



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΑΠΟΦΑΣΕΩΝ

**Χρήση του Σημασιολογικού Ιστού για τη διαχείριση
και την προαγωγή ψηφιακών βιβλίων με τη χρήση
κοινωνικών δικτύων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΚΑΣΤΡΙΝΑΚΗΣ Δ. ΓΕΩΡΓΙΟΣ

Επιβλέπων : Ασκούνης Δημήτριος
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2015



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΑΠΟΦΑΣΕΩΝ

Χρήση του Σημασιολογικού Ιστού για τη διαχείριση και την προαγωγή ψηφιακών βιβλίων με τη χρήση κοινωνικών δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΚΑΣΤΡΙΝΑΚΗ ΓΕΩΡΓΙΟΥ

Επιβλέπων : Ασκούνης Δημήτριος
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 26 Μαρτίου 2015.

(Υπογραφή)

.....
Ασκούνης Δημήτριος
Αναπληρωτής Καθηγητής
Ε.Μ.Π.

(Υπογραφή)

.....
Ψαρράς Ιωάννης
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Ασημακόπουλος Βασίλειος
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2015

.....

ΚΑΣΤΡΙΝΑΚΗΣ Δ. ΓΕΩΡΓΙΟΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Καστρινάκης Γεώργιος, 2015 .

Με επιφύλαξη παντός δικαιώματος. All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Βασικός σκοπός αυτής της διπλωματικής εργασίας είναι αρχικά να γίνει μια εισαγωγή στις βασικές έννοιες του σημασιολογικού ιστού (semantic web) και των τεχνολογιών που τον διέπουν και εν συνεχεία να εξετασθούν τα πλεονεκτήματα της υλοποίησης και αξιοποίησης τους στο πεδίο των βιβλίων καθώς και εν γένει η ανάγκη δημιουργίας και συντήρησης οντολογιών γύρω από αυτά, για την καλύτερη διαχείριση της γνώσης από τους υπολογιστές και τις μηχανές αναζήτησης. Θα επισημάνουμε τη σημασία των μέσων κοινωνικής δικτύωσης για την ενίσχυση της προσπάθειας εξάπλωσης του Σημασιολογικού Ιστού και την μαζική άντληση δεδομένων παρεχόμενων από τους χρήστες. Ακόμα θα αναφερθούμε σε νέες τεχνολογίες και καινοτόμες ιδέες που έχουν διατυπωθεί και υλοποιηθεί αλλά δεν έχει γίνει χρήση τους από το ευρύτερο κοινό μιάς και βρίσκονται ακόμη σε ερευνητικό – πειραματικό στάδιο.

Ξεκινώντας δημιουργήσαμε μια ιστοσελίδα αναζήτησης βιβλίων, ώστε να αντλούμε πληροφορίες, κριτικές και σχόλια αναγνωστών, τιμές και μέρη πώλησης για έναν συγκεκριμένο τίτλο βιβλίου, από άλλες ιστοσελίδες που παρέχουν αυτές τις πληροφορίες ελεύθερα. Με δεδομένη όλη την πληροφορία που μας επιστρέφεται από κάθε αναζήτηση, συνθέτουμε οντολογίες και αποθηκεύουμε όλη αυτή την πληροφορία με δομημένο τρόπο, ώστε να είναι επαναχρησιμοποιήσιμη και αναγνώσιμη από μια μηχανή αναζήτησης ή οποιοδήποτε υπολογιστικό σύστημα. Όλα αυτά επιτυγχάνονται αξιοποιώντας και εφαρμόζοντας τις αρχές του Σημασιολογικού Ιστού και των RDF δεδομένων.

Γίνεται ανάλυση των διαφόρων θεωρητικών και τεχνικών πτυχών του προβλήματος και της λύσης αυτού. Εξετάζονται σενάρια χρήσης και η σειρά με την οποία εκτυλίσσονται οι διάφορες ενέργειες προκειμένου να καταλήξουμε στο επιθυμητό αποτέλεσμα και τέλος αξιολογούμε το κατά πόσο έχουμε καλύψει τις απαιτήσεις του αρχικού προβλήματος που θέσαμε.

Λέξεις Κλειδιά:<< σημασιολογικός ιστός, RDF, οντολογία, βιβλίο, triples, SPARQL, query, δομημένη πληροφορία, διασυνδεδεμένα δεδομένα, ανοικτά δεδομένα, διαδίκτυο, αναγνωσιμότητα >>

Abstract

The main purpose of this thesis is to make a proper introduction to the fundamental concepts of semantic web and all the technologies that it utilizes. Afterwards, we will try to examine the benefits of the implementation and utilization of the semantic web in the field of books and the creation of ontologies describing them, so that computers and search engines maximize the value of the knowledge we currently have. Also we will make use of social media APIs in order to prove their significant contribution towards the overall goal of expanding the semantic web and pulling data with an automated manner.

We started from creating a search engine for book titles, in order to extract and manipulate information, reviews and comments from users, and finally offers from marketplaces, concerning the title that we entered. Then we compose all the needed ontologies and save all this data in a more structured way so that it is readable and reusable by a search engine and generally by any computing machine. In general, we utilized the principles of Semantic Web and RDF data.

Finally we examined some basic use case scenarios and summarized the outcome of our work.

Keywords: <<semantic web, RDF, ontology, book, structured data, readability, reusability, open data>>

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω όλους εκείνους που με βοήθησαν και με στήριξαν καθόλη τη διάρκεια της προσπάθειάς μου, καθώς και αυτούς που με εμπιστεύθηκαν από το ξεκίνημα κιόλας.

Ξεκινώντας θα ήθελα να ευχαριστήσω τον επιβλέποντα, Αναπληρωτή Καθηγητή ΕΜΠ κύριο Ασκούνη Δημήτριο που με την ενθάρρυνση του και την βοήθειά του κατάφερα να ολοκληρώσω αυτό το έργο.

Ξεχωριστά, θα ήθελα να ευχαριστήσω και τον Διδακτορικό φοιτητή Μιχάλη Πετυχάκη που με βοήθησε και με καθοδήγησε προκειμένου να ολοκληρωθεί επιτυχώς αυτή η διπλωματική εργασία.

Ακόμη θα ήθελα να ευχαριστήσω και όλους τους φίλους μου που συνέβαλαν σε αυτή τη διαδρομή, ο καθένας με το δικό του τρόπο.

Τέλος, θέλω να ευχαριστήσω την οικογένειά μου που με στήριξε και με ενθάρρυνε να συνεχίσω να προσπαθώ και να προοδεύω.

Ευχαριστώ πολύ.

Πίνακας περιεχομένων

1	Εισαγωγή.....	13
1.1	Πολιτιστική κληρονομιά – Βιβλία στο Διαδίκτυο.....	15
1.2	Το Πρόβλημα- Η Ιδέα.....	16
1.3	Δομή και Στόχοι της Εργασίας.....	18
1.4	Τα Επιχειρηματικά Πλεονεκτήματα της χρήσης του Σημασιολογικού Ιστού.....	19
1.5	Μέσα Κοινωνικής Δικτύωσης και Σημασιολογικός Ιστός.....	22
1.6	Οργάνωση κειμένου.....	23
2	Σχετικές Εργασίες.....	24
2.1	Σημασιολογικός Ιστός στη Μουσική - MusicBrainz.....	26
2.2	DBPEDIA.....	28
2.3	European Union Open Data	31
2.4	LinkedGeoData	33
3	Θεωρητικό υπόβαθρο.....	34
3.1	Διαδίκτυο και Δεδομένα.....	35
3.2	Διασυνδεδεμένα Δεδομένα (Linked Data).....	36
3.2.1	<i>Δημοσίευση Διασυνδεδεμένων Δεδομένων (Publishing Linked Data on the Web).....</i>	<i>38</i>
3.2.2	<i>Χρήση Διασυνδεδεμένων Δεδομένων (Consuming Linked Data on the Web).....</i>	<i>39</i>
3.3	Σημασιολογικός Ιστός (Semantic Web).....	42
3.4	Λεξικά (Vocabularies).....	43
3.5	Μορφές Κωδικοποίησης Δεδομένων(Serialization Formats).....	44
3.5.1	<i>RDF –Triples.....</i>	<i>44</i>
3.5.2	<i>Turtle.....</i>	<i>46</i>
3.5.3	<i>JSON & JSON-LD.....</i>	<i>46</i>
3.6	SPARQL.....	47
3.7	Linked Data Fragments.....	49

3.8	RDF Databases.....	50
4	Εφαρμογές με χρήση του Σημασιολογικού Ιστού.....	51
4.1	Αναζήτηση στο Σημασιολογικό Ιστό.....	52
4.1.1	<i>FALCONS</i>	54
4.1.2	<i>WATSON</i>	54
4.2	Περιήγηση στο Σημασιολογικό Ιστό (Semantic Web browsers).....	55
4.2.1	<i>Tabulator</i>	56
4.2.2	<i>Disco</i>	57
5	Περιγραφή Συστήματος.....	58
5.1	Διαγράμματα Χρήσης - Use Case Diagrams.....	58
5.2	Σχεδίαση Συστήματος - UML.....	61
5.3	Screenshots Χρήσης Ιστοσελίδας.....	65
6	Υλοποίηση Συστήματος.....	68
6.1	Λεπτομέρειες Υλοποίησης.....	68
6.1.1	<i>Υλοποίηση Ιστοσελίδας Αναζήτησης Βιβλίων</i>	69
6.1.2	<i>Υλοποίηση Οντολογιών-Διασυνδεδεμένων Δεδομένων</i>	72
6.2	Πλατφόρμες και προγραμματιστικά εργαλεία.....	73
6.2.1	<i>Django</i>	73
6.2.2	<i>PyCharm</i>	73
6.2.3	<i>Amazon API</i>	74
6.2.4	<i>Goodreads API</i>	74
6.2.5	<i>Bootstrap</i>	76
6.2.6	<i>RDFLib</i>	76
7	Επίλογος.....	78
7.1	Σύνοψη και συμπεράσματα.....	78
7.2	Μελλοντικές επεκτάσεις.....	80
8	Βιβλιογραφία.....	81

1

Εισαγωγή

Περιτριγυριζόμαστε από δεδομένα, δεδομένα από όλους τομείς της ζωής μας και σε διάφορες μορφές, καθεμία χρήσιμη για διαφορετικούς σκοπούς. Τα δεδομένα αυτά όμως αποκτούν αξία μόνο όταν γίνονται χρήσιμα για τη βελτίωση και τη διευκόλυνση των ανθρώπων ή των επιχειρήσεων[3]. Πολλοί οργανισμοί, επιχειρήσεις, εκδοτικοί οίκοι ακόμα και κυβερνήσεις κρατών επιλέγουν και έχουν ανοικτά προς όλους τα δεδομένα τους, συντελώντας με αυτόν τον τρόπο στον εμπλουτισμό και την αύξηση των δεδομένων που υπάρχουν στο διαδίκτυο. Με τη σειρά τους ιδιώτες και νέες επιχειρήσεις χρησιμοποιούν αυτά τα δεδομένα δημιουργώντας αξία και νέες δυνατότητες. Μέσα από αυτή τη διαδικασία ανταλλαγής και χρήσης δεδομένων, δημιουργούνται πολλαπλά ‘οικοσυστήματα’ τα οποία διαρκώς εξελίσσονται και δείχνουν έμπρακτα την ύπαρξη της ανάγκης πρόσβασης σε πληθώρα δεδομένων. Συνεπώς εγείρονται τα τρία παρακάτω ερωτήματα :

- 1) Ποιά είναι η καλύτερη μέθοδος πρόσβασης στα δεδομένα, προκειμένου να είναι πιο εύκολα επαναχρησιμοποιήσιμα ;
- 2) Πως είναι δυνατή η εύρεση συσχετισμένων δεδομένων μέσα στον τεράστιο όγκο των διαθέσιμων δεδομένων;
- 3) Πως μπορεί μια εφαρμογή να αντλεί δεδομένα από πολυάριθμες πηγές που μέχρι πρότινος ήταν άγνωστες;

Η απάντηση σε αυτά τα ερωτήματα, μπορεί να δοθεί με την επέκταση του υπάρχοντος διαδικτύου μέσα από τις αρχές των διασυνδεδεμένων δεδομένων (linked data).

