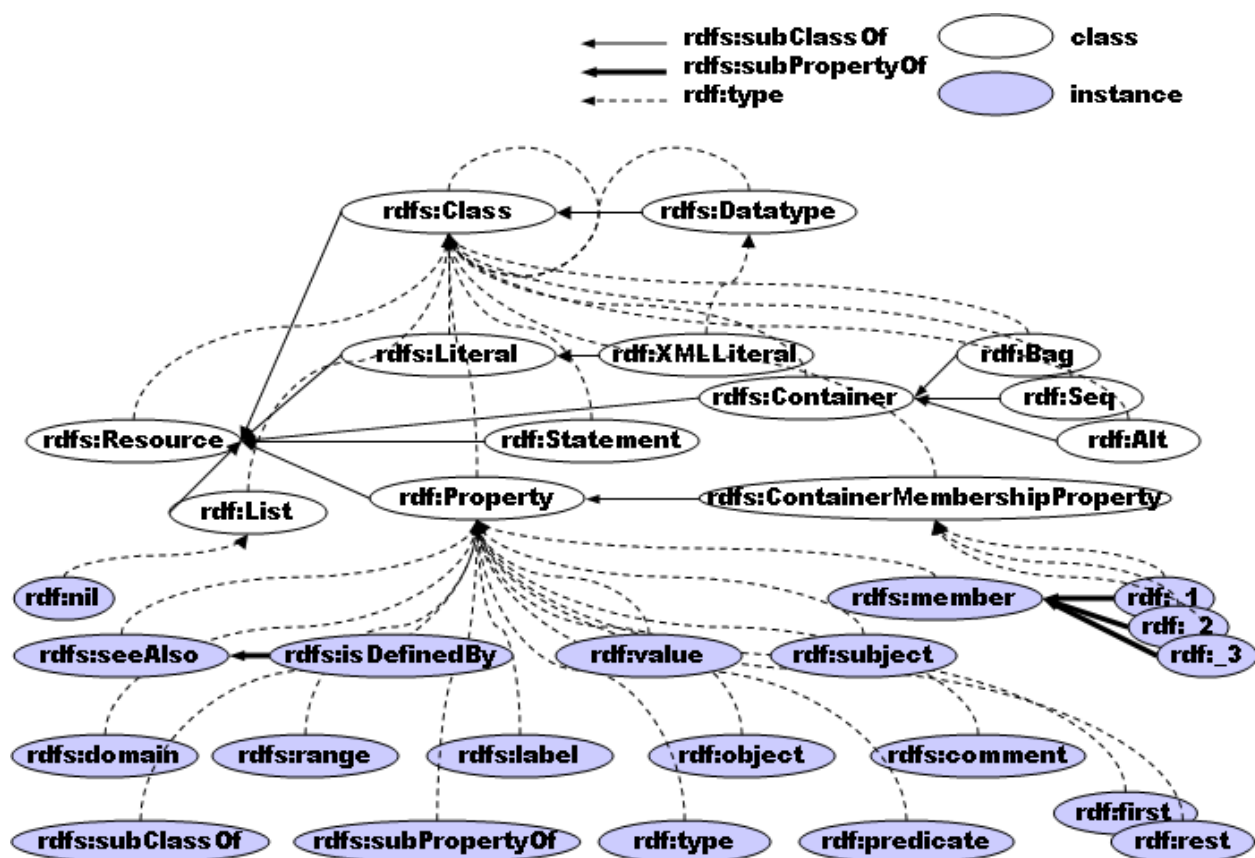


Εικόνα 4.2: Παράδειγμα τριπλετών

Στην εικόνα 4.2 δίνεται ένα παράδειγμα τριπλετών. Πιο συγκεκριμένα, το κατηγορημα «*Βρίσκεται στον*» συνδέει το υποκείμενο «*Πανεπιστήμιο Πειραιά*» με το αντικείμενο «*Πειραιάς*», οπότε από την εν λόγω τριπλέτα παίρνουμε την πληροφορία ότι το «*Πανεπιστήμιο Πειραιά Βρίσκεται στον Πειραιά*». Αντίστοιχα, η επόμενη τριπλέτα δηλώνει ότι το κατηγορημα «*έχει πληθυσμό*» συνδέει το υποκείμενο «*Πειραιάς*» με το αντικείμενο «*163.688*», δηλώνοντας ότι ο «*Πειραιάς έχει πληθυσμό 163.688*».

Τα οφέλη από τη χρήση του μοντέλου δεδομένων RDF στο πλαίσιο των Συνδεδεμένων Δεδομένων μπορούν να συνοψιστούν στα εξής (Heath & Bizer, 2011):

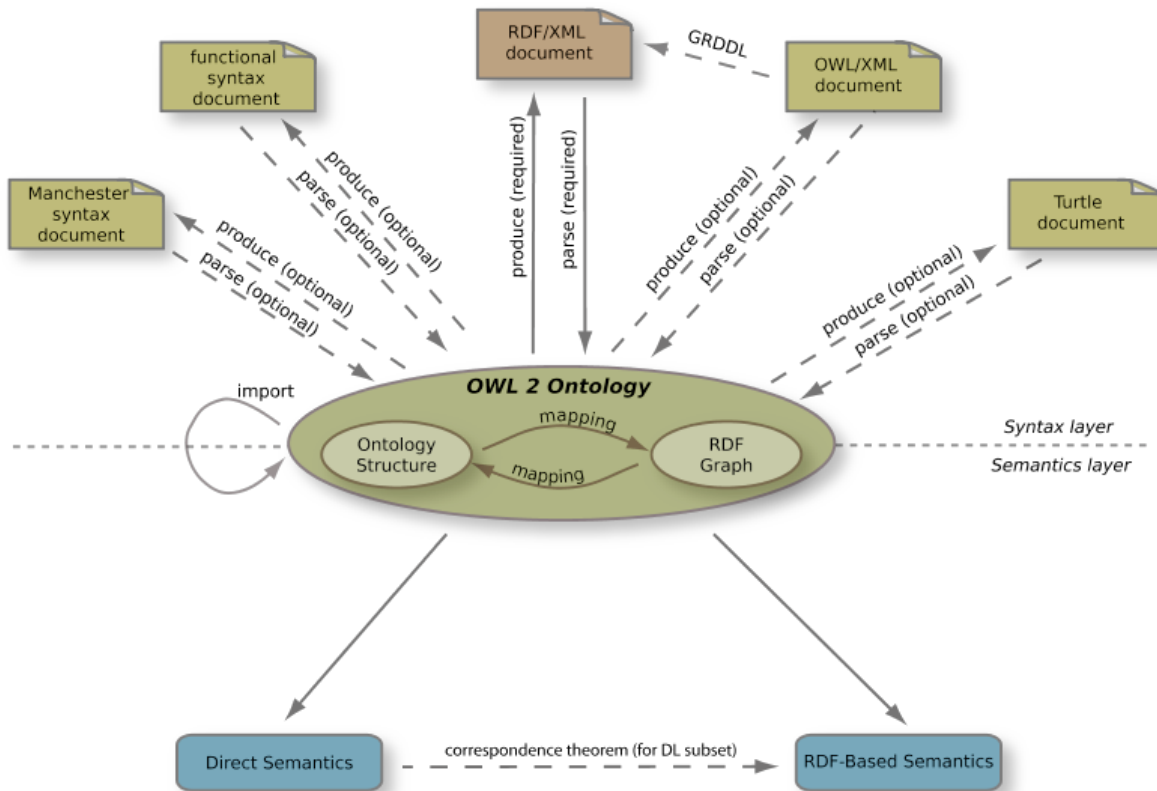
1. Με τη χρήση HTTP URIs ως παγκόσμιους μοναδικούς προσδιοριστές για τα δεδομένα καθώς και για τους όρους των λεξιλογίων, το μοντέλο δεδομένων έχει εγγενώς σχεδιαστεί για να χρησιμοποιείται σε παγκόσμια κλίμακα και να δίνει τη δυνατότητα σε οποιονδήποτε να αναφέρεται σε οτιδήποτε.
2. Οι πελάτες μπορούν να αναζητήσουν οποιοδήποτε URI σε ένα RDF γράφο στο διαδίκτυο για να ανακτήσουν επιπλέον πληροφορίες. Ως εκ τούτου, κάθε τριπλέτα RDF είναι μέρος του παγκόσμιου δικτύου των δεδομένων και μπορεί να χρησιμοποιηθεί ως σημείο εκκίνησης για την εξερεύνηση του συγκεκριμένου πεδίου δεδομένων.
3. Το μοντέλο δεδομένων δίνει τη δυνατότητα να οριστούν RDF σχέσεις ανάμεσα σε δεδομένα από διαφορετικές πηγές.
4. Οι πληροφορίες από δύο διαφορετικές πηγές μπορούν εύκολα να συνδυαστούν με τη συγχώνευση συνόλων τριπλετών σε ένα γράφο.
5. Το RDF επιτρέπει την αναπαράσταση της πληροφορίας, που εκφράζεται με τη χρήση ενός διαφορετικού σχήματος σε ένα γράφο, δηλαδή μπορεί να γίνει ανάμιξη όρων από διαφορετικά λεξιλόγια για την αναπαράσταση των δεδομένων.
6. Συνδυάζοντας το RDF με γλώσσες σχημάτων, όπως είναι το RDFS και η OWL, το RDF επιτρέπει τη χρήση όσο το δυνατόν μικρότερης δομής.



Εικόνα 4.3: Ιεραρχική δομή του RDFS⁶⁴

Λόγω του γεγονότος ότι το RDF είναι ένα γενικό μοντέλο δεδομένων, έχουν δημιουργηθεί το RDFS 1.1 (Brickley and Guha, 2014) και η OWL 2 (W3C OWL Working Group, 2012) ως η βάση για τη δημιουργία λεξιλογίων, τα οποία μπορούν να χρησιμοποιηθούν για να περιγράψουν οντότητες του πραγματικού κόσμου και πως αυτές οι οντότητες συνδέονται μεταξύ τους (Bizer, Heath & Berners-Lee, 2009, σ. 3). Τα λεξιλόγια αυτά είναι ουσιαστικά συλλογή από κλάσεις και ιδιότητες. Τα λεξιλόγια εκφράζονται σε μορφή RDF, χρησιμοποιώντας όρους από το RDFS (βλέπε εικόνα 4.3, για την ιεραρχική δομή του RDFS) και την OWL 2 (βλέπε εικόνα 4.4, για τη δομή της OWL 2), τα οποία παρέχουν διάφορα επίπεδα εκφραστικότητας στη μοντελοποίηση των τομέων ενδιαφέροντος.

⁶⁴ iswc2004.semanticweb.org, (2015). *SWCLOS: A Semantic Web Processor on Common Lisp Object System*. [online] Διαθέσιμο στο: <http://iswc2004.semanticweb.org/demos/32/> [Πρόσβαση: 08/07/2015].



Εικόνα 4.4: Δομή της OWL 2⁶⁵

Ως εκ τούτου, έχουν δημιουργηθεί διάφορα λεξιλόγια για την αναπαράσταση του πραγματικού κόσμου σε μορφή των Συνδεδεμένων Δεδομένων βάσει των αναγκών που προκύπτουν για την περιγραφή του εκάστοτε πεδίου. Οι Isaac et al. (2011) έχουν δημοσιεύσει μια λίστα με τα πιο κοινά λεξιλόγια που βασίζονται στο RDF και σχετίζονται με τις βιβλιοθήκες και γενικά με τα πολιτιστικά ιδρύματα (π.χ. Dublin Core, OAI-ORE, CIDOC-CRM, FRBR, Schema.org κ.ά.).

Στην επόμενη ενότητα παρουσιάζονται οι σειριοποιήσεις του RDF (RDF serializations).

4.4.3. Σειριοποίηση του RDF

Επειδή το RDF δεν περιγράφει ένα μορφότυπο δεδομένων αλλά ένα μοντέλο δεδομένων της μορφής τριπλετών υποκείμενο, κατηγορημα, αντικείμενο για να δημοσιευθεί στο διαδίκτυο θα πρέπει να σειριοποιηθεί χρησιμοποιώντας μια σύνταξη RDF. Αυτό σημαίνει ότι παίρνουμε τις τριπλέτες που αποτελούν ένα γράφο RDF και χρησιμοποιώντας μια συγκεκριμένη σύνταξη τις αποθηκεύουμε σε ένα αρχείο. Δύο μορφές σειριοποίησης έχουν καθιερωθεί από το W3C, το RDF/XML και το RDFa. Υπάρχουν

⁶⁵ Εικόνα από το W3C OWL Working Group (2012).

όμως, και άλλες μορφές σειριοποίησης που δεν είναι καθιερωμένες αλλά έχουν δημιουργηθεί για να εξυπηρετήσουν συγκεκριμένες ανάγκες, π.χ. Turtle, N-Triples, RDF/JSON κλπ. (Heath & Bizer, 2011).

4.4.3.1. RDF/XML

Η σύνταξη RDF/XML έχει καθιερωθεί από το W3C και χρησιμοποιείται ευρέως για τη δημοσίευση Συνδεδεμένων Δεδομένων στο διαδίκτυο. Ωστόσο, η σύνταξη αυτή είναι δύσκολο να αναγνωστεί και να γραφτεί από ανθρώπους και για αυτό το λόγο θα πρέπει να εξετάζονται και άλλες σειριοποιήσεις για τη διαχείριση και την επιμέλεια ροών εργασίας που εμπλέκουν ανθρώπινη συμμετοχή (Heath & Bizer, 2011). Η RDF/XML σύνταξη περιγράφεται αναλυτικά στο Schreiber & Raimond (2014). Στην εικόνα 4.5, παρουσιάζεται ένα παράδειγμα γράφου σε σύνταξη RDF/XML.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:schema="http://schema.org/"
  <rdf:Description rdf:about="http://example.org/bob#me">
    <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
    <schema:birthDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1990-07-04</schema:birthDate>
    <foaf:knows rdf:resource="http://example.org/alice#me"/>
    <foaf:topic_interest rdf:resource="http://www.wikidata.org/entity/Q12418"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.wikidata.org/entity/Q12418">
    <dcterms:title>Mona Lisa</dcterms:title>
    <dcterms:creator rdf:resource="http://dbpedia.org/resource/Leonardo_da_Vinci"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619">
    <dcterms:subject rdf:resource="http://www.wikidata.org/entity/Q12418"/>
  </rdf:Description>
</rdf:RDF>
```

Εικόνα 4.5: Παράδειγμα σύνταξης σε RDF/XML⁶⁶

4.4.3.2. RDFa

Η RDFa είναι μια μορφή σειριοποίησης που ενσωματώνει τριπλέτες RDF σε έγγραφα της HTML και της XML. Έτσι, τα ήδη υπάρχοντα περιεχόμενα της ιστοσελίδας μπορούν να σημανθούν με RDFa τροποποιώντας τον κώδικα στην HTML και επομένως να δημιουργηθούν δομημένα δεδομένα στο διαδίκτυο (Heath & Bizer, 2011). Στην εικόνα 4.6, παρουσιάζεται ένα παράδειγμα γράφου σε σύνταξη RDFa.

⁶⁶ Εικόνα από το Schreiber & Raimond (2014).

```

<body prefix="foaf: http://xmlns.com/foaf/0.1/
          schema: http://schema.org/
          dcterms: http://purl.org/dc/terms/">
<div resource="http://example.org/bob#me" typeof="foaf:Person">
  <p>
    Bob knows <a property="foaf:knows" href="http://example.org/alice#me">Alice</a>
    and was born on the <time property="schema:birthDate" datatype="xsd:date">1990-07-04</time>.
  </p>
  <p>
    Bob is interested in <span property="foaf:topic_interest"
    resource="http://www.wikidata.org/entity/Q12418">the Mona Lisa</span>.
  </p>
</div>
<div resource="http://www.wikidata.org/entity/Q12418">
  <p>
    The <span property="dcterms:title">Mona Lisa</span> was painted by
    <a property="dcterms:creator" href="http://dbpedia.org/resource/Leonardo_da_Vinci">Leonardo da Vinci</a>
    and is the subject of the video
    <a href="http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619">'La Joconde à Washington'</a>.
  </p>
</div>
<div resource="http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619">
  <link property="dcterms:subject" href="http://www.wikidata.org/entity/Q12418"/>
</div>
</body>

```

Εικόνα 4.6: Παράδειγμα σύνταξης σε RDFa⁶⁷

4.4.3.3. Turtle

Η Turtle είναι μια σειριοποίηση RDF δεδομένων σε μορφή απλού κειμένου (Beckett et al., 2014). Λόγω του γεγονότος ότι υποστηρίζει τα προθέματα των χώρων ονοματοδοσίας (namespaces) και διάφορες άλλες συντομεύσεις, η Turtle είναι η μορφή σειριοποίησης που επιλέγεται όταν κάποιος άνθρωπος θέλει να διαβάσει τριπλέτες RDF ή θέλει να τις γράψει στο χέρι (Heath & Bizer, 2011). Στην εικόνα 4.7, παρουσιάζεται ένα παράδειγμα γράφου σε σύνταξη Turtle.

⁶⁷ ο.π.

```
BASE <http://example.org/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX schema: <http://schema.org/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX wd: <http://www.wikidata.org/entity/>

<bob#me>
  a foaf:Person ;
  foaf:knows <alice#me> ;
  schema:birthDate "1990-07-04"^^xsd:date ;
  foaf:topic_interest wd:Q12418 .

wd:Q12418
  dcterms:title "Mona Lisa" ;
  dcterms:creator <http://dbpedia.org/resource/Leonardo_da_Vinci> .

<http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619>
  dcterms:subject wd:Q12418 .
```

Εικόνα 4.7: Παράδειγμα σύνταξης σε Turtle⁶⁸

4.4.3.4. N-Triples

Η σειριοποίηση σε μορφή N-Triples είναι ένα υποσύνολο της σύνταξης Turtle, πλην κάποιων χαρακτηριστικών, όπως προθέματα χώρων ονοματοδοσίας και άλλες συντομεύσεις. Το αποτέλεσμα είναι μια μορφή σειριοποίησης που βρίσκεται πλεονασμών, αφού όλα τα URIs θα πρέπει να περιγράφονται ολόκληρα σε κάθε τριπλέτα. Οπότε τα αρχεία που προκύπτουν με αυτό τον τρόπο σειριοποίησης μπορεί να είναι σχετικά μεγάλα σε σχέση με τα αρχεία που δημιουργούνται από τη σειριοποίηση σε μορφή Turtle ή ακόμα και σε RDF/XML. Ωστόσο, αυτό είναι παράλληλα και πλεονέκτημα γιατί δίνουν τη δυνατότητα να γίνεται γραμμική προσπέλαση του αρχείου, καθιστώντας τα ιδανικά για τη φόρτωση μεγάλων αρχείων δεδομένων που δε θα μπορούσαν να φορτωθούν στη μνήμη των κοινών υπολογιστών. Ο πλεονασμός καθιστά επίσης τα N-Triples πολύ δεκτικά στη συμπίεση, μειώνοντας έτσι την κίνηση στο διαδίκτυο όταν ανταλλάσσονται αρχεία. Αυτοί οι δύο παράγοντες έχουν καταστήσει τη σειριοποίηση N-Triples ως το «*ντε φάκτο*» πρότυπο για την ανταλλαγή μεγάλων αρχείων dump των Συνδεδεμένων Δεδομένων (Heath & Bizer, 2011). Στην εικόνα 4.8, παρουσιάζεται ένα παράδειγμα γράφου σε N-Triples σύνταξη.

⁶⁸ ο.π.

```
<http://example.org/bob#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person> .
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice#me> .
<http://example.org/bob#me> <http://schema.org/birthDate> "1990-07-04"^^<http://www.w3.org/2001/XMLSchema#date> .
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/topic_interest> <http://www.wikidata.org/entity/Q12418> .
<http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/title> "Mona Lisa" .
```

Εικόνα 4.8: Παράδειγμα σύνταξης σε N-Triples⁶⁹

4.4.3.5. RDF/JSON

Η σύνταξη RDF/JSON αναφέρεται στις προσπάθειες για την παροχή του RDF σε μορφή σειριοποίησης της JavaScript Object Notation – JSON⁷⁰. Η δημοσίευση RDF δεδομένων σε μορφή JSON τα καθιστά προσβάσιμα σε προγραμματιστές του διαδικτύου χωρίς να είναι απαραίτητη η εγκατάσταση επιπρόσθετου λογισμικού για την προσπέλαση και διαχείριση δεδομένων σε RDF (Heath & Bizer, 2011). Στην εικόνα 4.9, παρουσιάζεται ένα παράδειγμα γράφου σε σύνταξη RDF/JSON.