Η διπλωματική αυτή εργασία έχει ως στόχο να καταδείξει και να εξετάσει μια συγκεκριμένη όψη του προβλήματος της απουσίας ενιαίας μεθόδου δόμησης των δεδομένων που υπάρχουν διαθέσιμα στο διαδίκτυο. Η υπερισχύουσα λύση που διαφαίνεται για το πρόβλημα αυτό είναι η υλοποίηση και η συντήρηση του σημασιολογικού ιστού, ενώ η όψη του προβλήματος στην οποία και θα επικεντρωθούμε, είναι ο χώρος των βιβλίων και της πολιτιστικής κληρονομιάς που αυτά εμπεριέχουν.

Η ενιαία δόμηση της πληροφορίας στο διαδίκτυο θα επέφερε πολύ μεγάλες διευκολύνσεις για την ανάπτυξη και χρήση νέων τεχνολογιών σε πολλούς τομείς που κάνουν χρήση των ηλεκτρονικών υπολογιστών. Η ανάγκη δόμησης των δεδομένων που υπάρχουν στο διαδίκτυο όπως θα αναλυθεί και στη συνέχεια της εργασίας, πηγάζει από τις δυνατότητες και την ευελιξία που θα προέκυπταν αν τα δεδομένα υπήρχαν διαθέσιμα με τρόπο τέτοιο ώστε να είναι αναγνώσιμα και αξιοποιήσιμα με αυτόματο τρόπο από τους ηλεκτρονικούς υπολογιστές. Αυτή η ιδέα αποτελεί την ουσία του Σημασιολογικού Ιστού, για τον οποίο θα μιλήσουμε εκτενώς παρακάτω. Η πτυχή του προβλήματος η οποία θα μας απασχολήσει κυρίως στην παρούσα εργασία, είναι τα βιβλία και όλες οι πληροφορίες που σχετίζονται με αυτά και είναι προσβάσιμες μέσα από το διαδίκτυο. Η σημασία της εισαγωγής σημασιολογικής μεταπληροφορίας στα δεδομένα που αποθηκεύονται στο διαδίκτυο θα καταδειχθεί καθ' όλη την έκταση της εργασίας και συνοπτικότερα στο κεφάλαιο 3.2 .

1.1 Πολιτιστική κληρονομιά – Βιβλία στο Διαδίκτυο

Τα βιβλία αποτελούν και πάντα αποτελούσαν σημαντικό κομμάτι της πολιτιστικής μας κληρονομιάς και συνοδεύουν την ιστορία των ανθρώπων στο μεγαλύτερο κομμάτι της. Τα βιβλία αποτέλεσαν το πρώτο και βασικότερο μέσο αποθήκευσης και διαμοιρασμού της πληροφορίας που κατείχαν οι άνθρωποι. Τα βιβλία ακόμη και σήμερα είναι αναντικατάστατα γιατί προσφέρουν μια πληθώρα στοιχείων που μια οθόνη υπολογιστή δεν μπορεί να προσφέρει στον χρήστη. Πολλοί θεωρούν ότι όσο και να εξελιχθεί η τεχνολογία μας, τα βιβλία θα συνεχίσουν να χρησιμοποιούνται από τους ανθρώπους μαζικά. Τα βιβλία αποτελούν τον πιο εύκολα προσβάσιμο τρόπο διαχείρισης της πληροφορίας (αποθήκευση, μετάδοση κλπ), για όλους τους ανθρώπους. Ωστόσο για την καλύτερη διαχείριση, συντήρηση και εκμετάλλευση όλου αυτού του όγκου πληροφορίας μπορούν να μας βοηθήσουν οι νέες τεχνολογίες και πιο συγκεκριμένα οι ηλεκτρονικοί υπολογιστές και το διαδίκτυο.

Με την πρόοδο της τεχνολογίας και τη μαζική χρήση του διαδικτύου, άρχισαν να μεταφέρονται τα βιβλία στον ψηφιακό κόσμο και αποτελούν πλέον κομμάτι του ηλεκτρονικού ιστού. Συνεπώς τα βιβλία και η πληροφορία που εμπεριέχεται σε αυτά θα πρέπει να ακολουθεί τις τάσεις και τις εξελίξεις του διαδικτύου. Όπως θα αναλύσουμε και παρακάτω, μία από τις κυρίαρχες τάσεις στο διαδίκτυο είναι η εξάπλωση του σημασιολογικού ιστού (semantic web), οι έννοιες του οποίου θα αναλυθούν σε επόμενο κεφάλαιο, και γίνεται προσπάθεια όλο και περισσότερα από τα ήδη υπάρχοντα δεδομένα στο διαδίκτυο να αποκτήσουν σημασιολογική μεταπληροφορία. Συνεπώς το ίδιο θα πρέπει να συμβεί και με την ψηφιακή πληροφορία που υπάρχει για τα βιβλία. Προς αυτή την κατεύθυνση στοχεύει και η παρούσα διπλωματική εργασία, αρχικά την κατάδειξη της αξίας ενός ολοκληρωμένου σημασιολογικού ιστού, στη συνέχεια τις προσπάθειες και τις μεθόδους δημιουργίας του, το όραμα που οδηγεί αυτή την προσπάθεια και τέλος τη θέση που μπορεί πρέπει να διαδραματίσουν τα βιβλία στον ψηφιακό κόσμο με τη βοήθεια του Σημασιολογικού Ιστού και των κοινωνικών δικτύων στο διαδίκτυο.

1.2 Το Πρόβλημα- Η Ιδέα

Προκειμένου τα δεδομένα να είναι επαναχρησιμοποιήσιμα, βασική απαίτηση είναι να έχουν σωστή, τυποποιημένη και απόλυτα ορισμένη δομή προκειμένου να μπορούν να αναπτυχθούν εργαλεία και εφαρμογές για την αυτοματοποίηση της επεξεργασίας και διαχείρισης της πληροφορίας.

Ήδη από το 1960 είχε ειπωθεί και σκιαγραφηθεί η ιδέα της διασυνδεδεμένης γνώσης και της δόμησής της με χρηστικό και αποδοτικό τρόπο για την καλύτερη αξιοποίησή της από τους υπολογιστές για την εξυπηρέτηση των ανθρώπων.

Ο Tim Berners-Lee εφευρέτης του διαδικτύου (World Wide Web), ξεκίνησε το 1989 αναπτύσσοντας την ιδέα του ηλεκτρονικού ιστού και οραματίστηκε την μετεξέλιξη αυτού σε ένα σημασιολογικό δίκτυο με διασυνδεδεμένα δεδομένα. Ένα δίκτυο δεδομένων με σημασιολογική μεταπληροφορία, το οποίο θα είχε την ιδιότητα να παρέχει αναγνώσιμη για τους υπολογιστές γνώση. Αυτή η μορφή του διαδικτύου ονομάστηκε Σημασιολογικό Διαδίκτυο (semantic web) και κάνοντας χρήση κανόνων ταξινόμησης και οργάνωσης των αντικειμένων που περιέχει, επιτρέπει την μεταξύ τους διασύνδεση. Ακόμη όλα τα δεδομένα έχουν προσκολλημένη πάνω τους την απαραίτητη μεταπληροφορία έτσι ώστε να είναι εφικτό με έναν αλγόριθμο, ο υπολογιστής να μπορεί να “κατανοήσει” και να αξιοποιήσει την πληροφορία που περιέχεται καθώς και τις διασυνδέσεις που σχετίζονται μαζί της.

Ξεκινώντας λοιπόν, στο διαδίκτυο για τη δημιουργία και περιγραφή των δεδομένων μιας ιστοσελίδας χρησιμοποιείται η γλώσσα HTML, η οποία όμως όταν χρησιμοποιείται μόνη της έχει τον περιορισμό ότι δεν παρέχει ουσιαστική πληροφορία σχετικά με το περιεχόμενό – τα δεδομένα της στον υπολογιστή. Αυτό το κενό λοιπόν έρχονται να καλύψουν κάποιες γλώσσες όπως οι XML, RDF και OWL. Αυτές οι γλώσσες μας επιτρέπουν να περιγράψουμε οτιδήποτε περιέχει μία ιστοσελίδα με τρόπο κατανοητό για έναν υπολογιστή, προκειμένου να υλοποιήσουμε την ιδέα του σημασιολογικού ιστού και να αξιοποιήσουμε το μέγιστο των δυνατοτήτων του.

Με την περιγραφή των δεδομένων με τρόπο που μπορεί να “κατανοήσει” ένας υπολογιστής και με την παράλληλη σύνδεση των δεδομένων που υπάρχουν στο διαδίκτυο, κατασκευάζουμε το Σημασιολογικό Ιστό. Ο Σημασιολογικός Ιστός αποτελώντας ένα δίκτυο δεδομένων προσπελάσιμων και κατανοητών από τους υπολογιστές μας παρέχει πληθώρα δυνατοτήτων μέσω της αυτοματοποίησης πολλών λειτουργιών. Μία τέτοια λειτουργία θα μπορούσε να είναι η συγκεντρωτική αναζήτηση (aggregated search) ενός όρου, δηλαδή ο αυτόματος συμψηφισμός όλης της υπάρχουσας πληροφορίας γύρω από τον συγκεκριμένο όρο και η παροχή ολοκληρωμένης και ενημερωμένης γνώσης για αυτόν. Ακόμη οι αναζητήσεις μας στο διαδίκτυο θα μπορούσαν να είναι σημασιολογικές και όχι λεκτικές[1], δηλαδή μια αναζήτηση να μην βασίζεται στην λεκτική ταύτιση των όρων που έχουμε εισάγει αλλά στην ουσιαστική απάντηση ενός ερωτήματος. Εν ολίγοις οι δυνατότητες αναζήτησης που μας παρέχει ο σημασιολογικός ιστός μπορούν να θεωρηθούν παρόμοιες με αυτές που μας παρέχει μια σχεσιακή βάση δεδομένων, με κεντρική τη δυνατότητα τοποθέτησης ερωτημάτων – queries με πολλαπλά κριτήρια αναφορικά με κάποια ιδιότητα μιας οντότητας ή τις σχέσεις που τη διέπουν.

Ο Σημασιολογικός Ιστός έχει ξεκινήσει και αναπτύσσεται τα τελευταία χρόνια και επεκτείνεται σε ολοένα και νέους τομείς και κλάδους της ζωής μας. Οι κυριότερες δυσκολίες για την επέκταση και την ευρεία εφαρμογή του σημασιολογικού ιστού [1], είναι αρχικά η έκταση του και συνεπώς η δυσκολία μετατροπής του καθώς και οι μεγάλοι περιορισμοί που προκύπτουν από τον τεράστιο όγκο των συνολικών δεδομένων του, ενώ έπειτα η κοινή αποδοχή μιας νέας μορφής διαδικτύου αποτελεί και θέμα διεθνών σχέσεων και συμφωνιών μεταξύ των χωρών του κόσμου. Τελικά το πρόβλημα του σημασιολογικού ιστού μπορεί να χαρακτηριστεί περισσότερο ως κοινωνικό – επικοινωνιακό, παρά τεχνικό μιας και υπάρχουν διαθέσιμες οι απαραίτητες τεχνολογίες αλλά δεν είναι επαρκώς διαδεδομένες αυτή τη στιγμή.

Το πρόβλημα που εντοπίσαμε και προσπαθήσαμε να αναδείξουμε μέσα από τη διπλωματική αυτή εργασία είναι το γεγονός ότι , συγκεκριμένα η πληροφορία που υπάρχει στο διαδίκτυο σχετικά με τα βιβλία δεν αποτελεί ακόμη σημαντικό κομμάτι του Σημασιολογικού Ιστού.

Προς το παρόν δεν έχει γίνει κάποια μεγάλη και συντονισμένη προσπάθεια δόμησης και εμπλουτισμού της πληροφορίας γύρω από τα βιβλία, τους συγγραφείς και οτιδήποτε άλλο σχετικό και κοντινό με τον τομέα αυτό. Η ιδέα που θα παρουσιαστεί στην εργασία αυτή είναι μια προσπάθεια να συλλέξουμε και να οργανώσουμε ένα κομμάτι από την πληροφορία που υπάρχει στο διαδίκτυο σχετικά με τα βιβλία. Για την προσπάθεια αυτή θα παίξουν σημαντικό ρόλο τα μέσα κοινωνικής δικτύωσης του διαδικτύου σε συνδυασμό με το Σημασιολογικό Ιστό. Αυτό το επιχειρούμε μέσω μιας ιστοσελίδας αναζήτησης τίτλων βιβλίων, η οποία αντλεί

δεδομένα που υπάρχουν σε άλλες ιστοσελίδες και δημιουργώντας οντότητες για κάθε βιβλίο που συναντά. Κατασκευάζουμε λοιπόν δοδημένη και διασυνδεδεμένη πληροφορία την οποία αποθηκεύουμε σε μια εξειδικευμένη βάση δεδομένων για την επίτευξη του σκοπού μας. Συντελώντας με τον τρόπο αυτό στην επέκταση του σημασιολογικού ιστού και τη βελτίωση των υπηρεσιών που παρέχουν οι μηχανές αναζήτησης καθώς και την καλύτερη οργάνωση της υπάρχουσας πληροφορίας γύρω από τα βιβλία.

1.3 Δομή και Στόχοι της Εργασίας

Μέσα από την παρούσα διπλωματική εργασία, θα εξετασθούν σε θεωρητικό επίπεδο όλες οι εκφάνσεις, τα πλεονεκτήματα και οι τεχνολογίες σχετικά με τα διασυνδεδεμένα δεδομένα και το Σημασιολογικό Ιστό. Θα επισημανθεί η σημασία χρήσης των μέσων κοινωνικής δικτύωσης για την άντληση πληροφοριών. Ενώ σε πρακτικό – λειτουργικό επίπεδο θα παρουσιαστεί μέσα από την ιστοσελίδα που αναπτύχθηκε, “booksinfo”, πως μπορεί κάποιος να δημιουργήσει σημασιολογία και να δομήσει τα δεδομένα που αντλεί και χρησιμοποιεί από άλλες ιστοσελίδες.

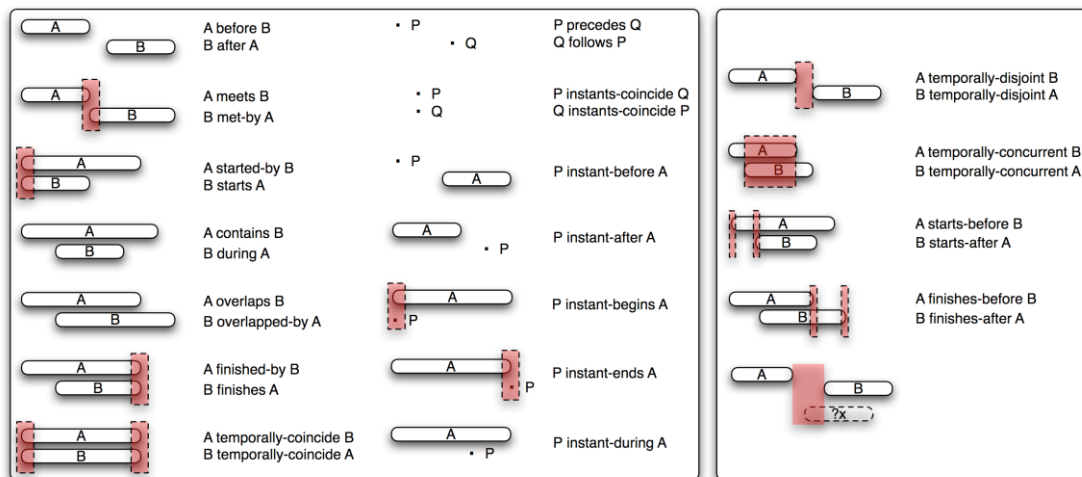
Υλοποιήσαμε λοιπόν μια ιστοσελίδα αναζήτησης βιβλίων που αντλεί πληροφορίες από τις διευθύνσεις www.goodreads.com, www.amazon.com και τις παρουσιάζει στο χρήστη ενώ παράλληλα δημιουργεί τις απαραίτητες οντότητες που περιγράφουν το βιβλίο που αναζητήθηκε και τις συνδέει μεταξύ τους, ώστε όλη η πληροφορία που βρέθηκε να γίνει πιο χρήσιμη και πιο ωφέλιμη για όλους τους χρήστες. Τέλος όλες οι παραπάνω οντότητες και η πληροφορία που περιέχουν αποθηκεύονται σε μία βάση δεδομένων, ώστε να χρησιμοποιούνται και σε μελλοντικές αναζητήσεις.

Κάνοντας χρήση των διεπαφών – APIs κοινωνικών δικτύων, όπως το amazon.com και το goodreads.com, κάναμε αναζήτηση στα δικά τους δεδομένα με βάση τις λέξεις κλειδιά που εισάγει ο χρήστης στη δική μας ιστοσελίδα. Έπειτα επεξεργαζόμαστε τα xml αρχεία που μας επιστρέφονται και εξάγουμε όλη την πληροφορία, την οποία στη συνέχεια ταιριάζουμε με τα αντίστοιχα πεδία με βάση το schema.org/Book ώστε να δημιουργήσουμε δομημένη πληροφορία με σημασιολογικό περιεχόμενο, την οποία τελικά αποθηκεύουμε στη βάση δεδομένων μας με μορφή τριάδων (RDF triples), η οποία θα επεξηγηθεί σε επόμενο κεφάλαιο.

1.4 Τα Επιχειρηματικά Πλεονεκτήματα της χρήσης του Σημασιολογικού Ιστού

Αναφέραμε διάφορα σχετικά με τα χαρακτηριστικά και τη μορφή του Σημασιολογικού Ιστού, ακόμη και για την έλλειψη που παρατηρήσαμε στον τομέα των δεδομένων γύρω από τα βιβλία , τώρα όμως θα εξετάσουμε δύο ερωτήματα, πρώτον ποιοί είναι οι λόγοι για τους οποίους θα ήθελε κάποιος να έχει τα δεδομένα του ανοιχτά στον Σημασιολογικό Ιστό και δεύτερον ποιά άλλα θετικά στοιχεία μπορεί να προσφέρει αυτός στη λειτουργία μιας επιχείρησης και στη βελτίωσή των λειτουργιών της.

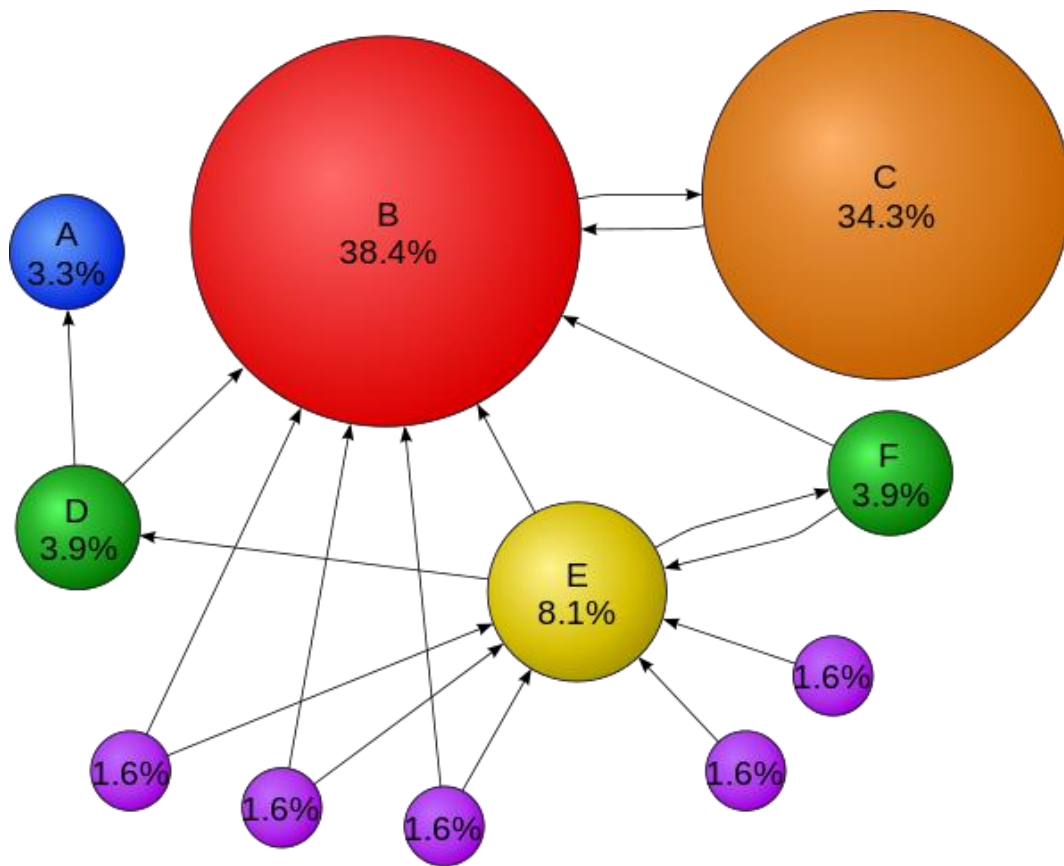
Θα ξεκινήσουμε δίνοντας απάντηση στο δεύτερο ερώτημα και πως ο σημασιολογικός ιστός και οι εφαρμογές του μπορούν να δώσουν λύσεις για τη βελτίωση της λειτουργίας των επιχειρήσεων[2]. Τα τελευταία χρόνια έχουν υλοποιηθεί πολλά προγράμματα λογισμικού που βοηθούν στον καλύτερο προγραμματισμό και διαχείριση των απαιτήσεων μιας επιχείρησης, όμως όλα έχουν περιορισμούς στην αυτοματοποίηση που προσφέρουν μιας και η μετάβαση από τον πραγματικό επιχειρηματικό κόσμο στον αλγοριθμικό, απαιτεί την απασχόληση κάποιου προγραμματιστή που θα θέσει τις σωστές παραμέτρους. Αυτό μεταφράζεται σαν αύξηση στο κόστος λειτουργίας της επιχείρησης και είναι κάτι που προφανώς θέλουμε να αποφύγουμε. Τη λύση σε αυτό το πρόβλημα γίνεται προσπάθεια για να δώσει ο συνδυασμός του σημασιολογικού ιστού και των υπηρεσιών διαχείρισης επιχειρηματικών διεργασιών (BPM). Αυτή η προσπάθεια έχει στηριχθεί στην ανάπτυξη μιας οντολογίας ονόματι COBRA(Core Ontology for Business pRocess Analysis), η οποία αποτελεί ένα επεκτάσιμο και παραμετροποιήσιμο πλαίσιο για την υποστήριξη της αυτοματοποίησης των επιχειρηματικών διεργασιών (BPA).