```
{
  "@context": "example-context.json",
  "@id": "http://example.org/bob#me",
  "@type": "Person",
  "birthdate": "1990-07-04",
  "knows": "http://example.org/alice#me",
  "interest": {
    "@id": "http://www.wikidata.org/entity/Q12418",
    "title": "Mona Lisa",
    "subject_of": "http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619",
    "creator": "http://dbpedia.org/resource/Leonardo_da_Vinci"
  }
}
```

Εικόνα 4.9: Παράδειγμα σύνταξης σε RDF/JSON⁷¹

Στη συνέχεια, θα παρουσιαστεί το SKOS (Miles & Bechhofer, 2009), το οποίο αποτελεί το κατ' εξοχήν RDF λεξιλόγιο που μπορεί να αναπαραστήσει ελεγχόμενα λεξιλόγια.

4.4.4. SKOS

Το SKOS (Issac & Summers, 2009) είναι μια πρωτοβουλία του W3C και παρέχει ένα μοντέλο για την αποτύπωση της βασικής δομής και του περιεχομένου σχημάτων εννοιών, όπως είναι οι θησαυροί, τα ταξινομικά σχήματα, οι λίστες θεματικών επικεφαλίδων και άλλα παρόμοια λεξιλόγια. Με τη χρήση του SKOS ένα σύστημα οργάνωσης γνώσης μπορεί να μετατραπεί σε μορφή ικανή να αναγνωστεί από υπολογιστές

⁶⁹ ο.π.

⁷⁰ Json.org, (2015). *JSON*. [online] Διαθέσιμο στο: <http://json.org/> [Πρόσβαση: 21/07/2015].

⁷¹ ο.π.

και να δημοσιευτεί στο διαδίκτυο. Είναι μια εφαρμογή του RDF, επομένως μπορεί να εκφραστεί ως τριπλέτες του RDF και να σειριοποιηθεί (Miles & Pérez-Agüera, 2007, σ. 70).

Παρέχει ένα απλό λεξιλόγιο που μπορεί να χρησιμοποιηθεί μαζί με την OWL ή ανεξάρτητα. Επομένως, το SKOS αντιπροσωπεύει ένα ενδιάμεσο βήμα ανάμεσα στο χαμηλό επίπεδο δομής που εμφανίζεται κυρίως στο διαδίκτυο και στον αυστηρό περιγραφικό φορμαλισμό των οντολογιών που ορίζονται με την OWL (Pastor-Sanchez, Mendez & Rodríguez-Muñoz, 2009).

Πιο συγκεκριμένα, το μοντέλο δεδομένων του SKOS είναι μια οντολογία που ορίζεται από την OWL Full. Βασιζόμενο στο RDF, το SKOS δομεί τα δεδομένα σε μορφή τριπλετών, τα οποία μπορούν να κωδικοποιηθούν σε οποιαδήποτε έγκυρη RDF σύνταξη (βλέπε ενότητα 4.4.3). Μπορεί να χρησιμοποιηθεί μαζί με την OWL για να αποτυπώσει επίσημα τις δομές της γνώσης ενός συγκεκριμένου τομέα, αφού το SKOS δεν είναι μια γλώσσα για την επίσημη αποτύπωση της γνώσης (Alistair et al., 2005).

Για το SKOS, ένα σύστημα οργάνωσης γνώσης εκφράζεται σε όρους εννοιών (concepts) δομημένες σε σχέσεις για τη δημιουργία σχημάτων εννοιών (concept schemes). Οι έννοιες και τα σχήματα εννοιών προσδιορίζονται από URIs. Οι έννοιες μπορούν να επισημανθούν (labelled) σε οποιαδήποτε γλώσσα. Μια έννοια μπορεί να έχει πολλές ετικέτες επισήμανσης, αλλά μόνο μία από αυτές τις γλώσσες μπορεί να καθιερωθεί ως προτιμώμενη ετικέτα (preferred label). Οι εναπομείναντες ετικέτες σχετίζονται με την έννοια ως εναλλακτικές ετικέτες (alternative labels). Στις έννοιες μπορούν να ανατεθούν ταξινομικοί ή προσδιοριστικοί κωδικοί μέσα σε ένα συγκεκριμένο εννοιολογικό σχήμα. Αυτοί οι συμβολισμοί (notations) δεν εκφράζονται σε φυσική γλώσσα, αλλά χρησιμοποιούν μνημονικούς ή παρόμοιους κωδικούς. Οι έννοιες μπορούν επίσης, να τεκμηριωθούν με τη χρήση διαφορετικών ειδών σημειώσεων, όπως είναι οι ορισμοί (definitions), σημειώσεις σκοπού (scope notes) κ.ά. Ακόμα, το SKOS καλύπτει την καθιέρωση συνδέσμων ανάμεσα στις έννοιες γνωστές ως σημασιολογικές σχέσεις (semantic relationships). Αυτές οι σχέσεις μπορεί να είναι ιεραρχικές ή συνειρμικές. Τέλος, οι δυνατότητες του SKOS συμπληρώνονται με την ικανότητα που έχει να αντιστοιχεί/συνδέει έννοιες από διαφορετικά σχήματα με τη χρήση ιεραρχικών, συνειρμικών ή ακριβώς ισοδύναμων σχέσεων (Pastor-Sanchez, Mendez & Rodríguez-Muñoz, 2009). Στη συνέχεια, θα παρουσιαστεί αναλυτικά το βασικό λεξιλόγιο του SKOS.

4.4.4.1. Έννοιες και σχήματα εννοιών

Μια έννοια αναπαριστά μια ιδέα, μια αντίληψη ή μια μονάδα σκέψης. Οι έννοιες συνδέονται με ένα URI ή ένα RDF προσδιοριστή. Οι έννοιες μπορούν επίσης να συνδεθούν με σχήματα εννοιών. Τα σχήματα εννοιών είναι η ομαδοποίηση μιας ή περισσότερων εννοιών SKOS, που χρησιμοποιούνται για να αναπαραστήσουν και να προσδιορίσουν ένα θησαυρό ή ένα ταξινομικό σχήμα. Ένα εννοιολογικό σχήμα μπορεί να έχει μια ή περισσότερες εννοιολογικές κορυφές (top concept), οι οποίες είναι επικεφαλής της ιεραρχικής δομής μέσα στο ίδιο εννοιολογικό σχήμα (Pastor-Sanchez, Mendez & Rodríguez-Muñoz, 2009). Στον πίνακα 4.1 παρουσιάζεται το λεξιλόγιο SKOS για έννοιες και σχήματα εννοιών.

Πίνακας 4.1: Λεξιλόγιο SKOS για έννοιες και σχήματα εννοιών

Λεξιλόγιο	Περιγραφή
skos:Concept	Έννοια
skos:ConceptScheme	Σχήμα εννοιών
skos:InScheme	Συμπερίληψη σε ένα σχήμα εννοιών
skos:hasTopConcept	Επικεφαλής έννοια

4.4.4.2. Λεξιλογικές ετικέτες

Οι λεξιλογικές ετικέτες (lexical labels) είναι συμβολοσειρές χαρακτήρων που αναπαριστούν εκφράσεις φυσικής γλώσσας. Το SKOS παρέχει προτιμώμενες, εναλλακτικές και κρυφές ετικέτες. Οι προτιμώμενες ετικέτες σχετίζονται με τις έννοιες για να αναπαραστήσουν όρους περιγραφών. Μόνο μια προτιμώμενη ετικέτα μπορεί να σχετίζεται με μια έννοια. Εναλλακτικές ετικέτες μπορούν να χρησιμοποιηθούν για την εκχώρηση μη προτιμώμενων εκφράσεων για μια έννοια. Αυτό είναι χρήσιμο για την αναπαράσταση ισοδύναμων, συνώνυμων και ακρώνυμων όρων. Με αυτό τον τρόπο εμπλουτίζεται το λεξιλόγιο και ενισχύεται η πρόσβαση στην έννοια, αυξάνοντας έτσι την επιτυχία της αναζήτησης από τον τελικό χρήστη. Οι κρυφές ετικέτες χρησιμοποιούνται συνήθως για να αποτυπώσουν ορθογραφικά λάθη και δεν είναι ορατές στους τελικούς χρήστες (Pastor-Sanchez, Mendez & Rodríguez-Muñoz, 2009). Στον πίνακα 4.2 παρουσιάζεται το λεξιλόγιο SKOS για λεξιλογικές ετικέτες.

Πίνακας 4.2: Λεξιλόγιο SKOS για λεξιλογικές ετικέτες

Λεξιλόγιο	Περιγραφή
skos:prefLabel	Προτιμώμενη ετικέτα
skos:altLabel	Μη-Προτιμώμενη ετικέτα
skos:hiddenLabel	Κρυμμένη ετικέτα

4.4.4.3. Σημασιολογικές σχέσεις

Οι σημασιολογικές σχέσεις στο SKOS είναι οι συνδέσεις μεταξύ των εννοιών, οι οποίες είναι συνυφασμένες με το νόημά τους. Το SKOS αναγνωρίζει δύο βασικές σημασιολογικές σχέσεις: α) ιεραρχικές και β) συνειρμικές. Οι πρώτες υπάρχουν ανάμεσα σε δύο έννοιες και υποδηλώνουν ότι η μια έννοια είναι πιο γενική από την άλλη. Οι συνειρμικές σχέσεις υποδηλώνουν ότι οι δύο έννοιες συνδέονται με κάποιο τρόπο, ωστόσο καμιά από τις δύο δεν είναι πιο γενική ή πιο ειδική από την άλλη (Pastor-Sanchez, Mendez & Rodríguez-Muñoz, 2009). Στον πίνακα 4.3 παρουσιάζεται το λεξιλόγιο SKOS για σημασιολογικές σχέσεις.

Πίνακας 4.3: Λεξιλόγιο SKOS για σημασιολογικές σχέσεις

Λεξιλόγιο	Περιγραφή
skos:semanticRelation	Σημασιολογική σχέση
skos:broader	Ευρύτερη έννοια
skos:narrower	Στενότερη έννοια
skos:related	Σχετική έννοια

4.4.4.4. Τεκμηρίωση

Στοιχεία τεκμηρίωσης επιτρέπουν την καταχώρηση ανεπίσημης πληροφορίας που αφορούν τις σχέσεις, τις έννοιες και τις ετικέτες καθώς και την εξέλιξή τους στη διάρκεια του χρόνου. Η τυπολογία των στοιχείων τεκμηρίωσης καλύπτει διάφορα είδη σημειώσεων, όπως είναι ο σκοπός, η ιστορία, οι ορισμοί, η έκδοση, τα παραδείγματα κ.ά. (Pastor-Sanchez, Mendez & Rodríguez-Muñoz, 2009). Στον πίνακα 4.4 παρουσιάζεται το λεξιλόγιο SKOS για την τεκμηρίωση.

Πίνακας 4.4: Λεξιλόγιο SKOS για τεκμηρίωση

Λεξιλόγιο	Περιγραφή
skos:note	Γενική σημείωση
skos:scopeNote	Σημείωση σκοπού
skos:changeNote	Σημείωση αλλαγής
skos:definition	Ορισμός
skos:historyNote	Ιστορική σημείωση
skos:editorialNote	Σημείωση έκδοσης
skos:example	Παράδειγμα

4.4.4.5. Σχέσεις αντιστοίχισης

Το SKOS δίνει τη δυνατότητα της αντιστοίχισης ανάμεσα σε έννοιες που συμπεριλαμβάνονται σε διαφορετικά εννοιολογικά σχήματα. Στον πίνακα 4.5 παρουσιάζονται αυτές οι σχέσεις.

Πίνακας 4.5: Λεξιλόγιο SKOS για σχέσεις αντιστοίχισης

Λεξιλόγιο	Περιγραφή
skos:mappingRelation	Σχέση αντιστοίχισης
skos:exactMatch	Ακριβής αντιστοίχιση
skos:closeMatch	Κοντινή αντιστοίχιση
skos:broadMatch	Ευρύτερη αντιστοίχιση
skos:narrowMatch	Στενότερη αντιστοίχιση

4.4.5. Πρόσβαση στα δεδομένα

Οι βιβλιοθήκες διατηρούν τα RDF δεδομένα τους σε πληροφοριακά συστήματα βασισμένα στα Συνδεδεμένα Δεδομένα που ονομάζονται αποθετήρια τριπλετών. Τα αποθετήρια τριπλετών είναι εξειδικευμένα συστήματα διαχείρισης βάσεων δεδομένων

για την αποθήκευση και την ανάκτηση δεδομένων στη μορφή RDF⁷². Επί του παρόντος, υπάρχουν πολλά αποθετήρια τριπλετών, που εξυπηρετούν διαφορετικές ανάγκες και απαιτήσεις⁷³. Μερικές ενδεικτικές μηχανές για τη δημιουργία αποθετηρίων τριπλετών είναι η AllegroGraph⁷⁴, η Virtuoso Universal Server⁷⁵, η Garlik 4store⁷⁶ κλπ. Τα αποθετήρια τριπλετών διευκολύνουν την πρόσβαση στο περιεχόμενό τους μέσω της χρήσης των τελικών σημείων SPARQL. Η SPARQL είναι μια γλώσσα ερωτημάτων που είναι ικανή να ανακτά και να διαχειρίζεται δεδομένα διαθέσιμα σε RDF (Harris and Seaborne, 2013).

Το περιεχόμενο των αποθετηρίων τριπλετών που σχετίζεται με μια καθιερωμένη οντότητα μπορεί επίσης να εξυπηρετηθεί άμεσα όταν το αντίστοιχο URI της εγγραφής προσπελαστεί (π.χ. μέσω της πληκτρολόγησης του URI στην γραμμή διευθύνσεων ενός προγράμματος περιήγησης στο διαδίκτυο). Αυτός ο μηχανισμός προσπέλασης/πρόσβασης, που συχνά καλείται "*URI dereferencing*", παρέχει μια λεπτή μετάβαση από την καθιερωμένη οντότητα που προσδιορίζεται από το συγκεκριμένο URI, σε μια σελίδα που περιέχει την πληροφορία για τη συγκεκριμένη καθιερωμένη οντότητα.

Ενώ σήμερα, υπάρχει μικρή συναίνεση για το τι συνιστά χρήσιμη πληροφορία που σχετίζεται με μια καθιερωμένη οντότητα, ο μηχανισμός για την πρόσβαση στη συγκεκριμένη πληροφορία μέσω της χρήσης του πρωτοκόλλου HTTP είναι καλά εδραιωμένος. Η τεχνολογία που εμπλέκεται είναι προχωρημένη και όλα τα είδη των εφαρμογών μπορούν να επωφεληθούν από τη διαθεσιμότητα της πληροφορίας των βιβλιοθηκών ως Συνδεδεμένα Δεδομένα. Η πρόσβαση σε ένα URI των Συνδεδεμένων Δεδομένων περιλαμβάνει τα ακόλουθα βήματα:

- α) το URI μιας καθιερωμένης οντότητας είναι προσβάσιμο ως μια κανονική ηλεκτρονική διεύθυνση. Η αιτούσα εφαρμογή μπορεί να ζητήσει δεδομένα σε

⁷² Rusher, J. (2010). *Rhetorical Device: Triple Store*. [online] Διαθέσιμο στο: <http://rhetoricaldevice.com/articles/TripleStore.html> [Πρόσβαση: 09/07/2015].

⁷³ W3.org, (2015). *LargeTripleStores - W3C Wiki*. [online] Διαθέσιμο στο: <http://www.w3.org/wiki/LargeTripleStores> [Πρόσβαση: 09/07/2015].

⁷⁴ Composer, T., Suite, S. and Studio, T. (2015). *AllegroGraph RDFStore Web 3.0's Database*. [online]. Διαθέσιμο στο: <http://franz.com/agraph/allegrograph/> [Πρόσβαση: 09/07/2015].

⁷⁵ Virtuoso.openlinksw.com, (2015). *OpenLink Virtuoso Home Page*. [online] Διαθέσιμο στο: <http://virtuoso.openlinksw.com/> [Πρόσβαση: 09/07/2015].

⁷⁶ 4store.org, (2015). *4store - Scalable RDF Storage*. [online] Διαθέσιμο στο: <http://4store.org/> [Πρόσβαση: 09/07/2015].

μορφή RDF που μπορούν να προσπελαστούν από μηχανές ή μια απλή ιστοσελίδα, μέσω της επιλογής πρόσβασης «*διαπραγμάτευσης περιεχομένου*» (content negotiation).

β) στις περισσότερες περιπτώσεις, μια απάντηση ανακατεύθυνσης (redirection) ("*303 See Other*") στέλνεται πίσω, που παρέχει μια νέα τοποθεσία για το περιεχόμενο που έχει ζητηθεί. Η νέα διεύθυνση αναφέρεται σε ένα έγγραφο που περιέχει δεδομένα σε μορφή RDF ή δεδομένα σε μορφή κειμένου σχετικά με την καθιερωμένη οντότητα.

γ) η αιτούσα εφαρμογή (ή ο περιηγητής στο διαδίκτυο του χρήστη) αποκτά πρόσβαση με τη σειρά της στη νέα τοποθεσία που παρέχεται για να ανακτήσει τα δεδομένα της καθιερωμένης οντότητας.

Το δεύτερο (β) βήμα που αναφέρθηκε νωρίτερα μπορεί να φαίνεται περιττό, ωστόσο είναι ένας τυποποιημένος τρόπος για την παροχή πρόσβασης σε μεγάλες συλλογές Συνδεδεμένων Δεδομένων όπως αυτές που υπάρχουν στις βιβλιοθήκες. Από εννοιολογικής άποψης, το URI του πρώτου (α) βήματος είναι ένας προσδιοριστής μιας οντότητας (της ζητούμενης καθιερωμένης οντότητας), όχι μια διεύθυνση ενός εγγράφου του διαδικτύου. Το έγγραφο αυτό εκχωρείται μέσω της απάντησης ανακατεύθυνσης. Η ανακατεύθυνση παρέχει μια δεύτερη διεύθυνση στο διαδίκτυο για ένα έγγραφο που περιέχει την πληροφορία για το αρχικό URI.

Εκτός από τα τελικά σημεία SPARQL και το URI dereferencing, οι βιβλιοθήκες σε μορφή των Συνδεδεμένων Δεδομένων παρέχουν πρόσβαση στα δεδομένα τους και μαζικά μέσω αρχείων RDF dump.

4.5. Συμπεράσματα

Στις μέρες μας, στο διαδίκτυο και στα περισσότερα περιβάλλοντα των βιβλιοθηκών η ανακάλυψη πηγών λαμβάνει χώρα κυρίως μέσω της συλλογής και της ευρετηρίασης των περιεχομένων των σελίδων. Τέτοιες υπηρεσίες αναζήτησης και ανάκτησης παρέχουν πολύ αποτελεσματικούς τρόπους για να μπορέσουν οι χρήστες να εντοπίσουν υλικό που τους ενδιαφέρει μέσα σε μεγάλα αποθετήρια, όπως είναι το σύνολο του διαδικτύου ή οι αναπαραστάσεις του έντυπου και του ηλεκτρονικού υλικού που αποτελεί τη συλλογή μιας βιβλιοθήκης. Αλλά ακόμα και με τους πιο εξελιγμένους αλγόριθμους σχετικότητας, η αναζήτηση και η ανάκτηση υλικού βασισμένη σε ευρετήρια δεν έχει τη δυνατότητα να οδηγήσει τους χρήστες σε πιθανώς σχετικά τεκμήρια ή

περιεχόμενο. Αυτό το πρόβλημα έρχονται να λύσουν οι τεχνολογίες του σημασιολογικού ιστού, σε συνδυασμό με τα αποθετήρια των Συνδεδεμένων Δεδομένων, οι οποίες υπόσχονται νέες σημαντικές δυνατότητες για την εκμετάλλευση των πληροφοριακών πηγών που υπάρχουν στο διαδίκτυο. Πιο συγκεκριμένα, ο σημασιολογικός ιστός και τα Συνδεδεμένα Δεδομένα παρέχουν την τεχνολογία ώστε το σημερινό διαδίκτυο να προσπεράσει τους τρέχοντες περιορισμούς του σε σχέση με τη χειρωνακτική κωδικοποίηση των υπερσυνδέσμων σε ένα σύστημα που βασίζεται στην εξερεύνηση μονοπατιών σχετικών πηγών μέσω νοηματικών συσχετίσεων που κωδικοποιούνται σε αποθετήρια δεδομένων.

Τα βασικά δομικά στοιχεία των Συνδεδεμένων Δεδομένων περιλαμβάνουν την κωδικοποίηση του περιεχομένου με τη χρήση του RDF. Το RDF εκφράζει τις σχέσεις ανάμεσα στις πληροφοριακές μονάδες μέσω τριαδικών συσχετίσεων. Οι ιστοσελίδες στο σημασιολογικό ιστό βασίζονται στην HTML, η οποία παρέχει ένα μέσο με το οποίο οι χρήστες μπορούν να αλληλεπιδράσουν και να πλοηγηθούν. Τα έγγραφα που παρέχονται μέσω του σημασιολογικού ιστού επιτρέπουν στους υπολογιστές να αναλύσουν τις σχέσεις ανάμεσα στα αντικείμενα δεδομένων, σχηματίζοντας έτσι τη βάση για πιο ισχυρά εργαλεία για την ανακάλυψη, την οπτικοποίηση και την εκμετάλλευση της πληροφορίας. Τα Συνδεδεμένα Δεδομένα χρησιμοποιούν μια σειρά από αποθετήρια τριπλετών RDF για να καταστήσουν ένα σώμα δεδομένων κατανοητό μέσα στο πλαίσιο του σημασιολογικού ιστού.

Οι βιβλιοθήκες συμβαδίζουν με τις έννοιες του σημασιολογικού ιστού και των Συνδεδεμένων Δεδομένων, ειδικά σε σχέση με τη διάθεση των δεδομένων τους σε μορφή ανοιχτή και προσβάσιμη από όσο το δυνατόν περισσότερους χρήστες. Επίσης, μπορούν να καταστήσουν τα ελεγχόμενα λεξιλόγια που έχουν δημιουργηθεί από αυτές μέσα στα χρόνια για την οργάνωση των πληροφοριών πιο λειτουργικά. Άρα, η μετάβαση των βιβλιοθηκών στα νέα δεδομένα που δημιουργούνται με το κίνημα των Συνδεδεμένων Δεδομένων είναι ένα φυσικό επακόλουθο στην περαιτέρω εξέλιξη και ανάπτυξή τους.

Μάλιστα, τα τελευταία χρόνια έχουν πραγματοποιηθεί πολλές πρωτοβουλίες και έργα από διάφορες βιβλιοθήκες, τα οποία στοχεύουν στην εκμετάλλευση των τεχνολογιών του σημασιολογικού ιστού και έχουν συμβάλει στην αναδιαμόρφωση και διάθεση πηγών που τις ενδιαφέρουν σε μορφή συμβατή με τα Συνδεδεμένα Δεδομένα. Μέσα σε αυτό το πλαίσιο, στο επόμενο κεφάλαιο, παρουσιάζονται προσπάθειες που έχουν γίνει για την παροχή θεματικής πληροφορίας διαφόρων παρόχων σε μορφή

συμβατή με τα Συνδεδεμένα Δεδομένα. Παράλληλα, παρουσιάζεται και μια υπηρεσία πλοήγησης θεματικών επικεφαλίδων βασισμένη στις τεχνολογίες των Συνδεδεμένων Δεδομένων που έχει εγκατασταθεί στην ψηφιακή βιβλιοθήκη (Διώνη).

5. Υπηρεσία πλοήγησης θεματικών επικεφαλίδων βασισμένη στα Συνδεδεμένα Δεδομένα

5.1. Ελεγχόμενα λεξιλόγια ως Συνδεδεμένα Δεδομένα

Στις μέρες μας, ένας ολοένα αυξανόμενος αριθμός σηματολογικής πληροφορίας της πολιτιστικής κληρονομιάς παρέχεται ως Συνδεδεμένα Δεδομένα. Στη συνέχεια, παρουσιάζεται μια επισκόπηση της σχετικής βιβλιογραφίας.

Τον περασμένο αιώνα, οι LCSH κυριάρχησαν στον τομέα των βιβλιοθηκών ως το κυρίαρχο ελεγχόμενο λεξιλόγιο για την ανάθεση θεματικών περιγραφών στις πηγές (Stone, 2000). Ως εκ τούτου, το γεγονός ότι από το 2008, οι LCSH δημοσιεύονται επίσης ως Συνδεδεμένα Δεδομένα⁷⁷ (Summers et al., 2008) είναι ενδεικτικό για τις προθέσεις των βιβλιοθηκών σχετικά με την υιοθέτηση των Συνδεδεμένων Δεδομένων. Ακολουθώντας τα βήματα των LCSH, το Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié – RAMEAU, η γαλλική προσέγγιση στην καταλογογράφηση θεματικών επικεφαλίδων είναι επίσης διαθέσιμο ως Συνδεδεμένα Δεδομένα⁷⁸ (Meij, Isaac & Zinn, 2010). Ένας ακόμα βασικός παράγοντας της βιβλιοθηκονομικής κοινότητας που έκανε μια προσπάθεια να εκθέσει την πληροφορία του στο διαδίκτυο ως Συνδεδεμένα Δεδομένα είναι το σύστημα DDC⁷⁹ (Mitchell & Panzer, 2013).

Κατά παρόμοιο τρόπο, η υπηρεσία Finnish Ontology Library Service ONKI⁸⁰ είναι μια ακόμα περίπτωση εφαρμογής των Συνδεδεμένων Δεδομένων στις βιβλιοθήκες. Επιπλέον, το Semantic Computing – SeCo Research Group⁸¹ έχει πολλά ενεργά προγράμματα που σχετίζονται με την κοινότητα των Συνδεδεμένων Δεδομένων, όπως είναι το National Semantic Web Ontology Project (FinnONTO) (Hynönen et al., 2008).

⁷⁷ Id.loc.gov, (2015). *Home - LC Linked Data Service (Library of Congress)*. [online] Διαθέσιμο στο: <http://id.loc.gov> [Πρόσβαση: 09/07/2015].

⁷⁸ Cs.vu.nl, (2015). *RAMEAU subject headings as SKOS linked data*. [online] Διαθέσιμο στο: <http://www.cs.vu.nl/STITCH/rameau/> [Πρόσβαση: 09/07/2015].

⁷⁹ Oclc.org, (2015). *Dewey Web Services*. [online] Oclc.org. Διαθέσιμο στο: <https://www.oclc.org/dewey/webservices.en.html> [Πρόσβαση: 09/07/2015].

⁸⁰ Onki.fi, (2015). *ONKI3*. [online] Διαθέσιμο στο: <http://onki.fi/en/browser/> [Πρόσβαση: 09/07/2015].

⁸¹ Seco.tkk.fi, (2015). *Semantic Computing Research Group (SeCo)*. [online] Διαθέσιμο στο: <http://www.seco.tkk.fi/> [Πρόσβαση: 22/07/2015].

Η Εθνική Βιβλιοθήκη της Σουηδίας δημιούργησε το σύστημα LIBRIS⁸² (Söderbäck & Malmsten, 2008) ως μια προσπάθεια να ενσωματώσει τεχνολογίες Ανοιχτών Συνδεδεμένων Δεδομένων στο σουηδικό βιβλιοθηκονομικό τομέα. Μάλιστα, το 2008 όλος ο κατάλογος στο LIBRIS δημοσιεύτηκε ως Συνδεδεμένα Δεδομένα, καθιστώντας τον ως τον πρώτο ενοποιημένο κατάλογο που είναι εξ' ολοκλήρου σε μορφή συμβατή με τα Συνδεδεμένα Δεδομένα.

Αξίζει να τονιστεί εδώ ότι όλες οι προαναφερθείσες προσπάθειες για τη δημοσίευση Συνδεδεμένων Δεδομένων στον τομέα των βιβλιοθηκών έχει βασιστεί στο λεξιλόγιο SKOS (Miles & Pérez-Agüera, 2007). Τέλος, η σημασία του κινήματος των Συνδεδεμένων Δεδομένων για την κοινότητα της πολιτιστικής κληρονομιάς δεν θα μπορούσε να περάσει απαρατήρητη από την Europeana⁸³ (Purday, 2009). Κατά συνέπεια, το data.europeana.eu⁸⁴ αποτελεί μέρος της συνεχιζόμενης προσπάθειας της Europeana να διαθέσει τα δεδομένα της ως Συνδεδεμένα Δεδομένα.

Μια άλλη κατηγορία προσπαθειών για συμμετοχή στο κίνημα των Συνδεδεμένων Δεδομένων από τα ιδρύματα μνήμης σχετίζεται με τους θησαυρούς και τους αντίστοιχους θεματικούς περιγραφείς. Υπό αυτό το πρίσμα, ο STW Thesaurus for Economics⁸⁵ (Neubert, 2009), ο Agricultural Thesaurus and Glossary – ATG⁸⁶ και ο AGROVOC LOD⁸⁷ είναι ενδεικτικά παραδείγματα θησαυρών σε μορφή των Συνδεδεμένων Δεδομένων, οι οποίοι παρέχουν τα δεδομένα τους ως dereferenceable URIs.

Στη συνέχεια αυτής της ενότητας, θα παρουσιαστούν υπηρεσίες βιβλιοθηκών, οι οποίες όχι μόνο δημοσιεύουν τις πληροφορίες τους ως Συνδεδεμένα Δεδομένα, αλλά και καταναλώνουν Συνδεδεμένα Δεδομένα που παρέχονται από τρίτους.

⁸² Libris.kb.se, (2015). *LIBRIS*. [online] Διαθέσιμο στο: <http://libris.kb.se/> [Πρόσβαση: 09/07/2015].

⁸³ Europeana, (2015). *Europeana - Homepage*. [online] Διαθέσιμο στο: <http://www.europeana.eu/portal/> [Πρόσβαση: 05/10/2015].

⁸⁴ Europeana Labs, (2015). *Europeana Linked Open Data - Europeana Labs*. [online] Διαθέσιμο στο: <http://labs.europeana.eu/api/linked-open-data/introduction/> [Πρόσβαση: 09/07/2015].

⁸⁵ Zbw.eu, (2015). *STW Thesaurus for Economics: Home*. [online] Διαθέσιμο στο: <http://zbw.eu/stw> [Πρόσβαση: 09/07/2015].

⁸⁶ Agclass.nal.usda.gov, (2015). *Agricultural Thesaurus and Glossary Home Page*. [online] Διαθέσιμο στο: <http://agclass.nal.usda.gov/agt.shtml> [Πρόσβαση: 09/07/2015].

⁸⁷ Aims.fao.org, (1980). *AGROVOC Multilingual Agricultural Thesaurus | Agricultural Information Management Standards (AIMS)*. [online] Διαθέσιμο στο: <http://aims.fao.org/vest-reg-istry/vocabularies/agrovoc-multilingual-agricultural-thesaurus> [Πρόσβαση: 09/07/2015].

Η πρώτη υπηρεσία που παρουσιάζεται είναι ο Polythematic Structured Subject Heading – PSH⁸⁸ (Mynarz, Kozuchova & Kamradkova, 2009). Ο PSH είναι ένα δίγλωσσο (δηλ. τσεχικά-αγγλικά) εργαλείο για την οργάνωση και την αναζήτηση τεκμηρίων μέσω θεμάτων. Οι τελικοί χρήστες του PSH διαχειρίζονται ένα κομψό, διαδραστικό περιβάλλον διεπαφής χρήστη για την περιήγηση στις θεματικές επικεφαλίδες ενός αποθετηρίου τριπλετών εκμεταλλεόμενοι τη συνδυαστική τους δομή (Paradakis et al., 2009), παρέχοντας παράλληλα συνδέσμους για τη DBpedia ή/και τις LCSH.

Άλλη μια υπηρεσία των Συνδεδεμένων Δεδομένων που πηγαίνει ένα βήμα παραπέρα από την απλή δημοσίευση Συνδεδεμένων Δεδομένων στην κοινότητα της πολιτιστικής κληρονομιάς είναι το Europeana Connect Media Annotation Prototype (Haslhofer et al., 2010). Σύμφωνα με τους συγγραφείς (Haslhofer et al., 2010), οι τελικοί χρήστες μπορούν να αλληλοεπιδρούν με ένα μηχανισμό σχολιασμών (annotation mechanism) πάνω στην κορυφή των ψηφιακών τεκμηρίων της Europeana, ενώ σχόλια εκτίθενται ως Συνδεδεμένα Δεδομένα.

Ολοκληρώνοντας, είναι εμφανές ότι οι βασικοί παράγοντες της κοινότητας της πολιτιστικής κληρονομιάς έχουν συνειδητοποιήσει τη σημασία και τα οφέλη των Συνδεδεμένων Δεδομένων και έχουν αρχίσει να δημιουργούν υπηρεσίες σε Συνδεδεμένα Δεδομένα. Ωστόσο, ολοκληρωμένες εφαρμογές Συνδεδεμένων Δεδομένων που να είναι κατανοητές από τους τελικούς χρήστες είναι ακόμα δύσκολο να εντοπιστούν στο διαδίκτυο.

Στη συνέχεια, θα παρουσιαστεί η προσέγγιση της μετατροπής μιας υπηρεσίας θεματικής πλοήγησης σε μορφή των Συνδεδεμένων Δεδομένων, όπως παρουσιάστηκε στο Paradakis & Kyriianos (2012).

5.2. Προτεινόμενη προσέγγιση στη δημιουργία υπηρεσίας θεματικής πλοήγησης βασισμένη στα Συνδεδεμένα Δεδομένα

Η προτεινόμενη υπηρεσία θεματικής πλοήγησης σε μορφή Συνδεδεμένων Δεδομένων είναι ικανή να ενσωματώνει πηγές που προέρχονται από την ψηφιακή βιβλιοθήκη (Διώνη) και από τη βάση άρθρων των NYT⁸⁹. Η ψηφιακή βιβλιοθήκη (Διώνη) αναφέρεται σε μια εγκατάσταση του DSpace που περιέχει περίπου 4.380 μεταπτυχιακές και

⁸⁸ Psh.ntkcz.cz, (2015). *Prohlížení PSH*. [online] Διαθέσιμο στο: <http://psh.ntkcz.cz/skos/en> [Πρόσβαση: 09/07/2015].

⁸⁹ Nytimes.com, (2015). *Breaking News, World News & Multimedia*. [online] Διαθέσιμο στο: <http://www.nytimes.com/> [Πρόσβαση: 10/07/2015].

διδακτορικές διατριβές πάνω σε θέματα οικονομικών, επιχειρήσεων, διοίκησης, πληροφορικής, ναυτιλιακών σπουδών και τραπεζικών. Παράλληλα, περιέχει 3.323 δίγλωσσες (δηλ. ελληνικά-αγγλικά) θεματικές επικεφαλίδες που περιγράφουν τη συλλογή⁹⁰. Στα τεκμήρια της υποκείμενης ψηφιακής βιβλιοθήκης εκχωρούνται ειδικές θεματικές επικεφαλίδες που ολοκληρώνονται με θεματικές επικεφαλίδες από το λεξιλόγιο ευρετηρίασης των NYT μέσω της χρήσης ενός αλγόριθμου ευθυγράμμισης (alignment), ο οποίος θα παρουσιαστεί αργότερα σε αυτή την ενότητα, και αποθηκεύονται ανάλογα ως Συνδεδεμένα Δεδομένα σε ένα αντίστοιχο αποθετήριο τριπλετών. Η ηλεκτρονική βάση δεδομένων των NYT αποτελείται από άρθρα που χρονολογούνται από το 1981. Σχετικά πρόσφατα, οι NYT δημιούργησαν ένα ευρετηριασμένο λεξιλόγιο, το οποίο είναι διαθέσιμο στο ευρύ κοινό ως Συνδεδεμένα Δεδομένα⁹¹. Αυτό το λεξιλόγιο περιέχει ένα μεγάλο αριθμό (περίπου 10.000) θεματικών επικεφαλίδων που περιγράφει τη συλλογή άρθρων των NYT καθώς και το Αρχείο τους, που περιέχει μεγάλο αριθμό φωτογραφιών, γραφικών, ηχητικών ντοκουμέντων, βίντεο κ.ά.

Η υπηρεσία όχι μόνο παρέχει τις θεματικές επικεφαλίδες ως Συνδεδεμένα Δεδομένα, αλλά παρέχει στους τελικούς χρήστες επιπλέον πηγές προερχόμενες από ένα απομακρυσμένο αποθετήριο (δηλ. τη βάση δεδομένων των NYT) μέσω της χρήσης του εργαλείου NYT API tool⁹². Η αύξηση των πληροφοριών επιτυγχάνεται με διαφανή τρόπο για τον τελικό χρήστη.

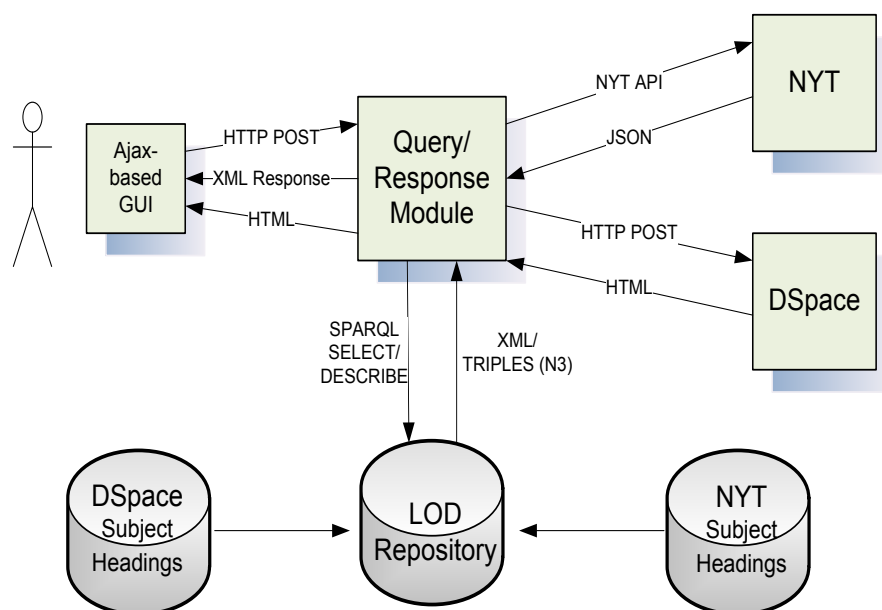
5.2.1. Αρχιτεκτονική

Όπως απεικονίζεται στην εικόνα 5.1, η προτεινόμενη υπηρεσία στηρίζεται σε ένα αποθετήριο Συνδεδεμένων Δεδομένων που αποτελείται από τις ευθυγραμμισμένες θεματικές επικεφαλίδες της υποκείμενης ψηφιακής βιβλιοθήκης που βασίζεται στο DSspace και στους NYT. Τα εν λόγω θέματα συνιστούν μια οντολογία, η οποία βασίζεται στις προδιαγραφές του SKOS.

⁹⁰ Οι εν λόγω αριθμοί αναφέρονται στον αριθμό των τεκμηρίων και των θεμάτων αντίστοιχα που περιείχε η ψηφιακή βιβλιοθήκη (Διώνη) κατά τη διάρκεια της εφαρμογής της προτεινόμενης μετατροπής της υπηρεσίας θεματικής πλοήγησης σε μορφή συμβατή με τα Συνδεδεμένα Δεδομένα (δηλ. Απρίλιος 2012).

⁹¹ Data.nytimes.com, (2015). *New York Times - Linked Open Data*. [online] Διαθέσιμο στο: <http://data.nytimes.com> [Πρόσβαση: 10/07/2015].

⁹² Developer.nytimes.com, (2015). *Times Developer Network - Welcome*. [online] Διαθέσιμο στο: <http://developer.nytimes.com/> [Πρόσβαση: 10/07/2015].



Εικόνα 5.1: Συνολική αρχιτεκτονική του συστήματος

Οι τελικοί χρήστες αλληλεπιδρούν με το αποθετήριο σε μορφή Συνδεδεμένων Δεδομένων μέσω ενός ταυτόχρονα δίγλωσσου γραφικού περιβάλλοντος βασισμένο στην τεχνολογία Asynchronous JavaScript and XML – AJAX⁹³. Το γραφικό περιβάλλον παρέχει τη δυνατότητα στους τελικούς χρήστες να ψάξουν για θέματα ή να πλοηγηθούν σε αυτά με απώτερο σκοπό να εντοπίσουν σχετικές μεταπτυχιακές και διδακτορικές διατριβές καθώς και άρθρα των NYT.

Η ροή της πληροφορίας ανάμεσα στα συστατικά της προτεινόμενης αρχιτεκτονικής εξυπηρετείται μέσω μιας αναλόγως σχεδιασμένης οντότητας ερωταποκρίσεων (Query/Response Module) που υλοποιήθηκε στη γλώσσα προγραμματισμού Python⁹⁴. Κατά συνέπεια, στην περίπτωση μιας αλληλεπίδρασης του χρήστη με το σύστημα, ένα αίτημα HTTP POST εκδίδεται και αντίστοιχα εξυπηρετείται από την οντότητα έτσι ώστε να παραχθούν κατάλληλες XML/HTML απαντήσεις, οι οποίες εν τέλει διαχειρίζονται από το γραφικό περιβάλλον.

Πιο συγκεκριμένα, για να ικανοποιήσει ένα ερώτημα από το γραφικό περιβάλλον, η οντότητα ερωταποκρίσεων απευθύνει ένα ερώτημα SPARQL στο αντίστοιχο τελικό σημείο του αποθετηρίου σε μορφή των Συνδεδεμένων Δεδομένων, το οποίο με τη

⁹³ W3schools.com, (2015). *AJAX Tutorial*. [online] Διαθέσιμο στο: <http://www.w3schools.com/ajax/default.asp> [Πρόσβαση: 03/10/2015].

⁹⁴ Python.org, (2015). *Welcome to Python.org*. [online] Διαθέσιμο στο: <https://www.python.org/> [Πρόσβαση: 05/10/2015].

σειρά του απαντάει με ένα SPARQL XML αποτέλεσμα. Όπως αναφέρθηκε προηγουμένως, το αποθετήριο σε μορφή των Συνδεδεμένων Δεδομένων αποτελείται από τις ευθυγραμμισμένες θεματικές επικεφαλίδες της ψηφιακής βιβλιοθήκης (Διώνη) και των NYT. Εάν η αντίστοιχη απάντηση περιέχει μια θεματική επικεφαλίδα που ανήκει στη ψηφιακή βιβλιοθήκη (Διώνη), τότε η οντότητα ερωταποκρίσεων δημιουργεί ένα άλλο ερώτημα, το οποίο απευθύνεται στην ψηφιακή βιβλιοθήκη (Διώνη) ως ένα HTTP POST ερώτημα. Εάν η αντίστοιχη απάντηση περιέχει μια θεματική επικεφαλίδα που ανήκει ταυτόχρονα και στους NYT, τότε η οντότητα ερωταποκρίσεων δημιουργεί ένα επιπλέον ερώτημα HTTP GET, το οποίο απευθύνεται στο εργαλείο API των NYT. Οι απαντήσεις από την ψηφιακή βιβλιοθήκη (Διώνη) και τους NYT παρέχονται ως απλή HTML και JSON αντίστοιχα.

5.2.2. Μοναδικά URIs στις υπηρεσίες θεματικής πλοήγησης

Όπως αναφέρεται από τους Bizer, Heath & Berners-Lee (2009, σ. 2), βασική προϋπόθεση για τη συμμετοχή στο κίνημα των Συνδεδεμένων Δεδομένων είναι η χρήση μοναδικών dereferenceable URIs. Επομένως, κάθε θεματική επικεφαλίδα της προτεινόμενης υπηρεσίας αντιστοιχεί σε ένα dereferenceable URI. Η πλειονότητα των θεματικών επικεφαλίδων που υπάρχουν στο αποθετήριο των Συνδεδεμένων Δεδομένων ανήκουν στο σύστημα των LCSH και προσδιορίζονται ανάλογα βάσει το χώρου ονοματοδοσίας των LCSH (π.χ. <http://id.loc.gov/authorities/subjects/sh94002414>) αντί να δημιουργηθεί ένα τοπικό URI. Μια τέτοια απόφαση πάρθηκε για να αποφευχθούν προβλήματα που μπορεί να δημιουργηθούν από τη χρήση του κατηγορήματος <owl:sameAs> (Halpin et al., 2010).

Επιπλέον, για τις ανάγκες της θεματικής ευρετηρίασης του υλικού της ψηφιακής βιβλιοθήκης (Διώνη), έχουν δημιουργηθεί επιπρόσθετες θεματικές επικεφαλίδες. Παρ' όλο που αυτές οι θεματικές επικεφαλίδες ακολουθούν τους κανόνες των LCSH, δεν υπάρχουν αυτούσιες στο λεξιλόγιο αυτό. Ως εκ τούτου, σε αυτές τις θεματικές επικεφαλίδες δόθηκαν νέα URIs. Αυτά τα URIs ανήκουν στο τοπικό χώρο ονοματοδοσίας της υπηρεσίας (δηλ. <http://id.lib.unipi.gr/authorities/subjects/>) και γίνονται dereference μέσω του τοπικού τελικού σημείου SPARQL.

5.2.3. Συνδεδεμένα Δεδομένα βασισμένα στο SKOS

Για τις ανάγκες της προτεινόμενης υπηρεσίας, αποφασίστηκε η υιοθέτηση του λεξιλογίου SKOS για τη δημιουργία του αποθετηρίου των Συνδεδεμένων Δεδομένων, το

οποίο αποτελείται από τις θεματικές επικεφαλίδες που ανήκουν στο χώρο ονοματοδοσίας των NYT, στο χώρο ονοματοδοσίας των LCSH και σε αυτές που ορίστηκαν τοπικά.

Πιο αναλυτικά, σύμφωνα με τις προδιαγραφές του SKOS, οι θεματικές επικεφαλίδες μοντελοποιούνται ως έννοιες (Isaac and Summers, 2009), και οργανώνονται ιεραρχικά σύμφωνα με τη συνδετική τους δομή (Papadakis et al., 2009), όπως αυτή ορίζεται από τις LCSH.

Ο πίνακας 5.1 συνοψίζει το λεξιλόγιο του SKOS που χρησιμοποιείται για να μοντελοποιηθούν οι σχέσεις ανάμεσα σε διαφορετικές έννοιες του υποκείμενου αποθετηρίου Συνδεδεμένων Δεδομένων:

Πίνακας 5.1: Χρήση λεξιλογίου SKOS

Θεματικές επικεφαλίδες	Κατηγορήμα SKOS
Broader Term	skos:broader
Narrower Term	skos:narrower
Related Term	skos:related
Use	skos:prefLabel
Use For	skos:altLabel
-	skos:exactMatch
-	skos:closeMatch

5.2.4. Διαδικασία ευθυγράμμισης

Για να ευθυγραμμιστούν οι θεματικές επικεφαλίδες που προέρχονται από την τοπική ψηφιακή βιβλιοθήκη και από τους NYT, αποφασίστηκε να χρησιμοποιηθεί η μεθοδολογία αντιστοίχισης οντολογιών που προτάθηκε στο Papadakis & Kyriianos (2011) και παρουσιάστηκε αναλυτικά στην ενότητα 3.7. Σύμφωνα με τη μεθοδολογία αυτή, δύο όροι θεωρούνται ισοδύναμοι εάν είναι ακριβώς ίδιοι. Σε αυτή την περίπτωση, η ορολογία του SKOS που χρησιμοποιείται για να δηλώσει τη σχέση αυτή είναι η <skos:exactMatch>.

Για παράδειγμα, η θεματική επικεφαλίδα "*Physics*" υπάρχει και στο χώρο ονομάτων των NYT και στο χώρο ονομάτων των LCSH. Επομένως, θα δημιουργηθεί η ακόλουθη τριπλέτα, σύμφωνα με τη σειριοποίηση N-Triples:

<<http://id.loc.gov/authorities/subjects/sh85101653>>

<<http://www.w3.org/2004/02/skos/core#exactMatch>>

<<http://data.nytimes.com/48662872284940183040>> .

Επιπλέον, υπάρχουν διάφορες περιπτώσεις όπου δύο θεματικές επικεφαλίδες θεωρούνται μερικώς ισοδύναμες:

- Μια θεματική επικεφαλίδα μπορεί να είναι στον πληθυντικό και η άλλη στον ενικό (π.χ. "*Bankruptcies*" στους NYT και "*Bankruptcy*" στη Διώνη)
- Παραλλαγή καταλήξεων: μερικοί όροι μπορεί να είναι διαφορετικοί με τη χρήση μερικών παραπάνω χαρακτήρων (π.χ. "*Iraq war (2003-)*" στους NYT και "*Iraq war, 2003-*" στη Διώνη)
- Μια θεματική επικεφαλίδα μπορεί να έχει διαφορετική σειρά λέξεων (π.χ. "*Colleges and universities*" στους NYT και "*Universities and colleges*" στη Διώνη)
- Μια θεματική επικεφαλίδα μπορεί να αντιστοιχεί σε ένα συνθετικό όρο και αντίστροφα (π.χ. "*Advertising and marketing*" στους NYT και "*Advertising*" και "*Marketing*" στη Διώνη).

Η ορολογία του SKOS που χρησιμοποιείται για να δηλώσει μια τέτοια σχέση είναι η <skos:closeMatch>. Ύστερα από την εφαρμογή του αλγόριθμου εντοπίστηκε ένας συνολικός αριθμός 207 κοινών θεματικών επικεφαλίδων ανάμεσα στους NYT και στις LCSH (δηλ. 111 "*closematch*" και 96 "*exactmatch*"). Τα κριτήρια που υποστηρίζουν τη μερική ισοδυναμία ανάμεσα στους NYT και τις LCSH βασίζονται στην κατά περίπτωση εφαρμογή εμπειρικών κανόνων που έχουν τεθεί από έμπειρους βιβλιοθηκονόμους. Η ποσοτικοποίηση των κριτηρίων αυτών έχει αποδειχθεί ότι είναι δύσκολο να επιβληθεί λόγω της έμφυτης ασάφειας των ομιλούμενων γλωσσών. Επομένως, ο προτεινόμενος αλγόριθμος βασίζεται μερικώς και στην ανθρώπινη παρέμβαση.

Στην επόμενη ενότητα, παρουσιάζεται το γραφικό περιβάλλον της προτεινόμενης υπηρεσίας.

5.2.5. Γραφικό περιβάλλον χρήσης

Οι τελικοί χρήστες αλληλεπιδρούν με την υπηρεσία των Συνδεδεμένων Δεδομένων μέσω του αντίστοιχα διαμορφωμένου γραφικού περιβάλλοντος. Το γραφικό περιβάλλον έχει δημιουργηθεί πάνω στην προηγούμενη του έκδοση, η οποία παρουσιάστηκε

στο Paradaakis et al. (2009) και αναλύεται διεξοδικά στην ενότητα 2.5.3. Πιο αναλυτικά, το γραφικό περιβάλλον αποτελείται από τα ακόλουθα συστατικά: α) το πεδίο αυτόματης συμπλήρωσης, β) το σύστημα πλοήγησης στις θεματικές επικεφαλίδες, και γ) τα αποτελέσματα από την ψηφιακή βιβλιοθήκη (Διώνη) και τους NYT.

α) Πεδίο αυτόματης συμπλήρωσης

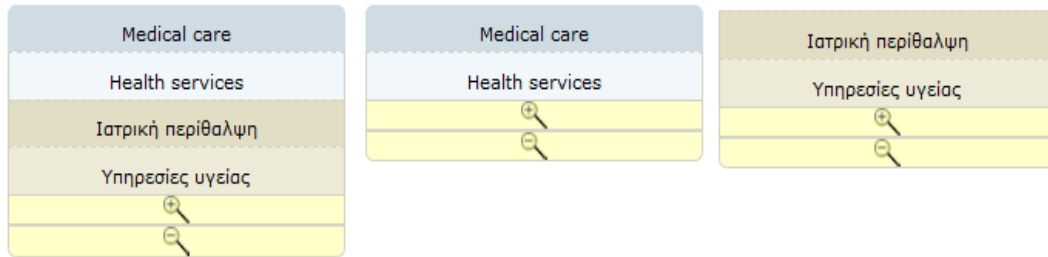
Οι τελικοί χρήστες αρχικά καλούνται να εκφράσουν την πληροφοριακή τους ανάγκη στα αγγλικά ή/και στα ελληνικά (εξαρτάται από την επιλογή γλώσσας που έχουν κάνει) στο πεδίο αυτόματης συμπλήρωσης (βλέπε εικόνα 5.2). Το πεδίο αυτό αντιδρά κατά την πληκτρολόγηση των χρηστών προτείνοντάς τους θεματικές επικεφαλίδες. Ένα τέτοιο συστατικό λειτουργεί ως ένα σημείο εισαγωγής στην υπηρεσία και βοηθά τους τελικούς χρήστες να εκφράσουν με προσυνδυασμένες λέξεις τις πληροφοριακές τους ανάγκες.



Εικόνα 5.2: Η δίγλωσση λειτουργικότητα του πεδίου αυτόματης συμπλήρωσης

β) Σύστημα πλοήγησης θεματικών επικεφαλίδων

Κατά την επιλογή μιας πρότασης, δημιουργείται ένα πλαίσιο που παρουσιάζει τη θεματική επικεφαλίδα που επιλέχθηκε κάτω από το πεδίο αυτόματης συμπλήρωσης (βλέπε εικόνα 5.3). Η δίγλωσση έκδοση του πλαισίου αποτελείται από την καθιερωμένη αγγλική ετικέτα της θεματικής επικεφαλίδας μαζί με την αντίστοιχη ελληνική μετάφραση και τους εναλλακτικούς όρους της επιλεγμένης θεματικής επικεφαλίδας στα αγγλικά και στα ελληνικά αντίστοιχα. Οι τελικοί χρήστες μπορούν να αποκλείσουν δυναμικά μία από τις δύο γλώσσες στο γραφικό περιβάλλον επιλέγοντας την κατάλληλη σημαία, όπως φαίνεται στην εικόνα 5.2.

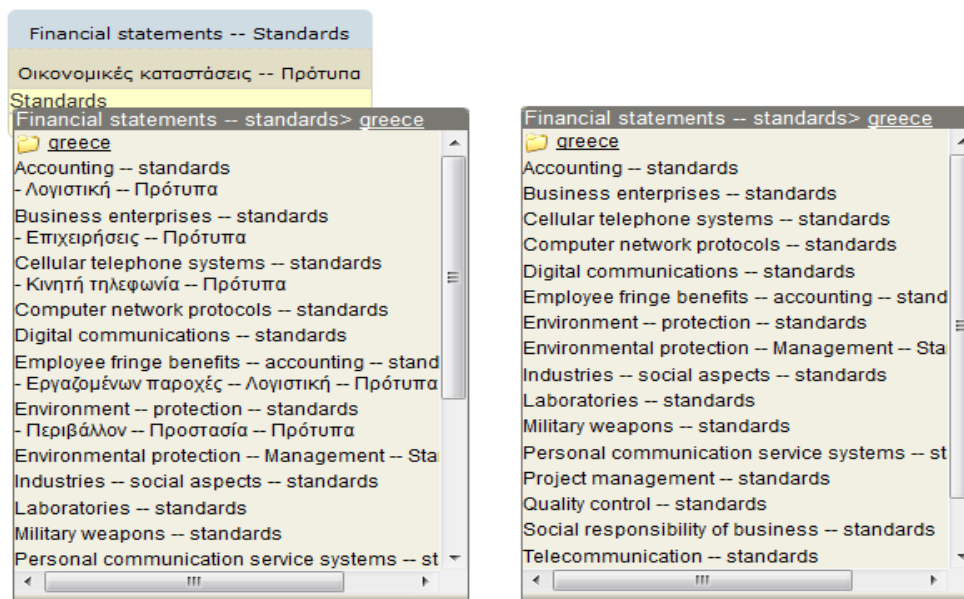


Εικόνα 5.3: Σχεδιασμένο πλαίσιο για το “Medical care” στις δύο γλώσσες, στα αγγλικά και στα ελληνικά

Κάτω από τις ετικέτες των όρων, υπάρχουν δύο μεταφορικά σύμβολα σε κίτρινο φόντο, τα οποία αντιστοιχούν στη λειτουργικότητα μεγέθυνσης ‘zoom in’ (+) και σμίκρυνσης ‘zoom out’ (-) της υπηρεσίας. Τα σήματα (+) και (-) ανακτούν τους στενότερους και ευρύτερους όρους της θεματικής επικεφαλίδας αντίστοιχα. Επομένως, κατά την επιλογή ενός εκ των δύο συμβόλων, ένα μενού περιεχομένων (context menu) δημιουργείται δίπλα στο σχετικό σήμα, το οποίο περιέχει τα αντίστοιχα θέματα.

Εάν ο τελικός χρήστης επιλέξει μια θεματική επικεφαλίδα που περιέχει και υποδιαίρεση, το πλαίσιο περιέχει εκτός από τα σήματα (+) και (-), και μια νέα σχέση που παίρνει το όνομά της από την υποδιαίρεση της θεματικής επικεφαλίδας που επιλέχθηκε προηγουμένως (δηλ. “Standards” στην εικόνα 5.4). Κατά την επιλογή της υποδιαίρεσης, το μενού περιεχομένων που σχηματίζεται περιέχει θεματικές επικεφαλίδες που μοιράζονται την ίδια υποδιαίρεση. Αυτή η λειτουργικότητα αναφέρεται στην εκτεταμένη συνδετική δομή, όπως περιεγράφηκε στην ενότητα 2.5.1.

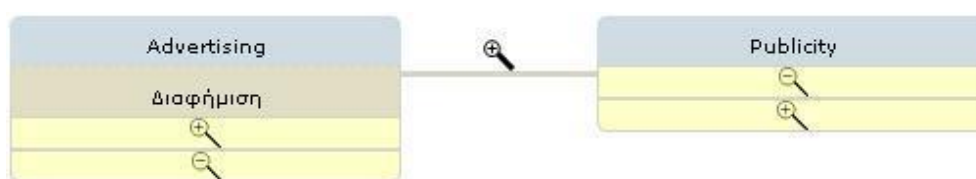
Επομένως, εάν υπάρχουν πολλές θεματικές επικεφαλίδες, οι οποίες μοιράζονται τις ίδιες υποδιαίρεσεις εκτός από αυτή που επιλέχθηκε στο προηγούμενο βήμα (π.χ. “Greece” στην εικόνα 5.4) στο πλαίσιο που σχεδιάστηκε, αυτές οι θεματικές επικεφαλίδες ομαδοποιούνται και παρουσιάζονται σε ξεχωριστό φάκελο, ο οποίος παίρνει το όνομά του από την κοινή υποδιαίρεση (βλέπε τη μπάρα πλοήγησης του μενού περιεχομένων στην εικόνα 5.4).



Εικόνα 5.4: Η θεματική επικεφαλίδα "Financial statements -- standards" και οι αντίστοιχες θεματικές επικεφαλίδες "standards" στις δύο γλώσσες και μόνο στα αγγλικά

Οι τελικοί χρήστες μπορούν να αποκτήσουν πρόσβαση στις θεματικές επικεφαλίδες που βρίσκονται μέσα στο φάκελο, επιλέγοντας αυτόν. Σε αυτή την περίπτωση, η μπάρα πλοήγησης πάνω από το μενού περιεχομένων δείχνει το μονοπάτι του χρήστη μέχρι την τρέχουσα τοποθεσία.

Όταν ο τελικός χρήστης επιλέξει μια θεματική επικεφαλίδα που βρίσκεται στο μενού περιεχομένων, ένα άλλο πλαίσιο δημιουργείται δίπλα στο πρώτο (βλέπε εικόνα 5.5). Τα δύο πλαίσια διασυνδέονται με μια γραμμή, η οποία περιέχει το σύμβολο που επιλέχθηκε κατά το προηγούμενο βήμα από τον τελικό χρήστη.

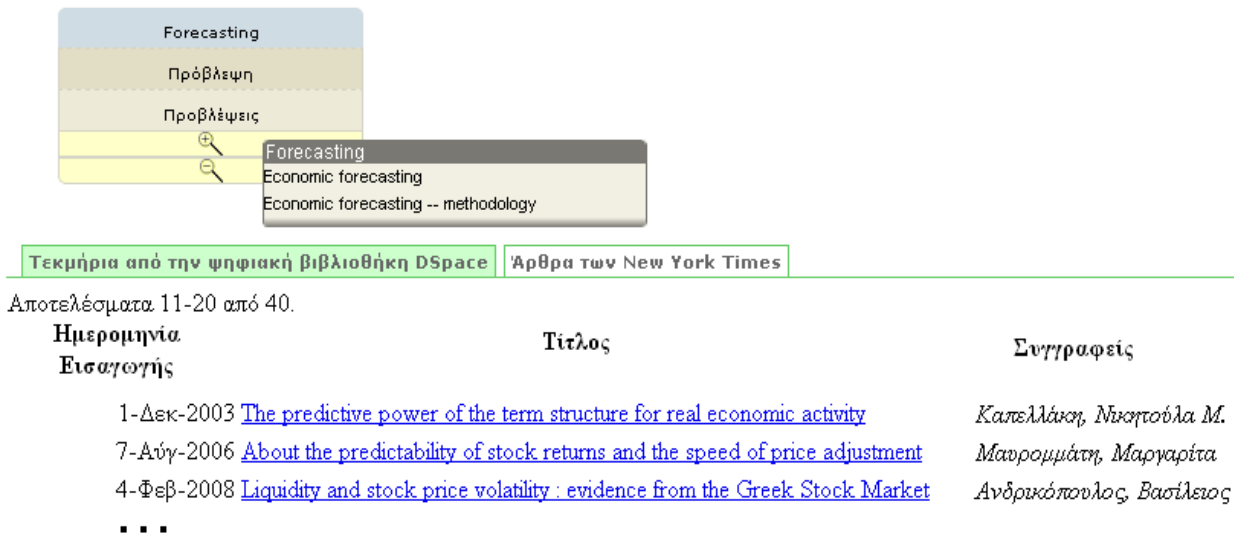


Εικόνα 5.5: Οι θεματικές επικεφαλίδες "Advertising" και "Publicity" σχετίζονται μέσω της σχέσης "NT"

γ) Αποτελέσματα αναζήτησης ψηφιακή βιβλιοθήκη (Διώνη) - NYT

Όταν επιλέγεται μια θεματική επικεφαλίδα από το μενού περιεχομένων, ένα ερώτημα, που περιέχει το συγκεκριμένο θέμα αποστέλλεται στην ψηφιακή βιβλιοθήκη και μια λίστα από τις αντίστοιχες πηγές παρουσιάζεται στους τελικούς χρήστες. Εάν η επιλεγμένη θεματική επικεφαλίδα σχετίζεται με μια ισοδύναμη ή μερικώς ισοδύναμη θεματική επικεφαλίδα των NYT, τότε ένα ακόμα ερώτημα αποστέλλεται στη βάση άρθρων των NYT και μια λίστα που περιέχει τα αντίστοιχα άρθρα παρουσιάζεται στους

τελικούς χρήστες ως μια νέα καρτέλα δίπλα στη λίστα αποτελεσμάτων της ψηφιακής βιβλιοθήκης (Διώνη) (βλέπε εικόνα 5.6). Τα άρθρα που εμφανίζονται κατατάσσονται ανάλογα με τη δημοτικότητα τους. Εν τέλει, οι τελικοί χρήστες μπορούν να κάνουν κλικ πάνω σε ένα αποτέλεσμα αναζήτησης έτσι ώστε να πλοηγηθούν στη συγκεκριμένη πηγή, ή να αναδιατυπώσουν το ερώτημά τους (χρησιμοποιώντας τη λειτουργικότητα του πλαισίου αυτόματης συμπλήρωσης ή επιλέγοντας μια άλλη θεματική επικεφαλίδα από τα σχεδιασμένα πλαίσια).



Εικόνα 5.6: Γραφικό περιβάλλον χρήστη με εμφάνιση και των σχετικών τεκμηρίων από την ψηφιακή βιβλιοθήκη (Διώνη)

Παράλληλα, θα πρέπει να σημειωθεί ότι οι χρήστες εκτός από το γραφικό περιβάλλον μπορούν να χρησιμοποιήσουν και το τελικό σημείο SPARQL της υπηρεσίας για την ανάκτηση αποτελεσμάτων. Πιο συγκεκριμένα, το αποθετήριο τριπλετών της υπηρεσίας δημιουργήθηκε με τη βοήθεια του λογισμικού 4store. Το 4store παρέχει ένα αποθετήριο τριπλετών και μια μηχανή ερωτημάτων και είναι σχεδιασμένο για να λειτουργεί σε συστήματα LINUX. Πιο συγκεκριμένα, το 4store περιλαμβάνει έναν SPARQL HTTP protocol server, ο οποίος μπορεί να απαντήσει σε SPARQL ερωτήματα χρησιμοποιώντας το πρωτόκολλο ερωτημάτων SPARQL HTTP. Το 4store είναι σχεδιασμένο για την αποθήκευση και ανάκτηση τριπλετών. Επομένως, όλες οι θεματικές επικεφαλίδες αναπαρίστανται ως τριπλέτες.

Στο τελικό σημείο SPARQL της υπηρεσίας (βλέπε εικόνα 5.7) ο χρήστης θέτοντας τα κατάλληλα ερωτήματα σε SPARQL μπορεί να πάρει τα αντίστοιχα αποτελέσματα σε μορφή τριπλετών.

SPARQL httpd server v1.1.4 test query

KB lcsh

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT * WHERE {
  ?s ?p ?o
} LIMIT 10
```

Soft limit

Εικόνα 5.7: Η αρχική οθόνη του τελικού σημείου SPARQL της υπηρεσίας

Για παράδειγμα, εάν ο χρήστης θέλει όλες τις θεματικές επικεφαλίδες που σχετίζονται με τη θεματική επικεφαλίδα "*Forecasting*" δεν έχει παρά να θέσει το ακόλουθο απλό ερώτημα στο τελικό σημείο SPARQL:

```
SELECT DISTINCT * WHERE {{
<http://id.loc.gov/authorities/subjects/sh85050485> ?p ?o .}
UNION {?s ?pre <http://id.loc.gov/authorities/subjects/sh85050485> .}}
```

Η υπηρεσία θα του δώσει την ακόλουθη απάντηση σε σειριοποίηση RDF/XML:

```
<?xml version="1.0"?>
<SPARQL xmlns="http://www.w3.org/2005/SPARQL-results#">
...
1. <results><result>
  <binding name="s">
    <uri>http://id.loc.gov/authorities/subjects/sh2009124595</uri></binding>
  <binding name="pre">
    <uri>http://www.w3.org/2004/02/skos/core#broader</uri></binding>
  </result>
2. <result>
  <binding name="s"> <uri>http://id.loc.gov/authorities/subjects/
sh85040814</uri></binding>
  <binding name="pre">
    <uri>http://www.w3.org/2004/02/skos/core#broader</uri></binding>
  </result>
3. <result>
```

```

    <binding name="p">
    <uri>http://www.w3.org/2004/02/skos/core#broader</uri></binding>
    <binding name="o"><uri>owl:Thing</uri></binding>
  </result>
4. <result>
    <binding name="p">
    <uri>http://www.w3.org/2004/02/skos/core#altLabel</uri></binding>
    <binding name="o"><literal xml:lang="GR">Προβλέψεις</literal></binding>
  </result>
5. <result>
    <binding name="p">
    <uri>http://www.w3.org/2004/02/skos/core#prefLabel</uri></binding>
    <binding name="o"><literal xml:lang="GR">Πρόβλεψη</literal></binding>
  </result>
6. <result>
    <binding name="p">
    <uri>http://www.w3.org/2004/02/skos/core#prefLabel</uri></binding>
    <binding name="o"><literal xml:lang="EN">Forecasting</literal></binding>
  </result></results></SPARQL>

```

Η επόμενη ενότητα αξιολογεί την προτεινόμενη υπηρεσία μέσω μιας αναλόγως σχεδιασμένης ανάλυσης έρευνας χρήστη. Η αξιολόγηση εστιάζει στη λειτουργικότητα και την αποτελεσματικότητα της προτεινόμενης υπηρεσίας. Αντιθέτως, η προστιθέμενη αξία του εμπλουτισμού με επιπλέον πληροφορία (δηλ. άρθρα των NYT), η οποία προέρχεται από την υπηρεσία δεν αξιολογείται. Όσον αφορά στα Συνδεδεμένα Δεδομένα, η συμβολή στην έρευνα έγκειται στη μεθοδολογία της ενοποίησης διαφορετικών συνόλων δεδομένων και όχι στην ποιότητα των συνόλων δεδομένων αυτή καθ' αυτή.

5.2.6. Αξιολόγηση

Ακολουθώντας τα βήματα της Borlund (2003), η οποία υποστήριξε την ιδέα ότι η αξιολόγηση ενός συστήματος Διαδραστικής Ανάκτησης Πληροφοριών (Interactive Information Retrieval – IIR system) εμπλέκει την προσομοίωση ενός σεναρίου εργασιών κατάστασης (simulated work task situation), καταλήξαμε ότι η διαδικασία της αξιολόγησης της προτεινόμενης υπηρεσίας θα πρέπει να αποτελείται από μια έρευνα

χρηστών που χρησιμοποιεί πραγματικά σενάρια εργασιών αναζήτησης και κατάλληλα σχεδιασμένα ερωτηματολόγια. Πιο αναλυτικά, οι στόχοι της συγκεκριμένης διαδικασίας αξιολόγησης ήταν τρεις: α) ο προσδιορισμός της αξίας από τη χρήση της προτεινόμενης υπηρεσίας ως ένας τρόπος εκμάθησης της συλλογής σε σχέση με τις συγκεκριμένες πληροφοριακές ανάγκες των χρηστών, β) η αξιολόγηση της ικανοποίησης των χρηστών σε σχέση με την προτεινόμενη υπηρεσία και γ) η αξιολόγηση της συνολικής απόδοσης και αποτελεσματικότητας.

Η έρευνα έλαβε χώρα στη βιβλιοθήκη του Πανεπιστημίου Πειραιά τον Ιούλιο του 2012. Δεκαεννιά άτομα συμμετείχαν στην έρευνα. Όλη η διαδικασία διήρκεσε περίπου 10 λεπτά για τον κάθε συμμετέχοντα και αποτελείτο από 4 φάσεις, οργανωμένες όπως περιγράφεται παρακάτω:

1. Ένα ερωτηματολόγιο πριν το σενάριο εργασιών⁹⁵, για να καταγραφούν η εμπειρία των συμμετεχόντων, το υπόβαθρό τους και τα δημογραφικά τους στοιχεία.
2. Εκπαίδευση στο σύστημα, κατά το οποίο οι συμμετέχοντες ήρθαν σε επαφή με τη λειτουργικότητα του συστήματος.
3. Τα σενάρια αναζήτησης. Ζητήθηκε από τους συμμετέχοντες να εκτελέσουν 4 σενάρια αναζήτησης αυξανόμενης δυσκολίας στην προτεινόμενη υπηρεσία. Τα δύο πρώτα ήταν στα αγγλικά και τα υπόλοιπα δύο στα ελληνικά. Τα σενάρια ήταν τα ακόλουθα: Προσπαθήστε να βρείτε σχετικές πληροφορίες για: α) "*Stress management*", β) "*Computer network protocols*", γ) «*Οικονομικά θέματα*» σχετικά με την «*Ελλάδα*» και την «*Ευρωπαϊκή Ένωση*» και δ) «*Ανάλυση παλινδρόμησης*».
4. Ένα ερωτηματολόγιο μετά το σενάριο εργασιών, για να καταγραφούν οι σκέψεις των συμμετεχόντων και η συνολική εντύπωσή τους για την προτεινόμενη υπηρεσία.

Οι επόμενες ενότητες παρουσιάζουν τα αποτελέσματα της διαδικασίας της αξιολόγησης.

⁹⁵ Τα ερωτηματολόγια που δόθηκαν πριν και μετά το σενάριο εργασιών καθώς και τα αποτελέσματα της έρευνας μπορούν να ανακτηθούν στο: <https://docs.google.com/open?id=0B-zgNdHCG27DSEY5R1dybGRVdkE>

5.2.6.1. Ανάλυση ερωτηματολογίου πριν το σενάριο εργασιών

Σύμφωνα με το ερωτηματολόγιο πριν το σενάριο εργασιών, 6 από τους 19 συμμετέχοντες ήταν προπτυχιακοί φοιτητές, 2 ήταν μεταπτυχιακοί φοιτητές, 2 ήταν υποψήφιοι διδάκτορες και 4 ήταν υπάλληλοι του πανεπιστημίου. Οι εναπομείναντες 5 συμμετέχοντες δεν είχαν κάποια άμεση σχέση με το Πανεπιστήμιο Πειραιά.

Όλοι οι συμμετέχοντες συμφώνησαν ότι είχαν μέτρια ως υψηλή εμπειρία στην αναζήτηση και περιήγηση πληροφορίας στο διαδίκτυο. Επιπλέον, 13 από τους συμμετέχοντες είχαν επισκεφθεί την ψηφιακή βιβλιοθήκη (Διώνη) τουλάχιστον μια φορά παλαιότερα. Τέλος, 18 από τους 19 συμμετέχοντες δήλωσαν ότι προτιμούν να χρησιμοποιούν τόσο την αγγλική όσο και την ελληνική γλώσσα όταν κάνουν αναζήτηση και περιήγηση πληροφορίας στο διαδίκτυο.

5.2.6.2. Ανάλυση ερωτηματολογίου μετά το σενάριο εργασιών

Σύμφωνα με τις απαντήσεις των συμμετεχόντων στο ερωτηματολόγιο μετά το σενάριο εργασιών, 90% βρήκαν την προτεινόμενη υπηρεσία πολύ εύκολη ή εύκολη στη χρήση. Επιπλέον, το 85% των συμμετεχόντων συμφώνησε ότι το πεδίο αυτόματης συμπλήρωσης τους βοήθησε για να ικανοποιήσουν τις πληροφοριακές τους ανάγκες. Ελαφρώς μικρότερο (δηλ. 68%) ήταν το ποσοστό των συμμετεχόντων που δήλωσαν ότι το συνολικό γραφικό περιβάλλον τους βοήθησε στην ικανοποίηση των πληροφοριακών τους αναγκών. Η μείωση στην ικανοποίηση των χρηστών αναφορικά με τη χρήση του πεδίου αυτόματης συμπλήρωσης σε σχέση με το συνολικό γραφικό περιβάλλον μπορεί να αποδοθεί στο γεγονός ότι στις μέρες μας, η λειτουργικότητα του πεδίου αυτόματης συμπλήρωσης έχει γίνει αναπόσπαστο κομμάτι πολλών ιστοσελίδων. Ως εκ τούτου, οι χρήστες του διαδικτύου είναι συνηθισμένοι σε μια τέτοια λειτουργικότητα. Από την άλλη μεριά, κάποιες μεταφορές που χρησιμοποιούνται από την προτεινόμενη υπηρεσία δεν είναι τόσο δημοφιλείς στο διαδίκτυο και, συνεπώς, θα μπορούσαν να αυξήσουν τη δυσκολία στη διαχείριση της υπηρεσίας, με αποτέλεσμα να επηρεαστεί η συνολική ικανοποίηση των χρηστών.

Επίσης, το 79% των συμμετεχόντων συμφώνησε ότι η προτεινόμενη υπηρεσία τους βοήθησε στην ολοκλήρωση των σεναρίων αναζήτησης γρήγορα.

Όσον αφορά στη διγλωσσία της προτεινόμενης υπηρεσίας, το 73% των συμμετεχόντων συμφώνησε ή συμφώνησε κατηγορηματικά ότι το ταυτόχρονα διγλωσσο γραφικό περιβάλλον (δηλ. αγγλικά-ελληνικά) τους βοήθησε στον εντοπισμό όλων των

σχετικών πληροφοριών της υποκείμενης συλλογής. Αυτό μπορεί να οφείλεται στο γεγονός ότι η προτεινόμενη υπηρεσία απευθύνει ταυτόχρονα ένα ερώτημα για την κάθε έκδοση του θέματος (δηλ. ελληνικά-αγγλικά, χρησιμοποίησε-χρησιμοποίησε για) και συγχωνεύει τα αντίστοιχα αποτελέσματα σε μια λίστα. Αντιθέτως, η παραδοσιακή αναζήτηση της ψηφιακής βιβλιοθήκης που βασίζεται στο DSpace απαιτεί διαφορετικά ερωτήματα για να φέρει όλα τα σχετικά αποτελέσματα (για περισσότερες λεπτομέρειες αναφορικά με την παραδοσιακή αναζήτηση βλέπε ενότητα 2.5.4.1). Για παράδειγμα, η αναζήτηση για το θέμα "*Regression analysis*" στην παραδοσιακή αναζήτηση αποδίδει 7 αποτελέσματα και η ελληνική του μετάφραση «*Ανάλυση παλινδρόμησης*» αποδίδει τα υπόλοιπα 10 από τα 17 αποτελέσματα, τα οποία θα είχαν ανακτηθεί σε περίπτωση που χρησιμοποιείτο η προτεινόμενη υπηρεσία για την αναζήτηση.