Εικόνα 1. Οντολογία COBRA.

Η απάντηση στο πρώτο ερώτημα είναι πολύ απλή και το μέγεθος της σημασίας της θα φανεί και θα τονισθεί και στη συνέχεια της εργασίας.

Η πληροφορία αποκτά αξία μόνο όταν μπορεί να την προσπελάσει κάποιος. Συνεπώς για μια πληροφορία είναι προφανές ότι αυξάνεται η αξία της αν υπάρχουν πολλαπλά μονοπάτια που οδηγούν σε αυτήν. Πάνω σε αυτήν τη λογική έχει στηριχθεί και ο βασικός αλγόριθμος υπολογισμού της αξία μιας ιστοσελίδας. Η αξία μιας ιστοσελίδας, προκύπτει άμεσα από τον αριθμό των ανακατευθύνσεων (links) που οδηγούν σε αυτήν και ο υπολογισμός αυτής βασίζεται πάνω στον αλγόριθμο PageRank και χρησιμοποιείται από την Google για την ταξινόμηση των αποτελεσμάτων που παρουσιάζει σε μια αναζήτηση στο Google Search. Μια αναπαράσταση των αποτελεσμάτων του αλγορίθμου PageRank, φαίνεται στην Εικόνα παρακάτω.



Εικόνα 2. Αναπαράσταση αλγορίθμου PageRank σε ένα απλό δίκτυο συνδεδεμένων αντικειμένων

Ο Σημασιολογικός Ιστός και οι σχέσεις που σχηματίζει ανάμεσα στις οντότητες που τον απαρτίζουν, δομούν ένα σύνθετα συνδεδεμένο δίκτυο πληροφοριών. Συνεπώς, η αξία του Σημασιολογικού Ιστού πηγάζει από την ευκολία προσπέλασης και εκμετάλλευσης των δεδομένων που περιέχει, οι πολλαπλές ανακατευθύνσεις ανάμεσα στα περιεχόμενά του και η ενιαία δόμησή τους τα κάνει πιο ελκυστικά για έναν άνθρωπο αλλά και για μια μηχανή αναζήτησης. Από τα παραπάνω λοιπόν, γίνεται εύκολα κατανοητό το γεγονός ότι μια ιστοσελίδα που έχει τα δεδομένα της στον Σημασιολογικό Ιστό αυξάνει την αξία της απέναντι σε ιστοσελίδες που έχουν τα δεδομένα τους αδόμητα.

Για να επανέλθουμε στον τομέα που θα μας απασχολήσει, μια αποθήκη βιβλίων που έχει ανεβασμένα τα δεδομένα της στον Σημασιολογικό Ιστό και ακολουθεί τους κανόνες προτυποποίησης που υπάρχουν είναι βέβαιο ότι θα αυξήσει άμεσα και έντονα την αξία των δεδομένων της έναντι σε μια άλλη που δεν έχει δομήσει τα δεδομένα της με αυτόν τον τρόπο. Περισσότεροι χρήστες θα οδηγηθούν στη σελίδα της, περισσότερες μηχανές αναζήτησης θα προβάλουν αποτελέσματα που θα οδηγούν σε αυτήν και ως τελική συνέπεια θα αυξηθεί και θα συνεχίσει να αυξάνεται η αξία της.

1.5 Μέσα Κοινωνικής Δικτύωσης και Σημασιολογικός Ιστός

Τα γενικής και ευρείας χρήσης μέσα κοινωνικής δικτύωσης όπως το Facebook και το Twitter, καθώς και αυτά που είναι πιο στοχευμένης χρήσης όπως το Goodreads που απευθύνεται σε χρήστες που διαβάζουν βιβλία, σε συνδυασμό με τις τεχνικές του Σημασιολογικού Ιστού μπορούν να διαδραματίσουν έναν πρωταγωνιστικό ρόλο στην ελεύθερη διακίνηση της γνώσης και τον εμπλουτισμό των ψηφιακών βιβλιοθηκών μας.

Η μεγάλη πλειοψηφία των μέσων κοινωνικής δικτύωσης, παρέχουν πλέον με ελεύθερο τρόπο τις διεπαφές τους για χρήση από οποιονδήποτε ενδιαφερόμενο. Αυτό δίνει μία πολύ βασική ευκολία σε όποιον προγραμματιστή επιθυμεί να κατασκευάσει μία εφαρμογή. Διότι του παρέχει πρόσβαση σε μια τεράστια γκάμα υπηρεσιών και δεδομένων τα οποία μπορεί να χρησιμοποιήσει προκειμένου να βελτιώσει δραματικά την εμπειρία των χρηστών της εφαρμογής του. Αυτές οι διεπαφές – APIs των διαφόρων μέσων κοινωνικής δικτύωσης θεωρούμε ότι μπορούν σε συνδυασμό με το Σημασιολογικό Ιστό να δώσουν μεγάλη ώθηση στην προσπάθεια εξάπλωσης του παγκόσμιου γράφου γνώσης των Διασυνδεδεμένων Δεδομένων.

Για αυτό το λόγο στην παρούσα διπλωματική εργασία επιλέξαμε σαν λύση για την άντληση δεδομένων να χρησιμοποιήσουμε τέτοιες διεπαφές. Αρχικά από το Goodreads, ένα μέσο κοινωνικής δικτύωσης με βασικό χαρακτηριστικό τα βιβλία και το Amazon, ένα ηλεκτρονικό κατάστημα που παρέχει δυνατότητα σχολιασμού και πρότασης προϊόντων σε άλλους χρήστες. Σε μελλοντική επέκταση της εφαρμογής που αναπτύχθηκε στα πλαίσια της διπλωματικής αυτής εργασίας θα μπορούσε να προστεθεί και η διεπαφή του Facebook,

προκειμένου να αντλούμε επιπλέον πληροφορίες, σχόλια και βαθμολογίες που αφορούν τίτλους βιβλίων.

1.6 Οργάνωση κειμένου

Εργασίες σχετικές με το αντικείμενο της διπλωματικής παρουσιάζονται στο Κεφάλαιο 2.

Το Κεφάλαιο 3 περιλαμβάνει το βασικό θεωρητικό υπόβαθρο που θα βοηθήσει τον αναγνώστη να κατανοήσει την τεχνολογία και το στόχο της εργασίας.

Στο Κεφάλαιο 4 εξετάζουμε διάφορα πιθανά σενάρια χρήσης μέσω διαγραμμάτων (use case diagrams).

Στο Κεφάλαιο 5 αναλύεται το σύστημα και το πως υλοποιήθηκαν όλα τα επιμέρους κομμάτια του.

Και τέλος στα Κεφάλαια 6 και 7 έχουμε τον επίλογο μαζί με κάποια συμπεράσματα και τις μελλοντικές επεκτάσεις της εργασίας μας και κλείνοντας την βιβλιογραφία που χρησιμοποιήθηκε.

2

Σχετικές εργασίες

Με αφορμή το οράμα του Tim Berners - Lee για την εξέλιξη του διαδικτύου σε διαδίκτυο δεδομένων αναγνώσιμων από τους υπολογιστές ,τα τελευταία χρόνια έχουν γίνει πολλές προσπάθειες. Οι προσπάθειες αυτές επικεντρώνονται σε πολλούς τομείς και στοχεύουν στην επέκταση του σημασιολογικού ιστού, μέσω της υπάρχουσας πληροφορίας. Η υπάρχουσα πληροφορία πρέπει να οργανωθεί και να εισαχθεί σε αυτήν όλη η απαραίτητη μεταπληροφορία (metadata) με τη βοήθεια του RDF (resource definition framework) και άλλων τεχνολογιών, προκειμένου να επιτευχθεί η οργάνωση των δεδομένων με δομή και ορολογία αναγνώσιμη από τους υπολογιστές.

Την τελευταία δεκατία περίπου έχει επισημανθεί ιδιαίτερα η αξία και η ουσιαστική χρηστικότητα που θα προκύψει από τη δόμηση και την εκμετάλλευση της πληροφορίας που υπάρχει διαθέσιμη στον παγκόσμιο Ιστό. Προς αυτό το όραμα και την προοπτική εξέλιξης έχουν ξεκινήσει και γίνονται πολλαπλές και οργανωμένες κινήσεις. Στο πλαίσιο αυτών των προσπαθειών έχουν δημιουργηθεί πολλές ομάδες – κινήσεις ανθρώπων, που εργάζονται με κοινό στόχο τη δημιουργία και την αξιοποίηση του σημασιολογικού ιστού.

Αξίζει να σημειωθεί και να επισημανθεί ξεχωριστά η σημασία μίας κίνησης με το όνομα Wikimedia. Η Wikimedia αποτελεί μία από τις πιο σημαντικές και οργανωμένες προσπάθειες προαγωγής του σημασιολογικού ιστού και της σημασίας του για την εκπαίδευση και την πρόσβαση του ανθρώπου στην παγκόσμια γνώση που κατέχουμε πλέον. Στη Wikimedia συμμετέχουν πολλά μη κερδοσκοπικά έργα όπως τα Wikipedia, Wikibooks, Wikinews, Wikiversity και Wikidata που απώτερο σκόπο έχουν την δωρεάν παροχή γνώσης σε παγκόσμια κλίμακα. Το κοινό όραμα όλων αυτών των έργων είναι η δημιουργία ενός ενιαίου, παγκόσμιου, συλλογικού και ελεύθερου δικτύου γνώσης, στο οποίο θα μπορεί ο καθένας να

συνεισφέρει. Στη δημιουργία αυτού του τεράστιου και δυναμικού δικτύου γνώσης, σημαντικό ρόλο έχουν παίξει τα διασυνδεδεμένα δεδομένα και ο σημασιολογικός ιστός. Στο δίκτυο αυτό, όλη η πληροφορία έχει αποτυπωθεί με τη μορφή οντοτήτων και συνδέσεων μεταξύ αυτών, με χρήση της ιδέας του σημασιολογικού ιστού και των διασυνδεδεμένων δεδομένων.