Όσον αφορά στην κατανόηση της προτεινόμενης υπηρεσίας, οι συμμετέχοντες κλήθηκαν να απαντήσουν στο ερώτημα εάν η εκπαίδευση ήταν απαραίτητη για να χρησιμοποιήσουν αποτελεσματικά την υπηρεσία. Περίπου οι μισοί απάντησαν ότι χρειαζόνταν την εκπαίδευση πριν χρησιμοποιήσουν την υπηρεσία. Το ποσοστό είναι ακόμα πιο υψηλό ανάμεσα σε ανθρώπους που δεν είχαν προηγούμενη εμπειρία με την αναζήτηση της ψηφιακής βιβλιοθήκης. Έχοντας, επίσης κατά νου τα παραπάνω μέτρια αποτελέσματα σχετικά με την ικανοποίηση των χρηστών από το παρεχόμενο γραφικό περιβάλλον, μπορεί με ασφάλεια να βγει το συμπέρασμα ότι το παρεχόμενο γραφικό περιβάλλον δεν είναι τόσο απλό και άμεσο όσο θα έπρεπε να ήταν.

Τελικά, για να εντοπίσουμε τη συνολική εντύπωση των συμμετεχόντων σχετικά με την προτεινόμενη υπηρεσία, τους ρωτήσαμε να μας πουν εάν θα ξαναχρησιμοποιούσαν την εφαρμογή για να ικανοποιήσουν τις πληροφοριακές τους ανάγκες. Σύμφωνα με τις απαντήσεις τους λοιπόν, το 74% από τους συμμετέχοντες συμφώνησε ότι θα χρησιμοποιούσε την προτεινόμενη υπηρεσία ξανά. Αυτό οδηγεί στο συμπέρασμα ότι η αποτελεσματικότητα της προτεινόμενης υπηρεσίας κυριάρχησε σε σχέση με τη μάλλον απότομη καμπύλη εκμάθησής της.

5.3. Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκε μια υπηρεσία θεματικής πλοήγησης βασισμένη στα Συνδεδεμένα Δεδομένα, ικανή να ενσωματώνει πηγές από δύο διαφορετικά αποθετήρια, δηλαδή την ψηφιακή βιβλιοθήκη (Διώνη) και τα άρθρα των NYT. Η ενσωμάτωση έγινε μέσω της ευθυγράμμισης των αντίστοιχων θεματικών επικεφαλίδων. Η

ευθυγράμμιση δεν θα ήταν δυνατή εάν οι θεματικές επικεφαλίδες κάθε αποθετηρίου δεν ήταν διαθέσιμες ως Συνδεδεμένα Δεδομένα.

Επιπλέον, η προτεινόμενη υπηρεσία χρησιμοποιεί ένα καινοτόμο σύστημα θεματικής πλοήγησης, το οποίο με μικρές τροποποιήσεις θα μπορούσε να εφαρμοστεί σε οποιοδήποτε αποθήκη δεδομένων που βασίζεται στο SKOS (Paradakis & Stefanidakis, 2008) (π.χ. άλλες ψηφιακές βιβλιοθήκες). Το βασικό χαρακτηριστικό ενός τέτοιου συστήματος είναι το γεγονός ότι καταφέρνει να αποκρύψει την ορολογία του συστήματος από τους τελικούς χρήστες.

Ένα ακόμα βασικό χαρακτηριστικό της προτεινόμενης υπηρεσίας είναι το ταυτόχρονα δίγλωσσο γραφικό περιβάλλον. Οι τελικοί χρήστες μπορούν να δουν τις πρωτότυπες θεματικές επικεφαλίδες, όπως έχουν οριστεί από τον αντίστοιχο αρχικό οργανισμό (σε αυτή την περίπτωση, τη Βιβλιοθήκη του Κογκρέσου) και τις μεταφράσεις τους ταυτόχρονα. Αυτό είναι ιδιαίτερα σημαντικό για ανθρώπους που μιλάνε διαφορετική γλώσσα σε σχέση με αυτή που έχει χρησιμοποιηθεί για να περιγραφούν τα τεκμήρια. Υπάρχουν περιπτώσεις όπου οι τελικοί χρήστες είναι εξοικειωμένοι με την πρωτότυπη έκδοση των θεματικών επικεφαλίδων και άλλες περιπτώσεις που οι τελικοί χρήστες είναι εξοικειωμένοι με τη μετάφραση αυτών. Το προτεινόμενο ταυτόχρονα δίγλωσσο γραφικό περιβάλλον αντιμετωπίζει και τις δύο περιπτώσεις ομοιόμορφα, χωρίς να απαιτείται ρητή αλληλεπίδραση από τον τελικό χρήστη.

Οι αρχές που διέπουν την προτεινόμενη υπηρεσία θα μπορούσαν να εφαρμοστούν και σε άλλες ψηφιακές βιβλιοθήκες της Ελλάδας με παρόμοιο περιεχόμενο με απώτερο στόχο να δημιουργηθεί μια ομόσπονδη ψηφιακή βιβλιοθήκη μεταπτυχιακών και διδακτορικών διατριβών αποτελούμενη από ψηφιακές βιβλιοθήκες μέλη που επικοινωνούν μεταξύ τους με τη χρήση τεχνολογιών των Συνδεδεμένων Δεδομένων.

Στη συνέχεια, θα παρουσιαστεί η δυνατότητα αντιστοίχισης ελεγχόμενων λεξιλογίων μέσω της σημασιολογικής ευθυγράμμισης όρων από τη χρήση γλωσσικών εργαλείων που βασίζονται στα Συνδεδεμένα Δεδομένα, με απώτερο σκοπό την εύρεση περισσότερων κοινών σημείων ανάμεσα σε δύο ελεγχόμενα λεξιλόγια.

6. Ευθυγράμμιση ελεγχόμενων λεξιλογίων μέσω της χρήσης υπηρεσιών σε Συνδεδεμένα Δεδομένα

6.1. Τεχνικές για την αντιστοίχιση ελεγχόμενων λεξιλογίων

Όπως έχει ήδη γίνει σαφές, το γεγονός ότι διαφορετικοί οργανισμοί συλλαμβάνουν τις ίδιες έννοιες με διαφορετικό τρόπο, έχει οδηγήσει στη δημιουργία παρόμοιων σημασιολογικά ελεγχόμενων λεξιλογίων που όμως είναι ετερογενή συντακτικά και γλωσσικά με αλληλοεπικαλυπτόμενα μέρη (Lin & Sandkuhl, 2008, σ. 341). Επομένως, η αναζήτηση και η ανάκτηση πληροφορίας ανάμεσα σε διαφορετικούς οργανισμούς έχει γίνει μια δύσκολη και κουραστική διαδικασία.

Σε μια προσπάθεια για να ξεπεραστούν αυτά τα θέματα και να έρθουν κοντά σημασιολογικά παρόμοια αλλά διαφορετικά ελεγχόμενα λεξιλόγια, διάφορα εργαλεία και μεθοδολογίες ταύτισης έχουν αναπτυχθεί (Euzenat & Shvaiko, 2007; Ivanova & Terziava, 2010). Τα εργαλεία αυτά επικεντρώνονται στην αναγνώριση σημασιολογικά όμοιων όρων που μπορεί να υπάρχουν ανάμεσα σε διαφορετικά λεξιλόγια. Οι πιο δημοφιλείς προσεγγίσεις στην επιδίωξη αυτής της αναζήτησης σημασιολογικά όμοιων όρων είναι: α) η λεξιλογική ομοιότητα (lexical similarity), η οποία συγκρίνει όρους σύμφωνα με τη σειρά των χαρακτήρων τους (Kwak & Yong, 2010, σ. 3) και β) η σημασιολογική ευθυγράμμιση (semantic alignment), η οποία βασίζεται σε σημασιολογικές τεχνικές για την αναγνώριση παρόμοιων όρων ανάμεσα σε δύο δομημένα λεξιλόγια.

Τα τελευταία χρόνια, ένας σημαντικός αριθμός συστημάτων και μεθοδολογιών έχει προταθεί⁹⁶ για την ευθυγράμμιση σημασιολογικά παρόμοιων ελεγχόμενων λεξιλογίων. Μέσα σε αυτό το πλαίσιο, η έρευνα που διενεργείται επικεντρώνεται ως επί το πλείστον στην εξεύρεση ομοιοτήτων σε επίπεδο στοιχείων (element-level).

6.2. Τεχνικές ταιριάσματος σε επίπεδο στοιχείων

Οι τεχνικές ταιριάσματος σε επίπεδο στοιχείων (element-level matching techniques) υπολογίζουν τις αντιστοιχίες με την ανάλυση των οντοτήτων (entities) ή των στιγμιότυπων των εν λόγω στοιχείων ξεχωριστά, αγνοώντας τις σχέσεις τους με άλλα

⁹⁶ Ontologymatching.org, (2015). *Ontology Matching*. [online] Διαθέσιμο στο: <http://ontologymatching.org/index.html> [Πρόσβαση: 10/07/2015].

στοιχεία ή τα στιγμιότυπά τους (Euzenat & Shvaiko, 2007). Όταν μιλάμε για ευθυγράμμιση θησαυρών ή/και θεματικών επικεφαλίδων, φαίνεται ότι η λεξιλογική ομοιότητα είναι η πιο κοινή προσέγγιση (Papadakis & Kyprianos, 2011, σ. 81). Η λεξιλογική προσέγγιση περιλαμβάνει έναν αριθμό τεχνικών, οι οποίες λαμβάνουν υπ' όψιν την απόσταση των χαρακτήρων (edit-distance) - δηλ. ο βαθμός ομοιότητας βασίζεται πάνω στον αριθμό των εισαγωγών, των διαγραφών και των αντικαταστάσεων που χρειάζονται για να μετατραπεί ο ένας όρος σε έναν άλλο - παραλλαγές προθεμάτων και επιθεμάτων (prefix and suffix variations), n-grams, κλπ. για να μετρήσουν τις ομοιότητες ανάμεσα στους αντίστοιχους όρους (Farooq, Arshad & Shah, 2010, σ. 70).

Η εμφάνιση του κινήματος των Συνδεδεμένων Δεδομένων (Bizer, Heath & Berners-Lee, 2009) υποκίνησε έναν αξιόλογο αριθμό παρόχων θησαυρών και θεματικών επικεφαλίδων (π.χ. LCSH, RAMEAU, BnF (Wenz et al., 2010), STW, GEMET, AGROVOC, κλπ.) να χρησιμοποιήσουν τις προαναφερθείσες τεχνικές για να εντοπίσουν ομοιότητες ανάμεσα στους όρους. Αυτές οι ομοιότητες εκφράζονται κοινώς μέσω της χρήσης του λεξιλογίου SKOS και γίνονται διαθέσιμοι στο κοινό μέσω της χρήσης τεχνολογιών των Συνδεδεμένων Δεδομένων.

Μέσα σε αυτό το πλαίσιο, το BLOOMS+ (Jain et al., 2011) χρησιμοποιεί σύνολα δεδομένων των Συνδεδεμένων Δεδομένων για να ευθυγραμμίσει τις οντολογίες που στηρίζουν τα αντίστοιχα ελεγχόμενα λεξιλόγια. Το BLOOMS+ χρησιμοποιεί τη Wikipedia⁹⁷ για την κατασκευή ενός συνόλου ιεραρχικών δέντρων κατηγοριών για κάθε κλάση για τις οντολογίες πηγή και στόχο. Κατόπιν, το BLOOMS+ καθορίζει ποιες κλάσεις ευθυγραμμίζει με τον υπολογισμό της ομοιότητας ανάμεσα στις κλάσεις της πηγής και του στόχου βασιζόμενο στα ιεραρχικά δέντρα κατηγοριών. Η ομοιότητα υπολογίζεται σύμφωνα με μια διαδοχική αρχιτεκτονική τριών ταιριασμάτων: ακριβές ταίριασμα (exact match), μερικό ταίριασμα (partial match) και σημασιολογική εκπομπή (semantic broadcast). Οι δύο πρώτες βασίζονται σε λεξιλογικές ομοιότητες, ενώ η τρίτη βασίζεται στη διάδοση ομοιοτήτων που έχουν υπολογιστεί από πριν (Pesquita et al., 2010).

Στην επόμενη ενότητα, προτείνεται μια τεχνική εξεύρεσης ομοιοτήτων σε επίπεδο στοιχείων, η οποία είναι ικανή να φέρει κοντά όρους από δυο σημασιολογικά παρόμοια ελεγχόμενα λεξιλόγια. Η προτεινόμενη μεθοδολογία προτείνει μια διαδικασία σημασιολογικής ευθυγράμμισης που βασίζεται σε γλωσσικά σύνολα δεδομένων σε μορφή των

⁹⁷ Wikipedia.org, (2015). *Wikipedia*. [online] Διαθέσιμο στο: <https://www.wikipedia.org/> [Πρόσβαση: 10/07/2015].

Συνδεδεμένων Δεδομένων, όπως είναι το WordNet⁹⁸ και η DBpedia⁹⁹. Σε αυτό το σημείο, θα πρέπει να τονιστεί ότι η προτεινόμενη προσέγγιση δεν απαιτεί καμία είδους δομή στα σύνολα δεδομένων που είναι υπό σύγκριση (π.χ. η ύπαρξη ενός σχήματος ή μιας οντολογίας).

6.3. Προτεινόμενη μεθοδολογία στη συγχώνευση ελεγχόμενων λεξιλογίων σε Συνδεδεμένα Δεδομένα

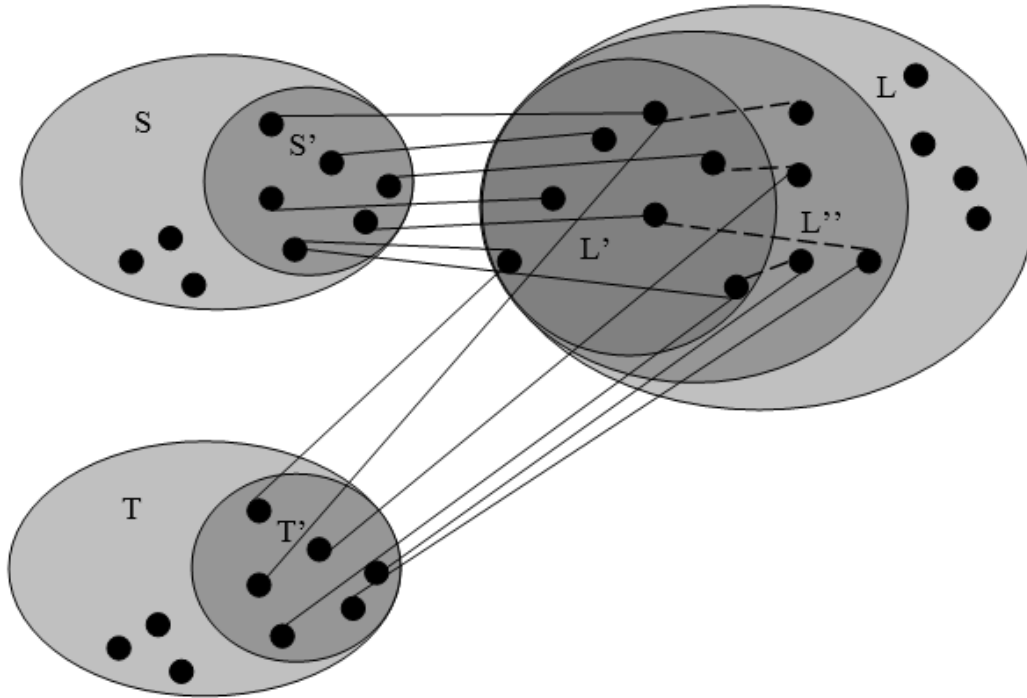
Σε αυτή την ενότητα, παρουσιάζεται μια μεθοδολογία που είναι ικανή να φέρνει κοντά σηματολογικά ισοδύναμους όρους από δύο ελεγχόμενα λεξιλόγια με τη χρήση εξωτερικών συνόλων δεδομένων σε μορφή Συνδεδεμένων Δεδομένων. Όπως αναφέρθηκε νωρίτερα, η προτεινόμενη προσέγγιση εκτελεί σηματολογική ευθυγράμμιση που δεν βασίζεται πάνω στην εγγενή δομή των αντίστοιχων λεξιλογίων. Αντίθετα, εκμεταλλεύεται εξωτερικά γλωσσικά εργαλεία που είναι διαθέσιμα στο σύννεφο των Συνδεδεμένων Δεδομένων. Αυτά τα σύνολα δεδομένων περιέχουν ρητά ορισμένες γλωσσικές συσχετίσεις (δηλ. συνώνυμα, υπώνυμα, μερώνυμα, ολώνυμα κλπ.) ανάμεσα στους όρους τους.

Δεδομένου ενός ελεγχόμενου λεξιλογίου S που λειτουργεί ως πηγή (Source) και ενός ελεγχόμενου λεξιλογίου T που λειτουργεί ως στόχος (Target), ο σκοπός της προσέγγισης είναι να ανακαλύψουμε ισοδύναμους όρους ανάμεσα στο S και το T . Για την επίτευξη αυτού του σκοπού, προτείνεται η ακόλουθη διαδικασία:

Αρχικά, κάθε όρος του S προσπαθεί να αντιστοιχηθεί με τους όρους από ένα εξωτερικό γλωσσικό σύνολο δεδομένων L (Linguistic dataset). Οι όροι του S που έχουν γλωσσικές ομοιότητες με ένα ή περισσότερους όρους από το L , διαμορφώνουν το σύνολο S' . Επομένως, και οι όροι του L που σχετίζονται με έναν ή περισσότερους όρους του S διαμορφώνουν το σύνολο L' . Κάθε όρος από το S' συνδέεται με τη σχέση "*isLexicalEquivalent*" με έναν ή περισσότερους όρους από το σύνολο L' (συνεχόμενη γραμμή στην εικόνα 6.1). Η σχέση "*isLexicalEquivalent*" ορίζεται ως συμμετρική και μεταβατική σχέση. Αυτό το αρχικό στάδιο χρησιμοποιεί λεξιλογικές μεθοδολογίες ταιριάσματος που ως επί το πλείστον συναντώνται στις τεχνικές εξεύρεσης ομοιοτήτων σε επίπεδο στοιχείων, όπως περιεγράφηκε στην ενότητα 6.2.

⁹⁸ Wordnet.princeton.edu, (2015). *About WordNet - WordNet*. [online] Διαθέσιμο στο: <https://wordnet.princeton.edu/> [Πρόσβαση: 10/07/2015].

⁹⁹ Wiki.dbpedia.org, (2015). *DBpedia*. [online] Διαθέσιμο στο: <http://wiki.dbpedia.org/> [Πρόσβαση: 10/07/2015].



Εικόνα 6.1: Προτεινόμενη μεθοδολογία

Κατόπιν, εξετάζεται εάν υπάρχουν όροι στο σύνολο L' και στο υπόλοιπο L που σχετίζονται μέσω μιας ρητά ορισμένης σχέσης ισοδυναμίας. Οι όροι του L που σχετίζονται με μερικούς ή με όλους τους όρους του L' , συγκροτούν το σύνολο L'' . Κάθε όρος από το L'' σχετίζεται με τη σχέση "*isSemanticEquivalent*" με ένα ή περισσότερους όρους από το σύνολο L' (διακεκομμένη γραμμή στην εικόνα 6.1). Αυτό το βήμα χρησιμοποιεί σημασιολογικές μεθοδολογίες ταιριάσματος που ως επί το πλείστον συναντώνται στα σύνολα δεδομένων σε μορφή Συνδεδεμένων Δεδομένων (περισσότερες λεπτομέρειες θα δοθούν στις επόμενες ενότητες).

Οι όροι που βρίσκονται στην ένωση των συνόλων L' και L'' αντιστοιχούνται με όρους από το T . Οι όροι του T που έχουν λεξιλογικές ομοιότητες με έναν ή περισσότερους όρους από τα σύνολα L' και L'' , συγκροτούν το σύνολο T' . Κάθε όρος του συνόλου T' σχετίζεται μέσω της σχέσης "*isLexicalEquivalent*" με ένα ή περισσότερους όρους από τα σύνολα L' και L'' .

Κατ' ουσίαν, "*isLexicalEquivalent*" και "*isSemanticEquivalent*" είναι εξειδικεύσεις της σχέσης "*isEquivalent*". Η πρώτη αναφέρεται στους όρους που βρίσκονται να είναι ισοδύναμοι μέσω τεχνικών λεξιλογικών ομοιοτήτων, ενώ η δεύτερη αναφέρεται σε όρους που βρίσκονται να είναι ισοδύναμοι μέσω της σημασιολογικής ευθυγράμμισης.

Επίσημοι ορισμοί

1. Το **S** είναι ένα σύνολο όρων στο σύνολο δεδομένων της πηγής (*Source*)
2. Το **T** είναι ένα σύνολο όρων στο σύνολο δεδομένων του στόχου (*Target*)
3. Το **L** είναι ένα σύνολο όρων στο γλωσσολογικό σύνολο δεδομένων (*Linguistic*)
4. Το **R** είναι η σχέση "*isEquivalent*" ανάμεσα στους όρους των δύο συνόλων δεδομένων
5. Το **R_L** είναι η σχέση "*isLexicalEquivalent*" ανάμεσα στους όρους των δύο συνόλων δεδομένων
6. Το **R_S** είναι η σχέση "*isSemanticEquivalent*" ανάμεσα στους όρους των δύο συνόλων δεδομένων
7. $xRy = \{(x, y) : xR_L y \vee xR_S y\}$
8. $S' = \{s \in S : \exists l \in L \wedge sR_L l\}$
9. $L' = \{l \in L : \exists s \in S \wedge lR_S s\}$
10. $L'' = \{l \in L : \exists x \in L' \wedge lR_S x\}$
11. $T' = \{t \in T : \exists x \in L' \wedge y \in L'' \wedge tR_L x \vee tR_L y\}$

Ο μεταβατικός χαρακτήρας της σχέσης "*isEquivalent*" υπονοεί ότι οι όροι στο σύνολο S' είναι ισοδύναμοι με τους όρους του συνόλου T' . Αυτή η διαδικασία στοχεύει στον εντοπισμό σημασιολογικά παρόμοιων όρων ανάμεσα στα λεξιλόγια S και T που δεν μπορούν δυνητικά να εντοπιστούν μέσω κοινών τεχνικών λεξιλογικής ομοιότητας.

Η επόμενη ενότητα καταδεικνύει την αποτελεσματικότητα της προτεινόμενης μεθοδολογίας μέσω της ανάπτυξης ενός σεναρίου στον πραγματικό κόσμο.

6.3.1. Εφαρμογή της μεθοδολογίας

Η προτεινόμενη μεθοδολογία εφαρμόζεται για να ανακαλύψουμε εμμέσως ισοδύναμους όρους ανάμεσα σε δύο ελεγχόμενα λεξιλόγια, δηλαδή τις θεματικές επικεφαλίδες της ψηφιακής βιβλιοθήκης (Διώνη) και των θεματικών επικεφαλίδων των άρθρων των NYT. Η συσχέτιση ανάμεσα στους όρους των λεξιλογίων πραγματοποιείται μέσω της χρήσης των συνόλων δεδομένων σε μορφή Συνδεδεμένων Δεδομένων της DBpedia και του WordNet.

Πιο συγκεκριμένα, η ψηφιακή βιβλιοθήκη (Διώνη) αποτελείται από 3.323 δίγλωσσες (ελληνικά-αγγλικά) θεματικές επικεφαλίδες που περιγράφουν τις πτυχιακές και διδακτορικές διατριβές σε διάφορα επιστημονικά πεδία (π.χ. οικονομικά, επιχειρήσεις, τραπεζικά, πληροφορική κ.ά.) μέσα σε μια εγκατάσταση του DSpace. Οι NYT παρέχουν

περίπου 10.000 θεματικές επικεφαλίδες¹⁰⁰, οι οποίες περιγράφουν τα υποκείμενα άρθρα. Και τα δύο λεξιλόγια δημοσιεύονται ως Συνδεδεμένα Δεδομένα.

Η DBpedia (Bizer et al., 2009) εξάγει δομημένη πληροφορία από τη Wikipedia και τη διαθέτει στο διαδίκτυο ως Συνδεδεμένα Δεδομένα. Η DBpedia περιέχει περισσότερες από 3.5 εκατομμύρια οντότητες¹⁰¹ και καλύπτει πολλούς τομείς. Κάθε οντότητα προσδιορίζεται από ένα μοναδικό URI, το οποίο είναι προσβάσιμο μέσω του τελικού σημείου SPARQL της DBpedia¹⁰². Η εγγενή φύση της DBpedia (δηλ. εγκυκλοπαίδεια) και η ευρεία αποδοχή της ως ένα γλωσσολογικό σύνολο δεδομένων, την καθιστά κατάλληλη για τις ανάγκες αυτής της έρευνας.

Το WordNet είναι μια μεγάλη λεξιλογική βάση δεδομένων της αγγλικής γλώσσας που διατίθεται ως Συνδεδεμένα Δεδομένα. Αποτελείται από ένα σύνολο synsets (π.χ. συνώνυμα). Κάθε synset υποδηλώνει μια ξεχωριστή έννοια και περιέχει όρους που αλληλοσυνδέονται μέσω της χρήσης εννοιολογικών-σημασιολογικών σχέσεων. Το WordNet μπορεί επίσης να προσπελαστεί μέσω του τελικού σημείου SPARQL¹⁰³ της υπηρεσίας.

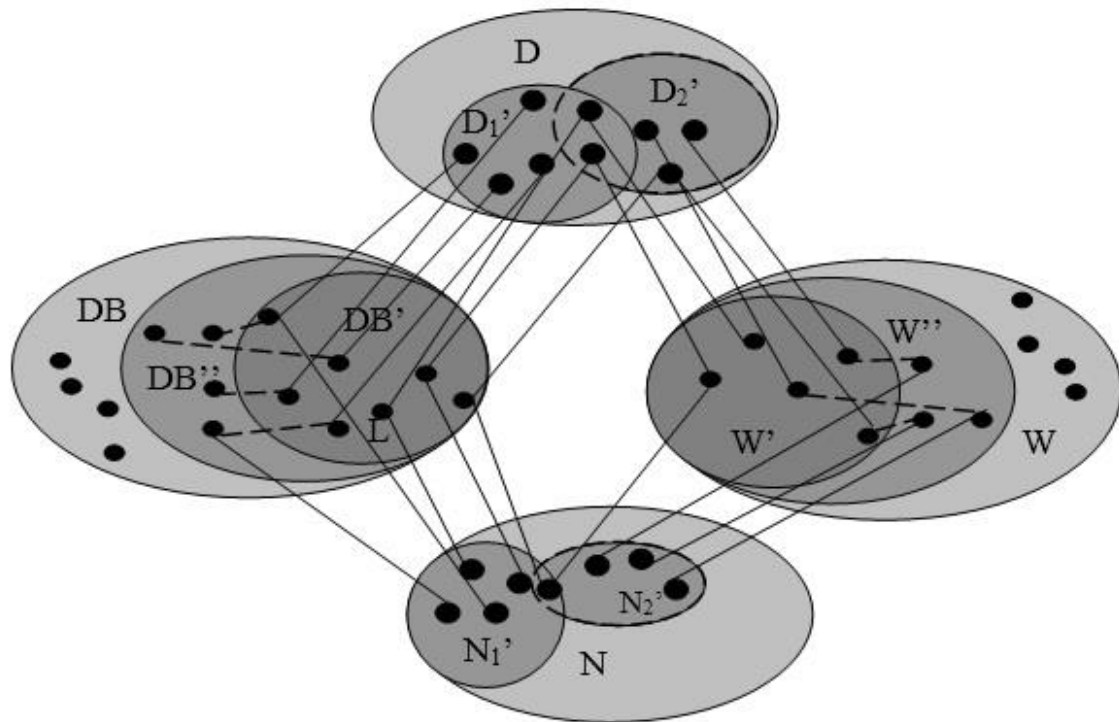
Όπως θα παρουσιαστεί αργότερα, και τα δύο σύνολα δεδομένων που αναφέρθηκαν νωρίτερα περιέχουν ισοδύναμες σχέσεις ανάμεσα στους όρους τους σε μορφή Συνδεδεμένων Δεδομένων (δηλ. "*wikiPageRedirects*" και "*containsWordsense*" για τη DBpedia και το WordNet αντίστοιχα) που είναι απαραίτητες για την ανάπτυξη της προτεινόμενης μεθοδολογίας.

¹⁰⁰ Οι εν λόγω αριθμοί αναφέρονται στον αριθμό των θεμάτων που περιείχε η ψηφιακή βιβλιοθήκη και οι NYT κατά τη διάρκεια της εφαρμογής της προτεινόμενης μεθοδολογίας (δηλ. Φεβρουάριος 2013).

¹⁰¹ Ο εν λόγω αριθμός αναφέρεται στον αριθμό των οντοτήτων της DBpedia κατά τη διάρκεια της εφαρμογής της προτεινόμενης μεθοδολογίας (δηλ. Φεβρουάριος 2013).

¹⁰² DBpedia.org, (2015). *Virtuoso SPARQL Query Editor*. [online] Διαθέσιμο στο: <http://dbpedia.org/sparql/> [Πρόσβαση: 10/07/2015].

¹⁰³ Wordnet.rkbexplorer.com, (2015). WordNet SPARQL Endpoint. [online] Διαθέσιμο στο: <http://wordnet.rkbexplorer.com/sparql/> [Πρόσβαση: 10/07/2015].



Εικόνα 6.2: Εφαρμογή της προτεινόμενης μεθοδολογίας

Η εικόνα 6.2 απεικονίζει την ανάπτυξη της προτεινόμενης μεθοδολογίας που παρουσιάστηκε στην εικόνα 6.1, η οποία οδηγεί στον προσδιορισμό των εμμέσως ισοδύναμων θεματικών επικεφαλίδων ανάμεσα στην ψηφιακή βιβλιοθήκη (Διώνη) και τους NYT με τη χρήση της DBpedia και του WordNet. Ακολουθώντας την επίσημη τυποποίηση της ενότητας 6.3, ένας αριθμός αντιστοιχίσεων ορίζεται:

1. Έστω ότι το σύνολο δεδομένων της πηγής S είναι το \mathbf{D} (δηλ. Διώνη)
2. Έστω ότι το σύνολο δεδομένων του στόχου T είναι το \mathbf{N} (δηλ. NYT)
3. Έστω ότι το γλωσσολογικό σύνολο δεδομένων A L να είναι \mathbf{DB} (δηλ. DBpedia) και,
4. Έστω ότι το γλωσσολογικό σύνολο δεδομένων B L να είναι \mathbf{W} (δηλ. WordNet)
5. Το \mathbf{D}_1' αντιστοιχεί στο S' , υποθέτοντας ότι το γλωσσολογικό σύνολο δεδομένων L είναι το \mathbf{DB} . Κατά παρόμοιο τρόπο, το \mathbf{D}_2' αντιστοιχεί στο S' , υποθέτοντας ότι το γλωσσολογικό σύνολο δεδομένων L είναι το \mathbf{W} .
6. Το \mathbf{DB}' και το \mathbf{DB}'' αντιστοιχούν στο L' και L'' αντίστοιχα, υποθέτοντας ότι το γλωσσολογικό σύνολο δεδομένων L είναι το \mathbf{DB} . Κατά παρόμοιο τρόπο, το \mathbf{W}' και το \mathbf{W}'' αντιστοιχεί στο L' και L'' αντίστοιχα, υποθέτοντας ότι το γλωσσολογικό σύνολο δεδομένων L είναι το \mathbf{W} .

7. Το \mathbf{N}_1' αντιστοιχεί στο T' , υποθέτοντας ότι το γλωσσολογικό σύνολο δεδομένων L είναι το DB . Κατά παρόμοιο τρόπο, το \mathbf{N}_2' αντιστοιχεί στο T' υποθέτοντας ότι το γλωσσολογικό σύνολο δεδομένων L είναι το W .

Στην επόμενη ενότητα, παρουσιάζονται οι λεπτομέρειες υλοποίησης της προαναφερθείσας μεθοδολογίας.

6.3.2. Υλοποίηση της μεθοδολογίας

Για να εντοπιστούν οι λεξιλογικές καθώς και οι σημασιολογικές ομοιότητες ανάμεσα σε όρους που ανήκουν σε διαφορετικά σύνολα δεδομένων χρησιμοποιήθηκε το Google Refine¹⁰⁴ (πλέον το Google Refine έχει αντικατασταθεί από το Open Refine¹⁰⁵, αλλά η λειτουργικότητά του παραμένει ακριβώς ίδια). Το Google Refine είναι ένα εργαλείο που έχει σχεδιαστεί για να διαχειρίζεται δεδομένα σε μορφή πίνακα (Maali, Cyganiak & Peristeras, 2011). Τα δεδομένα μπορούν να καθαριστούν, να σχηματιστούν από μία μορφή σε μια άλλη, να επεκταθούν με διαδικτυακές υπηρεσίες και να συνδεθούν με άλλες βάσεις δεδομένων. Παράλληλα, παρέχει ένα ισχυρό εργαλείο για τη «συμφιλίωση» (*"reconciliation"*) των δεδομένων με υπάρχουσες βάσεις γνώσης, δημιουργώντας με αυτόν τον τρόπο μία σύνδεση με την κοινότητα των Συνδεδεμένων Δεδομένων. Αυτό επιτυγχάνεται μέσω της χρήσης της επέκτασης (extension)¹⁰⁶ του RDF, η οποία δίνει τη δυνατότητα στους τελικούς χρήστες να συμφιλιώσουν τα δεδομένα τους μέσω της χρήσης τελικών σημείων SPARQL και αρχείων RDF dump.

Σε αυτό το σημείο θα πρέπει να τονιστεί ότι, όπως έχει ήδη αναφερθεί σε προηγούμενες ενότητες, τα Συνδεδεμένα Δεδομένα αναφέρονται σε ένα σύνολο αρχών και τεχνικών για τη δημοσίευση δομημένων δεδομένων στο διαδίκτυο (Heath, Bizer & Berners-Lee, 2009, σ.1). Ο τελικός στόχος του κινήματος των Συνδεδεμένων Δεδομένων είναι να καταστήσει τα δεδομένα ιδιαίτερα προσβάσιμα και χρήσιμα για τα ενδιαφερόμενα μέρη. Η ραχοκοκαλιά των τεχνολογιών των Συνδεδεμένων Δεδομένων είναι το λεξιλόγιο RDF, που αντιστοιχεί σε μια αναπαράσταση γράφου των οντοτήτων (δηλ. URIs και λεκτικά), τα οποία αποτελούν τέτοιου είδους δεδομένα. Τα γραφήματα RDF

¹⁰⁴ Code.google.com, (2015). *Google Refine, a power tool for working with messy data (formerly Freebase Gridworks) - Google Project Hosting*. [online] Διαθέσιμο στο: <https://code.google.com/p/google-refine/> [Πρόσβαση: 11/07/2015].

¹⁰⁵ Openrefine.org, (2015). *OpenRefine*. [online] Διαθέσιμο στο: <http://openrefine.org/> [Πρόσβαση: 11/07/2015].

¹⁰⁶ Refine.deri.ie, (2015). *GRefine RDF Extension*. [online] Διαθέσιμο στο: <http://refine.deri.ie/> [Πρόσβαση: 11/07/2015].

αποθηκεύονται ευρέως σε αποθετήρια τριπλετών των Συνδεδεμένων Δεδομένων (προσβάσιμα μέσω του σύννεφου των Συνδεδεμένων Δεδομένων). Η κυρίαρχη γλώσσα ερωτημάτων στα αποθετήρια τριπλετών των Συνδεδεμένων Δεδομένων είναι η SPARQL.

Πιο συγκεκριμένα, η υλοποίηση της προτεινόμενης μεθοδολογίας βασίζεται πάνω στα ακόλουθα έξι βήματα:

1. Οι θεματικές επικεφαλίδες της ψηφιακής βιβλιοθήκης (Διώνη) εισάγονται στο Google Refine.
2. Τα τελικά σημεία SPARQL της DBpedia και του WordNet δηλώνονται στο Google Refine ως υπηρεσίες συμφιλίωσης SPARQL.
3. Οι θεματικές επικεφαλίδες της ψηφιακής βιβλιοθήκης (Διώνη) αντιπαραβάλλονται με τις υπηρεσίες συμφιλίωσης της DBpedia και του WordNet. Η πρώτη διαδικασία αντιπαραβολής (δηλ. η αντιστοίχιση των συνόλων δεδομένων της ψηφιακής βιβλιοθήκης (Διώνη) και της DBpedia) καταλήγει σε ένα υποσύνολο της DBpedia, ενώ η δεύτερη διαδικασία αντιπαραβολής (δηλ. η αντιστοίχιση των συνόλων δεδομένων της ψηφιακής βιβλιοθήκης (Διώνη) και του WordNet) καταλήγει σε ένα υποσύνολο του WordNet.
4. Οι όροι των συνόλων του προηγούμενου βήματος (δηλ. τα υποσύνολα της DBpedia και του WordNet) εμπλουτίζονται με σημασιολογικά ισοδύναμους όρους που προέρχονται από τους υπόλοιπους όρους της DBpedia και του WordNet.
5. Οι θεματικές επικεφαλίδες από τους NYT εισάγονται στο Google Refine.
6. Οι θεματικές επικεφαλίδες των NYT αντιπαραβάλλονται με τους όρους που προκύπτουν από τα βήματα 3 και 4.

Στη συνέχεια, παρέχεται μια λεπτομερής περιγραφή κάθε βήματος:

Βήμα 1: Η λίστα των θεματικών επικεφαλίδων από την ψηφιακή βιβλιοθήκη (Διώνη) εισάγεται στο Google Refine ως λογιστικό φύλλο μιας στήλης. Κατά συνέπεια, το σύνολο D , όπως ορίστηκε στην ενότητα 6.3.1, περιέχει 3.323 όρους.

Βήμα 2: Η λειτουργικότητα του Google Refine μπορεί να επεκταθεί για τη διαχείριση δεδομένων RDF μέσω της επέκτασης RDF. Μια τέτοια επέκταση επιτρέπει τη δήλωση των τελικών σημείων SPARQL της DBpedia και του WordNet ως υπηρεσίες «συμφιλίωσης» (*"reconciliation services"*).

Βήμα 3: Το Google Refine συστήνει την υπηρεσία «*συμφιλίωσης*» ως μια υπηρεσία που είναι ικανή να εντοπίζει λεξιλογικές ομοιότητες ανάμεσα σε όρους που ανήκουν σε διαφορετικά σύνολα δεδομένων. Στην περίπτωση τυποποιημένων όρων, το Google Refine ξεκινά μια διαδραστική διαδικασία που επιτρέπει τον περιορισμό των συγκρίσεων σε συγκεκριμένο είδος/τύπο όρων. Μια τέτοια διαδικασία μειώνει δραματικά τη συνολική διάρκεια της διαδικασίας. Αφού οι θεματικές επικεφαλίδες της ψηφιακής βιβλιοθήκης (Διώνη) εισάγονται ως λογιστικό φύλλο μιας στήλης (δηλ. μη τυποποιημένοι όροι), το Google Refine θεωρεί ότι ανήκουν σε ένα ενιαίο ανώνυμο τύπο. Αντιθέτως, η DBpedia και το WordNet περιέχουν τυποποιημένους όρους. Επομένως, η διαδικασία συμφιλίωσης ανάμεσα στους όρους της ψηφιακής βιβλιοθήκης (Διώνη) και των DBpedia/WordNet περιορίζεται στον τύπο "*skos:Concept*".

Μετά την ολοκλήρωση της διαδικασίας συμφιλίωσης, ο τελικός χρήστης μπορεί να ελέγξει τις προτάσεις που παρέχονται από το Google Refine, και αντίστοιχα να δεχτεί ή να απορρίψει τις προτεινόμενες αντιστοιχίσεις. Στην προτεινόμενη προσέγγιση, ο τελικός χρήστης είναι ένας καλά εκπαιδευμένος θεματικός βιβλιοθηκονόμος, ο οποίος εκτελεί, επίσης, μια χειροκίνητη ποιοτική αξιολόγηση των προτάσεων. Αυτή η διαδικασία καταλήγει στην ανακάλυψη 1.574, 1-1 λεξιλογικά ισοδύναμων όρων ανάμεσα στην ψηφιακή βιβλιοθήκη (Διώνη) και στα DBpedia/WordNet. Πιο συγκεκριμένα, το DB' και το D₁' αποτελούνται το καθένα από 1.119 όρους και το W' και το D₂' αποτελούνται το καθένα από 455 όρους (βλέπε εικόνα 6.3).

Το ποσοστό επιτυχίας της διαδικασίας συμφιλίωσης εξαρτάται από τον αριθμό των λέξεων που αποτελούν μια θεματική επικεφαλίδα. Πράγματι, όσο ο αριθμός των λέξεων αυξάνει, τόσο το ποσοστό επιτυχίας της διαδικασίας «*συμφιλίωσης*» μειώνεται. Μια θεματική επικεφαλίδα που αποτελείται από μια ή δύο λέξεις έχει το μεγαλύτερο ποσοστό επιτυχίας. Το ποσοστό επιτυχίας της διαδικασίας συμφιλίωσης πέφτει δραματικά όταν η θεματική επικεφαλίδα αποτελείται από περισσότερες από τρεις λέξεις ή/και υποδιαιρέσεις. Αυτό οφείλεται στο γεγονός ότι κανένα από τα δύο λεξιλόγια (δηλ. DBpedia και WordNet) δεν περιέχουν όρους με υποδιαιρέσεις.

Πίνακας 6.1: Ποσοστό επιτυχίας της διαδικασίας συμφιλίωσης

Διώνη	DBpedia	WordNet
Θεματική επικεφαλίδα μιας λέξης	331 (29%)	297 (65%)

Θεματική επικεφαλίδα δύο λέξεων	658 (59%)	128 (28%)
Θεματική επικεφαλίδα τριών και περισσότερων λέξεων	130 (12%)	30 (7%)
Θεματική επικεφαλίδα με υποδιαίρεσεις	0	0
Σύνολο	1,119	455

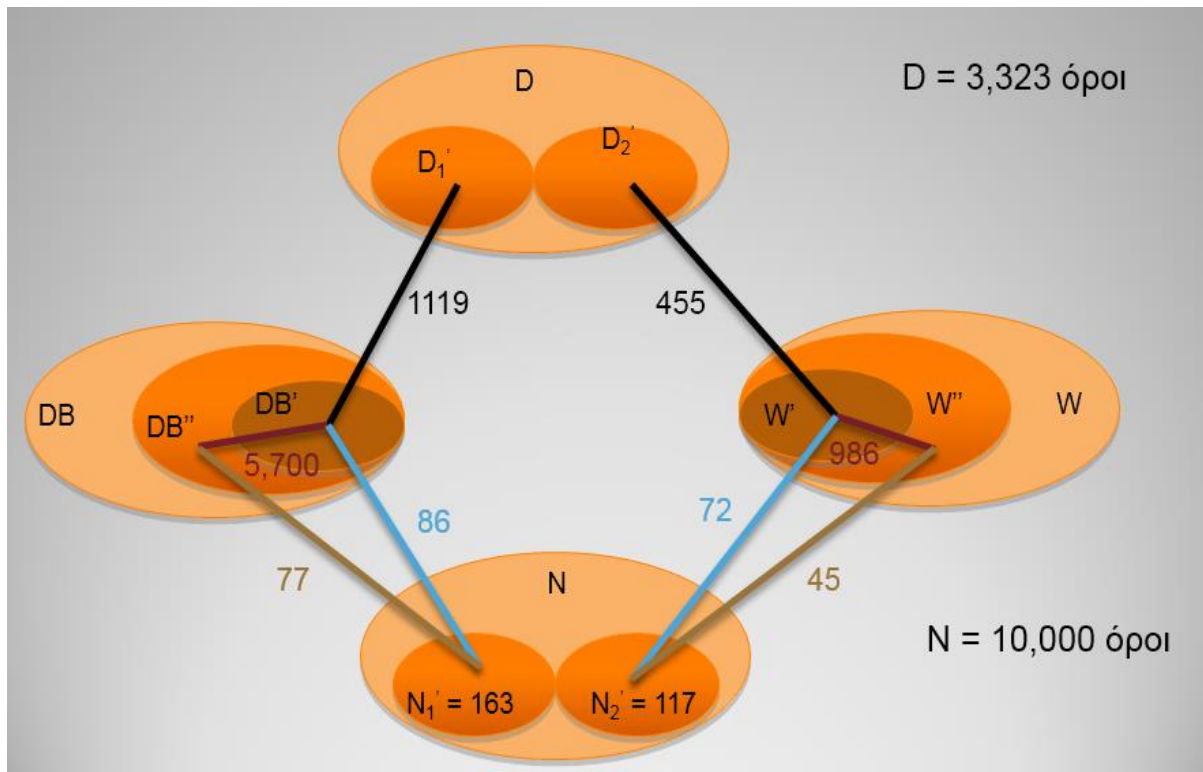
Βήμα 4: Ο εμπλουτισμός των παραπάνω συνόλων με σημασιολογικά ισοδύναμους όρους από τη DBpedia και το WordNet επιτυγχάνεται μέσω της χρήσης της λειτουργίας “*Add column By Fetching URL*” του Google Refine. Πιο συγκεκριμένα, το Google Refine παρέχει τη δυνατότητα δημιουργίας συνόλων που περιέχουν όρους, οι οποίοι σχετίζονται σημασιολογικά βάσει μιας συγκεκριμένης ιδιότητας. Για τις ανάγκες της προτεινόμενης προσέγγισης, επιλέχθηκαν οι “*dbpedia-owl:wikiPageRedirects*” και “*wordnet:containsWordsense*” από τη DBpedia και το WordNet αντίστοιχα. Κατά συνέπεια, το DB” αποτελείται από 5.700 όρους και το W” αποτελείται από 986 όρους.

Βήμα 5: Η λίστα των θεματικών επικεφαλίδων από τους NYT εισάγεται στο Google Refine ως αρχεία RDF dump¹⁰⁷. Επομένως, το σύνολο N περιέχει 10.000 όρους.

Βήμα 6: Τέλος, οι λεξιλογικά ισοδύναμοι όροι που προέρχονται από το βήμα 3 και οι σημασιολογικά ισοδύναμοι όροι που προέχονται από το βήμα 4 «*συμφιλιώνονται*» με τις θεματικές επικεφαλίδες των NYT για να ανακαλυφθούν πιθανές λεξιλογικές ομοιότητες. Αυτή η διαδικασία καταλήγει στα ακόλουθα 1-1 ταιριάσματα (βλέπε εικόνα 6.3):

1. Το N₁’ αποτελείται από 163 όρους και το N₂’ αποτελείται από 117 όρους.
2. Το DB’ αποτελείται από 86 όρους και το DB” αποτελείται από 77 όρους.
3. Το W’ αποτελείται από 72 όρους και το W” αποτελείται από 45 όρους.

¹⁰⁷ Data.nytimes.com, (2015). *New York Times - Linked Open Data*. [online] Διαθέσιμο στο: <http://data.nytimes.com/> [Πρόσβαση: 11/07/2015].



Εικόνα 6.3: Αποτελέσματα διαδικασίας

Στην επόμενη ενότητα, παρουσιάζεται μια συγκριτική αξιολόγηση ανάμεσα σε μια παραδοσιακή λεξιλογική μεθοδολογία εύρεσης ομοιοτήτων και σε αυτήν που προτείνεται σε αυτήν εδώ την ενότητα.

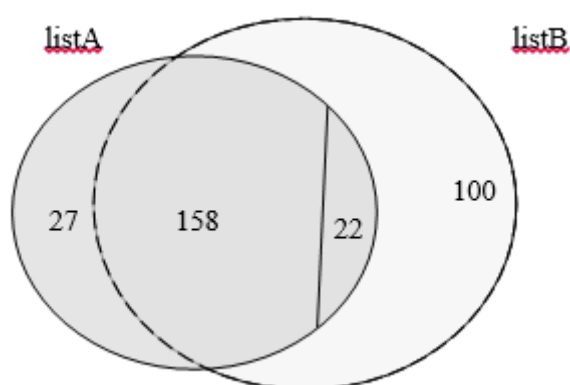
6.3.3. Συγκριτική αξιολόγηση

Ο τελικός στόχος της μεθοδολογίας που παρουσιάστηκε σε αυτό το κεφάλαιο είναι να ανακαλυφθούν ισοδύναμοι όροι ανάμεσα σε σημασιολογικά όμοια σύνολα δεδομένων. Για να αξιολογηθεί η αποτελεσματικότητα της προτεινόμενης προσέγγισης, αποφασίστηκε να συγκριθούν τα αποτελέσματα της εφαρμογής της συγκεκριμένης προσέγγισης με τα αποτελέσματα της εφαρμογής μιας προσέγγισης που βασίζεται στην ανακάλυψη μόνο λεξιλογικών ομοιοτήτων. Μια τέτοια απόφαση δικαιολογείται από το γεγονός της έλλειψης σχήματος στο σύνολο δεδομένων των NYT, που κάνει δυνατή την εφαρμογή οποιασδήποτε ευθυγράμμισης που βασίζεται σε οντολογίες.

Στο Kyrgianos & Paradakis (2012), αναπτύχθηκε μια μεθοδολογία για εξεύρεση λεξιλογικών ομοιοτήτων (παρουσιάστηκε στο Paradakis & Kyrgianos (2011)) με απώτερο στόχο την εξεύρεση ισοδύναμων όρων ανάμεσα στις θεματικές επικεφαλίδες της ψηφιακής βιβλιοθήκης (Διώνη) και των NYT. Η εφαρμογή των δύο μεθοδολογιών (δηλ. της εργασίας στο Kyrgianos & Paradakis (2012) και της προτεινόμενης μεθοδολογίας)

στα ίδια σύνολα δεδομένων εξυπηρετεί τη συγκριτική αξιολόγηση, η οποία βασίζεται πάνω στα αντίστοιχα αποτελέσματα αξιολόγησης των εφαρμογών.

Σύμφωνα με τα αποτελέσματα από την εφαρμογή της μεθοδολογίας στο Kyrgianos & Paradakis (2012), η οποία έχει παρουσιαστεί αναλυτικά στην ενότητα 5.2.4, η μεθοδολογία που βασίζεται στις λεξιλογικές ομοιότητες κατέληξε στη δημιουργία μιας λίστας με 207 ζευγάρια ισοδύναμων όρων (δηλ. listA) ανάμεσα στα σύνολα δεδομένων της ψηφιακής βιβλιοθήκης (Διώνη) και των NYT, ενώ η προτεινόμενη μεθοδολογία κατέληξε στη δημιουργία (βλέπε ενότητα 6.3.2, βήμα 6) μιας λίστας με 280 ζευγάρια ισοδύναμων όρων (δηλ. listB). Μια σύγκριση ανάμεσα στις δύο λίστες¹⁰⁸ αποκαλύπτει ότι η τομή τους αποτελείται από 180 κοινά ζευγάρια, 27 ζευγάρια διαθέσιμα μόνο στη λίστα listA και 100 διαθέσιμα μόνο στη λίστα listB (βλέπε εικόνα 6.4). Τα 22 από τα 180 κοινά ζευγάρια εμφανίζονται στη λίστα listB ως το προϊόν σημασιολογικής ευθυγράμμισης, ενώ τα υπόλοιπα 158 ζευγάρια εμφανίζονται στη λίστα listB ως το προϊόν της αντιστοίχισης με τη χρήση λεξιλογικών ομοιοτήτων.



Εικόνα 6.4: Αποτελέσματα συγκριτικής ανάλυσης

Τα αποτελέσματα της συγκριτικής ανάλυσης, όπως παρουσιάζονται στην εικόνα 6.4, αποκαλύπτουν ότι η απόδοση του Google Refine είναι λίγο πιο φτωχή εάν συγκριθεί με τον αλγόριθμο λεξιλογικών ομοιοτήτων που παρουσιάστηκε στο Paradakis & Kyrgianos (2011) (δηλ. 207 έναντι 158 επιτυχημένων ταιριασμάτων). Έτσι, φαίνεται ότι παρ' όλο που το Google Refine κάνει δυνατή την εφαρμογή πολύπλοκων τεχνικών

¹⁰⁸ Η ListA είναι διαθέσιμη στο: <https://docs.google.com/spreadsheet/ccc?key=0AuzgNdHCG27DdDRrMU1XMTVRcjJPZEg0cWowMDJxc3c#gid=0> [Πρόσβαση: 27/07/2015] και η ListB είναι διαθέσιμη στο: <https://docs.google.com/spreadsheet/ccc?key=0AuzgNdHCG27DdFZjRDNNLVRMRWg2TkJWZGFZNIBENXc#gid=0> [Πρόσβαση: 27/07/2015].

λεξιλογικής ομοιότητας στα δεδομένα από μη ειδικευμένους χρήστες, αυτή η ευκολία στη χρήση δεν έρχεται χωρίς το ανάλογο τίμημα.

Από την άλλη μεριά, η μεθοδολογία σημασιολογικής ευθυγράμμισης που βασίζεται στα σύνολα δεδομένων της DBpedia και του WordNet αύξησε σημαντικά των αριθμό των επιτυχημένων ταιριασμάτων ανάμεσα στην ψηφιακή βιβλιοθήκη (Διώνη) και τους NYT. Πιο συγκεκριμένα, όπως φαίνεται στην εικόνα 6.4, 100 νέα ζευγάρια ισοδύναμων όρων ανακαλύφθηκαν.

6.4. Συμπεράσματα

Τα αποτελέσματα της συγκριτικής αξιολόγησης είναι πολλά υποσχόμενα. Πιο συγκεκριμένα, η εμφάνιση ευρέως διαθέσιμων εργαλείων όπως το Google Refine καθιστά την εφαρμογή της προτεινόμενης μεθοδολογίας ως μια απλή και ξεκάθαρη διαδικασία, η οποία μπορεί να εφαρμοστεί και σε άλλες περιπτώσεις που έχουν ως στόχο την ανακάλυψη ισοδύναμων όρων σε διαφορετικά αλλά σημασιολογικά παρόμοια σύνολα δεδομένων. Ωστόσο, θα πρέπει να τονιστεί ότι υπάρχει ένας συμβιβασμός ανάμεσα στην ακρίβεια και τη λειτουργικότητα που παρέχεται αυτόματα, όταν κάποιος επιλέγει ένα «όλα-σε-ένα» προϊόν, όπως είναι το Google Refine αντί για εξειδικευμένες λύσεις.

Επιπλέον, η εφαρμογή της προτεινόμενης μεθοδολογίας εξυπηρετείται μέσω της χρήσης τεχνολογιών των Συνδεδεμένων Δεδομένων. Φαίνεται αρκετά λογικό να υποθέσουμε ότι η χρήση ακόμα περισσότερων γλωσσικών συνόλων δεδομένων από το σύννεφο των Συνδεδεμένων Δεδομένων θα μπορούσε να βελτιώσει την αποτελεσματικότητα της προτεινόμενης μεθοδολογίας.

Μέσα σε αυτό το πλαίσιο, θα μπορούσαν να «συμφιλιωθούν» θεματικές επικεφαλίδες της ψηφιακής βιβλιοθήκης (Διώνη) με άλλες υπηρεσίες των Συνδεδεμένων Δεδομένων, όπως είναι το RAMEAU, το SWD, η BNE, το LIBRIS κλπ.

Ολοκληρώνοντας και αυτό το κεφάλαιο, συμπεραίνουμε ότι έχει γίνει μια λεπτομερής μελέτη των δυνατοτήτων που παρέχουν τα ελεγχόμενα λεξιλόγια και πως αυτά μπορούν να συνδυαστούν για την επίτευξη διαλειτουργικότητας με απώτερο στόχο την πληρέστερη ικανοποίηση των πληροφοριακών αναγκών των χρηστών. Στο επόμενο κεφάλαιο, θα παρουσιαστούν οι δυνατότητες του MARC για να αποθηκεύσει πληροφορίες σε Συνδεδεμένα Δεδομένα και τι στην πραγματικότητα κάνουν οι εθνικές βιβλιοθήκες ανά τον κόσμο για να αποθηκεύσουν θεματική πληροφορία σε μορφή των Συνδεδεμένων Δεδομένων.

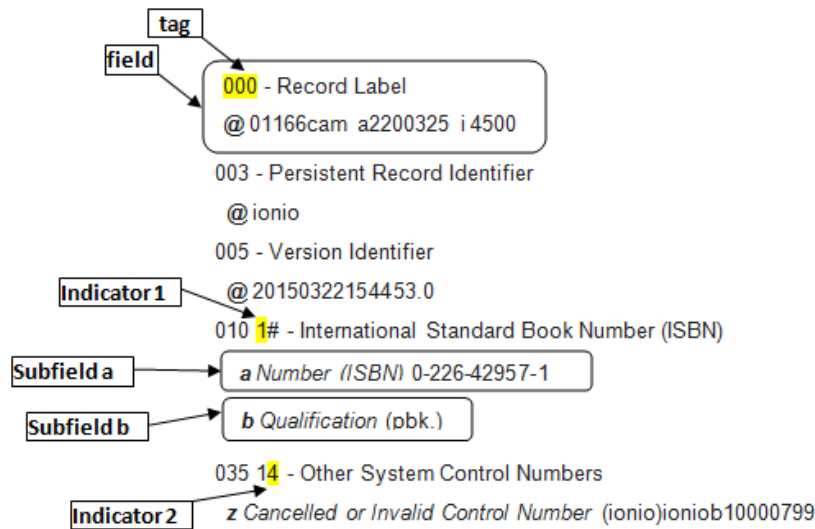
7. Συνδεδεμένα Δεδομένα, MARC και βιβλιοθήκες

7.1. MARC και βιβλιοθήκες

Τα δομημένα διαλειτουργικά δεδομένα δεν είναι ένα καινούριο ερευνητικό πεδίο για τις βιβλιοθήκες. Από τις αρχές της δεκαετίας του 1960, οι βιβλιοθήκες είχαν συμφωνήσει πάνω στο πρότυπο MARC για την τακτοποίηση των δεδομένων τους (Carlan, 2003, σ. 12). Το πρότυπο MARC προτείνει μια μορφή των δεδομένων που μπορεί να χρησιμοποιηθεί για την ανταλλαγή, τη χρήση και την ερμηνεία των βιβλιογραφικών δεδομένων και των καθιερωμένων θεμάτων ανάμεσα στις βιβλιοθήκες, ενισχύοντας με αυτό τον τρόπο τη διαλειτουργικότητα. Χρησιμοποιεί ένα σύστημα από αριθμούς, γράμματα και σύμβολα για να περιγράψει την εκάστοτε πληροφορία.

Κάθε εγγραφή σε μορφή MARC αποτελείται από πεδία (fields), ετικέτες (tags), δείκτες (indicators), υποπεδία (subfields), κώδικες υποπεδίων (subfield codes), και δείκτες περιεχομένου (content indicators)¹⁰⁹. Πιο συγκεκριμένα, κάθε θεματική εγγραφή στο MARC χωρίζεται λογικά σε πεδία (βλέπε εικόνα 7.1). Υπάρχουν διάφορα πεδία για τη δήλωση διαφορετικής περιγραφικής πληροφορίας μέσα σε μια εγγραφή MARC. Κάθε πεδίο σχετίζεται με ένα τριψήφιο αριθμό που ονομάζεται ετικέτα. Η ετικέτα προσδιορίζει το είδος των δεδομένων που ακολουθεί. Στη συνέχεια, οι δείκτες εξειδικεύουν ακόμα περισσότερο τα αντίστοιχα πεδία. Οι δείκτες αποτελούνται από δύο χαρακτήρες που ακολουθούν αμέσως μετά τις ετικέτες. Κανένας, ένας ή και οι δύο μπορούν να χρησιμοποιηθούν. Κάθε δείκτης παίρνει έναν αριθμό από το 0 έως το 9. Οι δείκτες ακολουθούνται από υποπεδία. Τα υποπεδία χαρακτηρίζονται από κωδικούς και οριοθέτες.

¹⁰⁹ Loc.gov, (2015). *Understanding MARC Bibliographic: Parts 1 to 6*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/umb/um01to06.html> [Πρόσβαση: 14/07/2015].



Εικόνα 7.1: Παράδειγμα εγγραφής σε MARC

Το γεγονός ότι το πρότυπο MARC χρησιμοποιείται τόσο πολλά χρόνια έχει συμβάλει στη δημιουργία σταθερών και πολύτιμων πληροφοριών στο χώρο των βιβλιοθηκών. Όμως, αυτή η πληροφορία είναι δύσκολο να μοιραστεί και να διατεθεί σε εξωτερικές οντότητες, πέραν των βιβλιοθηκών. Όπως έχει γίνει ήδη σαφές στις προηγούμενες ενότητες, για να αντιμετωπίσουν αυτό το πρόβλημα της διαλειτουργικότητας και για να διαθέσουν τα δεδομένα τους στο συνεχώς εξελισσόμενο διαδικτυακό περιβάλλον, οι βιβλιοθήκες τα τελευταία χρόνια άρχισαν να πειραματίζονται με τις τεχνολογίες των Συνδεδεμένων Δεδομένων. Στην επόμενη ενότητα, θα παρουσιαστούν τα πεδία του MARC που θα μπορούσαν να φιλοξενήσουν πληροφορία σε μορφή των Συνδεδεμένων Δεδομένων (Papadakis, Kyriianos & Stefanidakis, 2015).

7.2. MARC και Συνδεδεμένα Δεδομένα

Ο ίδιος ο χαρακτήρας των Συνδεδεμένων Δεδομένων, έγειρε το ενδιαφέρον της βιβλιοθηκονομικής κοινότητας από την αρχή. Τα τελευταία χρόνια, διάφορες βιβλιοθήκες ανά τον κόσμο έχουν γίνει ενεργά μέλη του σύννεφου των Συνδεδεμένων Δεδομένων¹¹⁰. Όπως έχει ήδη αναφερθεί, αυτό μπορεί να οφείλεται στο γεγονός ότι τα Συνδεδεμένα Δεδομένα βασίζονται πάνω σε κοινά διαδικτυακά πρότυπα (π.χ. HTTP, URI, κ.ά.). Επομένως, οι υπηρεσίες που προκύπτουν είναι εύκολο να συντηρηθούν και

¹¹⁰ Lod-cloud.net, (2015). *The Linking Open Data cloud diagram*. [online] Διαθέσιμο στο: <http://lod-cloud.net/> [Πρόσβαση: 15/07/2015].

να αναπτυχθούν (Baker et al., 2011). Επιπλέον, με τη χρήση των Συνδεδεμένων Δεδομένων εξυπηρετείται η ανταλλαγή πληροφοριών με τους χρήστες διαφορετικών επιστημονικών πεδίων και τομέων (Malmsten, 2009).

Σύμφωνα με τους Baker et al. (2011), το πρώτο πράγμα που πρέπει η βιβλιοθήκη να κάνει για να παρέχει τα δικά της δεδομένα ως Συνδεδεμένα Δεδομένα είναι να δημιουργήσει URIs (π.χ. σταθερούς στατικούς υπερσυνδέσμους (permalinks)) για τη συλλογή της.

Σε αυτή την ενότητα, οι προσδιοριστές που έχουν οριστεί τοπικά (locally-defined) αναφέρονται σε προσδιοριστές που έχουν δημιουργηθεί από τη βιβλιοθήκη που τους χρησιμοποιεί. Για παράδειγμα, ο αριθμός ελέγχου (π.χ. η τιμή του πεδίου 001 στο MARC) μιας θεματικής εγγραφής θεωρείται ως τοπικά ορισμένος προσδιοριστής, ενώ ο αριθμός ελέγχου άλλου συστήματος (π.χ. η τιμή του πεδίου 035 στο MARC) θεωρείται ως προσδιοριστής ορισμένος σε απομακρυσμένο σημείο (remotely-defined).

Στις επόμενες ενότητες, γίνεται μια έρευνα για να αξιολογηθεί η δυνατότητα συγκεκριμένων πεδίων για τη θεματική περιγραφή του MARC21 (Library of Congress, 2013) και του UNIMARC (Willer, 2009b) να φιλοξενήσουν URI που είναι ορισμένα είτε τοπικά είτε απομακρυσμένα. Τα πεδία που συμμετέχουν σε αυτή την αξιολόγηση είναι εξ' ορισμού ικανά να φιλοξενούν προσδιοριστές. Η επιλογή των δύο συγκεκριμένων προτύπων (δηλ. MARC21 και UNIMARC) έγινε λόγω του γεγονότος ότι είναι τα δύο βασικά πρότυπα που χρησιμοποιούνται από πολλές βιβλιοθήκες ανά τον κόσμο για την περιγραφή, τη δόμηση και την ανταλλαγή των δεδομένων τους. Πιο συγκεκριμένα, το MARC21¹¹¹, δημιουργήθηκε από τη συγχώνευση των προτύπων USMARC και Canadian MARC και χρησιμοποιείται από τις βιβλιοθήκες των ΗΠΑ και του Καναδά, καθώς και από άλλες βιβλιοθήκες ανά τον κόσμο, ενώ το UNIMARC αρχικά δημιουργήθηκε για να λειτουργήσει ως ένα ενδιάμεσο πρότυπο για να μειωθεί ο αριθμός των λογισμικών και των εργαλείων που χρειάζονται από μια βιβλιοθήκη για τη χρήση και τη μετατροπή των εγγραφών από διαφορετικές πηγές στο πρότυπο που χρησιμοποιείται από αυτή και χρησιμοποιείται ως επί το πλείστον από βιβλιοθήκες στην Ευρώπη (Galvao & Cordeiro, 2013, σ. 152).

¹¹¹ Loc.gov, (2015). *MARC STANDARDS (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/> [Πρόσβαση: 30/07/2015].

7.2.1. MARC21, UNIMARC και URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί τοπικά

Όσον αφορά στη δημιουργία URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί τοπικά, οι βιβλιοθήκες ήδη χρησιμοποιούν μοναδικούς προσδιοριστές (π.χ. Αριθμούς ελέγχου – Control Numbers) για τις εγγραφές τους και τις οντότητές τους στην καθημερινή λειτουργία τους. Αυτή η πληροφορία διατηρείται σε συγκεκριμένα πεδία του MARC21 και του UNIMARC. Στις επόμενες παραγράφους, συζητείται η σημασιολογική συμβατότητα αυτών των πεδίων με τα URIs των Συνδεδεμένων Δεδομένων. Η συζήτηση βασίζεται στον ορισμό του κάθε πεδίου, όπως αυτός ορίζεται από το πρότυπο (π.χ. MARC21 και UNIMARC) και τη σημασιολογική του ερμηνεία όπως αυτή συλλαμβάνεται από τους συγγραφείς (Papadakis, Kyprianos & Stefanidakis, 2015). Ο πίνακας 7.1 παρουσιάζει τα πεδία του MARC που φαίνονται κατάλληλοι υποψήφιοι για τη φιλοξενία URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί τοπικά.

Πίνακας 7.1: Πεδία του MARC που μπορούν να φιλοξενήσουν URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί τοπικά

Πεδία	MARC21	UNIMARC
001	Control Number	Record identifier
003	-	Persistent Record Identifier
024	Other Standard Identifier	-
856	Electronic Location and Access	Electronic Location and Access

Και το MARC21 και το UNIMARC χρησιμοποιούν το πεδίο 001 για να προσδιορίσουν μοναδικά τις καθιερωμένες εγγραφές τους. Ιδανικά, μια βιβλιοθήκη νέας γενιάς, βασισμένη στα Συνδεδεμένα Δεδομένα, θα μπορούσε να έχει όλες τις εγγραφές της προσδιορισμένες με permalinks που θα αποθηκεύονταν στο πεδίο 001. Στην πράξη όμως, οι βιβλιοθήκες χρησιμοποιούν αυτό το μη επαναλαμβανόμενο πεδίο για να αποθηκεύουν προσδιοριστές για τις εγγραφές τους που δημιουργούνται αυτόματα από το σύστημα και δεν είναι σε μορφή URI. Έτσι, οποιαδήποτε προσπάθεια και να έκανε μια βιβλιοθήκη για να αντικαταστήσει τους ήδη υπάρχοντες προσδιοριστές με URIs των Συνδεδεμένων Δεδομένων θα μπορούσε δυνητικά να δημιουργήσει προβλήματα λειτουργικότητας σε όλο το σύστημα. Επομένως, κάθε ένα από τα δύο κυρίαρχα πρότυπα

του MARC παρέχουν έναν αριθμό συμπληρωματικών πεδίων για την αποθήκευση μοναδικών προσδιοριστών, που θα μπορούσαν να φιλοξενήσουν και URIs των Συνδεδεμένων Δεδομένων.

Πιο συγκεκριμένα, το UNIMARC ορίζει το πεδίο 003 (Willer, 2009a) για την αποθήκευση permalinks για την τοπική καθιερωμένη εγγραφή. Αυτό το πεδίο είναι ιδανικό για τη φιλοξενία URIs των Συνδεδεμένων Δεδομένων που ορίζονται από την τοπική βιβλιοθήκη. Σε μια παλαιότερη έκδοση το UNIMARC παρείχε το πεδίο 009¹¹², το οποίο έδινε τη δυνατότητα στον καταλογογράφο να συμπληρώσει ένα προσδιοριστή διαφορετικό από αυτόν που είχε οριστεί από το υφιστάμενο σύστημα.

Το πεδίο 024 του MARC21 αναφέρεται σε ένα «Πρότυπο αριθμό ή κωδικό που σχετίζεται με την οντότητα που ορίζεται στο πεδίο 1XX, ο οποίος δεν μπορεί να τοποθετηθεί σε κάποιο άλλο πεδίο (π.χ. πεδίο 020 (International Standard Book Number - ISBN) και 022 (International Standard Serial Number - ISSN)). Η πηγή του πρότυπου αριθμού ή κωδικού προσδιορίζεται στο υποπεδίο \$2 (πηγή του αριθμού ή του κωδικού - Source of number or code)»¹¹³. Εάν τα URIs θεωρηθούν «κωδικοί», ο ορισμός που αναφέρθηκε προηγουμένως υπονοεί ότι το πεδίο 024 μπορεί να φιλοξενήσει URIs των Συνδεδεμένων Δεδομένων που ορίζονται τοπικά, με την προϋπόθεση ότι ο κωδικός του MARC που αναφέρεται στην εκάστοτε βιβλιοθήκη δηλώνεται στο υποπεδίο \$2.

Σε αυτό το σημείο, πρέπει να τονιστεί το γεγονός ότι το UNIMARC και το MARC21 ακολουθούν λίγο διαφορετική προσέγγιση σε σχέση με την οντότητα που δηλώνεται με προσδιοριστές στα πεδία 003 και 024. Πιο αναλυτικά, ο ορισμός του πεδίου 003 του UNIMARC αναφέρει ότι ο προσδιοριστής που δηλώνεται σε αυτό το πεδίο θα πρέπει να αναφέρεται σε μια καθιερωμένη εγγραφή, ενώ ο ορισμός του πεδίου 024 του MARC21 αναφέρει ότι ο προσδιοριστής που δηλώνεται σε αυτό το πεδίο θα πρέπει να αναφέρεται στην οντότητα που περιγράφεται από την εγγραφή. Ωστόσο, όσον αφορά

¹¹² Cerl.org, (2015). *UNIMARC and Cataloguing Rules (UNICAT) [CERL]*. [online] Διαθέσιμο στο: <http://www.cerl.org/resources/hpb/fileproc/unicat> [Πρόσβαση: 15/07/2015].

¹¹³ Ακριβής μετάφραση από το πρωτότυπο κείμενο: "Standard number or code associated with the entity named in the 1xx field which cannot be accommodated in another field (e.g., fields 020 (International Standard Book Number) and 022 (International Standard Serial Number)). The source of the standard number or code is identified in subfield \$2 (Source of number or code)". Loc.gov, (2015). *MARC 21 Format for Authority Data: 024: Other Standard Identifier (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/authority/ad024.html> [Πρόσβαση: 25/07/2015].

στα καθιερωμένα θέματα, οι βιβλιοθήκες δημιουργούν URIs των Συνδεδεμένων Δεδομένων που δεν κάνουν διαχωρισμό ανάμεσα στις εγγραφές και στις αντίστοιχες οντότητες.

Και τα δύο πρότυπα παρέχουν το πεδίο 856, το οποίο θα μπορούσε δυνητικά να φιλοξενήσει URIs των Συνδεδεμένων Δεδομένων που ορίζονται τοπικά. Εξ' ορισμού, το πεδίο 856 φιλοξενεί URLs ικανά να παρέχουν τοποθεσίες και πρόσβαση σε πληροφορία σχετική με την αντίστοιχη καθιερωμένη εγγραφή. Αυτό μπορεί, επίσης, να επαληθευτεί από το γεγονός ότι οι περισσότεροι δείκτες και υποπεδία που παρέχονται σε αυτό το πεδίο σχετίζονται ως επί το πλείστον με διαδικτυακά πρωτόκολλα πρόσβασης, που διαβιβάζουν πληροφορία σχετική με τη μηχανική των URIs και όχι τα URIs αυτά κάθε αυτά. Επομένως, αυτό το πεδίο δε θα πρέπει να χρησιμοποιείται για σκοπούς ταυτοποίησης. Αντιθέτως, θα πρέπει να χρησιμοποιείται για τον εντοπισμό και την πρόσβαση σε online πληροφορία σχετική με την εκάστοτε οντότητα. Από μια άλλη σκοπιά, μια βιβλιοθήκη που δημιουργεί τα δικά της URIs των Συνδεδεμένων Δεδομένων για τις καθιερωμένες εγγραφές της και ικανοποιεί τον δεύτερο κανόνα των Συνδεδεμένων Δεδομένων (δηλ. *"Χρησιμοποίησε HTTP URIs έτσι ώστε οι άνθρωποι να μπορούν να ψάξουν αυτά τα ονόματα"*), θα μπορούσε να χρησιμοποιήσει το πεδίο 003 (ή 024 στο MARC21) για λόγους ταυτοποίησης και το πεδίο 856 για πληροφοριακούς λόγους.

Η επόμενη ενότητα παρουσιάζει τα πεδία του MARC που μπορούν δυνητικά να φιλοξενήσουν προσδιοριστές των Συνδεδεμένων Δεδομένων που έχουν οριστεί σε απομακρυσμένα σημεία. Ξανά, τα συγκεκριμένα ευρήματα βασίζονται πάνω στα πρότυπα και τη σημασιολογία κάθε πεδίου, όπως αυτή συλλαμβάνεται από τους συγγραφείς.

7.2.2. MARC21, UNIMARC και URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί σε απομακρυσμένα σημεία

Το MARC στην παρούσα μορφή του μπορεί ρητά να φιλοξενήσει URIs που αναφέρονται σε προσδιοριστές των Συνδεδεμένων Δεδομένων που έχουν οριστεί σε απομακρυσμένα σημεία. Ο πίνακας 7.2 παρουσιάζει τα πεδία του MARC που φαίνονται κατάλληλοι υποψήφιοι για τη φιλοξενία URIs των Συνδεδεμένων Δεδομένων που προέρχονται από απομακρυσμένα σύνολα δεδομένων.

Πίνακας 7.2: Πεδία του MARC που μπορούν να φιλοξενήσουν URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί σε απομακρυσμένα σημεία

Πεδία	MARC21	UNIMARC
024	Other Standard Identifier	-
033	-	Other System Persistent Record Identifier
035	System Control Number	Other System Control Numbers
450	See From Tracing - Topical Term	Variant Access Point – Topical Subject
670	Source Data Found	-
750	Established Heading Linking Entry - Topical Term	Authorized Access Point in Other Language and/or Script – Topical Subject
810	-	Source Data Found
856	Electronic Location and Access	Electronic Location and Access

Στις επόμενες παραγράφους, παρουσιάζεται μια λεπτομερής συζήτηση για τη σημασιολογική συμβατότητα των παραπάνω πεδίων με τα απομακρυσμένα URIs των Συνδεδεμένων Δεδομένων.

Όπως αναφέρθηκε νωρίτερα, το πεδίο 024 του MARC21 αναφέρεται σε ένα «Πρότυπο αριθμό ή κωδικό που σχετίζεται με την οντότητα που ορίζεται στο πεδίο 1XX, ο οποίος δεν μπορεί να τοποθετηθεί σε κάποιο άλλο πεδίο (π.χ. πεδίο 020 (ISBN) και 022 (ISSN)). Η πηγή του πρότυπου αριθμού ή κωδικού προσδιορίζεται στο υποπεδίο \$2 (προέλευση του αριθμού ή του κωδικού - Source of number or code)»¹¹⁴. Σύμφωνα με τον παραπάνω ορισμό, το πεδίο 024 μπορεί να φιλοξενήσει ένα απομακρυσμένο URI των Συνδεδεμένων Δεδομένων, αν η προέλευση του URI παρέχεται στο υποπεδίο

¹¹⁴ Ακριβής μετάφραση από το πρωτότυπο κείμενο: "Standard number or code associated with the entity named in the 1xx field which cannot be accommodated in another field (e.g. fields 020 (International Standard Book Number) and 022 (International Standard Serial Number)). The source of the standard number or code is identified in subfield \$2 (Source of number or code)". Loc.gov, (2015). *MARC 21 Format for Authority Data: 024: Other Standard Identifier (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/authority/ad024.html> [Πρόσβαση: 25/07/2015].

§2. Οι πιθανές τιμές που μπορεί να πάρει το υποπεδίο §2 ορίζονται στο μητρώο "Standard Identifier Source Codes"¹¹⁵, το οποίο δηλώνει ότι αυτό το μητρώο «...εκχωρεί έναν κωδικό σε κάθε βάση δεδομένων ή δημοσίευση που καθορίζει ή περιέχει προσδιοριστές»¹¹⁶. Είναι προφανές ότι η προέλευση του κωδικού που εμφανίζεται στο μητρώο θα πρέπει να αναφέρεται σε οργανισμούς που ορίζουν ή περιέχουν προσδιοριστές. Ωστόσο, αυτή δεν είναι η περίπτωση των κωδικών προέλευσης των κωδικών των URI και των Uniform Resource Name – URN (Moats, 1997). Το URI (και το παράγωγο του URN) δεν παρέχει προσδιοριστές. Αντιθέτως, είναι ένα σχήμα σύνταξης, που χρησιμοποιείται από άλλα συστήματα που επιθυμούν να ορίσουν http-ready προσδιοριστές (Virtual Authority File – VIAF¹¹⁷, Digital Object Identifier System – DOI¹¹⁸, LCCN¹¹⁹, κλπ.). Εδώ θα πρέπει επίσης να αναφερθεί ότι το εν λόγω μητρώο δεν περιλαμβάνει κωδικό προέλευσης για τις καθιερώσεις και τα λεξιλόγια της Βιβλιοθήκης του Κογκρέσου¹²⁰. Επομένως, προς το παρόν¹²¹, εάν ένας καταλογογράφος αποφασίσει να χρησιμοποιήσει το πεδίο 024 για να αναφερθεί σε ένα URI των Συνδεδεμένων Δεδομένων που προέρχεται από τη Βιβλιοθήκη του Κογκρέσου (πχ. <http://id.loc.gov/>

¹¹⁵ Loc.gov, (2015). *Standard Identifier Source Codes: Source Codes for Vocabularies, Rules, and Schemes (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/standards/sourcelist/standard-identifier.html> [Πρόσβαση: 15/07/2015].

¹¹⁶ Ακριβής μετάφραση από το πρωτότυπο κείμενο: "... assigns a code to each database or publication that defines or contains the identifiers". Loc.gov, (2015). *Standard Identifier Source Codes: Source Codes for Vocabularies, Rules, and Schemes (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/standards/sourcelist/standard-identifier.html> [Πρόσβαση: 15/07/2015].

¹¹⁷ Vial.org, (2015). *VIAF – The Virtual International Authority File*. [online] Διαθέσιμο στο: <https://viaf.org/> [Πρόσβαση: 15/07/2015].

¹¹⁸ Doi.org, (2015). *Digital Object Identifier System*. [online] Διαθέσιμο στο: <http://www.doi.org/> [Πρόσβαση: 15/07/2015].

¹¹⁹ Loc.gov, (2015). *LC Control Number (Library of Congress)*. [online] Διαθέσιμο στο: http://www.loc.gov/marc/lccn_structure.html [Πρόσβαση: 15/07/2015].

¹²⁰ Id.loc.gov, (2015). *Home - LC Linked Data Service (Library of Congress)*. [online] Διαθέσιμο στο: <http://id.loc.gov> [Πρόσβαση: 15/07/2015].

¹²¹ Σύμφωνα με το έγγραφο συζήτησης (Discussion Paper) No. 2010-DP02: Encoding URIs for controlled values in MARC Records: «*To γραφείο Development and MARC Standards Office αναπτύσσει μια υπηρεσία μητρώου για ελεγχόμενες λίστες και με τον τρόπο αυτό θεσπίζονται URIs, τόσο για τους ίδιους τους καταλόγους όσο και για κάθε τιμή στη λίστα*». Ακριβής μετάφραση από το πρωτότυπο κείμενο: "Development and MARC Standards Office is developing a registry service for controlled lists and in so doing is establishing URIs both for the lists themselves and for each value on a list". Επομένως, τα πράγματα αναμένεται να αλλάξουν στο προσεχές μέλλον. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/marbi/2010/2010-dp02.html> [Πρόσβαση: 15/07/2015].

[authorities/subjects/sh85000411](#)), δεν μπορεί να καταλογογραφήσει και την προέλευση του URI στο υποπεδίο \$2¹²².

Το πεδίο 033 στο UNIMARC προστέθηκε πρόσφατα για να φιλοξενήσει «*τους σταθερούς προσδιοριστές των εγγραφών που έχουν αποκτηθεί από άλλες πηγές. Ο σταθερός προσδιοριστής αποδίδεται από τον οργανισμό, που δημιουργεί, χρησιμοποιεί ή εκδίδει την εγγραφή. Αυτός είναι ο σταθερός προσδιοριστής για τη βιβλιογραφική εγγραφή και όχι για την ίδια την οντότητα... Ο σταθερός προσδιοριστής που τοποθετείται στο πεδίο 033 είναι ισοδύναμος στο διαδίκτυο με τον αριθμό ελέγχου συστήματος (system control number) για την εγγραφή σε μια άλλη βάση που τοποθετείται στο πεδίο 035*»¹²³. Ο σκοπός αυτού του πεδίου είναι να φιλοξενήσει URIs, τα οποία μπορούν να αναζητηθούν από ανθρώπους. Αυτά τα URIs θα πρέπει να αντιστοιχούν σε καθιερωμένες εγγραφές, οι οποίες έχουν δημιουργηθεί, χρησιμοποιηθεί ή εκδοθεί από άλλους οργανισμούς και ακολούθως να έχουν εισαχθεί ως έχουν στον τοπικό οργανισμό. Επομένως, το πεδίο 033 δεν είναι κατάλληλο για τη φιλοξενία URIs των Συνδεδεμένων Δεδομένων που αναφέρονται σε καθιερωμένες εγγραφές που δεν είναι ακριβώς ίδιες με τις τοπικές καθιερωμένες εγγραφές. Σε αυτό το πλαίσιο, το πεδίο 033 είναι ιδανικό για τη φιλοξενία URIs των Συνδεδεμένων Δεδομένων για καθιερωμένες εγγραφές που βρίσκονται σε ένα συλλογικό κατάλογο (π.χ. Συλλογικός Κατάλογος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών – ΣΚΕΑΒ¹²⁴).

Το πεδίο 035 στο MARC21 (και στο UNIMARC) περιέχει έναν «*αριθμό ελέγχου για την εγγραφή σε ένα σύστημα άλλο σε σχέση με τον αριθμό ελέγχου που περιέχεται στο πεδίο 001, 010 ή 016. Για λόγους διαλειτουργικότητας, η τεκμηρίωση της δομής και της χρήσης του αριθμού ελέγχου συστήματος θα πρέπει να παρέχεται στα μέλη που ανταλλάσσουν δεδομένα μεταξύ τους από τον αρχικό οργανισμό. Κάθε έγκυρος*

¹²² Σύμφωνα με τη Βιβλιοθήκη του Κογκρέσου, τα permalinks για τις καθιερωμένες εγγραφές παρέχονται μέσω του [locn.loc.gov](#). Ωστόσο, αυτά τα permalinks δεν συμφωνούν με τον τρίτο κανόνα των Συνδεδεμένων Δεδομένων. Επομένως, αυτή η πληροφορία είναι πρακτικά μη προσβάσιμη από άλλες υπηρεσίες.

¹²³ Ακριβής μετάφραση από το πρωτότυπο κείμενο: "*the persistent identifier of records obtained from other sources. The persistent identifier is assigned by the agency, which creates, uses or issues the record. This is the persistent identifier for the bibliographic record, not for the entity itself... The persistent identifier carried in a 033 field is the equivalent on the web of the system control number for the record in another database carried in a 035 field*". Ifla.org, (2015). *IFLA -- UNIMARC Authorities, 3rd edition: UPDATES 2012*. [online] Διαθέσιμο στο: <http://www.ifla.org/publications/unimarc-authorities-3rd-edition-updates-2012> [Πρόσβαση: 30/07/2015].

¹²⁴ [Unioncatalog.gr](#), (2015). *Συλλογικός Κατάλογος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών*. [online] Διαθέσιμο στο: <http://www.unioncatalog.gr/> [Πρόσβαση: 25/07/2015].

αριθμός ελέγχου συστήματος και οποιοσδήποτε άλλος σχετικός αριθμός ελέγχου που έχει ακυρωθεί ή δεν είναι έγκυρος θα πρέπει να παρέχεται σε ξεχωριστό πεδίο 035»¹²⁵. Αυτό σημαίνει ότι το εν λόγω πεδίο φιλοξενεί προσδιοριστές για τη συγκεκριμένη καθιερωμένη εγγραφή που έχουν οριστεί από άλλους οργανισμούς. Πιθανές τιμές του πεδίου 035 θα πρέπει να υπάρχουν στο πεδίο 001 των άλλων συστημάτων. Επομένως, ένα URI των Συνδεδεμένων Δεδομένων θα μπορούσε να εισαχθεί στο πεδίο 035 ενός τοπικού συστήματος, με την προϋπόθεση ότι το ίδιο URI υπάρχει στο πεδίο 001 ενός απομακρυσμένου συστήματος.

Το πεδίο 450 (και τα αντίστοιχα 4XX πεδία) στο UNIMARC και το MARC21 περιέχει *«ένα διαφορετικό σημείο πρόσβασης ή μια θεματική κατηγορία σε κωδικοποιημένη μορφή ή σε μορφή κειμένου από διαφορετική πηγή»¹²⁶.* Το πεδίο 450 στην πραγματικότητα είναι μια παραλλαγή (ή μη προτιμώμενος όρος) της καθιέρωσης που φιλοξενείται στα πεδία 150 και 250 στο MARC21 και στο UNIMARC αντίστοιχα. Αυτό το πεδίο φιλοξενεί διαφορετικές, μη προτιμώμενες λεκτικές καθιέρώσεις που δεν έχουν ξεχωριστό προσδιοριστή. Επομένως, το πεδίο 450 δεν μπορεί να φιλοξενήσει URIs των Συνδεδεμένων Δεδομένων που προέρχονται από πηγές που έχουν οριστεί σε απομακρυσμένα σημεία.

Το πεδίο 670 στο MARC21 αναφέρεται σε μια *«παραπομπή για μια συμβουλευτική πηγή που φαίνεται να σχετίζεται με κάποιο τρόπο με την οντότητα που περιγράφεται με την καθιερωμένη εγγραφή ή με σχετικές οντότητες. Μπορεί επίσης, να περιέχει την πληροφορία που βρίσκεται στην πηγή»¹²⁷.* Οπότε, το πεδίο 670 παρέχει μια αναφορά

¹²⁵ Ακριβής μετάφραση από το πρωτότυπο κείμενο: *"control number for the record in a system other than the one whose control number is contained in field 001 (Control Number), 010 or 016. For interchange purposes, documentation of the structure and use of the system control number must be provided to exchange partners by the originating organization. Each valid system control number and any related canceled/invalid control number are contained in a separate 035 field"*. Loc.gov, (2015). *MARC 21 Format for Authority Data: 035: System Control Number (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/authority/ad035.html> [Πρόσβαση: 25/07/2015].

¹²⁶ Ακριβής μετάφραση από το πρωτότυπο κείμενο: *"a variant access point or a subject category in coded and/or textual form that is referred from"* (Willer, 2009b, σ. 193).

¹²⁷ Ακριβής μετάφραση από το πρωτότυπο κείμενο: *"citation for a consulted source in which information is found related in some manner to the entity represented by the authority record or related entities. May also include the information found in the source"*. Loc.gov, (2015). *MARC 21 Format for Authority Data: 670: Source Data Found (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/authority/ad670.html> [Πρόσβαση: 25/07/2015].

που έχει δημιουργηθεί αποκλειστικά για τη συγκεκριμένη καθιερωμένη οντότητα. Ακόμα και αν τα υποπεδία \$a (δηλ. \$a: Source Citation) και \$u (δηλ. \$u: Uniform Resource Identifier) χρησιμοποιούνταν (εσφαλμένα) για να φιλοξενήσουν το όνομα της πηγής στην οποία αναφέρεται καθώς και το αντίστοιχο permalink, δεν υπάρχει υποπεδίο για να δηλωθεί το είδος της σχέσης ανάμεσα στην καθιερωμένη οντότητα και στην πηγή στην οποία αναφέρεται. Επομένως, το πεδίο 670 δεν μπορεί να φιλοξενήσει URIs Συνδεδεμένων Δεδομένων πηγών που έχουν οριστεί σε απομακρυσμένα σημεία.

Το πεδίο 810 του UNIMARC είναι το αντίστοιχο του πεδίου 670 στο MARC21, το οποίο ορίζεται ως «*μία παραπομπή σε μια πηγή αναφοράς από όπου η πληροφορία για την επικεφαλίδα εντοπίστηκε. Το πρώτο 810 πεδίο που ορίζεται περιέχει την παραπομπή για την καταλογογράφηση του βιβλιογραφικού έργου για το οποίο η επικεφαλίδα δημιουργήθηκε*»¹²⁸. Όπως και προηγουμένως, το εν λόγω πεδίο είναι ικανό να φιλοξενεί μόνο γενικές αναφορές στην αντίστοιχη καθιερωμένη οντότητα. Επιπλέον, το πεδίο 810 δεν μπορεί να φιλοξενήσει URIs.

Το πεδίο 750 στο MARC21 (και τα αντίστοιχα πεδία 7XX) ορίζεται ως «*θεματικός όρος, ο οποίος είναι ισοδύναμος με τον θεματικό όρο στο πεδίο 150 ή στο πεδίο 180 της γενικής υποδιαίρεσης της επικεφαλίδας της ίδιας εγγραφής. Συνδέει επικεφαλίδες που βρίσκονται μέσα σε ένα σύστημα ή σε διαφορετικούς θησαυρούς ή καθιερωμένα αρχεία*»¹²⁹. Η σχέση ανάμεσα στις καθιερωμένες εγγραφές στα πεδία 150 και 750 (δηλ. τοπικές - απομακρυσμένες) ορίζεται ως ισοδύναμη.

Επιπλέον, το πεδίο 750 περιλαμβάνει ένα υποπεδίο \$0, το οποίο ορίζεται ως «*ο αριθμός ελέγχου συστήματος της σχετικής καθιερωμένης εγγραφής, ή ένας σταθερός/πρότυπος προσδιοριστής όπως είναι ο International Standard Name Identifier –*

¹²⁸ Ακριβής μετάφραση από το πρωτότυπο κείμενο: "a citation to a reference source when information about the heading was found. The first 810 field usually contains the citation for the bibliographic work for the cataloguing of which the heading has been established". (Willer, 2009b, σ. 267).

¹²⁹ Ακριβής μετάφραση από το πρωτότυπο κείμενο: "topical term that is equivalent to the 150 topical term or 180 general subdivision heading field of the same record. It links headings within a system or from different thesauri or authority files". Loc.gov, (2015). *MARC 21 Format for Authority Data: 750: Established Heading Linking Entry-Topical Term (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/authority/ad750.html> [Πρόσβαση: 25/07/2015].

ISNI¹³⁰. Πριν τον αριθμό ελέγχου ή τον προσδιοριστή προηγείται ο κατάλληλος κωδικός του MARC για τον οργανισμό (για μια σχετική καθιερωμένη εγγραφή) ή ο κωδικός Standard Identifier source code (για ένα τυποποιημένο αναγνωριστικό σχήματος), ο οποίος εσωκλείεται σε παρενθέσεις. Δείτε το 'MARC Code List for Organizations' για τη λίστα με τους κωδικούς των οργανισμών και το 'MARC Code List for Organizations' για τους τυποποιημένους κωδικούς αναγνωριστικών σχήματος»¹³¹. Επομένως, το υποπεδίο \$0 θα μπορούσε να φιλοξενήσει URIs Συνδεδεμένων Δεδομένων με την προϋπόθεση ότι το τυποποιημένο αναγνωριστικό σχήματος έχει προηγηθεί του URI.

Όπως αναφέρθηκε νωρίτερα, αυτός ο ορισμός δεν επιτρέπει την αναφορά URIs Συνδεδεμένων Δεδομένων της Βιβλιοθήκης του Κογκρέσου, αφού το μητρώο των Standard Identifier Source Codes δεν περιλαμβάνει έναν κωδικό για τις δημοφιλείς καθιερωμένες επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου¹³². Το πεδίο 750 περιλαμβάνει, επίσης, το υποπεδίο \$2, το οποίο ορίζεται ως «ο κωδικός MARC που προσδιορίζει το θησαυρό ή το καθιερωμένο αρχείο, το οποίο είναι η πηγή της επικεφαλίδας όταν στη θέση του δεύτερου δείκτη περιέχεται η τιμή 7. Κωδικός από: Subject Heading and Term Source Codes για το υποπεδίο \$2 στα πεδία 700-751»¹³³. Από τον ορισμό

¹³⁰ Isni.org, (2015). *International Standard Name Identifier*. [online] Διαθέσιμο στο: <http://www.isni.org/> [Πρόσβαση: 17/07/2015].

¹³¹ Ακριβής μετάφραση από το πρωτότυπο κείμενο: "the system control number of the related authority record, or a standard identifier such as an International Standard Name Identifier (ISNI). The control number or identifier is preceded by the appropriate MARC Organization code (for a related authority record) or the Standard Identifier source code (for a standard identifier scheme), enclosed in parentheses. See 'MARC Code List for Organizations' for a listing of organization codes and 'Standard Identifier Source Codes' for code systems for standard identifiers". Loc.gov, (2015). *MARC 21 Format for Authority Data: Appendix A: Control Subfields (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/authority/ecadcntf.html> [Πρόσβαση: 25/07/2015].

¹³² Μια αναφορά των LCSH ως κωδικό για τις θεματικές επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου δεν θα πρέπει να υπάρχει στο υποπεδίο \$0, αφού ένα τέτοιο σύστημα ορίζει λεκτικά (δηλ. θεματικές επικεφαλίδες) και όχι προσδιοριστές.

¹³³ Ακριβής μετάφραση από το πρωτότυπο κείμενο: "MARC code that identifies the thesaurus or authority file that is the source of the heading when the second indicator position contains value 7. Code from: Subject Heading and Term Source Codes for subfield \$2 in fields 700-751". Loc.gov, (2015). *MARC 21 Format for Authority Data: 7XX: Heading Linking Entries-General Information (Network Development and MARC Standards Office, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/marc/authority/ad7xx.html> [Πρόσβαση: 25/07/2015].

αυτό, είναι προφανές ότι το υποπεδίο \$2 μπορεί να αναφέρεται σε έναν οργανισμό που παρέχει την αντίστοιχη ετικέτα της καθιέρωσης (όχι τον προσδιοριστή)¹³⁴.

Ολοκληρώνοντας, φαίνεται ότι το πεδίο 750 είναι ικανό να φιλοξενεί URIs των Συνδεδεμένων Δεδομένων που αναφέρονται σε ισοδύναμες καθιερωμένες οντότητες που ορίζονται από απομακρυσμένα συστήματα.