Εικόνα 3. Έργα που συμμετέχουν στην Wikimedia

Παρακάτω θα αναλύσουμε τα πιο χαρακτηριστικά παραδείγματα σχετικών εργασιών, στο χώρο των διασυνδεδεμένων δεδομένων και του σημασιολογικού ιστού (semantic web).

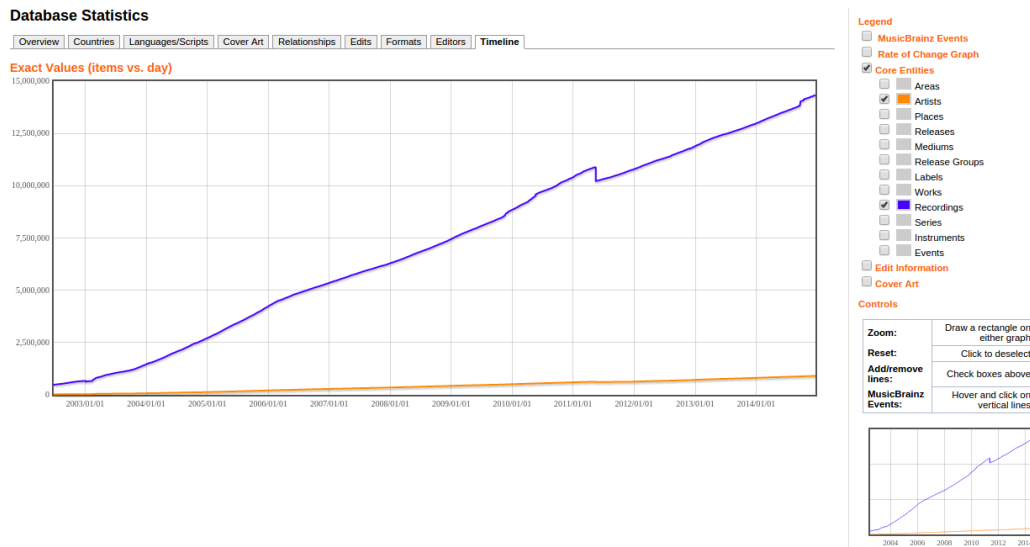
2.1 MusicBrainz

Το αντίστοιχο πρόβλημα ανάπτυξης του σημασιολογικού ιστού, στον εξειδικευμένο τομέα της μουσικής γίνεται προσπάθεια να καλυφθεί με τη δημιουργία του ιστοτόπου www.musicbrainz.org . Στο MusicBrainz έχει συγκεντρώθει μεγάλο κομμάτι της πληροφορίας που υπάρχει στο διαδίκτυο σχετικά με τη μουσική και τους καλλιτέχνες παγκοσμίως, συγκροτήματα και τα μέλη τους, τα άλμπουμ τους, τα τραγούδια τους και όποια άλλη πληροφορία μπορεί να υπάρχει διαθέσιμη καθώς και τις σχέσεις που απαρτίζουν και συνδέουν όλες αυτές τις πληροφορίες.

Το MusicBrainz αποτελεί μια διαδικτυακή ανοιχτή βάση δεδομένων για την μουσική, η οποία συντηρείται και επεκτείνεται από εθελοντές. Το MusicBrainz, μέσω της ιστοσελίδας του παρέχει στους χρήστες τη δυνατότητα αναζήτησης σχετικών με τη μουσική πληροφοριών με βάση πολλαπλά κριτήρια ή ακόμη και την αναζήτηση μουσικών κομματιών με βάση το ακουστικό τους αποτύπωμα (acoustic fingerprint).

Όπως φαίνεται στην εικόνα 2 παρακάτω, το MusicBrainz το έτος 2002 είχε γύρω στις 500.000 καταχωρημένες ηχογραφήσεις ενώ το έτος 2014 οι αντίστοιχες καταχωρήσεις έχουν φτάσει να είναι σχεδόν 15.000.000 . Μπορούμε λοιπόν να κατανοήσουμε τη συνεχή προσπάθεια, τις διαρκείς ανανεώσεις καθώς και τις νέες καταχωρήσεις που λαμβάνουν χώρα από την κοινότητα του MusicBrainz και το μέγεθος της συνεισφοράς του έργου τους στην ανάπτυξη και την προώθηση των ανοικτών διασυνδεδεμένων δεδομένων και του Σημασιολογικού Ιστού.

Τέλος πάνω στην δομημένη και οργανωμένη πληροφορία που έχει συγκεντρώσει το MusicBrainz, έχει αναπτυχθεί μια σειρά από εφαρμογές που οργανώνουν και προσθέτουν όλη την υπάρχουσα πληροφορία, η οποία είναι συνδεδεμένη με τα τραγούδια που είναι αποθηκευμένα στον υπολογιστή του χρήστη π.χ. τα ονόματα των καλλιτεχνών και των άλμπουμ , ακόμη υπάρχουν εφαρμογές όπως το MusicBrainz Android που παρέχουν όλες τις υπηρεσίες του MusicBrainz σε οποιαδήποτε συσκευή που τρέχει το λειτουργικό Android.



Εικόνα 4. Διάγραμμα με τις καταχωρημένες στο MusicBrainz ηχογραφήσεις από το 2002 μέχρι σήμερα

← → ↻ musicbrainz.org/search?query=coldplay&type=artist&limit=25&method=direct

MusicBrainz coldplay Artist Search

About Blog Products Search Documentation Contact Us Log In Create Account

Search Results

Found 3 results for "coldplay"

Score	Name	Sort Name	Type	Gender	Area	Begin	Begin Area	End	End Area
100	Coldplay	Coldplay	Group		United Kingdom	1996-09	London		
50	冷玩妹	Coldplay Sister	Person		Taiwan				
16	The Section Quartet	Section Quartet, The	Group		United States	1998	Los Angeles		

Alternatively, you may [add a new artist](#).

Query: coldplay

Type: Artist

Results per page: Up to 25

Search method:

- ☐ Indexed search
- ☐ Indexed search with [advanced query syntax](#)
- ☒ Direct database search

Please note:

Search indexes are updated every 3 hours, use the direct database search if you require up to the minute correct results.

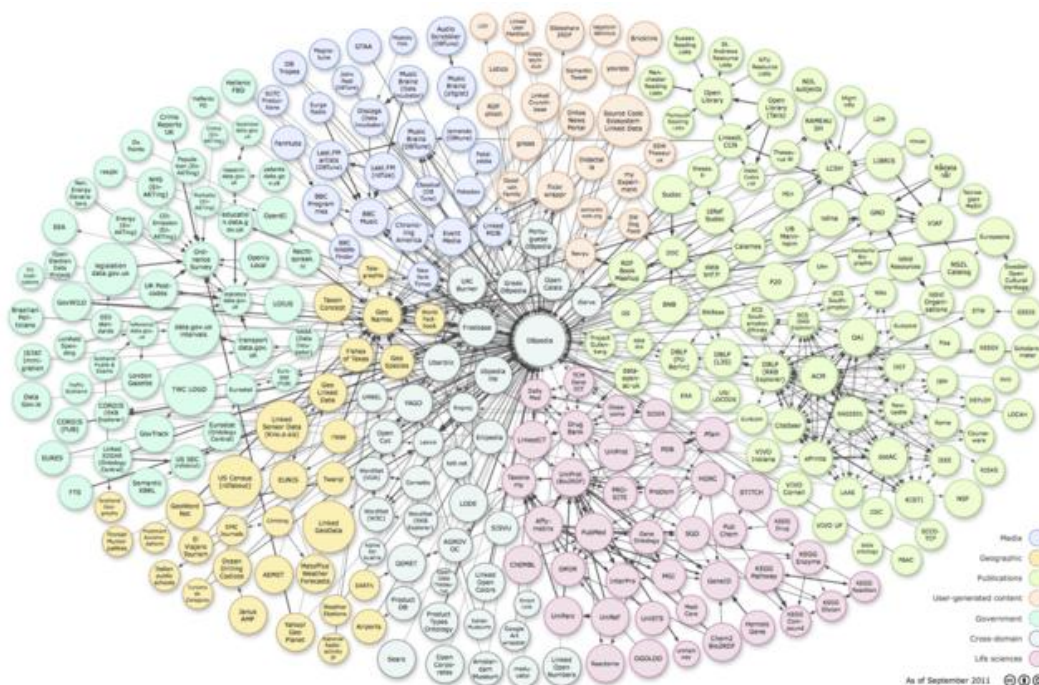
[Donate](#) | [Wiki](#) | [Forums](#) | [Bug Tracker](#) | [Twitter](#) | [Use beta site](#)

Cover Art provided by the [Cover Art Archive](#). Hosted by Digital West. Sponsored by: Google, OSUOL and others...

Εικόνα 5. Αναζήτηση Καλλιτέχνη στο MusicBrainz.org

2.2 DBpedia

Η DBpedia αποτελεί μια συλλογική προσπάθεια για την εξαγωγή και επαναχρησιμοποίηση δομημένης πληροφορίας από την Wikipedia. Στοχεύει στον μετασχηματισμό τις πληροφορίες που υπάρχει στην Wikipedia, προκειμένου να αποτελέσει κομμάτι του σημασιολογικού ιστού. Ως αποτέλεσμα θα προκύψει η ανάπτυξη καλύτερων και αποδοτικότερων μεθόδων αναζήτησης πληροφοριών μέσω της γνώσης που μας παρέχει η Wikipedia. Η DBpedia αποτελεί τον πυρήνα των διασυνδεδεμένων δεδομένων (Linked Data) και ολόκληρος ο σημασιολογικός ιστός ξεκίνησε να αναπτύσσεται γύρω από αυτήν. Στην παρακάτω εικόνα φαίνεται ένα διάγραμμα που αναπαριστά τις διάφορες ομάδες ελεύθερων διασυνδεδεμένων δεδομένων (Open Linked Data datasets) και τις μεταξύ τους συνδέσεις , στο κέντρο όλων των δεδομένων βρίσκεται η DBPedia. Αυτό το διάγραμμα έχει σχεδιαστεί για τα δεδομένα που υπήρχαν το Σεπτέμβριο του 2011, η σημερινή εικόνα αυτού του νέφους είναι πολύ πιο πολύπλοκη και εκτεταμένη.