Το πεδίο 750 στο UNIMARC είναι λίγο διαφορετικό από το ισοδύναμό του στο MARC21. Πιο συγκεκριμένα, το πεδίο αυτό ορίζεται ως *«ένα θεματικό καθιερωμένο σημείο πρόσβασης, το οποίο είναι σε άλλη γλώσσα ή μορφή γραφής σε σχέση με το σημείο πρόσβασης του πεδίου 250»*¹³⁵. Όπως και πριν, η σχέση μεταξύ των καθιερωμένων εγγραφών στο 250 και το 750 (δηλ. τοπικές - απομακρυσμένες) ορίζεται ως ισοδύναμη. Το πεδίο 750 περιέχει, επίσης, έναν αριθμό υποπεδίων προστιθέμενης αξίας. Πιο συγκεκριμένα, το υποπεδίο \$3 φιλοξενεί την αντίστοιχη απομακρυσμένη πηγή, το υποπεδίο \$2 φιλοξενεί τον κωδικό MARC του οργανισμού που ορίζει την απομακρυσμένη πηγή και το υποπεδίο \$8 φιλοξενεί τη γλώσσα τις αντίστοιχης πηγής. Σε αντίθεση με το ισοδύναμο πεδίο του MARC21, στο πεδίο 750 του UNIMARC ορίζεται ρητά ότι μπορεί να φιλοξενήσει URIs των Συνδεδεμένων Δεδομένων σε άλλη γλώσσα από αυτή που χρησιμοποιείται στο αντίστοιχο πεδίο 250. Επομένως, το πεδίο αυτό είναι κατάλληλο μόνο για τη φιλοξενία Συνδεδεμένων Δεδομένων απομακρυσμένων πηγών, τα οποία είναι σε διαφορετική γλώσσα.

Τέλος, όπως αναφέρθηκε νωρίτερα, το πεδίο 856 στο MARC21 (και στο UNIMARC) περιέχει *«την πληροφορία που είναι απαραίτητη για τον εντοπισμό ηλεκτρονικών τεκμηρίων. Το πεδίο μπορεί να χρησιμοποιηθεί σε μια καθιερωμένη εγγραφή για να παρέχει επιπρόσθετη πληροφορία διαθέσιμη ηλεκτρονικά για την οντότητα για την οποία η εγγραφή δημιουργήθηκε. Η πληροφορία αναγνωρίζει την ηλεκτρονική τοποθεσία που περιέχει το αντικείμενο ή από την οποία είναι διαθέσιμο. Περιέχει, επίσης, πληροφορία για το πώς να ανακτηθεί το αντικείμενο, ενώ η μέθοδος της πρόσβασης δηλώνεται από τη θέση του πρώτου δείκτη. Μπορεί να χρησιμοποιηθεί για τη δημιουργία*

¹³⁴ Μια αναφορά των LCSH ως κωδικό για τις θεματικές επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου δεν θα πρέπει να υπάρχει στο υποπεδίο \$2.

¹³⁵ Ακριβής μετάφραση από το πρωτότυπο κείμενο: *"a topical authorized access point or an authorized subject category access point that is in another language and/or script form of the 250 access point"*. (Willer, 2009b, σ. 259).

σημειώσεων που αφορούν τη λειτουργία της πρόσβασης»¹³⁶. Τέτοια πεδία αναφέρονται στα URLs που προσφέρουν πρόσβαση στην τοποθεσία της πληροφορίας που αφορά την καθιερωμένη εγγραφή.

Είναι προφανές ότι το πεδίο 856 είναι σχεδιασμένο για πρόσβαση σε τοποθεσία (δηλ. όχι για λόγους προσδιορισμού). Ωστόσο, η απουσία ενός πεδίου ικανού να φιλοξενήσει προσδιοριστές που έχουν δημιουργηθεί απομακρυσμένα για την ίδια καθιερωμένη οντότητα στο UNIMARC, ανάγκασε πολλές βιβλιοθήκες ανά τον κόσμο να χρησιμοποιήσουν το εν λόγω πεδίο για να καταγράψουν τον αντίστοιχο προσδιοριστή VIAF. Αυτό δεν συμβαίνει με το MARC21, όπου οι προσδιοριστές του VIAF μπορούν να φιλοξενηθούν στο πεδίο 024.

Στην επόμενη ενότητα, παρουσιάζονται οι πρακτικές που ακολουθούν στην πράξη οι εθνικές βιβλιοθήκες ανά τον κόσμο για να παρέχουν πληροφορίες βασισμένες στα Συνδεδεμένα Δεδομένα στους χρήστες τους. Αυτή η πληροφορία συγκεντρώθηκε από τη σχολαστική παρατήρηση της ιστοσελίδας της κάθε βιβλιοθήκης.

7.3. Εθνικές βιβλιοθήκες, καθιερωμένες εγγραφές, MARC και Συνδεδεμένα Δεδομένα

Οι βιβλιοθήκες που συμμετέχουν σε αυτή την έρευνα εμπλέκονται σε κάποιο βαθμό με το κίνημα των Συνδεδεμένων Δεδομένων. Σαν γενικό σχόλιο, θα μπορούσε να ειπωθεί ότι τα URIs που ορίζονται τοπικά βασίζονται πάνω σε τιμές συγκεκριμένων πεδίων του MARC που χρησιμοποιούνται για προσδιοριστές και αποθηκεύονται σε νέες μοντέρνες υπηρεσίες των Συνδεδεμένων Δεδομένων. Πιο συγκεκριμένα, η δημιουργία τέτοιων URIs ακολουθεί ένα συγκεκριμένο μοτίβο: «*Χρήση μιας συγκεκριμένης διαδρομής ως πρόθεμα και χρήση της τιμής από τα πεδία του MARC (π.χ. 001, 009 κλπ.) που παρουσιάστηκαν νωρίτερα ως κατάληξη*». Τα URIs που δημιουργούνται δεν αποθηκεύονται στον παραδοσιακό κατάλογο της βιβλιοθήκης (δηλ. τον OPAC). Τα URIs των Συνδεδεμένων Δεδομένων που ορίζονται σε απομακρυσμένα σημεία, σε κάποιες περιπτώσεις, αποθηκεύονται και στον παραδοσιακό κατάλογο της βιβλιοθήκης και στις

¹³⁶ Ακριβής μετάφραση από το πρωτότυπο κείμενο: *"the information required to locate an electronic item. The field may be used in an authority record to provide supplementary information available electronically about the entity for which the record was created. The information identifies the electronic location containing the item or from which it is available. It also contains information to retrieve the item by the access method identified in the first indicator position. It can be used to generate notes relating to mode of access"*. (Willer, 2009b, σ. 259).

νές υπηρεσίες που βασίζονται στα Συνδεδεμένα Δεδομένα. Στις επόμενες ενότητες, παρουσιάζεται μια λεπτομερής ανάλυση της κάθε βιβλιοθήκης ξεχωριστά.

7.3.1. Βιβλιοθήκη του Κογκρέσου

Η Βιβλιοθήκη του Κογκρέσου, όπως έχει ήδη αναφερθεί (βλέπε ενότητα 2.3.1) διατηρεί ένα αρχείο καθιερώσεων, που ονομάζεται Θεματικές Επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου (LCSH), το οποίο περιέχει τις καθιερωμένες εγγραφές σε MARC21. Εκτός από το πεδίο 001, το οποίο χρησιμοποιείται για να περιγράψει μοναδικά κάθε καθιερωμένη εγγραφή, η Βιβλιοθήκη του Κογκρέσου χρησιμοποιεί, επίσης, το πεδίο 010 ως τον Αριθμό Ελέγχου της Βιβλιοθήκης του Κογκρέσου (*"Library of Congress Control Number – LCCN"*). Αυτός είναι ένας μοναδικός αριθμός που φιλοξενεί προσδιοριστές από την υπηρεσία LCCN permalink¹³⁷. Πιο συγκεκριμένα, η Βιβλιοθήκη του Κογκρέσου δημιουργεί ένα permalink για κάθε καθιερωμένη εγγραφή προσαρτώντας ένα ειδικό πρόθεμα (δηλ. *"http://lcn.loc.gov/"*) στην αντίστοιχη καθιερωμένη εγγραφή. Για παράδειγμα, σύμφωνα με την προαναφερθείσα διαδικασία, το permalink που προκύπτει για την καθιερωμένη εγγραφή *"Accounting"* είναι *"http://lcn.loc.gov/sh85000411"*.

Το LCCN permalink δεν πρέπει να θεωρείται URI, το οποίο προσδιορίζει μοναδικά την αντίστοιχη καθιερωμένη εγγραφή μέσα στα πλαίσια του πεδίου εφαρμογής του κινήματος των Συνδεδεμένων Δεδομένων. Αντ' αυτού, για αυτό το σκοπό, η Βιβλιοθήκη του Κογκρέσου δημιούργησε την υπηρεσία *"id.loc.gov"* για να παρέχει πρόσβαση στα δεδομένα της μέσα από τη χρήση κοινών προτύπων και λεξιλογίων. Αυτή η υπηρεσία παρέχει αναλύσεις (resolvability) σε τιμές και λεξιλόγια με την ανάθεση URIs. Κάθε URI αποτελείται από ένα συγκεκριμένο πρόθεμα (δηλ. *"http://id.loc.gov/"*), ένα μονοπάτι που περιγράφει τα διάφορα σύνολα δεδομένων που περιέχονται στην υπηρεσία (π.χ. *"authorities/subjects"*, *"authorities/names"* κ.ά.), έναν κωδικό που είναι σύμφωνος με το κάθε σύνολο δεδομένων (π.χ. *"sh"* για τις θεματικές επικεφαλίδες, *"n"* για το Αρχείο Καθιερωμένων Ονομάτων (Name Authority File – NAF¹³⁸), κ.ά.) και την τιμή του αντίστοιχου 010 πεδίου. Για παράδειγμα, η καθιερωμένη εγγραφή

¹³⁷ Lcn.loc.gov, (2015). *Library of Congress LCCN Permalink: Frequently Asked Questions*. [online] Διαθέσιμο στο: <http://lcn.loc.gov> [Πρόσβαση: 15/07/2015].

¹³⁸ Id.loc.gov, (2015). *Library of Congress Names - LC Linked Data Service (Library of Congress)*. [online] Διαθέσιμο στο: <http://id.loc.gov/authorities/names.html> [Πρόσβαση: 30/07/2015].

“Accounting” προσδιορίζεται από το URI “<http://id.loc.gov/authorities/subjects/sh85000411>”.

Επιπλέον, η υπηρεσία Συνδεδεμένων Δεδομένων των LCSH παρέχει συνδέσμους απομακρυσμένων πηγών από άλλους παρόχους Συνδεδεμένων Δεδομένων, όπως είναι το RAMEAU, National Agricultural Library - NAL¹³⁹, GND κ.ά.

Τέλος, το 2011, η Βιβλιοθήκη του Κογκρέσου δημιούργησε την πρωτοβουλία Bibliographic Framework Initiative (Miller et al., 2012). Η πρωτοβουλία αυτή στοχεύει στη δημιουργία ενός νέου περιβάλλοντος για τη βιβλιογραφική περιγραφή στις βιβλιοθήκες. Το νέο μοντέλο ονομάζεται Bibliographic Framework – BIBFRAME¹⁴⁰ και αναμένεται να αντικαταστήσει το πρότυπο MARC και να καταστήσει τις συλλογές των βιβλιοθηκών διαθέσιμες ως μέρος του σημασιολογικού ιστού. Το BIBFRAME λειτουργεί, επίσης, ως ένα νέο μοντέλο ή οντολογία για την περιγραφή βιβλιογραφικών και καθιερωμένων δεδομένων.

7.3.2. Βρετανική Βιβλιοθήκη

Η Βρετανική Βιβλιοθήκη (British Library – BL)¹⁴¹ έχει ταυτίσει τις Βρετανικές καθιερωμένες εγγραφές, οι οποίες δημιουργήθηκαν για τη Βρετανική Εθνική Βιβλιογραφία (British National Bibliography – BNB) από το 1971-1987 και από το 1995 και μετά¹⁴², με τις LCSH. Η τρέχουσα εφαρμογή των LCSH στη Βρετανική Βιβλιοθήκη ακολουθεί τις αρχές, τις πολιτικές και τις οδηγίες που δίνονται από την έκδοση της Βιβλιοθήκης του Κογκρέσου “*Subject Cataloging Manual: Subject Headings*” (Library of Congress, 2007). Η Βρετανική Βιβλιοθήκη παλαιότερα χρησιμοποιούσε το UKMARC¹⁴³ ως πρότυπο για την καταλογογράφηση των καθιερωμένων εγγραφών (Hill, 2002), το οποίο δεν υποστηρίζεται πλέον και έχει αντικατασταθεί από το MARC21 (Δεκέμβριος 2008).

¹³⁹ Nal.usda.gov, (2015). *National Agricultural Library | United States Department of Agriculture*. [online] Διαθέσιμο στο: <http://www.nal.usda.gov> [Πρόσβαση: 15/07/2015].

¹⁴⁰ Bibframe.org, (2015). *Overview (Bibliographic Framework Initiative Technical Site - BIBFRAME.ORG)*. [online] Διαθέσιμο στο: <http://bibframe.org/> [Πρόσβαση: 15/07/2015].

¹⁴¹ Bl.uk, (2015). *THE BRITISH LIBRARY - The world's knowledge*. [online] Διαθέσιμο στο: <http://www.bl.uk/> [Πρόσβαση: 15/07/2015].

¹⁴² Bl.uk, (2015). *British Library: Metadata services Standards: Subject Access in British Library Bibliographic Records*. [online] Διαθέσιμο στο: <http://www.bl.uk/bibliographic/subject.html> [Πρόσβαση: 15/07/2015].

¹⁴³ Bl.uk, (2015). *British Library: Metadata services Standards UK Marc Manual*. [online] Διαθέσιμο στο: <http://www.bl.uk/bibliographic/ukmarc.html> [Πρόσβαση: 15/07/2015].

Η Βρετανική Βιβλιοθήκη προσφάτως δημιούργησε τη BNB ως Συνδεδεμένα Δεδομένα¹⁴⁴. Η υπηρεσία εκμεταλλεύεται την πληροφορία από το πεδίο 150 του MARC21 για τη δημιουργία URIs των Συνδεδεμένων Δεδομένων που ορίζονται τοπικά. Για παράδειγμα, η καθιερωμένη εγγραφή "*Accounting*" έχει το URI "<http://bnb.data.bl.uk/doc/concept/lcsh/Accounting>". Ένα τέτοιο URI των Συνδεδεμένων Δεδομένων δεν αποθηκεύεται μέσα στον κατάλογο της βιβλιοθήκης. Οι καθιερωμένες εγγραφές της Βρετανικής Βιβλιοθήκης παρέχουν συνδέσμους Συνδεδεμένων Δεδομένων σε απομακρυσμένες πηγές που προέρχονται από τις LCSH.

7.3.3. Εθνική Βιβλιοθήκη της Γαλλίας

Η Εθνική Βιβλιοθήκη της Γαλλίας (Bibliothèque Nationale de France – BNF)¹⁴⁵ χρησιμοποιεί το RAMEAU¹⁴⁶ για να ορίζει τις καθιερωμένες εγγραφές της. Το RAMEAU βασίζεται στο "*Subject Cataloging Manual: Subject Headings*" που εκδίδεται από τη Βιβλιοθήκη του Κογκρέσου (Library of Congress, 2007). Η Εθνική Βιβλιοθήκη της Γαλλίας χρησιμοποιεί το UNIMARC ως το πρότυπο για την καταλογογράφηση των καθιερωμένων εγγραφών. Το πεδίο 001 χρησιμοποιείται για να προσδιοριστεί μοναδικά μια καθιερωμένη εγγραφή. Εκτός από το πεδίο 001, η Εθνική Βιβλιοθήκη της Γαλλίας χρησιμοποιεί και το πεδίο 009 ως σταθερό προσδιοριστή της καθιερωμένης εγγραφής. Για παράδειγμα, η καθιερωμένη εγγραφή "*Etablissements religieux*" αντιστοιχεί στο "<http://catalogue.bnf.fr/ark:/12148/cb11954011h>".

Από τον Ιούλιο του 2011, οι καθιερωμένες εγγραφές του RAMEAU είναι διαθέσιμες ως Συνδεδεμένα Δεδομένα¹⁴⁷. Πιο συγκεκριμένα, το έργο data.bnf.fr στοχεύει στο να διαθέσει τα δεδομένα της Εθνικής Βιβλιοθήκης της Γαλλίας (δηλ. συγγραφείς, έργα, κ.ά.) (Wenz et al., 2013) ως μέρος του σηματολογικού ιστού. Το data.bnf.fr επιτρέπει: α) την πρόσβαση στις πηγές της Εθνικής Βιβλιοθήκης της Γαλλίας απευθείας μέσω μιας ιστοσελίδας και β) την πρόσβαση σε εξωτερικές πηγές από άλλους παρόχους Συνδεδεμένων Δεδομένων, όπως είναι η DBpedia, το VIAF κ.ά. Το μοναδικό URI μιας

¹⁴⁴ Bnb.data.bl.uk, (2015). *The British National Bibliography as Linked Open Data*. [online] Διαθέσιμο στο: <http://bnb.data.bl.uk> [Πρόσβαση: 15/07/2015].

¹⁴⁵ Bnf.fr, (2015). *Bibliothèque nationale de France*. [online] Διαθέσιμο στο: <http://www.bnf.fr/fr/acc/x.accueil.html> [Πρόσβαση: 15/07/2015].

¹⁴⁶ Rameau.bnf.fr, (2015). *BnF-RAMEAU: Accueil*. [online] Διαθέσιμο στο: <http://rameau.bnf.fr/> [Πρόσβαση: 15/07/2015].

¹⁴⁷ Data.bnf.fr, (2015). *BNF authorities linked data*. [online] Διαθέσιμο στο: <http://data.bnf.fr/> [Πρόσβαση: 15/07/2015].

καθιερωμένης εγγραφής προέρχεται από τον αρχικό προσδιοριστή της εκάστοτε εγγραφής. Πιο συγκεκριμένα, τα URIs που προκύπτουν περιέχουν τον προσδιοριστή Archival Resource Key – ARK¹⁴⁸ που βασίζεται στο πεδίο 009. Για παράδειγμα, η καθιερωμένη εγγραφή "*Etablissements religieux*" αντιστοιχεί στο URI των Συνδεδεμένων Δεδομένων: "<http://data.bnf.fr/ark:/12148/cb11954011h>".

7.3.4. Εθνική Βιβλιοθήκη της Γερμανίας

Η Εθνική Βιβλιοθήκη της Γερμανίας (Deutsche National Bibliothek – DNB)¹⁴⁹ χρησιμοποιεί το MARC21 για τις καθιερωμένες εγγραφές της. Η βιβλιοθήκη παρέχει, επίσης, το αρχείο Gemeinsame Normdatei – GND¹⁵⁰, το οποίο είναι το εθνικό αρχείο καθιερώσεων για την οργάνωση των προσωπικών ονομάτων, θεματικών επικεφαλίδων και συλλογικών οργάνων από τον κατάλογο. Χρησιμοποιείται κυρίως για την τεκμηρίωση σε βιβλιοθήκες και αρχεία. Το αρχείο GND (Haffner, 2012) άρχισε να λειτουργεί τον Απρίλιο του 2012 και ενσωματώνει το περιεχόμενο των ακόλουθων καθιερωμένων αρχείων, τα οποία σταμάτησαν να ανανεώνονται από εκεί και έπειτα: α) Αρχείο Καθιερωμένων Ονομάτων (Personennamendatei - PND), β) Αρχείο Καθιερωμένων Συλλογικών Οργάνων (Gemeinsame Körperschaftsdatei - GKD), γ) Αρχείο Καθιερωμένων Θεματικών Επικεφαλίδων (Schlagwortnormdatei - SWD) και δ) Αρχείο Ομοιόμορφων Τίτλων του Γερμανικού Μουσικού Αρχείου (Einheitssachtitel-Datei des Deutschen Musikarchivs - DMA-EST). Οι αρχικοί προσδιοριστές των καθιερωμένων εγγραφών μεταφέρθηκαν, επίσης, στο νέο ενοποιημένο κατάλογο. Για παράδειγμα, η καθιερωμένη εγγραφή από το SWD "*Management Accounting*" έχει το μοναδικό προσδιοριστή "4125415-6"¹⁵¹ και στο GND και στο SWD. Το URI που αντιστοιχεί στη συγκεκριμένη καθιέρωση είναι το "<http://d-nb.info/gnd/4125415-6>".

Η Εθνική Βιβλιοθήκη της Γερμανίας προγραμματίζει να προσφέρει μια υπηρεσία Συνδεδεμένων Δεδομένων που θα επιτρέπει στη σημασιολογική κοινότητα να χρησιμοποιεί όλο το υλικό των εθνικών βιβλιογραφικών δεδομένων, συμπεριλαμβανομένων

¹⁴⁸ Confluence.ucop.edu, (2015). *ARK - Curation - Confluence*. [online] Διαθέσιμο στο: <https://confluence.ucop.edu/display/Curation/ARK> [Πρόσβαση: 15/07/2015].

¹⁴⁹ Dnb.de, (2015). *Deutsche National Bibliothek - Home*. [online] Διαθέσιμο στο: <http://www.dnb.de/> [Πρόσβαση: 15/07/2015].

¹⁵⁰ Dnb.de, (2015). *Integrated Authority File (GND)*. [online] Διαθέσιμο στο: <http://www.dnb.de/EN/gnd> [Πρόσβαση: 15/07/2015].

¹⁵¹ Swb.bsz-bw.de, (2015). *SWB Online-Katalog - Start/welcome*. [online] Διαθέσιμο στο: <http://swb.bsz-bw.de/> [Πρόσβαση: 15/07/2015].

και των καθιερωμένων δεδομένων (German National Library, 2013). Τέλος, η βιβλιοθήκη παρέχει συσχετίσεις με το RAMEAU, τις LCSH και τον θησαυρό STW thesaurus of economics¹⁵².

7.3.5. Εθνική Βιβλιοθήκη της Ισπανίας

Η Εθνική Βιβλιοθήκη της Ισπανίας (Biblioteca Nacional de España – BNE)¹⁵³ χρησιμοποιεί το MARC21 για τις καθιερωμένες εγγραφές της. Πριν το MARC21, η βιβλιοθήκη χρησιμοποιούσε το IBERMARC¹⁵⁴. Η μεταφορά στο MARC21 αποφασίστηκε για να εξυπηρετηθεί η διεθνοποίηση και η τυποποίηση των βιβλιογραφικών εγγραφών, των αρχείων και των καθιερώσεων της Εθνικής Βιβλιοθήκης της Ισπανίας. Ο μοναδικός προσδιοριστής για κάθε καθιερωμένη εγγραφή παρέχεται στο πεδίο 001. Επιπλέον, κάθε καθιερωμένη εγγραφή έχει ένα πεδίο 670, το οποίο περιέχει την καθιερωμένη εγγραφή που προέρχεται από το RAMEAU, τις LCSH και το TMA¹⁵⁵. Για παράδειγμα, η καθιερωμένη εγγραφή "*Tecnología limpia*" είναι ίδια με την καθιερωμένη εγγραφή "*Green technology*" στις LCSH, "*Tecnologías blandas*" στο TMA και "*Technologie douce*" στο RAMEAU. Επιπλέον, υπάρχει ένα πεδίο 024, το οποίο φιλοξενεί το URI της αντίστοιχης LCSH.

Η Εθνική Βιβλιοθήκη της Ισπανίας ξεκίνησε ένα κοινό έργο, το οποίο εμπλέκει και το Ontology Engineering Group – OEG¹⁵⁶ για τον εμπλουτισμό του σημασιολογικού ιστού με βιβλιογραφικά δεδομένα από τον κατάλογο της. Το έργο ονομάζεται "*datos.bne.es*" (Vila-Suero, Villazón-Terrazas & Gómez-Pérez, 2013). Η υπηρεσία που παρέχεται περιέχει πληροφορία όχι μόνο από τις βιβλιογραφικές εγγραφές, αλλά και από τις καθιερωμένες εγγραφές. Κάθε URI ακολουθεί ένα συγκεκριμένο πρότυπο. Για παράδειγμα, το URI των Συνδεδεμένων Δεδομένων "*http://datos.bne.es/tema/XX544630.html*", που αναφέρεται στην καθιερωμένη εγγραφή "*Tecnología limpia*" αποτελείται από το πεδίο 001 (δηλ. "XX544630") ως κατάληξη και το

¹⁵² Zbw.eu, (2015). *STW Thesaurus for Economics: Home*. [online] Διαθέσιμο στο: <http://zbw.eu/stw/versions/latest/about> [Πρόσβαση: 15/07/2015].

¹⁵³ Bne.es, (2015). *Biblioteca Nacional*. [online] Διαθέσιμο στο: <http://www.bne.es/> [Πρόσβαση: 15/07/2015].

¹⁵⁴ Bne.es, (2015). *Normas Internacionales - Formato IBERMARC*. *Biblioteca Nacional de España*. [online] Διαθέσιμο στο: <http://www.bne.es/es/Inicio/Perfiles/Bibliotecarios/Procesos-tecnicos/NormasInternacionales/FormatoIBERMARC/> [πρόσβαση στο: 15/07/2015].

¹⁵⁵ Tesouro de Medio Ambientedel MOPU. [Madrid]: Ministerio de Obras Públicas y Urbanismo, 1990.

¹⁵⁶ Oeg-upm.net, (2015). *Ontology Engineering Group*. [online] Διαθέσιμο στο: <http://www.oeg-upm.net/> [Πρόσβαση: 15/07/2015].

“<http://datos.bne.es/tema/>” ως πρόθεμα. Η υπηρεσία Συνδεδεμένων Δεδομένων παρέχει συσχετίσεις με τις LCSH. Επιπλέον, παρέχει συσχετίσεις με το αρχείο καθιερωμένων ονομάτων που προέρχονται από το LIBRIS, το GND, τη DBpedia και το idRef¹⁵⁷.

7.3.6. Εθνική Βιβλιοθήκη της Σουηδίας

Η Εθνική Βιβλιοθήκη της Σουηδίας (Kungliga Biblioteket – KB)¹⁵⁸ χρησιμοποιεί το MARC21 για τις καθιερωμένες εγγραφές της. Η βιβλιοθήκη αποτελείται από δύο κύριους καταλόγους, που περιλαμβάνουν την πλειοψηφία της συλλογής της Εθνικής Βιβλιοθήκης της Σουηδίας, δηλαδή τη Regina¹⁵⁹ και τη Swedish Media Database – SMDB¹⁶⁰. Επίσης, παρέχει πρόσβαση στον εθνικό συλλογικό κατάλογο της Σουηδίας και αποτελεί ένα εργαλείο αναζήτησης με τίτλους από βιβλιοθήκες πανεπιστημίων, ερευνητικών κέντρων και δημόσιων υπηρεσιών. Το πεδίο που προσδιορίζει μοναδικά μια εγγραφή σε όλους τους καταλόγους είναι το 001. Η βιβλιοθήκη δεν παρέχει απευθείας πρόσβαση στο καθιερωμένο αρχείο.

Από το 2008, ο συλλογικός κατάλογος της Σουηδίας (δηλ. LIBRIS¹⁶¹) είναι διαθέσιμος ως Συνδεδεμένα Δεδομένα. Περιέχει συνδέσμους με τη Wikipedia, τη DBpedia, τα αρχεία καθιερώσεων της Βιβλιοθήκης του Κογκρέσου (ονόματα και θέματα) και το VIAF (Malmsten, 2009). Το πεδίο 001 χρησιμοποιείται για τη δημιουργία των URIs της υπηρεσίας των Συνδεδεμένων Δεδομένων. Για παράδειγμα, η καθιερωμένη εγγραφή “*Mödrar*”, που προσδιορίζεται από το “154863”, αντιστοιχεί στο ακόλουθο URI: “<http://libris.kb.se/auth/154863>”.

¹⁵⁷ Idref.fr, (2015). *IdRef - Identifiants et Référentiels pour l'Enseignement Supérieur et la Recherche*. [online] Διαθέσιμο στο: <http://www.idref.fr/autorites/autorites.html> [Πρόσβαση: 15/07/2015].

¹⁵⁸ Kb.se, (2015). *Startsida - Kungliga Biblioteket*. [online] Διαθέσιμο στο: <http://www.kb.se/> [Πρόσβαση: 15/07/2015].

¹⁵⁹ Kb.se, (2015). *Regina - Kungliga biblioteket*. [online] Διαθέσιμο στο: <http://www.kb.se/samlingarna/Regina> [Πρόσβαση: 15/07/2015].

¹⁶⁰ Smdb.kb.se, (2015). *SMDB / Svensk Mediedatabas (SMDB)*. [online] Διαθέσιμο στο: <http://smdb.kb.se/english/> [Πρόσβαση: 15/07/2015].

¹⁶¹ Libris.kb.se, (2015). *LIBRIS*. [online] Διαθέσιμο στο: <http://libris.kb.se> [Πρόσβαση: 15/07/2015].

7.3.7. Εθνική Βιβλιοθήκη της Ουγγαρίας

Τέλος, η Εθνική Βιβλιοθήκη της Ουγγαρίας (National Széchényi Library – NSL)¹⁶² χρησιμοποιεί το MARC21 για τις Ουγγρικές καθιερωμένες εγγραφές (Edelstein et al., 2013). Το πεδίο 001 χρησιμοποιείται για προσδιορίσει μοναδικά κάθε εγγραφή.

Για τη δημιουργία των URIs χρησιμοποιείται το πρόθεμα “<http://nektar.oszk.hu/auth/>” μαζί με την πληροφορία από το πεδίο 150. Για παράδειγμα, η καθιερωμένη εγγραφή “*elbeszélés*” αντιστοιχεί στο URI: “<http://nektar.oszk.hu/auth/elbeszélés>”. Η Εθνική Βιβλιοθήκη της Ουγγαρίας παρέχει συσχετίσεις με τη DBpedia και το VIAF.

Στην επόμενη ενότητα, παρουσιάζεται μια λεπτομερής περιγραφή της διαδικασίας που απαιτείται για τη δημοσίευση των καθιερωμένων πληροφοριών του καταλόγου μιας βιβλιοθήκης σε μορφή κατάλληλη για το σύννεφο των Συνδεδεμένων Δεδομένων καθώς και οι δυνατότητες που παρέχουν οι βιβλιοθήκες στους χρήστες για την πρόσβαση στα δεδομένα που είναι σε Συνδεδεμένα Δεδομένα.

7.4. Δημοσίευση καθιερωμένων εγγραφών του MARC ως Συνδεδεμένα Δεδομένα

Μέχρι τώρα, ερευνήθηκε η δυνατότητα των παραδοσιακών υπηρεσιών του OPAC να λειτουργήσουν ως αποθετήριο που να είναι ικανό να αποθηκεύσει URIs των Συνδεδεμένων Δεδομένων. Αυτή η ενότητα εστιάζει στην εκμετάλλευση αυτών των δεδομένων για τη δημιουργία νέων υπηρεσιών Συνδεδεμένων Δεδομένων από τις βιβλιοθήκες. Σε αυτό το πλαίσιο, όπως οποιοσδήποτε άλλος πάροχος που επιθυμεί να συμμετάσχει στο κίνημα των Συνδεδεμένων Δεδομένων, έτσι και οι βιβλιοθήκες θα πρέπει να συμφωνούν με συγκεκριμένες απαιτήσεις. Όπως ήδη αναφέρθηκε στην ενότητα 4.3, σύμφωνα με τον Berners-Lee (2006), τον ιδρυτή του κινήματος των Συνδεδεμένων Δεδομένων, υπάρχει ένας αριθμός κανόνων για τη δημοσίευση δεδομένων με τέτοιο τρόπο έτσι ώστε όλα τα δημοσιευμένα δεδομένα να είναι μέρος ενός μοναδικού παγκόσμιου χώρου δεδομένων.

Σε αυτό το πλαίσιο, οι κυριότερες εθνικές βιβλιοθήκες έχουν δημιουργήσει υπηρεσίες για αυτό το σκοπό πάνω από τους OPAC τους. Αυτές οι υπηρεσίες βασίζονται σε

¹⁶² Regi.oszk.hu, (2015). *Hungarian National Széchényi Library*. [online] Διαθέσιμο στο: http://regi.oszk.hu/index_hu.htm [Πρόσβαση: 15/07/2015].

ειδικά αποθετήρια δεδομένων που περιέχουν την πληροφορία των καθιερωμένων εγγραφών, οι οποίες είναι μοντελοποιημένες βάσει των κανόνων των Συνδεδεμένων Δεδομένων. Οι ενότητες που ακολουθούν περιγράφουν τα δομικά στοιχεία των εν λόγω υπηρεσιών.

7.4.1. Μοντελοποίηση δεδομένων

Σύμφωνα με τον 3^ο κανόνα των Συνδεδεμένων Δεδομένων (βλέπε ενότητα 4.3), οι πάροχοι δεδομένων που επιθυμούν να συμμετέχουν στο κίνημα των Συνδεδεμένων Δεδομένων θα πρέπει να δημοσιεύουν τα δεδομένα τους σε RDF. Το μοντέλο δεδομένων RDF είναι σχεδιασμένο για χρήση στα πλαίσια του διαδικτύου. Επιπλέον, όταν πρέπει να εκφραστούν οι σχέσεις ανάμεσα στις καθιερωμένες εγγραφές (δηλ. κατηγορήματα), ο τομέας των βιβλιοθηκών ως επί το πλείστον χρησιμοποιεί το SKOS (βλέπε ενότητα 4.4.4).

Στον πίνακα 7.3, παρουσιάζονται τα κατηγορήματα του λεξιλογίου του SKOS, τα οποία είναι σημασιολογικά ισοδύναμα με τα πεδία του MARC (Summers et al., 2008; Plassard, 2001) και χρησιμοποιούνται πιο συχνά στα καθιερωμένα αρχεία.

Πίνακας 7.3: Πεδία UNIMARC/MARC21 και τα σημασιολογικά ισοδύναμα κατηγορήματα του SKOS

Πεδία MARC21	Πεδία UNIMARC	Ερμηνεία πεδίου	Κατηγορήματα SKOS
001 or 010	001 or 009	Control Number	rdf:about
150	250	Topical Term	skos:prefLabel
450	450	See From Tracing	skos:altLabel
550 \$w (ind. 'g') ¹⁶³	550 \$5 (ind. 'g') ¹⁶⁴	See Also From Tracing	skos:broader
550 \$w (ind. 'h') ¹⁶⁵	550 \$5 (ind. 'h') ¹⁶⁶		skos:narrower
550 \$w (w/o ind. 'g' or 'h') ¹⁶⁷	550 \$w (w/o ind. 'g' or 'h') ¹⁶⁸		skos:related

¹⁶³ Να χρησιμοποιηθεί αυτή η ιδιότητα όταν το υποπεδίο \$w έχει δείκτη 'g'.

¹⁶⁴ Να χρησιμοποιηθεί αυτή η ιδιότητα όταν το υποπεδίο \$5 έχει δείκτη 'g'.

¹⁶⁵ Να χρησιμοποιηθεί αυτή η ιδιότητα όταν το υποπεδίο \$w έχει δείκτη 'h'.

¹⁶⁶ Να χρησιμοποιηθεί αυτή η ιδιότητα όταν το υποπεδίο \$5 έχει δείκτη 'h'.

¹⁶⁷ Να χρησιμοποιηθεί αυτή η ιδιότητα όταν το υποπεδίο \$w εμφανίζεται χωρίς του δείκτες 'g' ή 'h'.

¹⁶⁸ Να χρησιμοποιηθεί αυτή η ιδιότητα όταν το υποπεδίο \$5 εμφανίζεται χωρίς του δείκτες 'g' ή 'h'.

Για παράδειγμα, η ακόλουθη τριπλέτα δηλώνει ότι το URI "<http://id.loc.gov/authorities/subjects/sh85000411>" αντιστοιχεί στην καθιερωμένη οντότητα με την ετικέτα "*Accounting*":

```
<http://id.loc.gov/authorities/subjects/sh85000411>
```

```
skos:prefLabel
```

```
"Accounting"@EN .
```

Κατά παρόμοιο τρόπο, η ακόλουθη τριπλέτα δηλώνει ότι η καθιερωμένη οντότητα με την ετικέτα "*Accounting*" έχει στενότερη έννοια από την οντότητα που αντιστοιχεί στο URI "<http://id.loc.gov/authorities/subjects/sh85009477>":

```
< http://id.loc.gov/authorities/subjects/sh85000411 >
```

```
skos:narrower
```

```
<http://id.loc.gov/authorities/subjects/sh85009477> .
```

Έτσι, τα δεδομένα σε RDF που αντιστοιχούν σε ένα καθιερωμένο αρχείο μέσα στα πλαίσια μιας βιβλιοθήκης είναι ουσιαστικά ένα αποθετήριο τριπλετών που αποτελείται από τριπλέτες όμοιες με τα παραπάνω παραδείγματα.

7.4.2. Πρόσβαση στα δεδομένα

Η πρόσβαση στα δεδομένα των βιβλιοθηκών που παρέχονται σε μορφή Συνδεδεμένων Δεδομένων μπορεί να γίνει με τους ακόλουθους τρεις τρόπους: α) μέσω ενός τελικού σημείου SPARQL, β) μέσω αρχείων RDF dump και γ) μέσω URI dereferencing (για περισσότερες πληροφορίες βλέπε ενότητα 4.4.5).

Ο πίνακας 7.4 παρουσιάζει τον τρόπο με τον οποίο οι εθνικές βιβλιοθήκες ανά τον κόσμο δημοσιεύουν τα δεδομένα τους σε μορφή RDF¹⁶⁹.

¹⁶⁹ Αυτή η πληροφορία συνάγεται με την επίσκεψη των αντίστοιχων online υπηρεσιών των βιβλιοθηκών.

Πίνακας 7.4: Εθνικές βιβλιοθήκες και υπηρεσίες προς την κοινότητα των Συνδεδεμένων Δεδομένων

	Βιβλιοθήκη	Τελικό σημείο SPARQL	Αρχεία RDF Dump	Dereferencing
1	BL (BNB) ¹⁷⁰	✓		
2	BNF (RAMEAU) ¹⁷¹		✓	✓
3	DNB (GND) ^{172, 173}		✓	✓
4	BNE ¹⁷⁴	✓		
5	KB (LIBRIS) ¹⁷⁵	✓		
6	LoC (LCSH) ¹⁷⁶		✓	✓
7	NSL ¹⁷⁷	✓		

Περισσότερες πληροφορίες σχετικά με τις υπηρεσίες σε Συνδεδεμένα Δεδομένα που κάθε βιβλιοθήκη προσφέρει παρέχονται μέσω του datahub.io¹⁷⁸. Αυτή η υπηρεσία παρέχει ένα ελεύθερο, δημόσιο τομέα μητρώων όλων των παρόχων Συνδεδεμένων Δεδομένων στο διαδίκτυο μαζί με την πληροφορία για τον τρόπο που οι πάροχοι κάνουν διαθέσιμα τα δεδομένα τους. Το Datahub.io λειτουργεί ως ένας συλλέκτης όπου οι πάροχοι μπορούν να εγγραφούν τα δεδομένα τους και να τα διαθέσουν στην κοινότητα του διαδικτύου ως Συνδεδεμένα Δεδομένα. Επίσης, φέρνει κοντά συλλογές από συγκεκριμένους τομείς.

¹⁷⁰ Bnb.data.bl.uk, (2015). *SPARQL editor | The British National Bibliography*. [online] Διαθέσιμο στο: <http://bnb.data.bl.uk/flint-sparql> [Πρόσβαση: 15/07/2015].

¹⁷¹ Data.bnf.fr, (2015). *Semantic Web and Data Model*. [online] Διαθέσιμο στο: <http://data.bnf.fr/en/semanticweb> [Πρόσβαση: 15/07/2015].

¹⁷² Dnb.de, (2015). *Deutsche Nationalbibliothek - Linked Data Service*. [online] Διαθέσιμο στο: http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html [Πρόσβαση: 15/07/2015].

¹⁷³ Υπάρχει μόνο ένα πειραματικό SPARQL endpoint όσον αφορά στα αρχεία καθιερώσεων της DNB: Wiss-ki.eu, (2015). *SPARQL Endpoints for the Authority Files of the German National Library*. [online] Διαθέσιμο στο: <http://wiss-ki.eu/authorities/gnd/> [Πρόσβαση: 15/07/2015].

¹⁷⁴ Datos.bne.es, (2015). *BNE Virtuoso SPARQL Query Editor*. [online] Διαθέσιμο στο: <http://datos.bne.es/sparql> [Πρόσβαση: 15/07/2015].

¹⁷⁵ Libris.kb.se, (2015). *LIBRIS SPARQL Endpoint*. [online] Διαθέσιμο στο: <http://libris.kb.se/sparql> [Πρόσβαση: 15/07/2015].

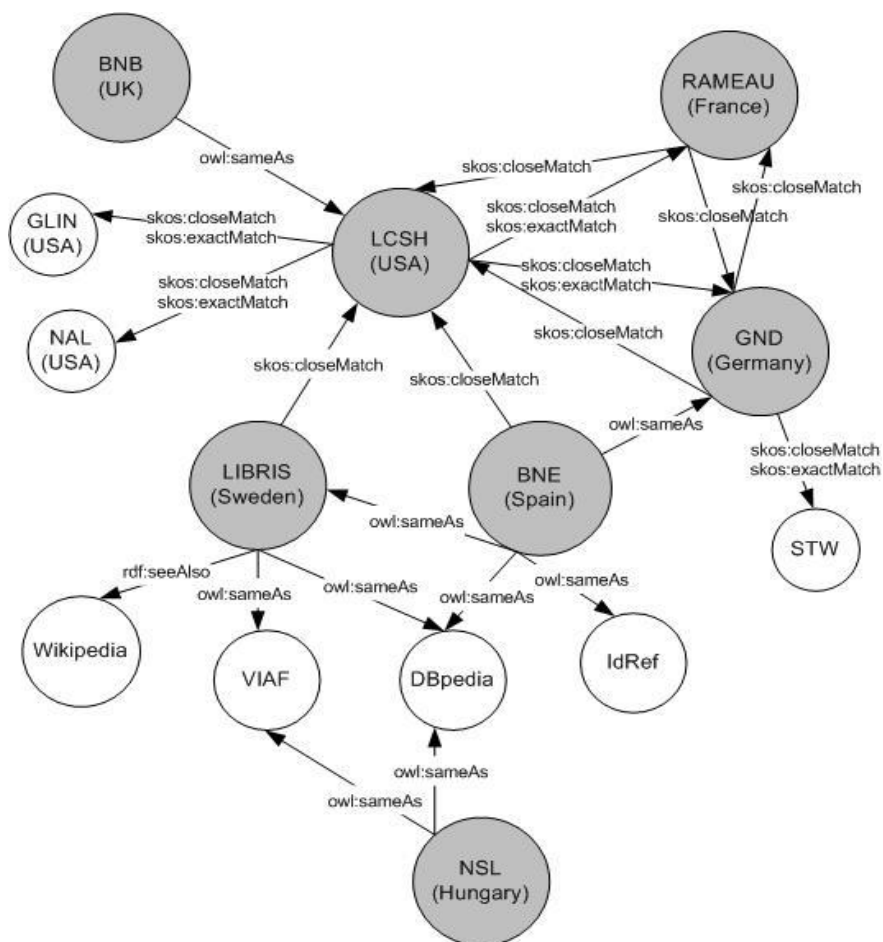
¹⁷⁶ Id.loc.gov, (2015). *LC Linked Data Service: Bulk downloads (Library of Congress)*. [online] Διαθέσιμο στο: <http://id.loc.gov/download/> [Πρόσβαση: 15/07/2015].

¹⁷⁷ Setaria.oszk.hu, (2015). *NSL Virtuoso SPARQL Query Form*. [online] Διαθέσιμο στο: <http://setaria.oszk.hu/sparql> [Πρόσβαση: 15/07/2015].

¹⁷⁸ Datahub.io, (2015). *Welcome - the Datahub.io*. [online] Διαθέσιμο στο: <http://datahub.io> [Πρόσβαση: 15/07/2015].

7.4.3. Διασύνδεση δεδομένων

Όπως αναφέρθηκε νωρίτερα στην ενότητα 4.3, ο τέταρτος κανόνας των Συνδεδεμένων Δεδομένων αναφέρει: «*Συμπεριέλαβε συνδέσμους σε άλλα URIs, έτσι ώστε και άλλα πράγματα να μπορούν να ανακαλυφθούν*». Με άλλα λόγια, οι υπηρεσίες των βιβλιοθηκών σε Συνδεδεμένα Δεδομένα θα πρέπει να περιέχουν τριπλέτες που αποτελούνται από URIs που έχουν οριστεί εξωτερικά από άλλες υπηρεσίες. Η εικόνα 7.2 απεικονίζει τη συνδεσιμότητα ανάμεσα στις σημαντικότερες υπηρεσίες Συνδεδεμένων Δεδομένων των βιβλιοθηκών. Οι εθνικές βιβλιοθήκες απεικονίζονται ως κόμβοι με γκρι χρώμα, ενώ οι άλλες υπηρεσίες Συνδεδεμένων Δεδομένων που σχετίζονται με τις βιβλιοθήκες απεικονίζονται με λευκό χρώμα¹⁷⁹. Επιπλέον, τα τόξα φέρουν πάνω τους το κατηγορήμα που χρησιμοποιείται για να διασυνδέσει πηγές από διαφορετικές υπηρεσίες.



Εικόνα 7.2: Διασύνδεση δεδομένων βιβλιοθηκών στο σύννεφο των Συνδεδεμένων Δεδομένων

¹⁷⁹ Αυτή η πληροφορία συνάγεται με την επίσκεψη των αντίστοιχων online υπηρεσιών των βιβλιοθηκών.

Η εικόνα 7.2 δείχνει ότι μόνο λίγα κατηγορήματα χρησιμοποιούνται για τη διασύνδεση καθιερωμένων εγγραφών ανάμεσα στις βιβλιοθήκες παγκοσμίως. Πιο συγκεκριμένα, 2 από τις 7 βιβλιοθήκες (δηλ. BNB και NSL) χρησιμοποιούν το κατηγορήμα *"owl:sameAs"* για να διασυνδέσουν τοπικές καθιερωμένες οντότητες με απομακρυσμένες. 4 από 7 βιβλιοθήκες (δηλ. LCSH, GND, RAMEAU και BNE) χρησιμοποιούν το λεξιλόγιο SKOS. Το RAMEAU και η BNE χρησιμοποιούν μόνο το κατηγορήμα *"skos:closeMatch"*, ενώ οι LCSH και η GND χρησιμοποιούν το κατηγορήμα *"skos:exactMatch"* όταν οι τοπικές και οι απομακρυσμένες οντότητες είναι ακριβώς οι ίδιες και το κατηγορήμα *"skos:closeMatch"*, όταν οι τοπικά ορισμένες και οι απομακρυσμένες οντότητες σχετίζονται «χαλαρά» (λεπτομερής ανάλυση έχει γίνει στην ενότητα 5.2.4). Εδώ θα πρέπει να τονιστεί επίσης, ότι το LIBRIS είναι η μόνη υπηρεσία Συνδεδεμένων Δεδομένων που χρησιμοποιεί το κατηγορήμα *"rdf:seeAlso"*, όταν αναφέρεται σε οντότητες που προέρχονται αποκλειστικά από τη Wikipedia, το κατηγορήμα *"skos:closeMatch"* όταν αναφέρεται σε οντότητες που προέρχονται αποκλειστικά από τις LCSH και το κατηγορήμα *"owl:sameAs"* όταν αναφέρεται σε οντότητες που προέρχονται από το VIAF και τη DBpedia. Τέλος, οι περισσότερες εθνικές βιβλιοθήκες (4 από 6) παρέχουν συνδέσμους προς τις LCSH.

7.5. Συμπεράσματα

Συνολικά θα μπορούσε να ειπωθεί ότι ο παραδοσιακός OPAC των βιβλιοθηκών και πιο συγκεκριμένα το κομμάτι των καθιερώσεων είναι μια πολύτιμη πηγή πληροφοριών στα συστήματα των βιβλιοθηκών. Η εμφάνιση του σημασιολογικού ιστού και πιο συγκεκριμένα του κινήματος των Συνδεδεμένων Δεδομένων παρέχει τη δυνατότητα της προώθησης της πρόσβασης στις καθιερωμένες εγγραφές με ένα πιο τυποποιημένο τρόπο.

Για αυτό το λόγο, οι καθιερωμένες εγγραφές του OPAC πρέπει να ενημερωθούν με πληροφορία των Συνδεδεμένων Δεδομένων. Μέσα αυτό το πλαίσιο, έγινε μια προσπάθεια να προσδιοριστούν τα πεδία του MARC που θα μπορούσαν δυνητικά να φιλοξενήσουν τέτοιου είδους πληροφορία. Από σημασιολογική άποψη, τα πιο κατάλληλα πεδία για να φιλοξενήσουν URIs των Συνδεδεμένων Δεδομένων που ορίζονται τοπικά είναι το πεδίο 003 στο UNIMARC και το πεδίο 024 στο MARC21. Όταν ασχολούμαστε με URIs των Συνδεδεμένων Δεδομένων που έχουν οριστεί σε απομακρυσμένα σημεία και πρέπει να γίνει αναφορά στον τοπικό OPAC, το πιο κατάλληλο πεδίο για να φιλοξενήσει τέτοια πληροφορία είναι το 7xx. Ωστόσο, στο UNIMARC ένα τέτοιο πεδίο δεν

είναι σηματολογικά συμβατό για τη φιλοξενία URIs των Συνδεδεμένων Δεδομένων που αναφέρονται σε οντότητες που έχουν οριστεί σε απομακρυσμένα σημεία και είναι γραμμένες στην ίδια γλώσσα με την τοπική οντότητα.

Στην πράξη, οι εθνικές βιβλιοθήκες ανά τον κόσμο (με εξαίρεση την Εθνική Βιβλιοθήκη της Ισπανίας) δεν έχουν ακόμα ενσωματώσει URIs των Συνδεδεμένων Δεδομένων στους OPAC τους. Αντ' αυτού, χρησιμοποιούν τον OPAC ως το πρωταρχικό αποθετήριο δεδομένων και αντίστοιχα φτιάχνουν υπηρεσίες σε μορφή Συνδεδεμένων Δεδομένων, που βασίζονται στην πληροφορία που προέρχεται από τον υποκείμενο OPAC.

Με την ολοκλήρωση και αυτού του κεφαλαίου έχει γίνει μια λεπτομερής παρουσίαση των δυνατοτήτων που παρέχουν τα Συνδεδεμένα Δεδομένα στον τομέα των βιβλιοθηκών και πιο συγκεκριμένα στην αποτύπωση της θεματικής τους πληροφορίας σε μορφή συμβατή με αυτά. Παράλληλα, έγινε προσπάθεια να βρεθούν σημεία σύγκλισης ανάμεσα στα παραδοσιακά εργαλεία που χρησιμοποιούνται από τις βιβλιοθήκες και τις νέες τεχνολογίες που παρέχονται μέσω των Συνδεδεμένων Δεδομένων. Μέσα σε αυτό το πλαίσιο, μελετήθηκαν τα πεδία του MARC που θα μπορούσαν να φιλοξενήσουν πληροφορία σε μορφή συμβατή με τα Συνδεδεμένα Δεδομένα και αντίστοιχα αυτή η πληροφορία να παρέχεται στους χρήστες μέσω των OPACs των βιβλιοθηκών.

Το επόμενο κεφάλαιο επικεντρώνεται στα μη ελεγχόμενα λεξιλόγια (δηλ. τις λαοταξινομίες) και κατά πόσο αυτά μπορούν να συμβάλλουν στην παροχή τρέχουσας πληροφορίας στους χρήστες βασισμένη στην υπάρχουσα συλλογή της βιβλιοθήκης μέσω του OPAC.

8. Λαοταξινομίες και βιβλιοθήκες

8.1. Το πρόβλημα της χρονικής διάστασης της πληροφορίας στις βιβλιοθήκες

Ανέκαθεν, οι βιβλιοθήκες στοχεύουν στο να παρέχουν στους χρήστες τους υλικό που είναι ενημερωμένο και σύμφωνο με τα ενδιαφέροντά τους. Παραδοσιακά, αυτό το στόχο οι βιβλιοθήκες τον πετυχαίνουν μέσω της χρήσης στατιστικών μεθόδων, όπως είναι η μέτρηση της κίνησης των βιβλίων (Agee, 2005, σ. 92; Knievel, Wicht & Silirigni Connaway, 2006, σ. 35) ή/και η ανάλυση των αρχείων καταγραφής (log files analysis) καταγράφοντας τις αναζητήσεις που κάνει ένας χρήστης στον OPAC και σε άλλες online υπηρεσίες της βιβλιοθήκης (Covey, 2002, σ. 34).

Αυτές οι μέθοδοι εστιάζουν στην ανάλυση των συναλλαγών των χρηστών κατά τη διάρκεια μιας συγκεκριμένης περιόδου και επομένως παρέχουν πληροφορίες για ένα χρονικό διάστημα στο παρελθόν. Με αυτό τον τρόπο, η διαδικασία λήψης αποφάσεων της βιβλιοθήκης αγνοεί τις τρέχουσες τάσεις και ανάγκες, αφού δεν αντικατοπτρίζονται ακόμη στις συναλλαγές των χρηστών με τον κατάλογο της βιβλιοθήκης. Επιπλέον, είναι δύσκολο να εντοπιστεί η περιστασιακή δημοτικότητα συγκεκριμένων τεκμηρίων, δηλαδή τεκμήρια που αφορούν τρέχοντα και επίκαιρα θέματα έχουν μεγαλύτερη ζήτηση σε συγκεκριμένες χρονικές περιόδους (Smith, 1999, σ. 308-309). Τέλος, με την επικράτηση του διαδικτύου, οι χρήστες των μηχανών αναζήτησης τείνουν να επηρεάζονται από την απλότητα της διαδικασίας αναζήτησης και ανάκτησης (Haigh, 2005, σ. 78), η οποία τους παρέχει εγκαίρως σχετική και ενημερωμένη πληροφορία. Επομένως, οι χρήστες περιμένουν και από τις βιβλιοθήκες να δράσουν κατά παρόμοιο τρόπο και να τους παρέχουν έγκαιρα και επίκαιρα αποτελέσματα σχετικά με τις πληροφοριακές τους ανάγκες και απαιτήσεις (Kumar, 2011, σ. 20-21).

Σε μια προσπάθεια να ξεπεραστούν τα προαναφερθέντα προβλήματα, οι βιβλιοθήκες θα πρέπει να αναπτύξουν νέες μεθόδους για να αντιμετωπίσουν το θέμα της δημοτικότητας και της χρονικής διάστασης της πληροφορίας και παράλληλα της αποτελεσματικής παροχής αυτής της πληροφορίας στους χρήστες. Η εμφάνιση της κοινωνικής δικτύωσης (social web) δείχνει προς αυτή την κατεύθυνση, αφού οι αντίστοιχες τεχνολογίες δίνουν την ευκαιρία στις βιβλιοθήκες να εστιάσουν στους χρήστες τους και στις ανάγκες τους (Anderson, 2007, σ. 36) με ένα πιο άμεσο τρόπο (Gerolimos &

Konsta, 2011; Walia & Gupta, 2012; Linh, 2008, σ. 632; Mahmood & Richardson, 2013, σ. 509).

8.2. Χρονική διάσταση της πληροφορίας (επικαιρότητα)

Η επικαιρότητα (timeliness) σχετίζεται άμεσα με την ποιότητα της πληροφορίας (information quality) και εξετάζει αν τα δεδομένα είναι επαρκώς ενημερωμένα και διαθέσιμα για χρήση (Goncalves et al., 2007, σ. 1425). Η πληροφορία έχει έναν κύκλο ζωής, ο οποίος εξαρτάται από το πόσο γρήγορα νέα πληροφορία καταφθάνει για να αντικαταστήσει την παλιά και από το πώς αυτή η πληροφορία παραδίδεται στους χρήστες έγκαιρα (Miller, 1996, σ. 80). Σύμφωνα με τους Pirino, Lee & Wang (2002, σ. 215) και Bovee, Srivastava & Mak (2001, σ. 317), η επικαιρότητα αποτελείται από την «ηλικία» (Age) και τη «μεταβλητότητα» (Volatility). Η ηλικία της πληροφορίας χρησιμοποιείται για να μετρήσει πόσο ενημερωμένη είναι η πληροφορία αναφορικά με το χρόνο που αυτή η πληροφορία καταγράφηκε (Bovee, Srivastava & Mak, 2001, σ. 317). Η μεταβλητότητα της πληροφορίας σχετίζεται με τη σταθερότητα της πληροφορίας, δηλαδή πόσο συχνά νέα πληροφορία καταφθάνει για να αντικαταστήσει την παλιά. Επομένως, όσο πιο μεταβλητή είναι η πληροφορία, τόσο πιο σύντομα αυτή η πληροφορία γίνεται ξεπερασμένη (Bovee, Srivastava & Mak, 2001, σ. 317). Στην επόμενη ενότητα, παρουσιάζεται η χρονική διάσταση της πληροφορίας στις βιβλιοθήκες.

8.2.1. Χρονική διάσταση της πληροφορίας και βιβλιοθήκες

Στο χώρο των βιβλιοθηκών η χρονική διάσταση της πληροφορίας μπορεί να χωριστεί σε δύο κύριες κατηγορίες ανάλογα με: α) το πόσο ενημερωμένη είναι η πληροφορία που παρέχεται από τη βιβλιοθήκη και β) την ταχύτητα παράδοσης της πληροφορίας, δηλαδή πόσο χρονικό διάστημα χρειάζεται η πληροφορία για να επεξεργαστεί και να παραδοθεί στο χρήστη (Weaver-Meyers & Stolt, 1997, σ. 24). Η πληροφορία μπορεί να είναι ακριβής και ολοκληρωμένη όταν εισέρχεται στη βιβλιοθήκη, αλλά μπορεί να αλλάξει με την πάροδο του χρόνου, καθιστώντας με αυτό τον τρόπο την υπάρχουσα πληροφορία παρωχημένη.

Όπως αναφέρθηκε προηγουμένως, αφού η επικαιρότητα συνδέεται άμεσα με την ηλικία της πληροφορίας, την καθιστά ένα πολύτιμο εργαλείο, το οποίο βοηθά στη διαδικασία αξιολόγησης των υπάρχουσών συλλογών της βιβλιοθήκης. Η αξιολόγηση της συλλογής (collection evaluation) είναι μια διαδικασία, που μπορεί να βοηθήσει τις βιβλιοθήκες να προσδιορίσουν το μέγεθος της συλλογής τους, τη θεματική κάλυψη

της συλλογής τους και σε τι βαθμό η συλλογή ανταποκρίνεται στις προσδοκίες των χρηστών (Agee, 2005, σ. 92).

Η επισκόπηση της βιβλιογραφίας αποκαλύπτει ότι οι υπάρχουσες τεχνικές αξιολόγησης της συλλογής της βιβλιοθήκης μπορούν να χωριστούν σε δύο γενικές κατηγορίες: α) τεχνικές βασισμένες στη συλλογή (collection-based techniques), οι οποίες εστιάζουν στο εάν η βιβλιοθήκη έχει αποκτήσει το απαραίτητο υλικό και β) τεχνικές βασισμένες στους χρήστες (user-based techniques), οι οποίες βοηθούν τις βιβλιοθήκες στη μέτρηση της ικανοποίησης των χρηστών από την υπάρχουσα συλλογή (Gessesse 2000, σ. 369). Σε αυτή την ενότητα, οι εν λόγω τεχνικές αξιολογούνται βάσει της δυνατότητάς τους να μας δώσουν επίκαιρη πληροφορία για τα τρέχοντα ενδιαφέροντα των χρηστών στο παρόν.

Προς αυτή την κατεύθυνση, η Crawley-Low (2002, σ. 310-311) προτείνει το διαδανεισμό (interlibrary loan) και την κίνηση των βιβλίων (circulation) ως υπηρεσίες που μπορούν να διευκολύνουν τις βιβλιοθήκες στη συγκέντρωση πληροφοριών για τη χρήση της συλλογής για συγκεκριμένες περιόδους του παρελθόντος. Επομένως, δεν μπορούν να μας δώσουν πληροφορίες για τα τρέχοντα ενδιαφέροντα των χρηστών.

Σε μια παρόμοια προσέγγιση, το Finnish Collection Map Consortium (Hyödynmaa, Ahlholm-Kannisto & Nurminen, 2010, σ. 45) προτείνει μια μέθοδο χαρτογράφησης της συλλογής (collection mapping) για την αξιολόγηση των συλλογών των βιβλιοθηκών. Η μεθοδολογία περιέχει τον προσδιορισμό του όγκου της θεματικής συλλογής, την περιγραφή της μέσης ηλικίας και τις γλώσσες της συλλογής, στατιστικά της χρήσης (usage statistics) και τη σάρωση των ραφιών (shelf-scanning). Τα στατιστικά χρήσης (π.χ. η κίνηση των βιβλίων, ο διαδανεισμός, η εσωτερική χρήση (in-house usage), το ποσοστό επιστροφής (turn-over rate) κλπ.) είναι εργαλεία που μπορούν να παρέχουν πληροφορίες βασισμένες στους χρήστες σχετικά με τη χρήση της συλλογής και αντίστοιχα για τις ανάγκες των χρηστών. Όπως προηγουμένως όμως, αυτές οι προσεγγίσεις μπορούν να δώσουν πληροφορίες για τη χρήση της συλλογής για συγκεκριμένες περιόδους στο παρελθόν. Επομένως, δεν μπορούν να μας δώσουν έγκαιρη πληροφόρηση για τις ανάγκες των χρηστών την παρούσα στιγμή.

Σε μια άλλη προσέγγιση, ο Agee (2005, σ. 92) προτείνει τρεις συγκεκριμένες μεθόδους για την αξιολόγηση των συλλογών: αξιολόγηση με επίκεντρο το χρήστη (user-centered evaluation), φυσική αξιολόγηση (physical assessment) και αξιολόγηση συγκεκριμένης θεματικής υποστήριξης (assessment of specific subject support). Πιο συγκεκριμένα, προτείνει την κίνηση των βιβλίων, το διαδανεισμό και τα κοινοπραξιακά

δεδομένα δανεισμού (consortial borrowing data) ως εργαλεία αξιολόγησης βασισμένα στους χρήστες. Επίσης, εμπλέκει τεχνικές αξιολόγησης βασισμένες στους χρήστες όταν χρησιμοποιεί ως μεθοδολογίες τη φυσική αξιολόγηση και την αξιολόγηση συγκεκριμένης θεματικής υποστήριξης. Πιο συγκεκριμένα, η φυσική αξιολόγηση μπορεί να δώσει πληροφορίες για τη σχέση που υπάρχει ανάμεσα στην ηλικία και την κατάσταση της συλλογής και συγκρίνει τις απαιτήσεις των χρηστών με τις διαθέσιμες πηγές. Η αξιολόγηση συγκεκριμένης θεματικής υποστήριξης μπορεί να λάβει υπ' όψιν της πρότυπα κίνησης των βιβλίων και τα αιτήματα των χρηστών. Αυτές οι προσεγγίσεις μπορούν επίσης, να αποδειχθούν χρήσιμα εργαλεία στη συλλογή πληροφοριών για τη χρήση της συλλογής σε συγκεκριμένες περιόδους του παρελθόντος, αλλά έχουν την τάση να καταναλώνουν πολύ χρόνο και πολλούς πόρους.

Σε μια λίγο πιο διαφορετική προσέγγιση, οι Borin & Yi (2008, σ. 139) προτείνουν έξι γενικούς δείκτες που πρέπει να λαμβάνονται υπ' όψιν για την αξιολόγηση της συλλογής (έντυπη και ηλεκτρονική): α) γενική χωρητικότητα (general capacity), β) ειδικά θεματικά πρότυπα (subject-specific standards), γ) επιστημονικές εκδόσεις (scholarly publishing), δ) χρήση (usage), ε) χρήστες (users) και στ) περιβαλλοντικοί παράγοντες/κοινωνικές δομές (environmental factors/social constructs). Όσον αφορά στο δείκτη που σχετίζεται με τους χρήστες, προτείνουν διάφορες μεθόδους για να τους μελετήσουν, δηλαδή έρευνες (surveys), ομάδες συζητήσεων (discussion groups), συνεντεύξεις (interviews) και πειράματα (experimental settings). Οι προαναφερθείσες μέθοδοι μπορούν να δώσουν επίκαιρη πληροφορία για τις προτιμήσεις των χρηστών, αλλά χρειάζονται πολλούς πόρους και πολύ καλό σχεδιασμό και μελέτη.

Ολοκληρώνοντας, είναι προφανές ότι οι βιβλιοθήκες έχουν πολλά εργαλεία στη διάθεσή τους για να αντιμετωπίσουν το πρόβλημα της χρονικής διάστασης της πληροφορίας. Αλλά τα περισσότερα από αυτά τα εργαλεία βοηθούν τις βιβλιοθήκες για να συγκεντρώσουν πληροφορίες για τους χρήστες τους και για τη χρήση της συλλογής για συγκεκριμένες περιόδους στο παρελθόν. Είναι επαρκή για την ανάπτυξη της συλλογής αλλά δεν μπορούν να αντιμετωπίσουν το πρόβλημα της παροχής τρέχουσας πληροφορίας στους χρήστες έγκαιρα. Επομένως, οι βιβλιοθήκες θα πρέπει να βρουν νέους τρόπους για την παροχή αυτής της πληροφορίας. Τα εργαλεία κοινωνικής δικτύωσης δείχνουν προς αυτή την κατεύθυνση αφού βοηθούν τις βιβλιοθήκες να μελετούν τους χρήστες με τρόπο πιο αποτελεσματικό, γρήγορο και άμεσο. Στην επόμενη ενότητα, εξετάζεται η κοινωνική δικτύωση ως ένα εργαλείο επικοινωνίας ανάμεσα στις βιβλιοθήκες και τους χρήστες τους.

8.2.2. Web 2.0, χρονική διάσταση της πληροφορίας και βιβλιοθήκες

Τα τελευταία χρόνια, το Web 2.0 έκανε την εμφάνισή του ως ένα εργαλείο επικοινωνίας ανάμεσα στις βιβλιοθήκες και τους χρήστες. Μπορεί να διευκολύνει την εξατομικευμένη ανταλλαγή πληροφοριών και τη συνεργασία κατά τρόπο έγκαιρο. Αναφέρεται στη «*συλλογή τεχνολογιών, όπως είναι τα blogs, τα wikis, τα RSS feeds, τα κοινωνικά δίκτυα κλπ. όπου οι χρήστες μπορούν να προσθέσουν, να μοιράσουν και να επεξεργαστούν το περιεχόμενο, δημιουργώντας ένα διαδικτυακό κοινωνικά δικτυωμένο περιβάλλον*» (Lwoga, 2013, σ. 288).

Σύμφωνα με τον Lwoga (2013, σ. 289), τα εργαλεία του Web 2.0 μπορούν να χωριστούν στις ακόλουθες τέσσερις κατηγορίες στον τομέα των βιβλιοθηκών: α) εργαλεία για την ανάκτηση πληροφοριών, δηλαδή εργαλεία για τη συγκέντρωση πληροφοριών από πηγές εξωτερικές προς τη βιβλιοθήκη (π.χ. blogs, wikis), β) εργαλεία για τη διάδοση της πληροφορίας, δηλαδή εργαλεία για τη διάθεση πληροφοριών από τις βιβλιοθήκες στους χρήστες (π.χ. RSS), γ) εργαλεία για την ανταλλαγή πληροφοριών, δηλαδή εργαλεία για τη διαδραστική επικοινωνία μεταξύ των βιβλιοθηκών και των χρηστών τους (π.χ. ανταλλαγή άμεσων μηνυμάτων (instant messaging), υπηρεσίες κοινωνικής δικτύωσης (Social Networking Services – SNS)), και δ) εργαλεία για την οργάνωση της γνώσης, δηλαδή εργαλεία για τη διευκόλυνση της αναπαράστασης του περιεχομένου και την επακόλουθη αναζήτηση και ανάκτησή του (π.χ. social tagging).

Γενικά, υπάρχουν διάφοροι λόγοι που οι βιβλιοθήκες χρησιμοποιούν τα εργαλεία κοινωνικής δικτύωσης. Ο κύριος λόγος είναι η διαφήμιση της βιβλιοθήκης και η δημοσιοποίηση νέων που αφορούν εκδηλώσεις και δραστηριότητες (Walia & Gupta, 2012). Ένα εργαλείο κοινωνικής δικτύωσης, το οποίο αρχίζει να κερδίζει την προσοχή των βιβλιοθηκών τα τελευταία χρόνια είναι το social tagging. Όπως αναφέρθηκε νωρίτερα, το tagging δίνει τη δυνατότητα στους χρήστες να προσθέσουν λέξεις κλειδιά στο διαμοιραζόμενο περιεχόμενο (Lawson, 2009, σ. 574). Πιο συγκεκριμένα, μια ετικέτα είναι μια μη ιεραρχική λέξη κλειδί ή όρος που εκχωρείται σε ένα κομμάτι των πληροφοριών (π.χ. μια ιστοσελίδα) (Spitieri, 2014). Το tagging μπορεί να διευκολύνει τις βιβλιοθήκες στη σύλληψη των αναγκών των χρηστών λόγω του γεγονότος ότι μπορεί να προσαρμοστεί γρήγορα στις αλλαγές που συμβαίνουν και στην ορολογία. Επιπλέον, το tagging επιτρέπει στις βιβλιοθήκες να συνεργαστούν με τους χρήστες τους για να τους παρέχουν βελτιωμένη θεματική πρόσβαση (Lawson, 2009, σ. 574). Στην επόμενη ενότητα, τα hashtags του Twitter μελετώνται ως μια παραλλαγή του social tagging.

8.2.3. Twitter και βιβλιοθήκες

Το Twitter είναι ένα εργαλείο του Web 2.0 για microblogging, το οποίο επιτρέπει στους χρήστες να μεταδώσουν σύντομες γραπτές ενημερώσεις για πράγματα που συμβαίνουν στη ζωή τους, όπως είναι το τι σκέφτονται, διαβάζουν ή βιώνουν. Μπορούν επίσης να μοιραστούν πληροφορίες, νέα και απόψεις (Zhao & Rosson, 2009, σ. 243). Εκτός από την ανταλλαγή πληροφοριών και την ενημέρωση της κατάστασης του χρήστη, το Twitter χρησιμοποιείται επίσης, για την εύρεση πληροφορίας. Πράγματι, οι χρήστες απευθύνονται στο Twitter όταν θέλουν να βρουν πληροφορία για συνταρακτικά νέα, γεγονότα σε πραγματικό χρόνο, πληροφορίες για ανθρώπους και θεματική πληροφορία (Teevan, Ramage & Morris, 2011, σ. 37; Lau, Li & Tjondronegoro, 2011). Αυτά τα γνωρίσματα καθιστούν το Twitter ένα εργαλείο που μπορεί να παρέχει επίκαιρη πληροφόρηση γρηγορότερα σε σχέση με οποιοδήποτε άλλο μέσο μαζικής ενημέρωσης (π.χ. τηλεόραση, ραδιόφωνο κλπ.).

Τα hashtags χρησιμοποιούνται από το Twitter για να ταξινομήσουν τα μηνύματα, να διαδώσουν ιδέες και επίσης για να προωθήσουν συγκεκριμένα θέματα και ανθρώπους (Cunha et al., 2011, σ. 58). Πιο συγκεκριμένα, τα hashtags μπορούν να κατηγοριοποιήσουν ένα tweet σύμφωνα με το θέμα του μηνύματος. *«Μπορούν να χρησιμοποιηθούν για να προσθέσουν θεματικό περιεχόμενο και μεταδεδομένα στις αναρτήσεις, αλλά επίσης για προώθηση και διαφήμιση. Με την προσθήκη απλά ενός συμβόλου δίεσης (#) πριν από μια συμβολοσειρά γραμμάτων, αριθμητικών ψηφίων ή σημεία υπογράμμισης (_), είναι δυνατόν να «επισημανθεί» (tagged) ένα μήνυμα, βοηθώντας άλλους χρήστες να βρουν τα tweets που έχουν κοινό θέμα. Τα hashtags επιτρέπουν στους χρήστες να δημιουργήσουν κοινότητες ανθρώπων που ενδιαφέρονται για το ίδιο θέμα καθιστώντας πιο εύκολο για αυτούς να εντοπίσουν και να μοιραστούν σχετική πληροφορία»* (Cunha et al., 2011, σ. 58).

Οι βιβλιοθήκες, ειδικά, μπορούν να επωφεληθούν από τη χρήση του Twitter, αφού μπορεί να λειτουργήσει ως ένα εργαλείο για διάφορους σκοπούς (Carscaddon & Charman, 2013; Aharony, 2010), ως ένα συνδετικό εργαλείο ανάμεσα στη βιβλιοθήκη και τους χρήστες της, ένα εργαλείο προώθησης προγραμμάτων και υπηρεσιών, ένα εργαλείο συζητήσεων με τους χρήστες, ένα πληροφοριακό εργαλείο για να ειδοποιούν τους χρήστες σχετικά με τις νέες προσθήκες στις συλλογές, ένα εργαλείο για την παροχή συνδέσμων σε διαδικτυακό περιεχόμενο που μπορεί να είναι σχετικό με τις πληροφοριακές ανάγκες των χρηστών κλπ. Πρόσφατα, μια άλλη ιδιότητα του Twitter, η οποία αρχίζει να κερδίζει την προσοχή και μπορεί να διευκολύνει τις βιβλιοθήκες

είναι η ανάλυση των tweets. Τα tweets μπορούν να παρέχουν πολύτιμη πληροφορία για συγκεκριμένα ονόματα, hashtags, λέξεις κλειδιά ή θέματα. Οι βιβλιοθήκες μπορούν να εκμεταλλευτούν αυτή την πληροφορία για να προσφέρουν εξατομικευμένη και έγκαιρη πληροφόρηση στους χρήστες για συγκεκριμένα θέματα, τα οποία απασχολούν την κοινότητα αυτή την περίοδο (Chang & Iyer, 2012, σ. 256-257).

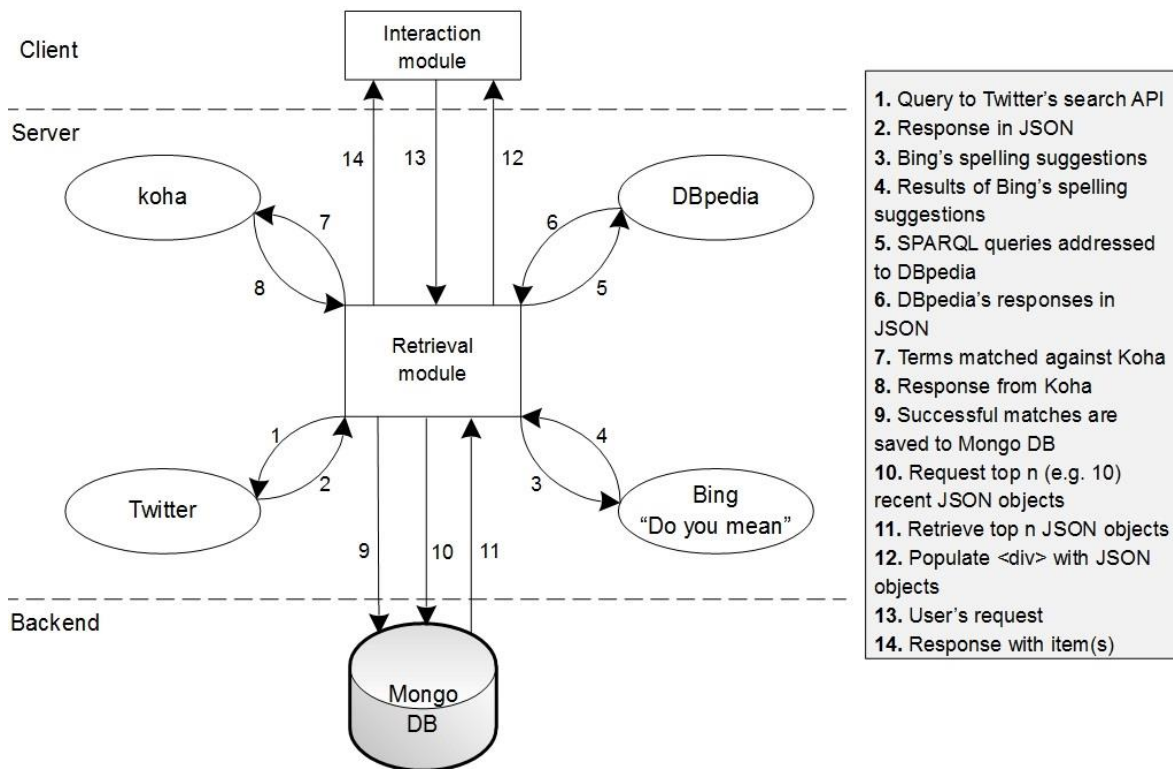
Επομένως, αυτό το κεφάλαιο εστιάζει στο Twitter όχι σαν εργαλείο για επικοινωνία με τους χρήστες αλλά ως εργαλείο που μπορεί να δώσει στις βιβλιοθήκες πολύτιμες πληροφορίες για τα θέματα της επικαιρότητας που απασχολούν την κοινωνία. Σε αυτό το πλαίσιο, τα hashtags των tweets αξιοποιούνται ως μέσο για τον προσδιορισμό των θεμάτων που οι χρήστες συζητούν επί του παρόντος σε ένα συγκεκριμένο τομέα. Τα εξαγόμενα hashtags μπορούν μετά να αντιστοιχηθούν στον κατάλογο της βιβλιοθήκης για να προτείνουν αντίστοιχα τεκμήρια. Αυτή η διαδικασία θα βοηθήσει τις βιβλιοθήκες να παρέχουν ενημερωμένη πληροφόρηση στους χρήστες τους σχετικά με τις τρέχουσες ανάγκες τους. Ως υλοποίηση της ιδέας, η επόμενη ενότητα προτείνει μια υπηρεσία που είναι ικανή να αντιστοιχεί hashtags σε τεκμήρια της βιβλιοθήκης έγκαιρα.

8.2.4. Υλοποίηση της ιδέας

Ένας τρόπος για την παροχή επίκαιρης πληροφορίας στους χρήστες μιας βιβλιοθήκης είναι μέσω της ενσωμάτωσης των δημοφιλών tweets στον OPAC της βιβλιοθήκης. Σε αυτή την ενότητα, ως υλοποίηση της ιδέας, παρουσιάζεται μια υπηρεσία, η οποία αντιστοιχεί δημοφιλή tweets στα τεκμήρια της συλλογής της βιβλιοθήκης. Πιο συγκεκριμένα, η προτεινόμενη υπηρεσία βασίζεται πάνω στην ακόλουθη ροή εργασίας: αρχικά, η υπηρεσία ανακαλύπτει τα πιο δημοφιλή hashtags σε ένα συγκεκριμένο τομέα (π.χ. πληροφορική). Ύστερα, αυτά τα hashtags ευθυγραμμίζονται σε ένα πιο επίσημο λεξιλόγιο μέσω της χρήσης της υπηρεσίας "*Do you mean*" του Bing¹⁸⁰. Το αντίστοιχο λεξιλόγιο που εξάγεται διοχετεύεται σε μια υπηρεσία του σηματολογικού ιστού, η οποία είναι ικανή να παρέχει σχετικούς όρους με τα αρχικά hashtags. Τέλος, το σύνολο από τους όρους που προκύπτουν μεταμορφώνεται σε αντίστοιχα διατυπωμένα ερωτήματα, τα οποία απευθύνονται στον OPAC της βιβλιοθήκης. Τα εν λόγω ερωτήματα εκτίθενται στο γραφικό περιβάλλον του OPAC ως προτάσεις τεκμηρίων για τους χρήστες της βιβλιοθήκης.

¹⁸⁰ Bing.com, (2015). *Bing*. [online] Διαθέσιμο στο: <https://www.bing.com/> [Πρόσβαση: 03/10/2015].

Όπως φαίνεται στην εικόνα 8.1, η αρχιτεκτονική της υπηρεσίας αποτελείται από 3 στρώματα/επίπεδα (layers): τον πελάτη (client), τον εξυπηρετητή (server) και το σύστημα υποστήριξης (backend). Το επίπεδο πελάτη αποτελείται από τη διαδραστική οντότητα της προτεινόμενης υπηρεσίας, η οποία είναι διαθέσιμη στους χρήστες. Το επίπεδο εξυπηρετητή βασίζεται στην οντότητα της ανάκτησης, η οποία λειτουργεί ως διαμεσολαβητής ανάμεσα στις ακόλουθες υπηρεσίες: το Twitter search API¹⁸¹, την υπηρεσία του Bing "Do you mean"¹⁸², τη DBpedia¹⁸³ και το Koha¹⁸⁴. Επιπλέον, διευκολύνει την επικοινωνία ανάμεσα στο επίπεδο πελάτη και στο επίπεδο του συστήματος υποστήριξης. Τέλος, το επίπεδο του συστήματος υποστήριξης αποτελείται από το αποθετήριο των δεδομένων της υπηρεσίας, το οποίο υλοποιείται ως μια βάση δεδομένων Mongo¹⁸⁵.



Εικόνα 8.1: Η συνολική αρχιτεκτονική της υπηρεσίας

¹⁸¹ Dev.twitter.com, (2015). *The Search API | Twitter Developers*. [online] Διαθέσιμο στο: <https://dev.twitter.com/rest/public/search> [Πρόσβαση: 14/07/2015].

¹⁸² Datamarket.azure.com, (2015). *Bing Search API | Microsoft Azure Marketplace*. [online] Διαθέσιμο στο: <http://datamarket.azure.com/dataset/bing/search> [Πρόσβαση: 14/07/2015].

¹⁸³ Dbpedia.org, (2015). *Virtuoso SPARQL Query Editor*. [online] Διαθέσιμο στο: <http://dbpedia.org/sparql> [Πρόσβαση: 14/07/2015].

¹⁸⁴ Koha-community.org, (2015). *Official Website of Koha Library Software*. [online] Διαθέσιμο στο: <http://koha-community.org/> [Πρόσβαση: 14/07/2015].

¹⁸⁵ Mongoddb.org, (2015). *MongoDB*. [online] Διαθέσιμο στο: <https://www.mongoddb.org/> [Πρόσβαση: 14/07/2015].

8.2.4.1. Λεπτομερής ανάλυση της προτεινόμενης αρχιτεκτονικής

Η λειτουργικότητα της προτεινόμενης υπηρεσίας πραγματοποιείται μέσα από ένα μοντέλο ροής εργασίας που διευκολύνει την επικοινωνία ανάμεσα στις διάφορες οντότητες της προαναφερθείσας αρχιτεκτονικής. Σε αυτή την ενότητα, παρουσιάζεται η ροή της πληροφορίας μέσα στην υπηρεσία.

Βήμα 1: Αρχικά, η οντότητα ανάκτησης (Retrieval module) απευθύνει ένα ερώτημα στο Search API του Twitter για τα δημοφιλή tweets σε μια συγκεκριμένη θεματική περιοχή (π.χ. πληροφορική).

Βήμα 2: Τα αντίστοιχα tweets παραδίδονται στην οντότητα ανάκτησης ως αντικείμενα της JSON. Η οντότητα απομονώνει τα hashtags από το περιβάλλον κείμενο που υπάρχει σε κάθε tweet.

Βήμα 3-4: Αφού η κοινωνική (crowdsourcing¹⁸⁶) φύση του Twitter υπονοεί ότι τα hashtags είναι πιθανόν να αντικατοπτρίζουν ένα ανεπίσημο λεξιλόγιο, η προτεινόμενη υπηρεσία ευθυγραμμίζει τα αντίστοιχα hashtags σε ένα πιο επίσημο λεξιλόγιο μέσω της χρήσης του search API του Bing (Movahedian & Khayyambashi, 2014). Ως εκ τούτου, για κάθε hashtag ένα ερώτημα απευθύνεται στην υπηρεσία Bing search API για να πάρουμε ορθογραφικές προτάσεις (για το υπόλοιπο της παρουσίασης της υπηρεσίας, οι ορθογραφικές προτάσεις θα ονομάζονται επίκαιροι όροι (trending terms)), λειτουργώντας ως φίλτρο "*Do you mean*" (π.χ. το #CloudComputing θα μεταμορφωθεί σε "*Cloud Computing*").

Βήμα 5-6: Οι επίκαιροι όροι μετατρέπονται ύστερα σε SPARQL ερωτήματα, τα οποία αντίστοιχα απευθύνονται στο τελικό σημείο SPARQL της DBpedia σε μια προσπάθεια να επεκταθούν σημασιολογικά τα hashtags με τις πιθανές κατηγορίες ή/και τους σχετικούς όρους. Τα αντίστοιχα αποτελέσματα λαμβάνονται από την οντότητα ανάκτησης σε μορφή JSON.

Βήμα 7-8: Οι επίκαιροι όροι καθώς και οι κατηγορίες στις οποίες ανήκουν και οι σχετικοί όροι αντιστοιχίζονται με το ευρετήριο τίτλων του OPAC της βιβλιοθήκης. Μια

¹⁸⁶ Crowdsourcing is "*the process of obtaining needed services, ideas, or content by soliciting contributions from a large group of people, and especially from an online community, rather than from traditional employees or suppliers*". Από το λεξικό: Merriam-webster.com, (2015). *Dictionary and Thesaurus / Merriam-Webster*. [online] Διαθέσιμο στο: <http://www.merriam-webster.com/> [Πρόσβαση: 14/07/2015].

τέτοια διαδικασία επιτυγχάνεται μέσω της χρήσης του SRU protocol¹⁸⁷ του Koha. Σε αυτό το σημείο θα πρέπει να τονιστεί ότι ιδανικά, οι επίκαιροι όροι θα πρέπει επίσης να αντιστοιχίζονται με το ευρετήριο θεμάτων της βιβλιοθήκης. Ωστόσο, αυτό δεν είναι εφικτό με τη συγκεκριμένη βιβλιοθήκη, η οποία διατηρεί το θεματικό της ευρετήριο στα ελληνικά.

Βήμα 9: Στην περίπτωση που έχουμε μια θετική απόκριση (δηλ. υπάρχουν τεκμήρια που σχετίζονται με το συγκεκριμένο όρο), ένα αντικείμενο σε μορφή JSON αποθηκεύεται στη Mongo DB, το οποίο αναπαριστά τις σχετικές βιβλιογραφικές εγγραφές που αντιστοιχούν στο συγκεκριμένο όρο. Κάθε αντικείμενο JSON αποτελείται από τα ακόλουθα πεδία:

- α) Λέξη κλειδί: ο όρος (δηλ. ο επίκαιρος όρος, η κατηγορία, ο σχετικός όρος) που αντιστοιχίζεται στον OPAC της βιβλιοθήκης.
- β) URL: το URL του αντίστοιχου ερωτήματος αναζήτησης του Koha. Κατόπιν αιτήσεως, επιστρέφεται το τεκμήριο ή τα τεκμήρια που αντιστοιχίζονται στον OPAC της βιβλιοθήκης.
- γ) Hashtag: το αρχικό hashtag που προκάλεσε την όλη διαδικασία.
- δ) Χρονοσφραγίδα (Timestamp).

Βήμα 10-11: Τα n (π.χ. 10) πιο πρόσφατα αντικείμενα JSON ζητούνται από τη βάση δεδομένων Mongo και αντίστοιχα ανακτώνται από την οντότητα ανάκτησης.

Βήμα 12: Τα ανακτηθέντα αντικείμενα JSON προωθούνται σε ένα τμήμα (division) στοιχείου της HTML (π.χ. <div>) μέσα στην αρχική σελίδα του OPAC της βιβλιοθήκης.

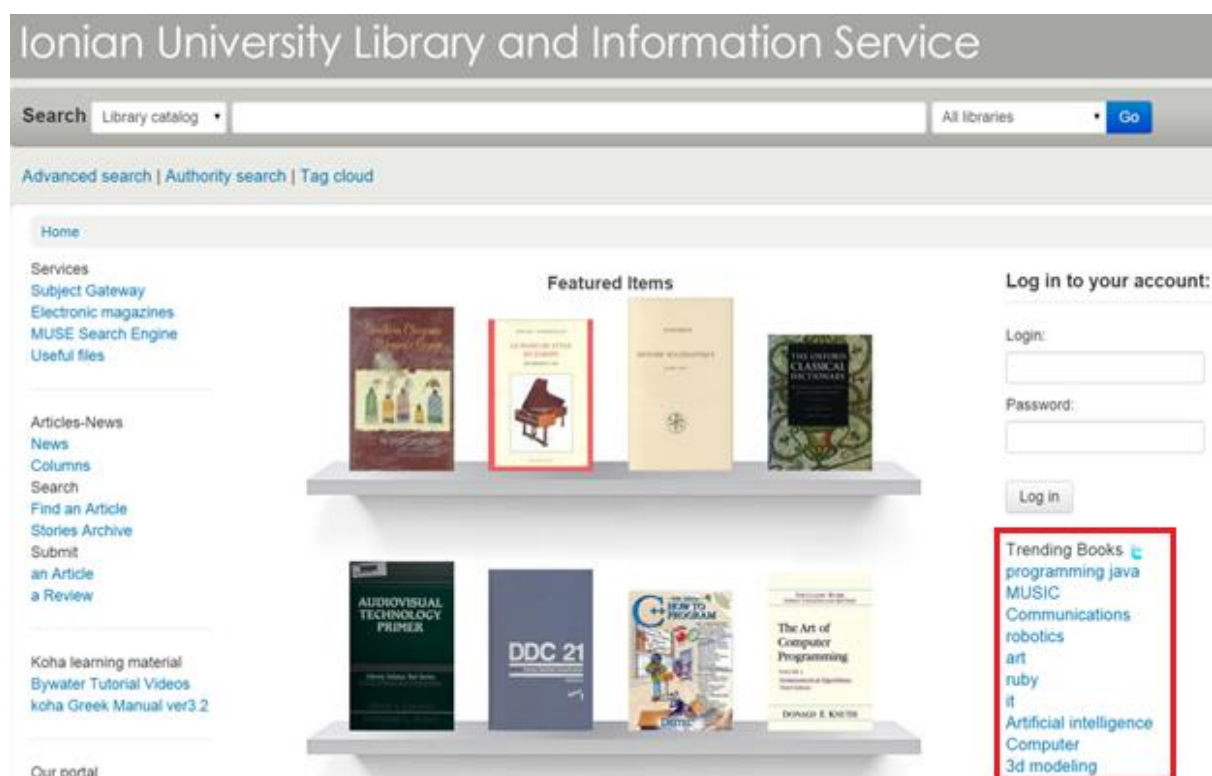
Βήμα 13-14: Κατά την αλληλεπίδραση του χρήστη με το τμήμα στοιχείου της HTML της προτεινόμενης υπηρεσίας, ένα αντίστοιχο ερώτημα αναζήτησης του Koha απευθύνεται στον OPAC της βιβλιοθήκης και η προκύπτουσα απάντηση παρουσιάζεται ως μια λίστα προτεινόμενων τεκμηρίων εντός του OPAC της βιβλιοθήκης.

Στην επόμενη ενότητα, παρουσιάζεται η εφαρμογή της προτεινόμενης μεθοδολογίας στον κατάλογο της ακαδημαϊκής βιβλιοθήκης του Ιονίου Πανεπιστημίου.

¹⁸⁷ Loc.gov, (2015). *SRU: Search/Retrieval via URL -- SRU, CQL and ZeeRex (Standards, Library of Congress)*. [online] Διαθέσιμο στο: <http://www.loc.gov/standards/sru/> [Πρόσβαση: 14/07/2015].

8.2.5. Εφαρμογή της προτεινόμενης υπηρεσίας

Η προτεινόμενη υπηρεσία εγκαταστάθηκε σε μια ακαδημαϊκή βιβλιοθήκη. Πιο συγκεκριμένα, η εγκατάσταση υλοποιήθηκε στον OPAC που είναι βασισμένος στο Koha της Βιβλιοθήκης και Κέντρου Πληροφόρησης του Ιονίου Πανεπιστημίου¹⁸⁸. Ο OPAC περιέχει τεκμήρια που σχετίζονται με τα τμήματα του Πανεπιστημίου (δηλ. Ιστορία, Ξένες γλώσσες, Μετάφραση και Διερμηνεία, Μουσικές σπουδές, Αρχαιονομία, Επιστήμη της πληροφόρησης και Μουσειακές σπουδές, Πληροφορική και Οπτικοακουστικές τέχνες). Η προτεινόμενη υπηρεσία εστιάζει στα τεκμήρια που σχετίζονται με το τμήμα Πληροφορικής.



Εικόνα 8.2: Αρχική σελίδα του OPAC της Βιβλιοθήκης και Κέντρου Πληροφόρησης του Ιονίου Πανεπιστημίου

Όπως φαίνεται στην εικόνα 8.2, η προτεινόμενη υπηρεσία αναπαρίσταται ως ένα τμήμα στοιχείου της HTML που περιέχει τα επίκαιρα θέματα στο κάτω δεξιό μέρος (κόκκινη μαρκαρισμένη περιοχή) της αρχικής σελίδας του OPAC.

¹⁸⁸ Koha Online Catalog, (2015). *Ionian University Library and Information Service*. [online] Διαθέσιμο στο: <http://83.212.169.55/> [Πρόσβαση: 15/07/2015].

8.2.5.1. Προβλήματα κατά την εφαρμογή

Αρχικά, ο στόχος ήταν να ασχοληθούμε με τα πιο δημοφιλή hashtags στο Twitter, ασχέτως του πεδίου στο οποίο ανήκουν. Ωστόσο, ήταν δύσκολο να αντιστοιχήσουμε τα γενικού περιεχομένου hashtags στον OPAC της βιβλιοθήκης (ανιχνεύθηκε πολύς «*θόρυβος*»), οπότε αποφασίστηκε να περιορίσουμε το σκοπό της προτεινόμενης υπηρεσίας σε ένα συγκεκριμένο πεδίο που εκπροσωπείται στη βιβλιοθήκη. Αφού το Ιόνιο Πανεπιστήμιο έχει τμήμα Πληροφορικής, η υπηρεσία ανίχνευε tweets σχετικά με την πληροφορική μέσω της χρήσης ενός ερωτήματος αναζήτησης που περιείχε την λέξη κλειδί "*programming*". Στην επόμενη ενότητα, παρουσιάζεται μια μικρή αξιολόγηση της προτεινόμενης υπηρεσίας.

8.2.6. Αξιολόγηση της υπηρεσίας

Η αξιολόγηση της υπηρεσίας έχει ως στόχο τον προσδιορισμό σε ποιο βαθμό τα δημοφιλή tweets επηρεάζουν τους χρήστες της βιβλιοθήκης στην επιλογή τεκμηρίων. Για να έρθουμε σε αυτό το συμπέρασμα, αποφασίστηκε να συγκρίνουμε τον αριθμό των αιτημάτων για τεκμήρια προερχόμενα από την προτεινόμενη υπηρεσία με τον αριθμό των αιτημάτων για τεκμήρια προερχόμενα από τις υπόλοιπες υπηρεσίες ανάκτησης της βιβλιοθήκης (όπως απεικονίζονται στο γραφικό περιβάλλον, βλέπε εικόνα 8.2). Ένας ακόμα στόχος της αξιολόγησης ήταν να ερευνήσουμε εάν η προτεινόμενη υπηρεσία βοήθα τους χρήστες στην επιλογή τεκμηρίων σχετικά με τα επίκαιρα θέματα από τον OPAC της βιβλιοθήκης. Για την επίτευξη αυτών των στόχων, διεξήχθη το ακόλουθο πείραμα.

8.2.6.1. Το πείραμα

Το πείραμα έλαβε χώρα στα πλαίσια του μαθήματος «*Τεχνολογίες Πληροφόρησης*» που προσφέρεται από το τμήμα Πληροφορικής. Πιο συγκεκριμένα, 10 μεταπτυχιακοί φοιτητές συμμετείχαν στο πείραμα (4 γυναίκες και 6 άντρες). Οι ηλικίες ποίκιλαν από 25 έως 45 και όλοι παρουσίαζαν ικανοποιητική εμπειρία στη χρήση υπηρεσιών κοινωνικής δικτύωσης και OPAC βιβλιοθηκών. Σύμφωνα με το πείραμα, ζητήθηκε από τους συμμετέχοντες να επισκεφτούν την αρχική σελίδα της βιβλιοθήκης σε διάστημα 3 ημερών (δηλ. 12-15 Μαΐου 2015) και να επιλέξουν 5 τεκμήρια που βρήκαν ελκυστικά στη θεματική περιοχή του μαθήματος. Ύστερα, οι συμμετέχοντες κλήθηκαν να αποστείλουν ένα email με τα τεκμήρια που επέλεξαν.