Εικόνα 6. Διάγραμμα Νέφους Ελεύθερων Διασυνδεδεμένων Δεδομένων(Open Linked Data datasets)

Η Wikipedia περιέχει άρθρα που αποτελούνται κυρίως από κείμενο, περιλαμβάνουν όμως πλέον και τα infoboxes που περιέχουν δομημένη πληροφορία η οποία αρχικά βοηθούσε στην καλύτερη και συνοπτικότερη παρουσίαση βασικών πληροφοριών. Στη συνέχεια και με την εμφάνιση της DBpedia τα infoboxes βοήθησαν στην αυτόματη εξαγωγή, συλλογή και δόμηση της πληροφορίας ώστε να είναι εκμεταλλεύσιμη για την άμεση απάντηση των ερωτημάτων (queries) που υποβάλλονται στην DBpedia. Η DBpedia λοιπόν παρέχει τη δυνατότητα υποβολής σύνθετων ερωτημάτων προς τη Βάση της, η οποία δομημένη σε τριάδες, για την εξεύρεση πληροφορίας ή των σχέσεων που απαρτίζουν τις διάφορες πληροφορίες. Τα ερωτήματα αυτά υποβάλλονται με SPARQL, την οποία θα αναλύσουμε σε επόμενο κεφάλαιο.

Γύρω από την DBpedia έχει αναπτυχθεί μια πληθώρα εργαλείων που προσφέρουν στους χρήστες πολλαπλές δυνατότητες. Τα πιο σημαντικά από αυτά είναι τα παρακάτω :

- DBpedia Live : Εργαλείο για την σύγχρονη υποβολή ερωτημάτων (queries) στην πληροφορία που εξάγεται από την Wikipedia μέσω των SPARQL endpoints .

← → live.dbpedia.org/sparql

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#) | [iSPARQL](#)

Default Data Set Name (Graph IRI)

Query Text
select distinct ?Concept where {[[] a ?Concept]} LIMIT 100

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:
Execution timeout: milliseconds (values less than 1000 are ignored)

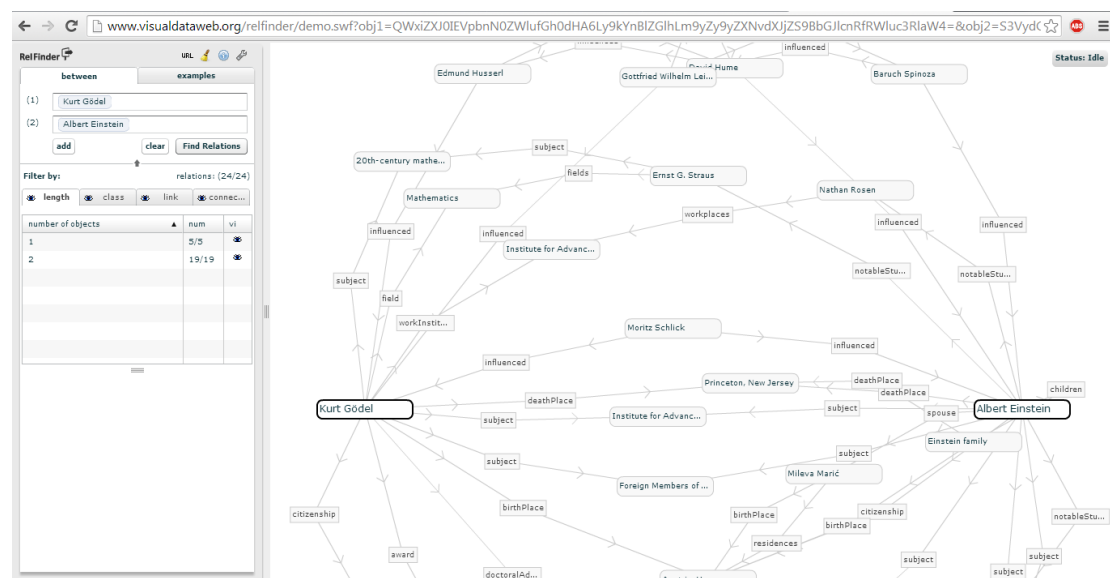
Options:
☒ **Strict checking of void variables**

(The result can only be sent back to browser, not saved on the server, see [details](#))

Copyright © 2014 [OpenLink Software](#)
Virtuoso version 07.10.3211 on Linux (x86_64-redhat-linux-gnu), Single Server Edition

Εικόνα 7. Υποβολή query μέσω του DBpedia Live

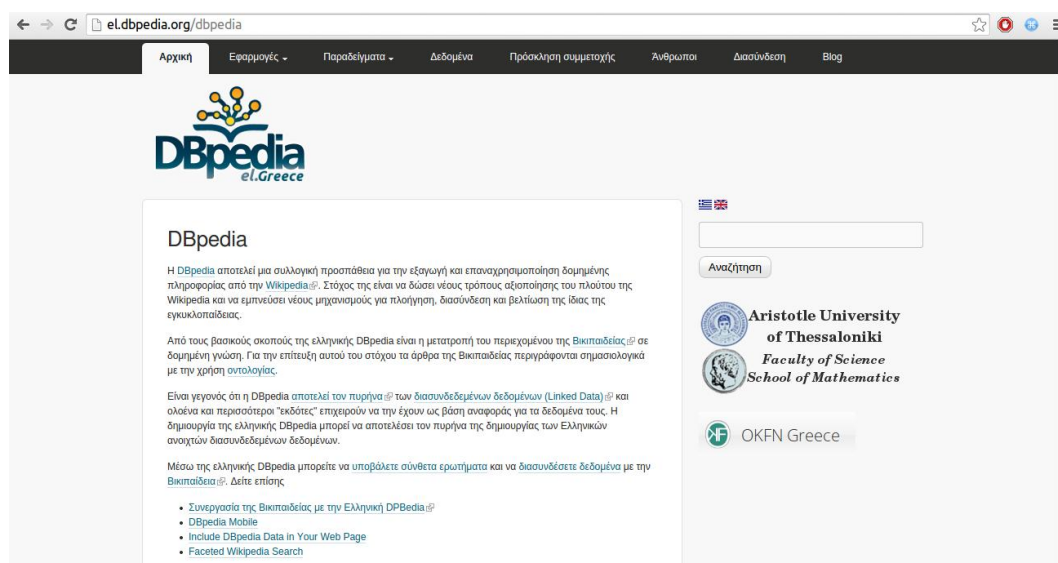
- DBpedia Mapping Tool : Εργαλείο που βοηθάει τους χρήστες να δημιουργούν αντιστοιχίσεις μεταξύ των προτύπων της Wikipedia και των οντολογιών της DBpedia, μέσα από ένα γραφικό περιβάλλον εργασίας που εκτελείται στον browser.
- DBpedia RelFinder : Web εφαρμογή που εξάγει και οπτικοποιεί σχέσεις μεταξύ αντικειμένων που έχουν περιγραφεί με δεδομένα σε RDF (αναπαράσταση δεδομένων στην DBpedia). Εκμεταλλεύεται την πληροφορία που παρέχει ο σημασιολογικός ιστός προκειμένου να εξερευνήσει μεγάλα σύνολα δεδομένων και να εξάγει τις διασυνδέσεις που υπάρχουν ανάμεσα στις διάφορες οντότητες. Έχει τη δυνατότητα άντλησης δεδομένων από οποιαδήποτε βάση διασυνδεδεμένων δεδομένων. Οπτικοποιεί τις σχέσεις των αντικειμένων με τη μορφή κατευθυνόμενου γράφου, παρέχει τη δυνατότητα επιλογής SPARQL endpoint για την υποβολή query σε συγκεκριμένη βάση διασυνδεδεμένων δεδομένων και τέλος ο χρήστης μπορεί να φιλτράρει τα επιστρεφόμενα αποτελέσματα.



Εικόνα 8. DBpedia RelFinder – (Albert Einstein, Kurt Godel)

Τέλος η αντίστοιχη ελληνική έκδοση της DBpedia, el.dbpedia.org, ξεκίνησε να αναπτύσσεται ως έργο μεταπτυχιακού μαθήματος του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, με βάση την ελληνική Wikipedia και με συνδέσεις προς τα αντίστοιχα άρθρα της αγγλικής Wikipedia. Η ελληνική DBpedia φτιάχνετε με πρότυπο την αγγλική DBpedia και τις

υπηρεσίες που παρέχει αυτή και παρότι βρίσκεται σε αρκετά πρώιμο στάδιο είναι μία από τις πρώτες προσπάθειες που γίνονται εξαιρουμένης της αγγλικής έκδοσης.



Εικόνα 9. Αρχική σελίδα της ελληνικής Dbpedia, έκδοση πρώτη

2.3 European Union Open Data

Τα διασυνδεδεμένα δεδομένα έχουν χρησιμοποιηθεί και από κρατικούς φορείς για την κοινοποίηση στοιχείων και τη διασφάλιση της διαφάνειας διαφόρων διαδικασιών. Τέτοιο εγχείρημα είναι και το EU Open Data, το συγκεκριμένο έργο δίνει στους χρήστες τη δυνατότητα να αναζητήσουν, να συνδέσουν και να “κατεβάσουν” δεδομένα για οποιαδήποτε χρήση εμπορική ή μη. Η αναζήτηση γίνεται μέσω ενός καταλόγου μεταδεδομένων, ο οποίος μας δίνει πρόσβαση σε στοιχεία ευρωπαϊκών οργανισμών και υπηρεσιών. Για τις αναζητήσεις δεδομένων εκμεταλλευόμαστε τεχνολογίες του σημασιολογικού ιστού, για παράδειγμα στον κατάλογο μεταδεδομένων μπορούμε να ανατρέξουμε την πληροφορία που μας ενδιαφέρει με απλή αναζήτηση όρων (Data Tab) ή υποβάλλο κάποιον SPARQL query (Linked Data Tab) . Όλα τα δεδομένα υπάρχουν σε 23 γλώσσες και οι χρήστες έχουν τη δυνατότητα να υποβάλλουν αίτημα ώστε να γίνει διαθέσιμη κάποια πληροφορία τους ενδιαφέρει αλλά ακόμη δεν υπάρχει και τέλος.