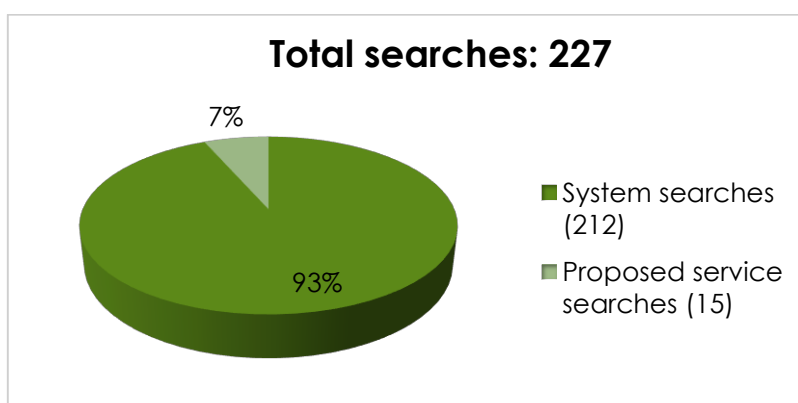
Όπως φαίνεται στην εικόνα 8.2, το γραφικό περιβάλλον της αρχικής σελίδας του OPAC της βιβλιοθήκης αντιμετωπίζει ομοιόμορφα όλες τις παρεχόμενες υπηρεσίες ανάκτησης, συμπεριλαμβανομένης και της υπηρεσίας που παρουσιάστηκε στην προηγούμενη ενότητα (βλέπε ενότητα 8.2.4). Πιο συγκεκριμένα, η αρχική σελίδα παρέχει πρόσβαση στις ακόλουθες υπηρεσίες: α) τη γενική επιλογή αναζήτησης (general search option), β) την προηγμένη επιλογή αναζήτησης (advanced search option), γ) τα προτεινόμενα τεκμήρια (featured items) και δ) την προτεινόμενη υπηρεσία. Το σύνολο των αλληλεπιδράσεων με το σύστημα αποθηκεύτηκαν σε αντίστοιχα αρχεία καταγραφής συμβάντων (transaction log files).

8.2.6.2. Αποτελέσματα

Στο τέλος του πειράματος, οι συμμετέχοντες επέλεξαν 40 μοναδικά τεκμήρια (αυτός ο αριθμός προκύπτει από τα αντίστοιχα emails) σχετικά με τη θεματική περιοχή του μαθήματος.

Σύμφωνα με την ανάλυση των αρχείων καταγραφής συμβάντων¹⁸⁹, 227 αιτήματα HTTP GET απευθύνθηκαν στο σύστημα από τις διάφορες υπηρεσίες της αρχικής σελίδας. Αυτά τα αιτήματα κατέληξαν είτε σε μοναδικά τεκμήρια (31 αιτήματα) ή σε λίστες αποτελεσμάτων αναζήτησης, οι οποίες περιέχουν κυμαινόμενο αριθμό τεκμηρίων (196 αιτήματα). Ως εκ τούτου, συμπεραίνεται ότι οι συμμετέχοντες εξέτασαν τα τεκμήρια/λίστες και επέλεξαν 40 μοναδικά τεκμήρια που τους προσέλκυσαν περισσότερο.

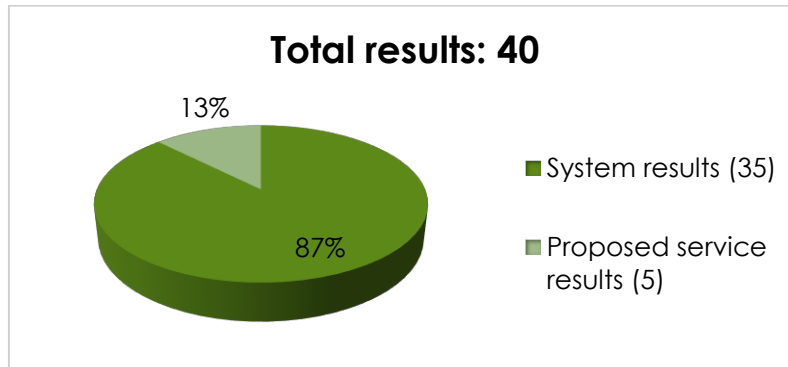
Δεκαπέντε από τα 227 αιτήματα HTTP GET προήλθαν από την προτεινόμενη υπηρεσία ενώ τα υπόλοιπα (δηλ. 212) από τις άλλες υπηρεσίες του συστήματος (βλέπε εικόνα 8.3).



Εικόνα 8.3: Αναζητήσεις που έγιναν στο σύστημα

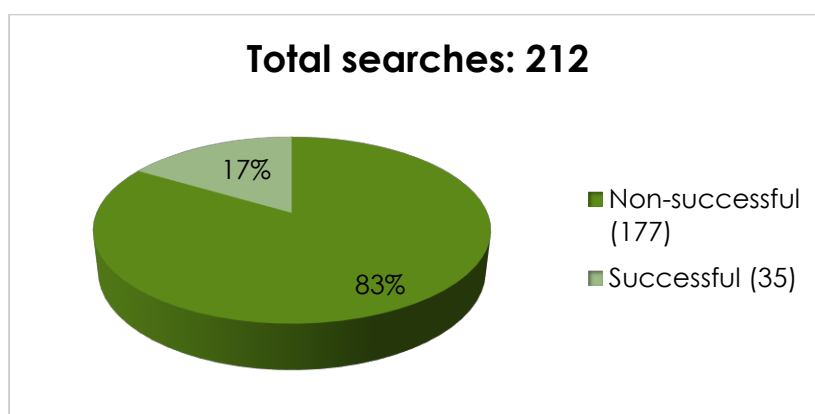
¹⁸⁹ Τα αντίστοιχα αρχεία καταγραφής συμβάντων καθώς και η πρόσβαση στους πίνακες της βάσης δεδομένων Mongo είναι διαθέσιμα στο: <http://aimashup.org/timeliness>

Επιπλέον, 5 από τα 40 τεκμήρια που επιλέχθηκαν από τους συμμετέχοντες βρέθηκαν επίσης στα αποτελέσματα των αναζητήσεων από τα 15 αιτήματα που προήλθαν από την προτεινόμενη υπηρεσία. Συνεπώς, τα 35 από τα 40 τεκμήρια προήλθαν από τις υπόλοιπες υπηρεσίες ανάκτησης του συστήματος (βλέπε εικόνα 8.4).

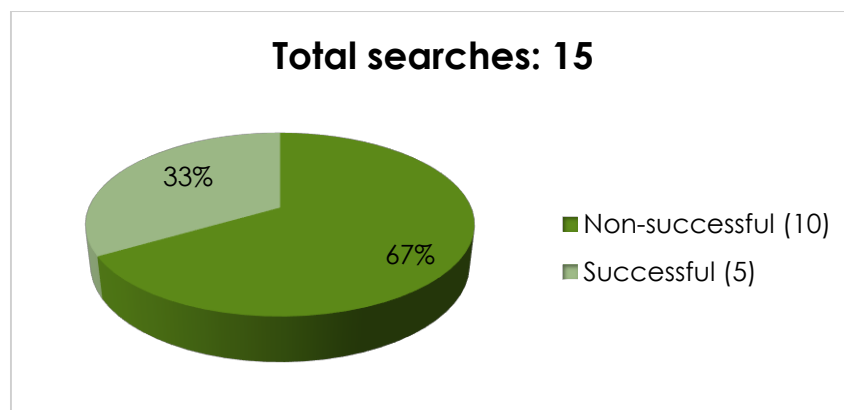


Εικόνα 8.4: Αποτελέσματα του συστήματος και της προτεινόμενης υπηρεσίας

Από μια συγκεκριμένη σκοπιά, ο αντίκτυπος της προτεινόμενης υπηρεσίας στην εύρεση ενδιαφερόντων τεκμηρίων που σχετίζονται με ένα δημοφιλές hashtag είναι μάλλον χαμηλή (δηλ. 15 από τα 227 αιτήματα προήλθαν από την προτεινόμενη υπηρεσία). Ωστόσο, μια περαιτέρω ανάλυση των δεδομένων υποδεικνύει ότι ο αντίκτυπος των προτάσεων που προέρχονται από την προτεινόμενη υπηρεσία είναι σημαντικά υψηλότερος από τον αντίκτυπο των προτάσεων που προέρχονται από τις υπόλοιπες υπηρεσίες ανάκτησης του συστήματος. Πιο συγκεκριμένα, οι συμμετέχοντες επέλεξαν 5 τεκμήρια προερχόμενα από τη λίστα των 15 τεκμηρίων που παρέχονταν από την προτεινόμενη υπηρεσία (δηλ. 33%, βλέπε εικόνα 8.6). Αντιθέτως, οι συμμετέχοντες επέλεξαν 35 τεκμήρια προερχόμενα από τη λίστα των 212 αιτημάτων που προτάθηκαν από το υπόλοιπο σύστημα (δηλ. 17%, βλέπε εικόνα 8.5).



Εικόνα 8.5: Αποτελέσματα συστήματος



Εικόνα 8.6: Αποτελέσματα προτεινόμενης υπηρεσίας

8.3. Συμπεράσματα

Ένα συμπέρασμα που μπορεί να εξαχθεί με ασφάλεια από το παραπάνω πείραμα είναι το γεγονός ότι η ενσωμάτωση πληροφοριών που προέρχονται από εξωτερικές υπηρεσίες κοινωνικής δικτύωσης με τη βιβλιοθήκη είναι εφικτή.

Εξ' αιτίας περιορισμών στη χρήση των τρίτων εφαρμογών που χρησιμοποιήθηκαν από την προτεινόμενη υπηρεσία, η ανάλυση των αρχείων συμβάντων αναφέρεται σε αιτήματα που απευθύνονταν στις διάφορες υπηρεσίες κάθε 1,5 ώρα για μια χρονική περίοδο 13,5 ωρών. Πιο συγκεκριμένα, μέσα σε αυτή τη χρονική περίοδο, καταγράφηκαν 951 μη διακριτά (non-distinct) hashtags. Η προτεινόμενη υπηρεσία κατάφερε να αντιστοιχήσει τα 29 από αυτά σε 35 μη κενά (non-empty) ερωτήματα αναζήτησης στον OPAC της βιβλιοθήκης. Δέκα από τα 35 ερωτήματα αναζήτησης προήλθαν από τη DBpedia. Ως εκ τούτου, είναι προφανές ότι ακόμα και σε μια εξειδικευμένη περιοχή ενδιαφέροντος (δηλ. πληροφορική), θα υπάρξουν πολλά μη σχετικά (irrelevant) tweets μέχρις ότου να βρεθεί ένα tweet που να αντιστοιχεί σε ένα τεκμήριο της βιβλιοθήκης. Αυτό μπορεί να οφείλεται στο γεγονός ότι το λεξιλόγιο των tweets διαφέρει σημαντικά από τη γλώσσα ευρετηρίασης που χρησιμοποιείται από τη βιβλιοθήκη. Επιπλέον, η σημασιολογική ανάλυση που παρείχε η DBpedia βελτίωσε σημαντικά την απόδοση της προτεινόμενης υπηρεσίας.

Ολοκληρώνοντας, η αξιολόγηση της προτεινόμενης υπηρεσίας δείχνει ξεκάθαρα ότι οι βιβλιοθήκες μπορούν να επωφεληθούν πολύ από τη χρήση υπηρεσιών, όπως το Twitter στους OPAC τους. Είναι επίσης προφανές ότι σημασιολογικές διαδικτυακές υπηρεσίες θα μπορούσαν να παίξουν ένα σημαντικό ρόλο στην ενίσχυση της συνολικής απόδοσης των υπηρεσιών που βασίζονται σε πληροφορία εξωτερική της βιβλιο-

θήκης. Μελλοντική εργασία, θα μπορούσε να είναι η εστίαση στη μείωση του «*θορύβου*» των tweets και στην εξεύρεση τρόπων στην εφαρμογή της προτεινόμενης μεθοδολογίας σε ολόκληρη τη συλλογή της βιβλιοθήκης με χρήση των κατάλληλων υπηρεσιών.

Με την ολοκλήρωση αυτού του κεφαλαίου, έγινε μια συνολική παρουσίαση των δυνατοτήτων που προσφέρουν τα ελεγχόμενα και μη λεξιλόγια και οι τεχνολογίες των Συνδεδεμένων Δεδομένων στην παροχή ποιοτικότερων και πληρέστερων αποτελεσμάτων αναζήτησης στους τελικούς χρήστες.

9. Συμπεράσματα

9.1. Ανακεφαλαίωση

Κύρια επιδίωξη της παρούσας διατριβής είναι η δημιουργία καινοτόμων υπηρεσιών και μεθοδολογιών προστιθέμενης αξίας για την εκμετάλλευση των παραδοσιακών εργαλείων θεματικής επεξεργασίας υλικού. Οι δράσεις αυτής της διατριβής βασίζονται στις τεχνολογίες του σημασιολογικού ιστού.

Αρχικά, η παρούσα διατριβή ασχολείται με τη θεματική ευρετηρίαση στις βιβλιοθήκες και τα ιδρύματα μνήμης εν γένει. Πιο συγκεκριμένα, παρουσιάζονται διεξοδικά τα παραδοσιακά εργαλεία (δηλ. τα ελεγχόμενα λεξιλόγια) που χρησιμοποιούν εδώ και πολλές δεκαετίες οι βιβλιοθήκες για να ευρετηριάσουν το υλικό τους. Τα ελεγχόμενα λεξιλόγια μπορούν να χωριστούν σε τρεις βασικές κατηγορίες: τις θεματικές επικεφαλίδες, τους θησαυρούς και τα ταξινομικά σχήματα. Αναλύονται τα σημεία που συγκλίνουν και αποκλίνουν τα εν λόγω εργαλεία και εξετάζονται οι δυνατότητες που παρέχει το κάθε ένα από αυτά. Κατόπιν, περιγράφονται οι σύγχρονες τάσεις στη θεματική ανάκτηση πληροφοριών και παρουσιάζεται μια μεθοδολογία που εκμεταλλεύεται τη σημασιολογική εκφραστικότητα των LCSH. Πιο συγκεκριμένα, η μεθοδολογία εκμεταλλεύεται τη σημασιολογική εκφραστικότητα των LCSH, όπως αυτή ορίζεται μέσω της απλής συνδετικής δομής (δηλ. τους ευρύτερους, στενότερους και σχετικούς όρους) και την επεκτείνει μέσω της χρήσης της εκτεταμένης συνδετικής δομής, όπου ορίζεται ως η σύνδεση των θεματικών επικεφαλίδων μέσω των τυχόν κοινών υποδιαίρεσεων που μπορεί να έχουν. Έπειτα, η προκύπτουσα εκτεταμένη συνδετική δομή αντιστοιχίζεται σε ένα κατάλληλα σχεδιασμένο οντολογικό σχήμα, το οποίο μπορεί να εφαρμοστεί σε οποιαδήποτε ψηφιακή βιβλιοθήκη που περιέχει θεματικούς περιγραφείς σύμφωνα με τις οδηγίες των LCSH. Ύστερα, προκειμένου να αξιοποιηθεί ο σημασιολογικός πλούτος της υποκείμενης οντολογίας, παρουσιάζεται ένα κατάλληλο γραφικό περιβάλλον διεπαφής χρήστη, το οποίο δίνει τη δυνατότητα στους χρήστες να εξοικειωθούν με την ορολογία που υπάρχει στις LCSH και τους καθοδηγεί να επιλέξουν το θέμα που ταιριάζει καλύτερα στις πληροφοριακές τους ανάγκες. Η καθοδηγούμενη πλοήγηση βασίζεται στις σχέσεις ανάμεσα στις θεματικές επικεφαλίδες που προέρχονται από την εκτεταμένη συνδετική δομή που αναφέρθηκε προηγουμένως. Η εν λόγω οντολογία βρίσκει εφαρμογή στην ψηφιακή βιβλιοθήκη (Διώνη), η οποία βασίζεται στο DSpace, ως επέκταση της λειτουργικότητας της παραδοσιακής θεματικής αναζήτησης που προσφέρεται από αυτήν. Μέσα σε αυτό το πλαίσιο, δίνεται η δυνατότητα στους

χρήστες να αναζητήσουν και να πλοηγηθούν στις θεματικές επικεφαλίδες που περιγράφουν την υποκείμενη συλλογή. Τέλος, η υπηρεσία αξιολογείται μέσω της συγκριτικής αξιολόγησης της παραδοσιακής θεματικής λειτουργικότητας που προσφέρεται από την ψηφιακή βιβλιοθήκη (Διώνη) και της θεματικής λειτουργικότητας που προσφέρεται από την προτεινόμενη υπηρεσία με την εκτέλεση σεναρίων χρήσης.

Στη συνέχεια, για να πραγματοποιηθεί μια εμπειριστατωμένη παρουσίαση των θεματικών εργαλείων που χρησιμοποιούνται από τις βιβλιοθήκες παρουσιάζονται και τα μη ελεγχόμενα λεξιλόγια και πιο συγκεκριμένα οι λαοταξινομίες που είναι ιδιαίτερα δημοφιλείς ανάμεσα στους χρήστες του διαδικτύου. Η έρευνα ολοκληρώνεται με μια ανάλυση των θετικών και των αρνητικών σημείων κατά τη χρήση των ελεγχόμενων και μη λεξιλογίων στη θεματική ευρετηρίαση.

Παράλληλα, μέσα από την αναλυτική παρουσίαση των εργαλείων θεματικής ευρετηρίασης, γίνεται κατανοητή η ανάγκη που υπάρχει για τη συγχώνευση ή αντιστοίχισή τους, αφού μέσα στα χρόνια έχουν δημιουργηθεί διάφορα συστήματα ελεγχόμενων λεξιλογίων για να εξυπηρετήσουν τις ανάγκες συγκεκριμένων τομέων και κοινοτήτων χρηστών. Αυτά τα συστήματα τις περισσότερες φορές δεν είναι συμβατά μεταξύ τους, οδηγώντας σε προβλήματα διαλειτουργικότητας ανάμεσα σε διαφορετικούς οργανισμούς και τομείς γνώσης, με αποτέλεσμα η γνώση και οι πληροφορίες που υπάρχουν σε έναν οργανισμό να μένουν εντός των τειχών του. Έτσι, κατά τη διάρκεια των ετών πολλές προσπάθειες έγιναν για να μπορέσουν να συγχωνευθούν αυτά τα λεξιλόγια μεταξύ τους με απώτερο σκοπό να εξυπηρετηθεί η επικοινωνία και η ανταλλαγή πληροφοριών ανάμεσα σε διαφορετικές υπηρεσίες και βιβλιοθήκες. Μέσα σε αυτό το πλαίσιο, παρουσιάζεται μια μεθοδολογία, η οποία είναι ικανή να ενσωματώνει τα ήδη υπάρχοντα ευρετήρια βασισμένα στις LCSH με θησαυρούς μέσα στα πλαίσια των δραστηρίων συστημάτων ανάκτησης πληροφοριών. Με αυτό τον τρόπο, τα εγγενή πλεονεκτήματα των θησαυρών αξιοποιούνται χωρίς να χρειάζεται να ευρετηριαστεί ξανά η υποκείμενη συλλογή της βιβλιοθήκης. Για να επιτευχθεί αυτό, μια μεθοδολογία αντιστοίχισης/ταύτισης παρουσιάζεται, η οποία είναι ικανή να παρέχει μια κοινή δομή και για τα θέματα των LCSH και για τα θέματα των θησαυρών. Η δομή εκφράζεται ως μια SKOS οντολογία, η οποία μπορεί να χρησιμοποιηθεί από κατάλληλα συστήματα ανάκτησης πληροφοριών βασισμένα σε θέματα. Ως απόδειξη της ορθότητας της ιδέας, η προτεινόμενη μεθοδολογία ταύτισης εφαρμόζεται στο σύστημα θεματικής ανάκτησης πληροφοριών της ψηφιακής βιβλιοθήκης (Διώνη) που βασίζεται στις LCSH και τους περιγραφείς του θησαυρού ILO. Η προκύπτουσα οντολογία αποθηκεύεται σε ένα

κατάλληλο σύστημα βάσης δεδομένων. Οι τελικοί χρήστες αλληλεπιδρούν με την υποκείμενη οντολογία μέσω ενός αντίστοιχα σχεδιασμένου γραφικού περιβάλλοντος διεπαφής. Για να αξιολογηθεί η αποτελεσματικότητα της προτεινόμενης προσέγγισης γίνεται σύγκριση της αρχικής έκδοσης του συστήματος με την προκύπτουσα έκδοση της κεντρικής συνιστώσας του συστήματος ανάκτησης πληροφοριών βασισμένο στα θέματα, δηλαδή την κοινή δομή.

Μάλιστα, στις μέρες μας, με την έκρηξη του διαδικτύου και τη γρήγορη και άμεση διάδοση πληροφοριών που αυτό προσφέρει, η ανάγκη για διαλειτουργικότητα ανάμεσα στους διάφορους οργανισμούς γίνεται ολοένα και περισσότερο επιτακτική. Μέσα σε αυτό το πλαίσιο, μελετώνται οι τεχνολογίες του σημασιολογικού ιστού και ειδικότερα τα Συνδεδεμένα Δεδομένα. Πιο συγκεκριμένα, με την υιοθέτηση αυτών των τεχνολογιών δίνεται η δυνατότητα στις βιβλιοθήκες να αντιμετωπίσουν προβλήματα διαλειτουργικότητας που τις βασανίζουν εδώ και πολλές δεκαετίες και να κάνουν ισχυρή την παρουσία τους στο διαδίκτυο. Στη συνέχεια, παρουσιάζονται αναλυτικά τα βασικά συστατικά που απαιτούνται για τη συμμετοχή στο κίνημα των Συνδεδεμένων Δεδομένων. Ιδιαίτερη βαρύτητα δίνεται στο SKOS, το οποίο είναι το βασικό εργαλείο για την περιγραφή ελεγχόμενων λεξιλογίων στο κίνημα των Συνδεδεμένων Δεδομένων.

Έχοντας τα παραπάνω κατά νου, δίνεται το παράδειγμα μιας παραδοσιακής εφαρμογής θεματικής πλοήγησης και αναζήτησης που βασίζεται στην ψηφιακή βιβλιοθήκη (Διώνη) και πως μπορεί αυτή να μετατραπεί σε μορφή συμβατή με τα Συνδεδεμένα Δεδομένα, ακολουθώντας πιστά τους κανόνες και τις οδηγίες για τη δημιουργία τέτοιων υπηρεσιών. Η προτεινόμενη υπηρεσία όχι μόνο εκθέτει τη θεματική της πληροφορία στην ευρύτερη κοινότητα των Συνδεδεμένων Δεδομένων, αλλά καταφέρνει επίσης να παρέχει σχετικές πηγές που προέρχονται από απομακρυσμένα αποθετήρια (δηλ. άρθρα των NYT) μέσω της ευθυγράμμισης των εν λόγω πληροφοριών με ένα διαφανή και ξεκάθαρο τρόπο για τον τελικό χρήστη. Η προτεινόμενη υπηρεσία αξιολογείται ανάλογα μέσω της χρήσης μιας έρευνας χρηστών.

Κατόπιν, παρουσιάζεται μια ακόμα δυνατότητα που προσφέρουν τα Συνδεδεμένα Δεδομένα και είναι αυτή της αντιστοίχισης δύο διαφορετικών ελεγχόμενων λεξιλογίων μέσω της χρήσης γλωσσικών εργαλείων για την εύρεση περισσότερων κοινών όρων ανάμεσα σε αυτά. Πιο συγκεκριμένα, προτείνεται μια μεθοδολογία που στοχεύει στο συγκεκριμένο σημασιολογικά παρόμοιων αλλά διαφορετικών ελεγχόμενων λεξιλογίων μέσω της χρήσης σημασιολογικής ευθυγράμμισης των υποκείμενων όρων. Η σημασιο-

λογική ευθυγράμμιση πραγματοποιείται με την εκμετάλλευση τεχνολογιών των Συνδεδεμένων Δεδομένων για την ταύτιση σημασιολογικά ισοδύναμων όρων ανάμεσα σε διαφορετικά ελεγχόμενα λεξιλόγια και δεν βασίζεται στην εγγενή δομή των αντίστοιχων λεξιλογίων. Αντίθετα, εκμεταλλεύεται εξωτερικά γλωσσικά εργαλεία που είναι διαθέσιμα στο σύννεφο των Συνδεδεμένων Δεδομένων. Αυτά τα σύνολα δεδομένων περιέχουν ρητά ορισμένες γλωσσικές συσχετίσεις (δηλ. συνώνυμα, υπώνυμα, μερώνυμα, ολώνυμα κλπ.) ανάμεσα στους όρους τους. Η προτεινόμενη προσέγγιση βρίσκει εφαρμογή στις θεματικές επικεφαλίδες της ψηφιακής βιβλιοθήκης (Διώνη) και στις θεματικές επικεφαλίδες των NYT μέσω της χρήσης της DBpedia και του WordNet. Για να αξιολογηθεί η αποτελεσματικότητα της προτεινόμενης προσέγγισης, συγκρίνονται τα αποτελέσματα της εφαρμογής της συγκεκριμένης προσέγγισης με τα αποτελέσματα της εφαρμογής μιας προσέγγισης που βασίζεται στην ανακάλυψη μόνο λεξιλογικών ομοιοτήτων.

Αφού έγινε μια συνολική παρουσίαση της χρήσης των ελεγχόμενων και μη λεξιλογίων από τις βιβλιοθήκες με απώτερο σκοπό την καλύτερη και αποδοτικότερη παροχή αποτελεσμάτων αναζήτησης στους τελικούς χρήστες για την κάλυψη των πληροφοριακών τους αναγκών, στο κεφάλαιο 7 γίνεται μια προσπάθεια σύγκλισης ανάμεσα στα παραδοσιακά εργαλεία που χρησιμοποιούν οι βιβλιοθήκες για να αποθηκεύουν τα θεματικά τους ευρετήρια (π.χ. MARC21, UNIMARC) και στις τεχνολογίες του σημασιολογικού ιστού. Πιο συγκεκριμένα, εξετάζονται τα πεδία του MARC που θα μπορούσαν να φιλοξενήσουν URIs. Μάλιστα, η έρευνα χωρίζεται σε δύο επίπεδα: α) στον έλεγχο των πεδίων του MARC που θα μπορούσαν να φιλοξενήσουν URIs που έχουν οριστεί τοπικά από ένα συγκεκριμένο οργανισμό και β) στον έλεγχο των πεδίων του MARC που θα μπορούσαν να φιλοξενήσουν URI που έχουν οριστεί σε απομακρυσμένα σημεία από κάποιον άλλο οργανισμό. Αφού εξετάστηκαν διεξοδικά τα πεδία μέσω της μελέτης των ορισμών που δίνονται από το εκάστοτε πρότυπο (π.χ. MARC21, UNIMARC), γίνεται έλεγχος στις εθνικές βιβλιοθήκες ανά τον κόσμο που έχουν τη μεγαλύτερη ενασχόληση με τα Συνδεδεμένα Δεδομένα για να διαπιστωθεί κατά πόσο επιλέγουν να προσθέσουν πληροφορία των Συνδεδεμένων Δεδομένων στις ήδη υπάρχουσες υπηρεσίες τους ή προτιμούν να δημιουργούν καινούριες υπηρεσίες βασισμένες αποκλειστικά στις νέες τεχνολογίες.

Έπειτα, η διατριβή ασχολείται με ένα άλλο σημαντικό κομμάτι που απασχολεί τη βιβλιοθηκονομική κοινότητα: αυτό της χρονικής διάστασης της πληροφορίας και πως μπορούν οι βιβλιοθήκες να παρέχουν έγκαιρη και έγκυρη πληροφόρηση στους χρήστες

τους. Υπό αυτό το πρίσμα, μελετώνται τα παραδοσιακά εργαλεία που χρησιμοποιούν οι βιβλιοθήκες για να αντιμετωπίσουν το πρόβλημα αυτό (π.χ. κίνηση των βιβλίων, ανάλυση των αρχείων καταγραφής, διαδανεισμός κλπ.). Η έρευνα καταλήγει στο συμπέρασμα ότι τα περισσότερα από τα αυτά τα εργαλεία είναι χρήσιμα για την ανάπτυξη της συλλογής αλλά δεν μπορούν να αντιμετωπίσουν το πρόβλημα της παροχής σύγχρονης και επίκαιρης πληροφόρησης στους χρήστες. Προκειμένου να αντιμετωπιστεί αυτό το ζήτημα, εξετάζονται οι λαοταξινομίες και πιο συγκεκριμένα τα hashtags του Twitter και πως αυτά μπορούν να χρησιμοποιηθούν στον κατάλογο της βιβλιοθήκης για την παροχή τρέχουσας και επίκαιρης πληροφόρησης στους χρήστες. Πιο αναλυτικά, μια υπηρεσία αναπτύσσεται, η οποία είναι ικανή να αναγνωρίζει τα επίκαιρα θέματα που συλλαμβάνονται από τα hashtags του Twitter. Τα επίκαιρα θέματα ύστερα, επεκτείνονται μέσω της χρήσης των κατάλληλων τεχνολογιών του σημασιολογικού ιστού. Κατόπιν, τα θέματα που προκύπτουν απευθύνονται στον κατάλογο μιας βιβλιοθήκης για τον εντοπισμό σχετικών τεκμηρίων και την πρόταση αυτών στους χρήστες. Με αυτό τον τρόπο παρουσιάζονται στους χρήστες τεκμήρια που σχετίζονται με τα επίκαιρα θέματα έγκαιρα. Η υπηρεσία βρίσκει εφαρμογή στον κατάλογο της Βιβλιοθήκης και Κέντρου Πληροφόρησης του Ιονίου Πανεπιστημίου. Η προτεινόμενη υπηρεσία αξιολογείται μέσω της σύγκρισης του αριθμού των αιτημάτων για τεκμήρια που προέρχονται από την προτεινόμενη υπηρεσία με τον αριθμό των αιτημάτων για τεκμήρια που προέρχονται από τις υπόλοιπες υπηρεσίες ανάκτησης της βιβλιοθήκης.

9.2. Πιθανή μελλοντική εργασία

Το αντικείμενο της διατριβής είναι ένας ερευνητικός χώρος όπου απασχολεί τις βιβλιοθήκες και τα ιδρύματα μνήμης εδώ και πολλές δεκαετίες. Η ανάγκη για την παροχή αποτελεσματικών υπηρεσιών στους χρήστες που θα έχουν ως στόχο την ενίσχυση της αναζήτησης και της ανάκτησης υλικού για την κάλυψη των πληροφοριακών τους αναγκών, θα είναι πάντα μια βασική προτεραιότητα. Μέσα σε αυτό το πλαίσιο, τα προαναφερθέντα ευρήματα της παρούσας διατριβής θα μπορούσαν σε πολλές περιπτώσεις να δώσουν έναυσμα για περαιτέρω μελέτη και δημιουργία υπηρεσιών προστιθέμενης αξίας ή ακόμα και τη βελτίωση των προτεινόμενων υπηρεσιών και μεθοδολογιών για την παροχή ακόμα πιο ικανοποιητικών αποτελεσμάτων όσον αφορά στην αναζήτηση και ανάκτηση πληροφοριών. Η πιθανή μελλοντική εργασία, θα μπορούσε να βασιστεί στους ακόλουθους 5 τομείς όπου επικεντρώθηκε και η έρευνα της συγκεκριμένης διατριβής.

9.2.1. Εκμετάλλευση της σημασιολογικής εκφραστικότητας των LCSH

Η μεθοδολογία που προτάθηκε και εκμεταλλεύεται την άμεση και έμμεση σημασιολογική εκφραστικότητα των θεματικών επικεφαλίδων που συμμορφώνονται με τις οδηγίες δημιουργίας των LCSH (βλέπε κεφάλαιο 2), μπορεί να συνεισφέρει στη δημιουργία πιο αποδοτικών υπηρεσιών θεματικής ανάκτησης πληροφοριών στις ψηφιακές βιβλιοθήκες. Η ευρεία χρήση των LCSH σε πολλές ψηφιακές βιβλιοθήκες ανά τον κόσμο, καθιστά τα ευρήματα της διεξαγόμενης έρευνας εφαρμόσιμα σε έναν αριθμό ιδρυμάτων που διατηρούν συστήματα ψηφιακών βιβλιοθηκών.

Η μελλοντική εργασία θα μπορούσε να προσανατολιστεί στην εξερεύνηση άλλων σημασιολογικά πλούσιων πληροφοριακών δομών που παράγονται από τους βιβλιοθηκονόμους κατά τη διάρκεια της εργασίας τους. Ο απώτερος στόχος αυτής της μεθοδολογίας είναι να συμβάλλει στην ανάπτυξη ψηφιακών βιβλιοθηκών επόμενης γενιάς, που θα είναι ικανές να εκμεταλλεύονται τον πλούτο της γνώσης, ο οποίος είναι συσσωρευμένος στις παραδοσιακές βιβλιοθήκες.

9.2.2. Μεθοδολογία για τη συγχώνευση ελεγχόμενων λεξιλογίων

Η μεθοδολογία που προτάθηκε για τη συγχώνευση ελεγχόμενων λεξιλογίων (βλέπε κεφάλαιο 3) είναι ικανή να ενσωματώνει το ευρέως χρησιμοποιούμενο ελεγχόμενο λεξιλόγιο των LCSH με ένα θησαυρό. Η ενσωμάτωση καταλήγει στη δημιουργία μιας κοινής δομής, η οποία εκφράζεται ως μια οντολογία που περιέχει μαζί τους όρους των LCSH και τους περιγραφείς του θησαυρού ακολουθώντας τις προδιαγραφές του SKOS. Η οντολογία θα μπορούσε να χρησιμοποιηθεί από διάφορα συστήματα ανάκτησης πληροφοριών βασισμένα σε θέματα, χωρίς να χρειάζεται να ξαναχτιστεί το υπάρχον θεματικό ευρετήριο. Η προτεινόμενη προσέγγιση θα μπορούσε επίσης να εφαρμοστεί σε άλλες συλλογές πέρα από τον τομέα των βιβλιοθηκών, με την προϋπόθεση ότι αυτές οι συλλογές χρησιμοποιούν ελεγχόμενα λεξιλόγια που διαθέτουν κάποιο είδος συνδετικής δομής.

Τέλος, αναμένεται ότι οι πιθανές εξελίξεις στην έρευνα του σημασιολογικού ιστού θα μπορούσαν να ωφεληθούν από την προτεινόμενη προσέγγιση. Αυτό οφείλεται στο γεγονός ότι η προτεινόμενη προσέγγιση βασίζεται σε πρότυπα του σημασιολογικού ιστού, όπως είναι οι προδιαγραφές SKOS. Επομένως, η μελλοντική εργασία θα μπορούσε να στοχεύει στην χρήση OWL reasoners ικανούς να προσθέσουν νέα ταξινομικά σχήματα στην κοινή δομή.

9.2.3. Μεθοδολογία μετατροπής θεματικών ευρετηρίων σε μορφή των Συνδεδεμένων Δεδομένων

Η προτεινόμενη προσέγγιση για τη μετατροπή μιας υπηρεσίας θεματικής πλοήγησης σε μορφή συμβατή με τα Συνδεδεμένα Δεδομένα (βλέπε κεφάλαιο 5), θα μπορούσε να αποτελέσει το παράδειγμα για τη μετατροπή και άλλων ψηφιακών βιβλιοθηκών σε μορφή συμβατή με τα Συνδεδεμένα Δεδομένα, οι οποίες θα μπορούσαν να συνδέονται μεταξύ τους. Έπειτα, ο απώτερος στόχος μιας τέτοιας προσπάθειας θα μπορούσε να ήταν η δημιουργία μιας ομόσπονδης ψηφιακής βιβλιοθήκης μεταπτυχιακών και διδακτορικών διατριβών αποτελούμενη από ψηφιακές βιβλιοθήκες-μέλη που επικοινωνούν μεταξύ τους με τη χρήση τεχνολογιών των Συνδεδεμένων Δεδομένων.

9.2.4. Μεθοδολογία για την αντιστοίχιση σημασιολογικά παρόμοιων ε- λεγχόμενων λεξιλογίων μέσω τεχνολογιών των Συνδεδεμένων Δε- δομένων

Η μεθοδολογία για την αντιστοίχιση σημασιολογικά παρόμοιων ελεγχόμενων λεξιλογίων (βλέπε κεφάλαιο 6) θα μπορούσε να αποτελέσει πρότυπο για την αντιστοίχιση συναφών λεξιλογίων. Μάλιστα, η χρήση ευρέως διαθέσιμων εργαλείων, όπως το Google Refine, καθιστά την εφαρμογή της προτεινόμενης μεθοδολογίας ως μια απλή και ξεκάθαρη διαδικασία, η οποία μπορεί να εφαρμοστεί και σε άλλες περιπτώσεις που έχουν ως στόχο την ανακάλυψη ισοδύναμων όρων σε διαφορετικά αλλά σημασιολογικά όμοια σύνολα δεδομένων και μάλιστα από ανθρώπους που δεν έχουν ιδιαίτερες τεχνολογικές γνώσεις. Η εφαρμογή της προτεινόμενης μεθοδολογίας εξυπηρετείται μέσω της χρήσης τεχνολογιών των Συνδεδεμένων Δεδομένων. Φαίνεται αρκετά λογικό να υποθέσουμε ότι η χρήση ακόμα περισσότερων γλωσσικών συνόλων δεδομένων από το σύννεφο των Συνδεδεμένων Δεδομένων θα μπορούσε να βελτιώσει την αποτελεσματικότητα της προτεινόμενης μεθοδολογίας. Επομένως, η μελλοντική εργασία θα μπορούσε να έχει ως στόχο την αντιστοίχιση των θεματικών επικεφαλίδων της ψηφιακής βιβλιοθήκης (Διώνη) με άλλες υπηρεσίες των Συνδεδεμένων Δεδομένων, όπως είναι το RAMEAU, το SWD, η BNE και το LIBRIS κ.ά.

9.2.5. Τα hashtags στην αντιμετώπιση του προβλήματος της χρονικής διάστασης της πληροφορίας

Η διαδικασία αντιστοίχισης των hashtags με τεκμήρια της βιβλιοθήκης (βλέπε κεφάλαιο 8) είναι μια πρωτότυπη ιδέα, η οποία θα μπορούσε δυνητικά να προσφέρει πολλά στον τομέα των βιβλιοθηκών, αφού επιτυγχάνει με απλό και άμεσο τρόπο να αφουγκράζεται την επικαιρότητα και τις τρέχουσες ανάγκες των χρηστών. Το βασικό συμπέρασμα που μπορεί να εξαχθεί από το πείραμα είναι το γεγονός ότι είναι εφικτή η ενσωμάτωση πληροφοριών που προέρχονται από εξωτερικές υπηρεσίες κοινωνικής δικτύωσης. Βέβαια από τα αποτελέσματα του πειράματος είναι προφανές ότι ακόμα και σε μια εξειδικευμένη περιοχή ενδιαφέροντος, θα υπάρξουν πολλά μη σχετικά tweets μέχρις ότου να βρεθεί ένα tweet που να αντιστοιχεί σε ένα τεκμήριο της βιβλιοθήκης. Αυτό βέβαια μπορεί να οφείλεται εν μέρει και στο γεγονός ότι το λεξιλόγιο των tweets διαφέρει σημαντικά από τη γλώσσα ευρετηρίασης που χρησιμοποιείται από τη βιβλιοθήκη. Επιπλέον, η χρήση γλωσσικών λεξιλογίων (π.χ. DBpedia, WordNet κλπ.) θα μπορούσε να βελτιώσει σημαντικά την απόδοση της προτεινόμενης υπηρεσίας. Μελλοντική εργασία, θα μπορούσε να αναφέρεται στη μείωση του «*θορύβου*» των tweets και στην εξεύρεση τρόπων για την εφαρμογή της προτεινόμενης μεθοδολογίας σε ολόκληρη τη συλλογή μιας βιβλιοθήκης με τη χρήση των κατάλληλων υπηρεσιών.

Βιβλιογραφία

- Agee, J. (2005). Collection Evaluation: a Foundation for Collection Development. *Collection Building*, 24 (3), pp. 92-95.
- Aharony, N. (2010). Twitter Use in Libraries: an Exploratory Analysis. *Journal of Web Librarianship*, 4 (4), pp. 333-350.
- Aitchison, J. and Clarke, S. (2004). The Thesaurus: a Historical Viewpoint, with a Look to the Future. *Cataloging & Classification Quarterly*, 37 (3-4), pp. 5-21.
- ALCTS/CCS/SAC/Subcommittee on Metadata and Classification. (1999). Final Report. [online] Διαθέσιμο στο: http://www.ala.org/alcts/resources/org/cat/subjectdata_record [Πρόσβαση: 08/07/2015].
- Alemu, G., Stevens, B., Ross, P. and Chandler, J. (2012). Linked Data for Libraries. *New Library World*, 113 (11/12), pp. 549-570.
- Alistair, M., Matthews, B., Beckett, D., Brickley, D., Wilson, M. and Rogers, N. (2005). SKOS: a Language to Describe Simple Knowledge Structures for the Web. In: *XTech 2005: XML, the Web and Beyond*. [online] Διαθέσιμο στο: <http://ids.snu.ac.kr/w/images/f/f1/SC18.pdf> [Πρόσβαση: 08/09/2015].
- Anderson, J.D. and Hofmann, M.A. (2006). A Fully Faceted Syntax for Library of Congress Subject Headings. *Cataloging & Classification Quarterly*, 43 (1), pp. 7-38.
- Anderson, P. (2007). What is Web 2.0? Ideas, Technologies and Implications for Education. *JISC Technology and Standards Watch*. [online] Διαθέσιμο στο: <http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf> [Πρόσβαση: 25/05/2015].
- Antelman, K., Lynema, E. and Pace, A.K. (2006). Toward a Twenty-first Century Library Catalog. *Information Technology and Libraries*, 25 (3), pp. 128-139.
- Archer, P., Goedertier, S. and Loutas, N. (2012). *Deliverable: D7.1.3 - Study on Persistent URIs, with Identification of Best Practices and Recommendations on the Topic for the MSs and the EC*. European Union: Interoperability Solutions for Eu-

- European Public Administrations. [online] Διαθέσιμο στο: https://joinup.ec.europa.eu/sites/default/files/D7.1.3%20-%20Study%20on%20persistent%20URIs_0.pdf [Πρόσβαση: 08/07/2015].
- Asghar, A. Revie, C. and Ghowdhury, G. (2002). Thesaurus-enhanced Search Interfaces. *Journal of Information Science*, 28 (2), pp. 111-122.
- Auray, N. (2007). Folksonomy: the New Way to Serendipity. *Communication & Strategies*, 65, pp. 67-89.
- Baader, F. (2007). *The Description Logic Handbook*. Cambridge: Cambridge University Press.
- Baker, T., Bermes, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., Panzer, M., Schneider, J. and Singer, R. (2011). *Library Linked Data Incubator Group Final Report. W3C Incubator Group Report. World Wide Web Consortium*. [online] Διαθέσιμο στο: <http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025/> [Πρόσβαση: 15/07/2015].
- Bates, M. (1998). Indexing and Access for Digital Libraries and the Internet: Human Database and Domain Factors. *Journal of the American Society for Information Science*, 49 (13), pp. 1186-1205.
- Bates, M. (2002). The Cascade of Interactions in the Digital Library Interface. *Information Processing & Management*, 38, pp. 381-400.
- Bechhofer, S. and Goble, C. (2001). Thesaurus Construction through Knowledge Representation. *Data & Knowledge Engineering*, 37, pp. 25-45.
- Beckett, D., Berners-Lee, T., Prud'hommeaux, E., Carothers, G. and Lex Machina Inc. (2014). *RDF 1.1 Turtle Terse RDF Triple Language W3C Recommendation 25 February 2014*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/2014/REC-turtle-20140225/> [Πρόσβαση: 09/08/2015].
- Bennett, D. E. (2006). Immaculate Catalogues, Indexes and Monsters Too.... *Ariadne Online Journal*, 49. [online] Διαθέσιμο στο: <http://www.ariadne.ac.uk/issue49/cig-2006-rpt/> [Πρόσβαση: 09/08/2015].

- Berners-Lee, T. (1996). *Universal Resource Identifiers - Axioms of Web Architecture*. [online] Διαθέσιμο στο: <http://www.w3.org/DesignIssues/Axioms.html> [Πρόσβαση: 07/07/2015].
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001). The Semantic Web. *Scientific American*, 284, pp. 34-43.
- Berners-Lee, T. (2006). *Linked Data - Design Issues*. [online] Διαθέσιμο στο: <http://www.w3.org/DesignIssues/LinkedData.html> [Πρόσβαση: 07/07/2015].
- Berners-Lee, T., Hall, W., Hendler, J.A., O' Hara, K., Shadbolt, N. and Weitzner, D.J. (2006). A Framework for Web Science. *Now the Essence of Knowledge: Foundations and Trends in Web Science*, 1 (1), pp. 1-130. <DOI: 10.1561/1800000001>
- Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. and Hellmann, S. (2009). DBpedia - A Crystallization Point for the Web of Data. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 7, pp. 154-165.
- Bizer, C., Heath, T. and Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5 (3), pp.1-22.
- Bodoff, D. and Kambil, A. (1998). Partial Coordination. I. The Best of Pre-coordination and Post-coordination. *Journal of the American Society for Information Science*, 49 (14), pp. 1254-1269.
- Borin, J. and Yi, H. (2008). Indicators for Collection Evaluation: a New Dimensional Framework. *Collection Building*, 27 (4), pp. 136-143.
- Borlund, P. (2003). The IIR Evaluation Model: a Framework for Evaluation of Interactive Information Retrieval Systems. *Information Research*, 8 (3). [online] Διαθέσιμο στο: <http://www.informationr.net/ir/8-3/paper152.html> [Πρόσβαση: 10/07/2015].
- Bovee, M., Srivastava, R.P. and Mak, B. (2001). A Conceptual Framework and Belief-function Approach to Assessing Overall Information Quality. In: *Proceedings of International Conference on Information Quality (ICIQ 2001)*.

- Brickley, D. and Guha, R.V. (2014). *RDF Schema 1.1: W3C Recommendation 25 February 2014*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/rdf-schema/> [Πρόσβαση: 08/07/2015].
- Calhoun, K. (2006). *The Changing Nature of the Catalog and its Integration with Other Discovery Tools*. Final Report. Washington, DC: Library of Congress. [online] Διαθέσιμο στο: <http://www.loc.gov/catdir/calhoun-report-final.pdf> [Πρόσβαση: 07/07/2015].
- Caplan, P. (2003). *Metadata Fundamentals for All Librarians*. Chicago, IL: American Library Association.
- Carscaddon, L. and Chapman, K. (2013). Twitter as a Marketing Tool for Libraries. In: Thomsett-Scott, B.C. (Ed.), *Marketing with Social Media: A LITA Guide*. Chicago: American Library Association.
- Chamis, A.Y. (1991). *Vocabulary Control and Search Strategies in Online Searching*. Westport, Conn.: Greenwood Press.
- Chan, L.M. (1995). *Library of Congress Subject Headings: Principles and Application*. Englewood, Co.: Libraries Unlimited.
- Chang, H.C. and Iyer, H. (2012). Trends in Twitter Hashtag Applications: Design Features for Value-added Dimensions to Future Library Catalogues. *Library Trends*, 61 (1), pp. 248-258.
- Chaplan, M.A. (1995). Mapping Laborline Thesaurus Terms to Library of Congress Subject Headings: Implications for Vocabulary Switching. *Library Quarterly*, 65 (1), pp. 39-61.
- Cleveland, D. and Cleveland, A. (2013) 4th ed. *Introduction to Indexing and Abstracting*. Santa Barbara, California: ABC-CLIO, LLC.
- Covey, D.T. (2002). *Usage and Usability Assessment: Library Practices and Concerns*. Digital Library Federation, CLIR, Washington, DC.
- Crawley-Low, J.V. (2002). Collection Analysis Techniques Used to Evaluate a Graduate-level Toxicology Collection. *Journal of Medical Libraries Association*, 90 (3), pp. 310-316.

- Cunha, E., Magno, G., Comarcel, G., Almeida, V. Goncalves, M.A. and Benevenuto, F. (2011). Analyzing the Dynamic Evolution of Hashtags on Twitter: a Language-based Approach. In: *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, pp. 58-65.
- Dean, R.J. (2004). FAST: Development of Simplified Headings for Metadata. *Cataloguing & Classification Quarterly*, 39 (1-2), pp. 331-352.
- Doerr, M. (2001). Semantic Problems of Thesaurus Mapping. *Journal of Digital Information*, 1 (8). [online] Διαθέσιμο στο: <https://journals.tdl.org/jodi/index.php/jodi/article/view/31/32> [Πρόσβαση: 08/07/2015].
- Edelstein, J., Galla, L., Li-Madeo, C., Marden, J., Rhonemus, A. and Whyssel, N. (2013). *Linked Open Data for Cultural Heritage: Evolution of an Information Technology*. [online] Διαθέσιμο στο: <http://www.whysel.com/papers/LIS670-Linked-Open-Data-for-Cultural-Heritage.pdf> [Πρόσβαση: 15/07/2015].
- Euzenat, J. and Shvaiko, P. (2007). *Ontology Matching*. Berlin, Heidelberg: Springer-Verlag.
- Euzenat, J. and Shvaiko, P. (2013). 2nd ed. *Ontology Matching*. Berlin, Heidelberg: Springer-Verlag.
- Fallside, D. and Walmsley, P. (2004). *XML Schema Part 0: Primer Second Edition: W3C Recommendation 28 October 2004*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/xmlschema-0/> [Πρόσβαση: 17/07/2015].
- Farooq, A., Arshad, M.J. and Shah, A. (2010). A layered Approach for Similarity Measurement between Ontologies. *Journal of American Science*, 6 (12), pp. 69-77.
- Francesconi, E., Faro, S. and Marinai, E. (2008). A Framework for Semantic Mapping between Thesauri. In: *Proceedings of the 2nd International Conference on Theory and Practice of Electronic Governance – ICEGOV'08, vol. 351* (pp. 251-257). New York: ACM.
- Galvao, R.M. and Cordeiro, M.I. (2013). UNIMARC - Understanding the Past to Envision Future. *IFLA Journal*, 39, pp. 151-161.