Τα δεδομένα στα οποία μπορεί κανείς να έχει πρόσβαση, ανήκουν στους παρακάτω τομείς:

- Οικονομίας
- Εκπαίδευσης
- Εργασίας
- Περιβάλλοντος
- Επιστήμης

Popular terms

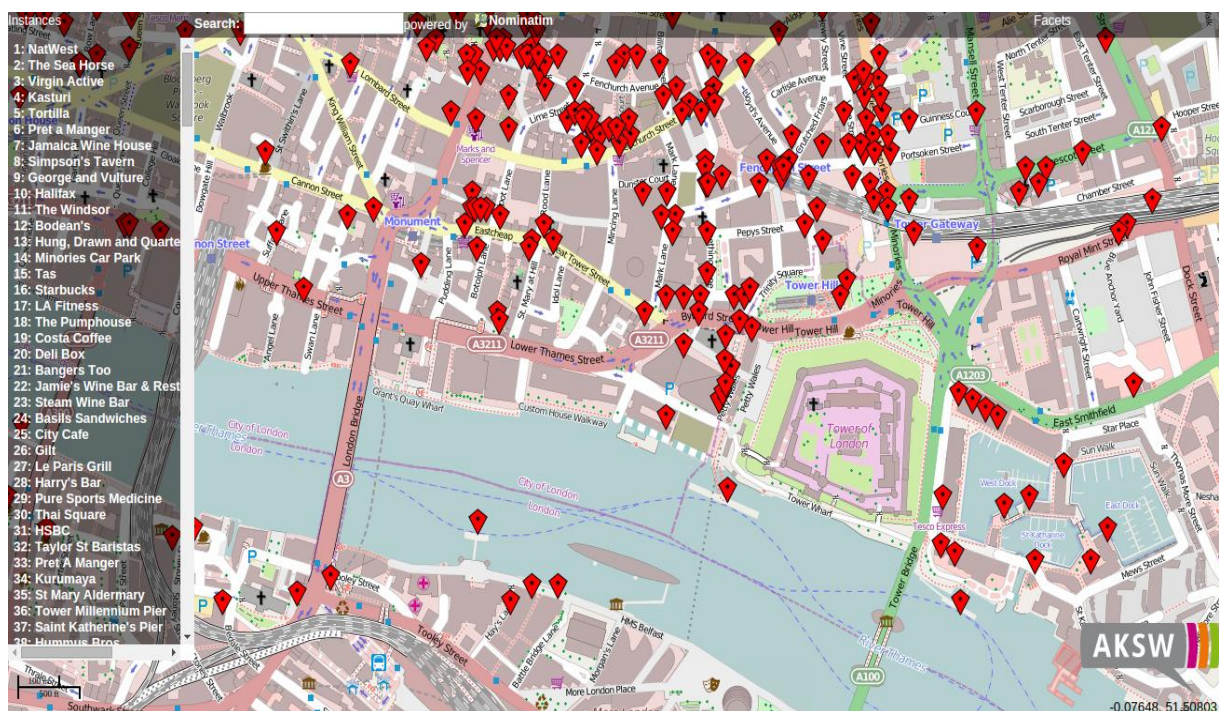


Εικόνα10. Δημοφιλέστεροι όροι αναζήτησης

2.4 LinkedGeoData

Το LinkedGeoData είναι ένα έργο που προσπαθεί να εισάγει μια “χωρική” συνιστώσα στο σημασιολογικό ιστό. Για να το επιτύχει αυτό παίρνει πληροφορίες από το OpenStreetMap, το οποίο είναι μια ανοιχτή συλλογή πληροφοριών που συνδέονται με κάποια τοποθεσία πάνω στον παγκόσμιο χάρτη και παρέχεται από οποιονδήποτε χρήστη του διαδικτύου που επιθυμεί να συνεισφέρει. Τις πληροφορίες λοιπόν αυτές τις αποθηκεύει με δομημένο τρόπο, σύμφωνα με το RDF μοντέλο, και δημιουργεί έναν διαδραστικό και γεμάτο πληροφορία παγκόσμιο άτλα, προσβάσιμο από τον καθένα. Το γεγονός ότι όλη η πληροφορία είναι δομημένη, την καθιστά άμεσα αξιοποιήσιμη από υπολογιστικά συστήματα και εφαρμογές για την εξαγωγή χρήσιμων και σημαντικών αποτελεσμάτων που σε διαφορετική περίπτωση θα απαιτούσαν πολύ παραπάνω χρόνο και πολυπλοκότητα.

Για παράδειγμα μπορούμε να εξάγωγουμε σημαντικά στατιστικά, δημογραφικά και κοινωνικά συμπεράσματα για οποιαδήποτε συγκεκριμένη γεωγραφική τοποθεσία με πολύ εύκολο και αποδοτικό τρόπο.



Εικόνα 11. *LinkedGeoData*, Λονδίνο ποταμός Τάμεσης

3

Θεωρητικό υπόβαθρο

Για την κατανόηση του αντικειμένου της διπλωματικής εργασίας αλλά και για την υλοποίηση του συστήματος αναζήτησης βιβλίων και δημιουργίας δομημένων RDF δεδομένων, χρειάστηκε να έρθουμε σε επαφή με πολλές διαφορετικές τεχνολογίες.

Οι σημαντικότερες μεθοδολογίες – τεχνικές καθώς και κάποια κομμάτια από τη θεωρία που τις υποστηρίζει θα αναλυθούν παρακάτω, ώστε να γίνει πιο εύκολη η κατανόηση του εγχειρήματος από τον αναγνώστη.

Αυτές οι τεχνολογίες έχουν να κάνουν είτε με το χειρισμό και τη δόμηση των δεδομένων, με στόχο το σημασιολογικό εμπλουτισμό τους, είτε με τις μεθόδους αναζήτησης τέτοιων δεδομένων μέσα σε μια Βάση.

3.1 Διαδίκτυο και Δεδομένα

Προκειμένου τα δεδομένα να είναι επαναχρησιμοποιήσιμα, βασική απαίτηση είναι να έχουν σωστή, τυποποιημένη και απόλυτα ορισμένη δομή προκειμένου να μπορούν να αναπτυχθούν εργαλεία και εφαρμογές για την αυτοματοποίηση της επεξεργασίας και διαχείρισης της πληροφορίας.

Στο κλασσικό - υπάρχον διαδίκτυο, η πλειοψηφία των ιστοσελίδων έχουν κάποιας μορφής δομή, όμως η ίδια η γλώσσα περιγραφής τους έχει σχεδιασθεί για τη δόμηση αρχείων κειμένου και όχι δεδομένων. Κάνοντας χρήση της HTML για τη δημιουργία μιας ιστοσελίδας, δυσχεραίνεται η διαδικασία εξαγωγής δομημένης πληροφορίας μέσω αυτοματοποιημένων εφαρμογών λογισμικού μιάς και τα δεδομένα βρίσκονται αναμειγμένα με το κείμενο.

Για να αντιμετωπιστεί το παραπάνω πρόβλημα, έγινε μια προσπάθεια ανάπτυξης μικροδομών (microformats) [3], οι οποίες επέτρεπαν τη δημοσίευση δεδομένων σε μια ιστοσελίδα με παράλληλη δήλωση του τύπου αυτών και ορίζοντας την προκαθορισμένη μορφή με την οποία αυτά θα γραφτούν. Οι μικροδομές όμως παρουσίαζαν πολλούς περιορισμούς, αρχικά οι τύποι δεδομένων που ήταν διαθέσιμοι δεν κάλυπταν όλα τα πιθανά δεδομένα, ενώ και για τους διαθέσιμους τύπους δεδομένων τα πεδία και οι ιδιότητες που τα χαρακτήριζαν ήταν περιορισμένα. Συνεπώς οι μικροδομές αποδεικνύεται ότι δεν μπορούν να καλύψουν την ανάγκη μετάδοσης και διαχείρισης ολόκληρου του φάσματος των δεδομένων μέσα στο διαδίκτυο.

Στη συνέχεια στο προσκήνιο ήρθε μια πιο γενική και ευρεία αντιμετώπιση για τα δεδομένα στο διαδίκτυο, οι διεπαφές των διαδικτυακών εφαρμογών (Web APIs). Οι διεπαφές των διαδικτυακών εφαρμογών παρέχουν τη δυνατότητα ανταλλαγής δομημένων δεδομένων μέσω αιτημάτων και ερωτήσεων με βάση το πρωτόκολλο HTTP. Μεγάλοι διαδικτυακοί τόποι ανέπτυξαν και προσέφεραν διεπαφές για τα δεδομένα τους, μερικά παραδείγματα είναι τα Amazon Product Advertising API, Flickr API και Goodreads API. Μια συνεχώς ενημερωμένη λίστα όλων των διαθέσιμων διεπαφών βρίσκεται στο www.programmableweb.com. Γίνεται κατανοητό ότι οι διεπαφές είναι πολύ χρηστικές για την εξαγωγή δομημένης πληροφορίας, όμως η ύπαρξη ξεχωριστής διεπαφής για κάθε διαφορετική συλλογή δεδομένων, κάνει την ανάπτυξη προγραμμάτων για το συνδυασμό πολλαπλών διαφορετικών πηγών δεδομένων μία ιδιαίτερα κοπιαστική και επίπονη διαδικασία.

Καταλήγουμε λοιπόν πως ούτε οι μικροδομές αλλά ούτε και οι διεπαφες διαδικτυακών εφαρμογών μπορούν να επιτύχουν την επιθυμητή μορφή του διαδικτύου και την ελεύθερη διασύνδεση όλων των δεδομένων του. Όπως θα δείξουμε παρακάτω, ο τρόπος με τον οποίο προσπαθούμε να επιτύχουμε αυτό το αποτέλεσμα, δηλαδή το διαδίκτυο των δεδομένων (Web of Data) ή αλλιώς σημασιολογικό ιστό (Semantic Web), είναι με τις αρχές των διασυνδεδεμένων δεδομένων και τη χρήση του RDF(Resource Description Framework) για την περιγραφή των δεδομένων και των μεταξύ τους σχέσεων. Ουσιαστικά αυτό που επιδιώκουμε είναι η διασύνδεση αντικειμένων - οντοτήτων και όχι απλών κειμένων

3.2 *Linked Data*

Διασυνδεδεμένα δεδομένα είναι μια μεθοδολογία δημοσίευσης και διασύνδεσης δομημένων δεδομένων με βάση τέσσερις κανόνες, τους οποίους εισήγαγε ο εφευρέτης του διαδικτύου Tim Berners-Lee [3]. Οι κανόνες αυτοί αναλύονται παρακάτω και είναι οι εξής :

- Κάνουμε χρήση URIs (Uniform Resource Identifiers) για την ονομασία των διαφόρων αντικειμένων.
- Επιπρόσθετα κάνουμε χρήση HTTP URIs για την ονομασία των διαφόρων αντικειμένων, προκειμένου οι χρήστες να οδηγούνται σε σελίδες με όλες τις σχετικές πληροφορίες για την κατανόηση του αναφερόμενου αντικειμένου.
- Τα παραπάνω HTTP URIs, πρέπει να οδηγούν σε χρήσιμες πληροφορίες διατυπωμένες σε μία από τις τυποποιημένες γλώσσες αναπαράστασης δομημένης πληροφορίας, όπως για παράδειγμα οι RDF και SPARQL που θα αναλυθούν παρακάτω.
- Δημιουργούμε συνδέσεις μεταξύ των διαφόρων αντικειμένων, με χρήση των URI τους προκειμένου να έχουμε την επιδιωκόμενη διασύνδεση δεδομένων στο διαδίκτυο.

Προκειμένου να κατανοήσει κανείς τις παραπάνω πρακτικές – κανόνες, θα πρέπει να έχει καταλάβει την αρχιτεκτονική που διέπει το κλασσικό διαδίκτυο. Οι βασικοί πυλώνες του υπάρχοντος διαδικτύου είναι τα URI ως αναγνωριστικά, το πρωτόκολλο HTTP ως μέσο ανταλλαγής δεδομένων και δημιουργίας αιτημάτων, η HTML ως βασική γλώσσα περιγραφής του διαδικτύου και των επιμέρους ιστοσελίδων και τέλος την ύπαρξη υπερσυνδέσμων

(hyperlinks) για τη διασύνδεσή τους. Οι υπερσύνδεσμοι αποτελούν το μέσο με το οποίο μπορεί ένας χρήστης να μεταπηδά από τα δεδομένα ενός server στα δεδομένα κάποιου άλλου. Ακόμη μέσω των υπερσυνδέσμων κινούνται οι μηχανές αναζήτησης και μας προσφέρουν αποτελέσματα για εξειδικευμένα ερωτήματα. Συνεπώς οι υπερσύνδεσμοι είναι αυτοί που ολοκληρώνουν το διαδίκτυο σε ένα παγκόσμιο δίκτυο πληροφοριών, από τα επιμέρους διασυνδεδεμένα κείμενα, συνεπώς η αρχιτεκτονική του διαδικτύου δείχνει να έχει την κατάλληλη δομή για να υποστηρίξει το παγκόσμιο διαδίκτυο δομημένων και διασυνδεδεμένων δεδομένων. Για το λόγο αυτό τα διασυνδεδεμένα δεδομένα μαζί με τις αρχές και τους κανόνες που τα διέπουν, έρχονται να πατήσουν πάνω στην υπάρχουσα αρχιτεκτονική του διαδικτύου για να προκύψει ο σημασιολογικός ιστός και η ανταλλαγή δεδομένων σε παγκόσμια κλίμακα.