- German National Library. (2013). *The Linked Data Service of the German National Library: Modeling of Bibliographic Data*. Leipzig, Frankfurt am Main: Deutsche Nationalbibliothek.
- Gerolimos, M. and Konsta, R. (2011). Services for Academic Libraries in New Era. *D-Lib Magazine*, 17 (7/8). <DOI: 10.1045/july2011-gerolimos>
- Gessesse, K. (2000). Collection Development and Management in the Twenty-first Century with Special Reference to Academic Libraries: an Overview. *Library Management*, 21 (7), pp. 365-372.
- Giunchiglia, F., Soergel, D., Maltese, V. and Bertacco, A. (2009). Mapping Large-scale Knowledge Organization Systems. In: *Proceedings of the 2nd International Conference on the Semantic Web and Digital Libraries (ICSD 2009)*. Germany: Springer.
- Goncalves, M.A., Moreira, B.L., Fox, E.A. and Watson, L.T. (2007). What is a Good Digital Library? - A Quality Model for Digital Libraries. *Information Processing & Management*, 43, pp. 1416-1437.
- Gonzales, B. (2014). Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record. *Information Technology and Libraries*, 33 (4), pp. 10-22.
- Gordon, A.S. and Domeshek, E.A. (1998). Déjà vu: a Knowledge-rich Interface for Retrieval in Digital Libraries. In: *Proceedings of the 1998 International Conference on Intelligent User Interfaces*, San Francisco, California: ACM Press, pp. 127-134.
- Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition*, 5 (2), pp. 199-220.
- Haffner, A. (2012). *GND ontology*. Leipzig, Frankfurt am Main: Deutsche Nationalbibliothek. [online] Διαθέσιμο στο: http://d-nb.info/standards/element_set/gnd2012-06-30 [Πρόσβαση: 15/07/2015].
- Haigh, S. (2005). Library Catalogue Users are Influenced by Trends in Web Searching Search Strategies. *Evidence Based Library and Information Practice*, 1 (3), pp. 77-79.

- Hannemann, J. and Kett, J. (2010). Linked Data for Libraries. In: *Proceedings of the World Library and Information Congress of the Int'l Federation of Library Associations and Institutions (IFLA)*.
- Halpin, H., Hayes, P., McCusker, J., McGuinness, D. and Thompson, H. (2010). When Owl:sameAs Isn't the Same: an Analysis of Identity in Linked Data. In: *Proceedings of the 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, LNCS 6496*, pp. 305-320.
- Harper, C.A. (2006). Authority Control for the Semantic Web. Encoding Library of Congress Subject Headings in SKOS. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications, Mexico, Manzanillo*.
- Harper, C.A. and Tillett, B.B. (2007). Library of Congress Controlled Vocabularies and their Application to the Semantic Web. *Cataloging & Classification Quarterly*, 43 (3-4), pp. 47-68.
- Harpring, P. (2010). *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*. Los Angeles, California: Getty Research Institute.
- Harris, S. and Seaborne, A. (2013). *SPARQL 1.1 Query Language: W3C Recommendation 21 March 2013*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/sparql11-query/> [Πρόσβαση: 09/07/2015].
- Haslhofer, B., Momeni, E., Gay, M. and Simon, R. (2010). Augmenting Europeana Content with Linked Data Resources. In: *Proceedings of the 6th International Conference on Semantic Systems, Graz, Austria, ACM*.
- Heath, T. and Bizer, C. (2011). *Linked data: Evolving the Web into a Global Data Space*. San Rafael, California: Morgan & Claypool.
- Hickson, I., Berjon, R., Faulkner, S., Leithead, T., Doyle Navara, E., O' Connor, E. and Pfeiffer, S. (2014). *HTML5: a Vocabulary and Associated APIs for HTML and XHTML: W3C Recommendation 28 October 2014*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/html/> [Πρόσβαση: 08/07/2015].

- Hill, R.W. (2002). *Changing the Record: a Concise Guide to the Differences between the UKMARC and MARC21 Bibliographic Formats*. West Yorkshire: The British Library. [online] Διαθέσιμο στο: http://www.bl.uk/bibliographic/pdfs/marc_change.pdf [Πρόσβαση: 15/07/2015].
- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. and Rudolph, S. (2012). *OWL 2 Web Ontology Language Primer (Second Edition): W3C Recommendation 11 December 2012*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/owl2-primer/> [Πρόσβαση: 17/07/2015].
- Hyödynmaa, M., Ahlholm-Kannisto, A. and Nurminen, H. (2010). How to Evaluate Library Collections: a Case Study of Collection Mapping. *Collection Building*, 29 (2), pp. 43–49.
- Hyvönen, E., Viljanen, K., Tuominen, J. and Seppälä, K. (2008). Building a National Semantic Web Ontology and Ontology Service Infrastructure--the FinnONTO Approach. In: *Proceedings of the ESWC 2008, Springer, Tenerife, Spain, June 1-5*. [online] Διαθέσιμο στο: <http://www.seco.tkk.fi/publications/2008/hyvonen-et-al-building-2008.pdf> [Πρόσβαση: 09/07/2015].
- Hyvönen, E. (2012). *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. San Rafael, Calif.: Morgan & Claypool Publishers.
- Isaac, A., Schlobach, S., Mattheizing, H. and Zinn, C. (2008). Integrated Access to Cultural Heritage Resources through Representation and Alignment of Controlled Vocabularies. *Library Review*, 57 (3), pp. 187-199.
- Isaac, A. and Summers, E. (2009). *SKOS Simple Knowledge Organization System Primer: W3C Working Group Note 18 August 2009*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/skos-primer/> [Πρόσβαση: 08/07/2015].
- Isaac, A., Waites, W., Young, J. and Zeng, M. (2011). *Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets - W3C Incubator Group Report*. [online] Διαθέσιμο στο: <http://www.w3.org/2005/Incubator/llid/XGR-llid-vocabdataset-20111025/> [Πρόσβαση: 08/07/2015].

- Ivanova, T. and Terzieva, V. (2010). *Ontology Mapping Tools, Methods and Approaches - Analytical Survey*. [online] Διαθέσιμο στο: https://research.uni-sofia.bg/bitstream/10506/633/1/S3T2010_23-T.Ivanova%2BV.Terzieva_Ontology_Mapping_Tools.pdf [Πρόσβαση: 10/07/2015].
- Jacobs, I. and Walsh, N. (2004). *Architecture of the World Wide Web, Volume One: W3C Recommendation 15 December 2004*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/webarch/> [Πρόσβαση: 08/07/2015].
- Jahns, Y. (2012). *Guidelines for Subject Access in National Bibliographies*. Berlin: De Gruyter Saur.
- Jain, P., Yeh, P., Verma, K., Vasquez, R., Damova, M., Hitzler, P. and Sheth, A. (2011). Contextual Ontology Alignment of LOD with an Upper Ontology: a Case Study with Proton. In: *Proceedings of the 8th Extended Semantic Web Conference, ESWC 2011*, Heraklion, Crete, Greece, May 29-June 2, LNCS 6643, pp. 80-92.
- Kifer, M. and Boley, H. (2013). *RIF Overview (Second Edition): W3C Working Group Note 5 February 2013*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/rif-overview/> [Πρόσβαση: 30/07/2015].
- Klyne, G. and Carroll, J.J. (2004). *Resource Description Framework (RDF): Concepts and Abstract Syntax - W3C Recommendation*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/rdf-concepts/> [Πρόσβαση: 08/07/2015].
- Kniewel, J.E., Wicht, H. and Silipigni Connaway, L. (2006). Use of Circulation Statistics and Interlibrary Loan Data in Collection Management. *College & Research Libraries*, 67 (1), pp. 35-49.
- Krug, S. (2000). *Don't Make Me Think!: a Common Sense Approach to Web Usability*. Indianapolis, USA: New Riders.
- Krzysztof, J., Hitzler, P., Adams, B., Kolas, D. and Vardeman, C.II. (2014). Five Stars of Linked Data Vocabulary Use. Editorial. *Semantic Web Journal*, pp. 1-4. *Semantic Web - Interoperability, Usability, Applicability*. Amsterdam: IOS Press Journal.
- Kumar, S. (2011). Effect of Web Searching on the OPAC: a Comparison of Selected University Libraries. *Library Hi Tech News*, 28 (6), pp. 14-21.

- Kwak, J. and Yong, H. (2010). Ontology Matching Based on Hypernym, Hyponym, Holonym and Meronym Sets in WordNet. *International Journal of Web & Semantic Technology*, 1 (2), pp. 1-14.
- Κυπριανός, Κ. και Παπαδάκης, Ι. (2012). Καθιστώντας μια Υπηρεσία Θεματικής Πλοήγησης στο Διαδίκτυο Συμβατή με τις Τεχνολογίες των Συνδεδεμένων Δεδομένων. Στο: *21^ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών*, Αθήνα, 18-19 Οκτωβρίου 2012. [online] Διαθέσιμο στο: http://leo.hua.gr/palc2012/sites/default/files/pictures/Kyprianos_papadakis_palc2012.pdf [Πρόσβαση: 07/07/2015].
- Kyprianos, K. and Papadakis, I. (2012). Providing LOD-based Functionality in Digital Libraries. In: *Proceedings of the 6th Metadata and Semantics Research*, MTSR2012, Cádiz, Spain, November 28-30, 2012, pp.13-24.
- Lagergren, E. and Over, P. (1998). Comparing Interactive Information Retrieval Systems across Sites: the TREC-6 Interactive Track Matrix Experiment. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 164-172). New York, NY: ACM.
- Lancaster, F.W. and Smith, L.C. (1983). *Compatibility Issues Affecting Information Systems and Services*, (PGI-83/WS/23). Paris: General Information Programme and UNISIST, UNESCO.
- Lau, C.H., Li, Y. and Tjondronegoro, D. (2011). Microblog Retrieval Using Topical Features and Query Expansion. In: *Text REtrieval Conference*.
- Lawson, K.G. (2009). Mining Social Tagging Data for Enhanced Subject Access for Readers and Researchers. *The Journal of Academic Librarianship*, 35 (6), pp. 574-582.
- Lee, D.H. and Schleyer, T. (2010). A Comparison of MeSH Terms and CiteULike Social Tags as Metadata for the Same Items. In: *IHI '10: Proceedings of the 1st ACM International Health Informatics Symposium* (pp. 445-448). New York, NY: ACM.
- Liang, A., Sini, M., Chang C., Li, S., Lu, W.N., He, C. and Keizer, J. (2006). The Mapping Schema from Chinese Agricultural Thesaurus to AGROVOC. *New Review of Hypermedia and Multimedia – NRHM*, Taylor & Francis, 12 (1), pp. 51-62.

- Library of Congress (2007). 30th ed. *Library of Congress Subject Headings*. Washington, DC: Library of Congress.
- Library of Congress (2013). *MARC21 Format for Authority Data*. (update no. 17). [online] Διαθέσιμο στο: <http://www.loc.gov/marc/authority/ecadhome.html> [Πρόσβαση: 15/07/2015].
- Lin, F. and Sandkuhl, K. (2008). A Survey of Exploiting WordNet in Ontology Matching. In: *IFIP International Federation for Information Processing, 276; Artificial Intelligence and Practice II*; Max Bramer. Boston: Springer, pp. 341–350.
- Linh, N.C. (2008). A Survey of the Application of Web 2.0 in Australian University Libraries. *Library Hi Tech*, 26 (4), pp. 630-653.
- Lwoga, E.T. (2013). Measuring the Success of Library 2.0 Technologies in the African Context: the Suitability of the DeLone and McLean's Model. *Campus-Wide Information Systems*, 30 (4), pp. 288-307.
- Maali, F., Cyganiak, R. and Peristeras, V. (2011). Re-using Cool URIs: Entity Reconciliation against LOD Hubs. In: *Proceedings of the Linked Data on the Web Workshop, LDOW2011*, March 29, 3.
- Macgregor, G. and McCulloch, E. (2006). Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review*, 55 (5), pp. 291-300.
- Mahmood, K. and Richardson, J.V. Jr. (2013). Impact of Web 2.0 Technologies on Academic Libraries: a Survey of ARL Libraries. *The Electronic Library*, 31 (4), pp. 508-520.
- Malmsten, M. (2009). Exposing Library Data as Linked Data. In: *IFLA Satellite Pre-conference Sponsored by the Information Technology Section Emerging Trends*.
- Markham, J.W. and Avery, B. (1998). LCSH and the ASFIS Thesaurus: an Update. In: *IAMSLIC Conference Proceedings*. [online] Διαθέσιμο στο: <https://darchive.mblwhoilibrary.org/handle/1912/1903> [Πρόσβαση: 07/07/2015].
- Marshall, L. (2003). Specific and Generic Subject Headings: Increasing Subject Success to Library Materials. *Cataloging & Classification Quarterly*, 36 (2), pp. 59–87.

- McIlwaine, I.C. (2007). *The Universal Decimal Classification: a Guide to its Use*. The Hague: UDC Consortium.
- Meij, L., Isaac, A. and Zinn, C. (2010). A Web-based Repository Service for Vocabularies and Alignments in the Cultural Heritage Domain. In: *Proceedings of the 7th ESWC 2010*. Heraklion, Greece.
- Miles, A., Mathews, B., Wilson, M. and Brickley, D. (2005). SKOS Core: Simple Knowledge Organisation for the Web. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications*, September 12-15, Madrid, Spain, pp. 5-13.
- Miles, A. and Pérez-Agüera, J. (2007). SKOS: Simple Knowledge Organisation for the Web. *Cataloging & Classification Quarterly*, 43 (3-4), pp. 69-83.
- Miles, A. and Bechhofer, S. (2009). *SKOS Simple Knowledge Organization System Reference: W3C Recommendation 18 August 2009*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/skos-reference/> [Πρόσβαση: 08/07/2015].
- Miller, G.A. (1995). WordNet: a Lexical Database for English. *Communications of the ACM*, 38 (11), 39-41.
- Miller, H. (1996). The Multiple Dimensions of Information Quality. *Information Systems Management*, 13 (2), pp. 79-82.
- Miller, E., Ogbuji, U., Mueller, V. and MacDougall, K. (2012). *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services*. Washington, DC: Library of Congress.
- Mitchell, J.S. and Panzer, M. (2013). Dewey Linked Data: Making Connections with Old Friends and New Acquaintances. *Italian Journal of Library, Archives and Information Science*, 4 (1), pp. 177-199.
- Moats, R. (1997). *URN Syntax, RFC 2141* [online] Διαθέσιμο στο: <https://www.ietf.org/rfc/rfc2141.txt> [Πρόσβαση: 06/10/2015].
- Movahedian, H. and Khayyambashi, M.R. (2014). Folksonomy-based User Interest and Disinterest Profiling for Improved Recommendations: an Ontological Approach. *Journal of Information Science*, 40 (5), pp. 594-610.

- Mynarz, J., Kozuchova, K. and Kamradkova, K. (2009). *Polythematic Structured Subject Heading System & Creative Commons*. Praha: Státní Technická Knihovna.
- Neubert, J. (2009). Bringing the "Thesaurus for Economics" on to the Web of Linked Data. *Linked Data on the Web (LDOW) 2009. Workshop at the WWW 2009 Conference*, 20 April 2009, Madrid, Spain.
- Noruzi, A. (2006). Folksonomies: (Un)Controlled Vocabulary? *Knowledge Organization*, 33(4), pp. 199-203.
- Olson, T. and Strawn, G. (1997). Mapping the LCSH and MeSH Systems. *Information Technology and Libraries*, 16 (1), pp. 5-19.
- Paepcke, A., Chang, C., Winograd, T. and García-Molina, H. (1998). Interoperability for Digital Libraries Worldwide. *Communications of the ACM*, 41 (4), pp. 33-42.
- Papadakis, I., Stefanidakis, M. and Tzali, A. (2007). Semantic Navigation in a Library Catalogue. In: *Proceedings of Metadata and Semantics Research Conference – MTSR 2007*, Corfu, Greece, pp. 130-134.
- Papadakis, I. and Stefanidakis, M. (2008). Visualizing Ontologies on the Web. In: G. Tsihrintzis, M. Virvou, R. Howlett and L. Jain, ed., *New Directions in Intelligent Interactive Multimedia*, 1st ed. Studies in Computational Intelligence, 142. Berlin: Springer Berlin Heidelberg, pp. 303-311. [online] Διαθέσιμο στο: <http://www.springer.com/us/book/9783540681267> [Πρόσβαση: 10/07/2015].
- Papadakis, I., Kyprianos, K., Mavropodi, R. and Stefanidakis, M. (2009). Subject-based Information Retrieval within Digital Libraries Employing LCSHs. *D-Lib Magazine*, 15 (9/10). <DOI: 10.1045/september2009-papadakis>
- Papadakis, I. and Kyprianos, K. (2011). Merging Controlled Vocabularies for More Efficient Subject-Based IR Systems. *International Journal of Knowledge Management*, 7 (3), pp. 74-90.
- Papadakis, I., Kyprianos, K. and Stefanidakis, M. (2015). Linked Data URIs and Libraries: the Story so Far. *D-Lib Magazine*, 21 (5/6). <DOI: 10.1045/may2015-papadakis>

- Παρασκευόπουλος, Ι.Ν. (1993). *Μεθοδολογία Επιστημονικής Έρευνας*. Τόμος Α. Αθήνα: [χ.ο.].
- Pastor-Sanchez, J.A., Mendez, F.J.M. and Rodríguez-Muñoz, J.V. (2009). Advantages of Thesaurus Representation Using the Simple Knowledge Organization System (SKOS) Compared with Proposed Alternatives. *Information Research*, 14 (4). [online] Διαθέσιμο στο: <http://www.informationr.net/ir/14-4/paper422.html> [Πρόσβαση: 09/07/2015].
- Pesquita, C., Stroe, C., Cruz, I.F. and Couto, F.M. (2010). BLOOMS on Agreement-maker: Results for OAEI 2010. In: *Proceedings of the ISWC Workshop Ontology Matching - OM-2010*, pp. 137-145.
- Pipino, L.L., Lee, Y.W. and Wang, R.Y. (2002). Data Quality Assessment. *Communications of the ACM*, 45 (4), pp. 211-218.
- Plassard, M.F. (2001). Authority Control in an International Environment: the UNIMARC Format for Authorities. In: the *2nd Workshop on Authority Control among Chinese, Korean and Japanese Languages Held at National Institute of Informatics (NII) in Cooperation with National Diet Library*, 28-29 March 2001. [online] Διαθέσιμο στο: <http://www.nii.ac.jp/publications/CJK-WS/2-1Plassard.pdf> [Πρόσβαση: 15/07/2015].
- Prud'hommeaux, E. and Seaborne, A. (2008). *SPARQL Query Language for RDF - W3C Recommendation*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/rdf-sparql-query/> [Πρόσβαση: 08/07/2015].
- Purday, J. (2009). Think Culture: Europeana.eu from Concept to Construction. *The Electronic Library*, 27 (6), pp. 919-937.
- San Segundo, R. (2009). Using MARC Classification Format for UDC and Mappings to other KO Systems for an Enriched Authority File. In: *Proceedings UDC Seminar 2009 Classification at a Crossroads: Multiple Directions to Usability*, pp. 285-291.
- Sauermann, L. and Cyganiak, R. (2008). *Cool URIs for the Semantic Web: W3C Interest Group Note*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/cooluris/> [Πρόσβαση: 08/07/2015].

- Schatz, B.R., Johnson, E.H., Cochrane, P.A. and Chen, H. (1996). Interactive Term Suggestion for Users of Digital Libraries: Using Subject Thesauri and Co-occurrence Lists for Information Retrieval. In: *Proceedings of the First ACM International Conference on Digital Libraries*, pp. 126-133.
- Schreiber, G. and Raimond, Y. (2014). *RDF 1.1 Primer: W3C Working Group Note 24 June 2014*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/> [Πρόσβαση: 09/07/2015].
- Schwartz, C. (2008). Thesauri and Facets and Tags, Oh My! A Look at Three Decades in Subject Analysis. *Library Trends*, 56 (4), pp. 830-842.
- Sinclair, J. and Cardew-Hall, M. (2008). The Folksonomy Tag Cloud: When is it Useful? *Journal of Information Science*, 34 (15), pp. 15-29.
- Singh, S. (2011). *The Theory and Practice of the Dewey Decimal Classification System*. New Delhi: Isha Books.
- Smith, I.M. (1999). What do we Know about Public Library Use? *Aslib Proceedings*, 51 (9), pp. 302-314.
- Söderbäck, A. and Malmsten, M. (2008). LIBRIS - Linked Library Data. *Nodalities: the Magazine of the Semantic Web*, 5 (11/12).
- Spiteri, L.E. (2007). Structure and Form of Folksonomy Tags: the Road to the Public Library Catalogue. *Webology*, 4 (2), article 41.
- Spitieri, L. (2014). User-generated Metadata: Boon or Bust for Indexing and Controlled Vocabularies? [online] Διαθέσιμο στο: <http://www.slideshare.net/cleese6/usergenerated-metadata-boon-or-bust-for-indexing-and-controlled-vocabularies> [Πρόσβαση: 15/11/2014].
- Stone, A.T. (2000). The LCSH Century: a Brief History of the Library of Congress Subject Headings, and Introduction to the Centennial Essays. *Cataloging & Classification Quarterly*, 29 (1/2), pp. 1-16.
- Summers, E., Isaac, A., Redding, C. and Krech, D. (2008). LCSH, SKOS and Linked Data. In: *Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications (DCMI '08)*. Dublin Core Metadata Initiative, pp. 25-33.

- Svenonius, E. (1986). Unanswered Questions in the Design of Controlled Vocabularies. *Journal of the American Society for Information Science*, 37 (5), pp. 331-340.
- Svenonius, E. (2003). 2nd ed. Design of Controlled Vocabularies. *Encyclopedia of Library and Information Science*, vol. 2. New York, NY: Marcel Dekker, Inc.
- Tan, P.J. and Clarke, D. (2007). Audience-centric Taxonomy: Using Taxonomies to Support Heterogeneous User Communities. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications 2007*, pp. 123-127.
- Teevan, J., Ramage, D. and Morris, M.R. (2011). #TwitterSearch: a Comparison of Microblog Search and Web Search. In: *Proceedings of the fourth ACM International Conference on Web Search and Data Mining (WSDM '11)*, pp. 35-44.
- Thomas, M., Caudle, D. and Schmitz, C. (2009). To Tag or not to Tag? *Library Hi Tech*, 27 (3), pp. 411-434.
- Tordai, A., Omelayenko, B. and Schreiber, G. (2007). Thesaurus and Metadata Alignment for a Semantic E-culture Application. In: *Proceedings of the 4th International Conference on Knowledge Capture – K-CAP '07* (pp. 199-200). New York: ACM.
- Trant, J. (2008). Studying Social Tagging and Folksonomy: a Review and Framework. *Journal of Digital Information*, Special Issue on Digital Libraries and User-Generated Content.
- Tuominen, J., Frosterus, M., Viljanen, K. and Hyvonen, E. (2008). ONKI-SKOS - Publishing and Utilizing Thesauri in the Semantic Web. In: *Proceedings of the 13th Finnish Artificial Intelligence Conference STeP 2008: AI and Machine Consciousness*. Finland: Semantic Computing Research Group (SeCo).
- van Kesteren, A. and Sam, R. (2014). *URL: W3C Working Draft, 9 December 2014*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/url/> [Πρόσβαση: 08/07/2015].
- Vila-Suero, D., Villazón-Terrazas, B. and Gómez-Pérez, A. (2013). datos.bne.es: a Library Linked Data Dataset. *Semantic Web. IOS Press*, 4 (3), pp. 307-313.
- Vizine-Goetz, D. (2002). Classification Schemes for Internet Resources Revisited. *Journal of Internet Cataloguing*, 5 (4), pp. 5-18.

- Vizine-Goetz, D., Hickey, C., Houghton, A. and Thompson, R. (2004). Vocabulary Mapping for Terminology Services. *Journal of Digital Information*, 4 (4). [online] Διαθέσιμο στο: <https://journals.tdl.org/jodi/index.php/jodi/article/view/114/113> [Πρόσβαση: 08/07/2015].
- W3C OWL Working Group, (2012). *OWL 2 Web Ontology Language Document Overview (Second Edition): W3C Recommendation 11 December 2012*. [online] Διαθέσιμο στο: <http://www.w3.org/TR/owl2-overview/> [Πρόσβαση: 08/07/2015].
- Walia, P.K. and Gupta, M. (2012). Application of Web 2.0 Tools by National Libraries. *Webology*, 9 (2).
- Weaver-Meyers, P. and Stolt, W. (1997). Delivery Speed, Timeliness and Satisfaction: Patrons' Perceptions about ILL Service. *Journal of Library Administration*, 23 (1-2), pp. 23-42.
- Wenz, R., Di Mascio, A., Michel, V. and Simon, A. (2013). Publishing Bibliographic Records on the Web of Data: Opportunities for the BnF (French National Library). In: *Extended Semantic Web Conference – ESWC 2013*, Montpellier, France. [online] Διαθέσιμο στο: <http://eswc-conferences.org/sites/default/files/papers2013/simon.pdf> [Πρόσβαση: 10/07/2015].
- White, H. (2013). Examining Scientific Vocabulary: Mapping Controlled Vocabularies with Free Text Keywords. *Cataloging & Classification Quarterly*, 51 (6), pp. 655-674.
- Whitehead, C. (1990). Mapping LCSH into Thesauri: the AAT Model. In: *Beyond the Book, Extending MARC for Subject Access* (pp. 81-96). Boston, Massachusetts: G.K. Hall.
- Will, L. (2012). The ISO 25964 Data Model for the Structure of an Information Retrieval Thesaurus. *Bulletin of the American Society for Information Science and Technology*, 38 (4), pp. 48-51.
- Willer, M. (2009a). Third Edition of UNIMARC Manual: Authorities Format: how does it Implement Concepts from the FRAD Model and IME ICC Statement of International Cataloguing Principles? In: *World Library and Information Congress: 75th IFLA General Conference and Council, 23-27 August 2009, Milan, Italy*. [online]

Διαθέσιμο στο: <http://www.nlc.gov.cn/newen/fl/iflanlc/iclc/IFLAds/201012/P020101210594734166966.pdf> [Πρόσβαση: 15/07/2015].

Willer, M. (ed.) (2009b). 3rd ed. *UNIMARC Manual: Authorities Format*. München: K. G. Saur.

Williamson, N.J. (1996). Standards and Rules for Subject Access. *Cataloging & Classification Quarterly*, 21 (3/4), pp. 155-176.

Yi, K. and Chan, L.M. (2009). Linking folksonomy to Library of Congress Subject Headings: an Exploratory Study. *Journal of Documentation*, 65 (6), pp. 872-900.

Zhao, D. and Rosson, M.B. (2009). How and Why People Twitter: the Role that Microblogging Plays in Informal Communication at Work. In: *Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP '09)*, pp. 243-252.

Αρκτικόλεξα – Ακρωνύμια

ATG	Agricultural Thesaurus and Glossary
ATT	Art & Architecture Thesaurus
BIBFRAME	Bibliographic Framework
BL	British Library
BNB	British National Bibliography
BNE	Biblioteca Nacional de España
BNF	Bibliothèque nationale de France
BT	Broader Term
CSS	Cascading Style Sheets
DDC	Dewey Decimal Classification
DL	Description Logic
DMA-EST	Einheitssachtitel-Datei des Deutschen Musikarchivs
DNB	Deutsche National Bibliothek
DOI	Digital Object Identifier System
DTD	Document Type Definition
GKD	Gemeinsame Körperschaftsdatei
GND	Gemeinsame Normdatei
GUI	Graphical User Interface
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
IIR	Interactive Information Retrieval
ILO	International Labor Organization
ISBN	International Standard Book Number
ISO	International Standards Organization
ISSN	International Standard Serial Number
KB	Kungliga Biblioteket
KOS	Knowledge Organization System
LCC	Library of Congress Classification
LCCN	Library of Congress Control Number
LCSH	Library of Congress Subject Headings
LoC	Library of Congress
LOD	Linked Open Data
MARC	MAchine-Readable Cataloguing

MeSH	Medical Subject Headings
NAF	Name Authority File
NAL	National Agricultural Library
NALT	National Agricultural Library Thesaurus
NSL	National Széchényi Library
NT	Narrower Term
NYT	New York Times
OEG	Ontology Engineering Group
OWL	Web Ontology Language
PND	Personennamendatei
PSH	Polythematic Structured Subject Heading
	Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié
RAMEAU	
RDF	Resource Description Framework
RDFS	RDF Vocabulary Definition Language
RIF	Rule Interchange Format
RT	Related Term
SDW	Schlagwortnormdatei
SKOS	Simple Knowledge Organization System
SMDB	Swedish Media Database
SNS	Social Networking Services
SPARQL	SPARQL Protocol and RDF Query Language
UDC	Universal Decimal Classification
UF	Used For
	United Nations Educational, Scientific and Cultural Organization
UNESCO	
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
VIAF	Virtual Authority File
W3C	World Wide Web Consortium
WoD	Web of Data
XML	Extensible Markup Language
ΕΛΟΤ	Ελληνικός Οργανισμός Τυποποίησης
ΣΚΕΑΒ	Συλλογικός Κατάλογος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών

Απόδοση ξένων όρων

Ελληνική απόδοση

Ακρίβεια

Ακριβές ταίριασμα

Αλληγορίες

Ανακατεύθυνση

Ανάκτηση

Ανάλυση "κόστους-οφέλους"

Ανάλυση αρχείων καταγραφής

Ανάλυση της αναφοράς

Ανιχνευσιμότητα

Ανταλλαγή άμεσων μηνυμάτων

Αντικείμενο

Αξιολόγηση με επίκεντρο το χρήστη

Αξιολόγηση συγκεκριμένης θεματικής υποστήριξης

Αξιολόγηση συλλογής

Αποθετήρια τριπλετών

Απόσταση των χαρακτήρων

Αριθμός ελέγχου

Αριθμός ελέγχου συστήματος

Αρχεία καταγραφής συναλλαγών

Βιβλιοθήκη του Κογκρέσου

Βρετανική Βιβλιοθήκη

Βρετανική Εθνική Βιβλιογραφία

Γενική επιλογή αναζήτησης

Γενική χωρητικότητα

Γεωγραφικές υποδιαιρέσεις

Γλώσσα Σήμανσης Υπερκειμένου

Γλωσσάρια

Γλωσσικό σύνολο δεδομένων

Γνωρίσματα

Γραφικό περιβάλλον διεπαφής χρήστη

Δείκτης

Αγγλικός όρος

Precision

Exact match

Metaphors

Redirection

Recall

Cost-benefit analysis

Log files analysis

Dereference

Findability

Instant messaging

Object

User-centered evaluation

Assessment of specific subject support

Collection evaluation

Triplestores

Edit distance

Control number

System control number

Transaction log files

Library of Congress

British Library

British National Bibliography

General search option

General capacity

Geographic subdivisions

HyperText Markup Language

Glossaries

Linguistic dataset

Attributes

Graphical user interface

Indicator

Δείκτης περιεχομένου	Content indicator
Δεκαδική Ταξινόμηση Dewey	Dewey Decimal Classification
Διαδανεισμός	Interlibrary loan
Διαδραστική Ανάκτηση Πληροφοριών	Interactive Information Retrieval
Διακομιστής	Server
Διαπραγμάτευση περιεχομένου	Content negotiation
Δίκτυο των δεδομένων	Web of Data
Εθνική Βιβλιοθήκη της Γαλλίας	Bibliothèque nationale de France
Εθνική Βιβλιοθήκη της Γερμανίας	Deutsche National Bibliothek
Εθνική Βιβλιοθήκη της Ισπανίας	Biblioteca Nacional de España
Εθνική Βιβλιοθήκη της Ουγγαρίας	National Széchényi Library
Εθνική Βιβλιοθήκη της Σουηδίας	Kungliga Biblioteket
Ειδικά θεματικά πρότυπα	Subject-specific standards
Εκτεταμένη συνδετική δομή	Extended syndetic structure
Ελεγχόμενο λεξιλόγιο	Controlled vocabulary
Ενδιάμεσο λογισμικό	Middleware
Ενιαία Αναγνωριστικά Πόρων	Uniform Resource Identifiers
Ενιαίοι Εντοπιστές Πόρων	Uniform Resource Locators
Έννοιες	Concepts
Εννοιολογική κορυφή	Top concept
Επέκταση	Extension
Επίκαιροι όροι	Trending terms
Επικαιρότητα	Timeliness
Επίπεδο στοιχείων	Element-level
Επιστημονικές εκδόσεις	Scholarly publishing
Έρευνα	Survey
Εσωτερική χρήση	In-house usage
Ετικέτα	Label
Ευρύτερος όρος	Broader term
Ηλικία	Age
Θεματικές Επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου	Library of Congress Subject Headings
Θεματικές υποδιαίρεσεις	Topical subdivisions
Θησαυροί	Thesauri
Ιδιότητες	Properties

Ιεραρχική σχέση	Hierarchical relation
Ισοδυναμία	Equivalence
Κατανοητά από τις μηχανές	Machine-understandable
Κατηγορήμα	Predicate
Κίνηση των βιβλίων	Circulation
Κλάσεις	Classes
Κοινοπραξιακά δεδομένα δανεισμού	Consortial borrowing data
Κοινωνικές δομές	Social constructs
Κοινωνική δικτύωση	Social web
Κύρια επικεφαλίδα	Main heading
Κώδικας υποπεδίου	Subfield code
Λαοταξινομίες	Folksonomies
Λεξιλογικές ετικέτες	Lexical labels
Λεξιλογική ομοιότητα	Lexical similarity
Λίστες θεματικών επικεφαλίδων	Subject headings lists
Μενού περιεχομένων	Context menu
Μερικό ταίριασμα	Partial match
Μεταβλητότητα	Volatility
Μη διακριτά	Non-distinct
Μη ελεγχόμενο λεξιλόγιο	Uncontrolled vocabulary
Μηχαναγνώσιμη	Machine-readable
Ομάδες συζητήσεων	Discussion group
Ομόγραφα	Homographs
Οντότητα ανάκτησης	Retrieval module
Οντότητα ερωταποκρίσεων	Query/Response module
Ορισμοί	Definitions
Ορολογίες	Terminologies
Παγκόσμια Δεκαδική Ταξινόμηση	Universal Decimal Classification
Παγκόσμιος ιστός	World wide web
Παραλλαγές προθεμάτων και επιθεμάτων	Prefix and suffix variations
Πεδίο	Field
Πειράματα	Experimental settings
Πελάτης	Client
Περιβαλλοντικοί παράγοντες	Environmental factors
Περιγραφική λογική	Description logic

Πηγή	Source
Ποιότητα της πληροφορίας	Information quality
Ποσοστό επιστροφής	Turn-over rate
Προηγμένη επιλογή αναζήτησης	Advanced search option
Που έχουν οριστεί σε απομακρυσμένα σημεία	Remotely-defined
Που έχουν οριστεί τοπικά	Locally-defined
Προσομοίωση σεναρίου εργασιών κατάστασης	Simulated work task situation
Προ-συνδυσασμένα συστήματα	Pre-coordinated systems
Προτεινόμενα τεκμήρια	Featured items
Προτιμώμενοι όροι	Preferred terms
Πρωτόκολλο Μεταφοράς Υπερκειμένου	HyperText Transfer Protocol
Σάρωση ραφιών	Shelf scanning
Σημασιολογικές σχέσεις	Semantic relations
Σημασιολογική εκπομπή	Semantic broadcast
Σημασιολογική ευθυγράμμιση	Semantic alignment
Σημασιολογικός ιστός	Semantic web
Σημειώσεις σκοπού	Scope notes
Σταθερός στατικός υπερσύνδεσμος	Permalink
Στατιστικά χρήσης	Usage statistics
Στενότερος όρος	Narrower term
Στιγμιότυπα	Instances
Στοιχεία	Elements
Στόχος	Target
Στρώμα	Layer
Συμβολισμοί	Notations
Συμφιλίωση	Reconciliation
Συνδεδεμένα Δεδομένα	Linked Open Data
Συνδεδετική δομή	Syndetic structure
Συνειρμική σχέση	Associative relation
Συνεντεύξεις	Interviews
Σύστημα οργάνωσης γνώσης	Knowledge Organization System
Σύστημα υποστήριξης	Backend
Σχετικός όρος	Related term

Σχήμα εννοιών	Concept scheme
Ταξινόμηση της Βιβλιοθήκης του Κογκρέ- σου	Library of Congress Classification
Ταξινομίες	Taxonomies
Ταξινομικά σχήματα	Classification schemes
Τελικά σημεία SPARQL	SPARQL endpoints
Τεχνικές βασισμένες στη συλλογή	Collection-based techniques
Τεχνικές βασισμένες στους χρήστες	User-based techniques
Τεχνικές ταιριάσματος σε επίπεδο στοι- χείων	Element-level matching techniques
Τριπλέτες	Triples
Υπηρεσίες κοινωνικής δικτύωσης	Social Networking Services
Υποδιαιρέσεις	Subdivisions
Υποδιαιρέσεις μορφής	Form subdivisions
Υποκείμενο	Subject
Υποπεδίο	Subfield
Υπηρεσίες κοινωνικής δικτύωσης	Social Networking Services
Υστερο-συνδυσασμένα συστήματα	Post-coordinated systems
Φυσική αξιολόγηση	Physical assessment
Χαρτογράφηση συλλογής	Collection mapping
Χρήση	Usage
Χρησιμοποίησε για	Used for
Χρήστες	Users
Χρονολογικές υποδιαιρέσεις	Chronological subdivisions
Χρονοσφραγίδα	Timestamp

Αγγλικός όρος

Advanced search option
 Age

 Assessment of specific subject support
 Associative relation
 Attributes
 Backend
 Biblioteca Nacional de España
 Bibliothèque nationale de France
 British Library
 British National Bibliography
 Broader term
 Chronological subdivisions
 Circulation
 Classes
 Classification schemes
 Client
 Collection evaluation
 Collection mapping
 Collection-based techniques
 Concept scheme
 Concepts
 Consortial borrowing data
 Content indicator
 Content negotiation
 Context menu
 Control number
 Controlled vocabulary
 Cost-benefit analysis
 Definitions
 Dereference
 Description logic
 Deutsche National Bibliothek
 Dewey Decimal Classification

Ελληνική απόδοση

Προηγμένη επιλογή αναζήτησης
 Ηλικία
 Αξιολόγηση συγκεκριμένης θεματικής υποστήριξης
 Συνειρμική σχέση
 Γνωρίσματα
 Σύστημα υποστήριξης
 Εθνική Βιβλιοθήκη της Ισπανίας
 Εθνική Βιβλιοθήκη της Γαλλίας
 Βρετανική Βιβλιοθήκη
 Βρετανική Εθνική Βιβλιογραφία
 Ευρύτερος όρος
 Χρονολογικές υποδιαιρέσεις
 Κίνηση των βιβλίων
 Κλάσεις
 Ταξινομικά σχήματα
 Πελάτης
 Αξιολόγηση συλλογής
 Χαρτογράφηση συλλογής
 Τεχνικές βασισμένες στη συλλογή
 Σχήμα εννοιών
 Έννοιες
 Κοινοπραξιακά δεδομένα δανεισμού
 Δείκτης περιεχομένου
 Διαπραγμάτευση περιεχομένου
 Μενού περιεχομένων
 Αριθμός ελέγχου
 Ελεγχόμενο λεξιλόγιο
 Ανάλυση "κόστους-οφέλους"
 Ορισμοί
 Ανάλυση της αναφοράς
 Περιγραφική λογική
 Εθνική Βιβλιοθήκη της Γερμανίας
 Δεκαδική Ταξινόμηση Dewey

Discussion group	Ομάδες συζητήσεων
Edit distance	Απόσταση των χαρακτήρων
Element-level	Επίπεδο στοιχείων
Element-level matching techniques	Τεχνικές ταιριάσματος σε επίπεδο στοιχείων
Elements	Στοιχεία
Environmental factors	Περιβαλλοντικοί παράγοντες
Equivalence	Ισοδυναμία
Exact match	Ακριβές ταίριασμα
Experimental settings	Πειράματα
Extended syndetic structure	Εκτεταμένη συνδετική δομή
Extension	Επέκταση
Featured items	Προτεινόμενα τεκμήρια
Field	Πεδίο
Findability	Ανιχνευσιμότητα
Folksonomies	Λαοταξινομίες
Form subdivisions	Υποδιαιρέσεις μορφής
General capacity	Γενική χωρητικότητα
General search option	Γενική επιλογή αναζήτησης
Geographic subdivisions	Γεωγραφικές υποδιαιρέσεις
Glossaries	Γλωσσάρια
Graphical user interface	Γραφικό περιβάλλον διεπαφής χρήστη
Hierarchical relation	Ιεραρχική σχέση
Homographs	Ομόγραφα
HyperText Markup Language	Γλώσσα Σήμανσης Υπερκειμένου
HyperText Transfer Protocol	Πρωτόκολλο Μεταφοράς Υπερκειμένου
In-house usage	Εσωτερική χρήση
Indicator	Δείκτης
Information quality	Ποιότητα της πληροφορίας
Instances	Στιγμιότυπα
Instant messaging	Ανταλλαγή άμεσων μηνυμάτων
Interactive Information Retrieval	Διαδραστική Ανάκτηση Πληροφοριών
Interlibrary loan	Διαδανεισμός
Interviews	Συνεντεύξεις
Knowledge Organization System	Σύστημα οργάνωσης γνώσης

Kungliga Biblioteket	Εθνική Βιβλιοθήκη της Σουηδίας
Label	Ετικέτα
Layer	Στρώμα
Lexical labels	Λεξιλογικές ετικέτες
Lexical similarity	Λεξιλογική ομοιότητα
Library of Congress	Βιβλιοθήκη του Κογκρέσου
Library of Congress Classification	Ταξινόμηση της Βιβλιοθήκης του Κογκρέσου
Library of Congress Subject Headings	Θεματικές Επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου
Linguistic dataset	Γλωσσικό σύνολο δεδομένων
Linked Open Data	Συνδεδεμένα Δεδομένα
Locally-defined	Τοπικά ορισμένος
Log files analysis	Ανάλυση αρχείων καταγραφής
Machine-readable	Μηχαναγνώσιμη
Machine-understandable	Κατανοητά από τις μηχανές
Main heading	Κύρια επικεφαλίδα
Metaphors	Αλληγορίες
Middleware	Ενδιάμεσο λογισμικό
Narrower term	Στενότερος όρος
National Széchényi Library	Εθνική Βιβλιοθήκη της Ουγγαρίας
Non-distinct	Μη διακριτά
Notations	Συμβολισμοί
Object	Αντικείμενο
Partial match	Μερικό ταίριασμα
Permalink	Σταθερός στατικός υπερσύνδεσμος
Physical assessment	Φυσική αξιολόγηση
Post-coordinated systems	Υστερο-συνδυασμένα συστήματα
Pre-coordinated systems	Προ-συνδυασμένα συστήματα
Precision	Ακρίβεια
Predicate	Κατηγορημα
Preferred terms	Προτιμώμενοι όροι
Prefix and suffix variations	Παραλλαγές προθεμάτων και επιθεμάτων
Properties	Ιδιότητες

Query/Response module	Οντότητα ερωταποκρίσεων
Recall	Ανάκτηση
Reconciliation	Συμφιλίωση
Redirection	Ανακατεύθυνση
Related term	Σχετικός όρος
Remotely-defined	Εξ' αποστάσεως ορισμένος
Retrieval module	Οντότητα ανάκτησης
Scholarly publishing	Επιστημονικές εκδόσεις
Scope notes	Σημειώσεις σκοπού
Semantic alignment	Σημασιολογική ευθυγράμμιση
Semantic broadcast	Σημασιολογική εκπομπή
Semantic relations	Σημασιολογικές σχέσεις
Semantic web	Σημασιολογικός ιστός
Server	Διακομιστής
Shelf scanning	Σάρωση ραφιών
	Προσομοίωση σεναρίου εργασιών κα- τάστασης
Simulated work task situation	τάστασης
Social constructs	Κοινωνικές δομές
Social Networking Services	Υπηρεσίες κοινωνικής δικτύωσης
Social web	Κοινωνική δικτύωση
Source	Πηγή
SPARQL endpoints	Τελικά σημεία SPARQL
Subdivisions	Υποδιαιρέσεις
Subfield	Υποπεδίο
Subfield code	Κώδικας υποπεδίου
Subject	Υποκείμενο
Subject headings lists	Λίστες θεματικών επικεφαλίδων
Subject-specific standards	Ειδικά θεματικά πρότυπα
Survey	Έρευνα
Syndetic structure	Συνδετική δομή
System control number	Αριθμός ελέγχου συστήματος
Target	Στόχος
Taxonomies	Ταξινομίες
Terminologies	Ορολογίες
Thesauri	Θησαυροί

Timeliness	Επικαιρότητα
Timestamp	Χρονοσφραγίδα
Top concept	Εννοιολογική κορυφή
Topical subdivisions	Θεματικές υποδιαιρέσεις
Transaction log files	Αρχεία καταγραφής συναλλαγών
Trending terms	Επίκαιροι όροι
Triples	Τριπλέτες
Triplestores	Αποθετήρια τριπλετών
Turn-over rate	Ποσοστό επιστροφής
Uncontrolled vocabulary	Μη ελεγχόμενο λεξιλόγιο
Uniform Resource Identifiers	Ενιαία Αναγνωριστικά Πόρων
Uniform Resource Locators	Ενιαίοι Εντοπιστές Πόρων
Universal Decimal Classification	Παγκόσμια Δεκαδική Ταξινόμηση
Usage	Χρήση
Usage statistics	Στατιστικά χρήσης
Used for	Χρησιμοποίησε για
User-based techniques	Τεχνικές βασισμένες στους χρήστες
User-centered evaluation	Αξιολόγηση με επίκεντρο το χρήστη
Users	Χρήστες
Volatility	Μεταβλητότητα
Web of Data	Δίκτυο των δεδομένων
World wide web	Παγκόσμιος ιστός

Ευρετήριο

- 4store, 104
- DBpedia, 116
- Google Refine, 118
 - υπηρεσία συμφιλίωσης, 120
- LCSH, 18, 24
- MARC, 125
 - πεδία, 125
- MARC21, 127
- N-Triples, 83
- OWL, 7, 79
- RDF, 7, 77
 - γράφος, 77
 - σύνταξη, 80
- RDF/JSON, 84
- RDF/XML, 81
- RDFa, 81
- RDFS, 7, 79
- SKOS, 8, 57, 62, 84
- SPARQL, 7
- Turtle, 82
- Twitter, 158
- UNIMARC, 127
- URI, 75, 76
 - dereference, 76
- Web 2.0, 157
- WordNet, 116
- XML, 6
- Αξιολόγηση της συλλογής, 154
- Αποθετήρια τριπλετών, 88
- Διαλειτουργικότητα, 4, 51
- Εκτεταμένη συνδετική δομή, 31
- Ελεγχόμενα λεξιλόγια, 3, 16, 51
- Επικαιρότητα. *Βλέπε* Χρονική διάσταση της πληροφορίας
- Hashtags, 158
- Θεματικές Επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου. *Βλέπε* *LCSH*
- Θεματική ευρετηρίαση, 15
- Θησαυροί, 21, 25
- Λαοταξινομίες, 4, 46
- Λεξιλογική ομοιότητα, 111, 112
- Μη ελεγχόμενα λεξιλόγια, 45
- Οντολογίες, 7
- Πολυσημία, 29
- Προ-συνδυασμένη ευρετηρίαση, 20
- Σημασιολογική ευθυγράμμιση, 111
- Σημασιολογικός ιστός, 5
- Συνδεδεμένα Δεδομένα, 8, 72
- Συνωνυμία, 29
- Ταξινομικά σχήματα, 26
- Τεχνικές ταιριάσματος σε επίπεδο στοιχείων, 111
- Υστερο-συνδυασμένη ευρετηρίαση, 23
- Χρονική διάσταση της πληροφορίας, 154