Όπως αναφέρθηκε και παραπάνω, τα διασυνδεδεμένα δεδομένα αποτελούν μια αναπαράσταση της γνώσης, δηλαδή η υπάρχουσα γνώση μπορεί να αναπαρασταθεί με διασυνδεδεμένα δεδομένα. Πιο συγκεκριμένα, αυτή η αναπαράσταση της πληροφορίας, βασίζεται στην ιδέα ότι κάθε δεδομένο μπορεί να αποτελέσει μία οντότητα με συγκεκριμένες ιδιότητες – χαρακτηριστικά και σχέσεις που το συνδέουν με τις υπόλοιπες οντότητες μιας Βάσης. Με αυτόν τον τρόπο θα σταματήσει να υπάρχει η ίδια πληροφορία πολλαπλές φορές ή ακόμα και αν υπάρχει, θα είναι γνωστό ότι αποτελεί την ίδια πληροφορία με διαφορετική απλά διατύπωση. Για παράδειγμα ο όρος Παρίσι θα αναπαριστάται από την οντότητα Παρίσι με χαρακτηριστικά :Πόλη, Πρωτεύουσα και σχέσεις : Πρωτεύουσα της Γαλλίας, Πληθυσμός ... , Βρίσκεται στην Ευρώπη, Βιωτικό επίπεδο ... , κλπ. Σε γενικές γραμμές η ιδέα των διασυνδεδεμένων δεδομένων μπορεί να οδηγήσει στην καλύτερη οργάνωση και αξιοποίηση της υπάρχουσας γνώσης με αποτέλεσμα την καλύτερη, αποδοτικότερη εκμετάλλευσή της και την ευκολότερη διαχείρισή της στο μέλλον. Ο μεγαλύτερος και πιο συχνά χρησιμοποιούμενος όγκος πληροφοριών, βρίσκεται στο διαδίκτυο και αυτός είναι ο λόγος για τον οποίο τα διασυνδεδεμένα δεδομένα έχουν ταυτιστεί εν μέρει με τον σημασιολογικό ιστό. Η τήρηση των αρχών των διασυνδεδεμένων δεδομένων και η αποτύπωση της γνώσης με τη μορφή αυτή, οδηγεί αυτόματα στη δημιουργία ενός γράφου που αναπαριστά όλα τα δεδομένα και τις μεταξύ τους σχέσεις. Η ήδη αναπτυγμένη θεωρία γράφων και οι πολλαπλοί αλγόριθμοι που έχουν καθιερωθεί για τις διάφορες ενέργειες πάνω σε ένα γράφο, π.χ. αναζήτηση, ταξινόμηση, ανανέωση, διαγραφή , επιτρέπουν η διαχείριση δεδομένων πάνω σε γράφο να γίνεται πολύ αποδοτική.

Ο μονοσήμαντος ορισμός μιας οντότητας γίνεται με τη χρήση των URIs ή των URLs, δηλαδή κάθε αντικείμενο έχει και ένα μοναδικό αναγνωριστικό. Αυτό το αναγνωριστικό μπορεί και είναι επιθυμητό να οδηγεί σε κάποια ιστοσελίδα στην οποία και να περιγράφεται το ίδιο το

αντικείμενο μαζί με τις ιδιότητες και τις σχέσεις που το διέπουν ή απλά να του προσδίδει μια μοναδική ταυτότητα. Ως αντικείμενα δεν θεωρούνται μόνο οι πληροφορίες που υπάρχουν σε ιστοσελίδες, αλλά κάθε πραγματικό αντικείμενο που έχει φυσική ή όχι υπόσταση για παράδειγμα άνθρωποι, χώρες, ταινίες, εταιρείες ή ακόμα και τα διάφορα είδη σχέσεων που μπορεί να υπάρχουν μεταξύ τους για παράδειγμα εργάζεται_σε, είναι_φίλος_με .

Εν τέλει καταλήγουμε ότι με τη χρήση των διασυνδεδεμένων δεδομένων πάνω στο υπάρχον διαδίκτυο, μπορούμε να επιτύχουμε ένα διαδίκτυο δεδομένων το οποίο διέπουν κοινοί κανόνες ορισμού και δομής με αποτέλεσμα την εύκολη δημιουργία ενιαίων προγραμμάτων – εφαρμογών για την επεξεργασία, ανταλλαγή και αξιοποίηση των δεδομένων που υπάρχουν στο σύνολο του παγκόσμιου ιστού. Πάνω στο διαδίκτυο αυτής της μορφής έχουν αναπτυχθεί ειδικοί περιηγητές διασυνδεδεμένων δεδομένων (Linked Data browsers) οι οποίοι οδηγούν το χρήστη σε δεδομένα από διαφορετικές πηγές ακολουθώντας συνδέσμους RDF, ακόμη υπάρχουν μηχανές αναζήτησης διασυνδεδεμένων δεδομένων (Linked Data Search engines) που υποστηρίζουν την σύνθετη αναζήτηση σε όλο τον παγκόσμιο ιστό των διασυνδεδεμένων δεδομένων.

3.2.1 Δημοσίευση Διασυνδεδεμένων Δεδομένων (Publishing Linked Data on the Web)

Για την υποστήριξη της σωστής δημοσίευσης και προώθησης διασυνδεδεμένων δεδομένων [3], αναπτύχθηκε ένα απλό και πρακτικό σύστημα αξιολόγησης της ποιότητας των δεδομένων που παρέχει κάποιος ανοικτά για χρήση στο διαδίκτυο. Αυτό το σύστημα ονομάζεται «Σχήμα Αξιολόγησης 5 Αστεριών» - «Five – Star Rating Scheme», σύμφωνα με αυτό υπάρχουν συγκεκριμένες προϋποθέσεις που πρέπει να ικανοποιούν τα δεδομένα ώστε να βαθμολογηθεί και με τα αντίστοιχα αστερία. Τα κριτήρια αυτά είναι τα παρακάτω:

- 1 αστέρι : Τα δεδομένα είναι ανεβασμένα στο διαδίκτυο, ανοικτά για χρήση και σε οποιαδήποτε μορφή.
- 2 αστέρια : Ό,τι ισχύει για το 1 αστέρι και επιπλέον όμως τα δεδομένα έχουν μορφή κατανοητή και αξιοποιήσιμη από ένα υπολογιστικό σύστημα, π.χ. σε αρχείο Excel και όχι φωτογραφία ενός πίνακα δεδομένων.

- 3 αστέρια : Ότι ισχύει για τα 2 αστέρια επιπλέον όμως η μορφή του αρχείου να είναι ανοικτής άδειας, δηλαδή όχι Microsoft Excel αλλά CSV ή OpenOffice Calc.
- 4 αστέρια : Τα δεδομένα πληρούν τις παραπάνω προδιαγραφές και επιπλέον χρησιμοποιούν κάποιο από τα τυποποιημένα ανοικτά μοντέλα της W3C, όπως τα RDF και SPARQL, προκειμένου να μπορεί κάποιος άλλος χρήστης να αναγνωρίσει και να δημιουργήσει διασυνδέσεις με τα δεδομένα αυτά.
- 5 αστέρια : Ισχύουν όλα τα παραπάνω και επιπλέον τα δεδομένα αυτά εμπεριέχουν εξερχόμενους συνδέσμους προς άλλα ανοικτά δεδομένα, προκειμένου να επιτυγχάνεται μεγαλύτερη συνοχή μέσα στο συνολικό δίκτυο των Διασυνδεδεμένων Δεδομένων.

Με την σταδιακή συμπλήρωση και των πέντε αστεριών για όλα τα δεδομένα που υπάρχουν στο διαδίκτυο, θα έχουμε και αντίστοιχη σταδιακή μετάβαση από το υπάρχον διαδίκτυο στο Σημασιολογικό Διαδίκτυο που θα εμπεριέχει Διασυνδεδεμένα Δεδομένα που θα ικανοποιούν τις παραπάνω προδιαγραφές.

3.2.2 Χρήση Διασυνδεδεμένων Δεδομένων (Consuming Linked Data on the Web)

Η ιδέα των Διασυνδεδεμένων Δεδομένων καθώς και οι αρχές που τα διέπουν είχαν αρχικά προταθεί το 2006, συνεπώς σήμερα υπάρχει ένα πολύ στιβαρό υπόβαθρο και ήδη πολλοί μεγάλοι οργανισμοί και εταιρείες διαθέτουν ανοικτά τα αρχεία τους σύμφωνα με τις αρχές αυτές. Μερικά τέτοια παραδείγματα είναι το BBC, το Reuters, οι New York Times και η Κυβέρνηση των ΗΠΑ. Τα δεδομένα που υπάρχουν αγγίζουν ένα πολύ ευρύ φάσμα, από ταινίες, μουσική και ραδιόφωνο, μέχρι επιστημονικές δημοσιεύσεις, στατιστικά δεδομένα, γεωγραφικά δεδομένα και στοιχεία κρατών και εταιρειών.

Ο συνδυασμός και η άθροιση δεδομένων από διαφορετικούς παρόχους γύρω από το ίδιο αντικείμενο, είναι μια αρκετά απλή διαδικασία που οδηγεί σε μια πιο ολοκληρωμένη εικόνα και σε πιο σωστά αποτελέσματα σε κάθε είδους αναζήτηση.

Έχουν αναπτυχθεί πολλές μέθοδοι για την τοποθέτηση και εκτέλεση δομημένων, σύνθετων ερωτημάτων (queries) πάνω σε δομές που περιέχουν Διασυνδεδεμένα Δεδομένα. Τα δεδομένα προέρχονται από πολλαπλές πηγές, διανεμημένες τόσο από φυσικής άποψης, διαφορετικές γεωγραφικές τοποθεσίες, όσο και από τεχνικής άποψης, ο καθένας μπορεί να

προσθέσει και να διαθέσει τα δεδομένα του στο διαδίκτυο. Αυτή η πολυμορφία των πηγών των δεδομένων δημιουργεί μια σειρά προβλημάτων – προκλήσεων, για την επιστημονική κοινότητα που ασχολείται με την προαγωγή και τη βελτίωση των Διασυνδεδεμένων Δεδομένων για χρήση στο διαδίκτυο. Αυτά τα προβλήματα στη χρήση και στην αξιοποίηση των Διασυνδεδεμένων Δεδομένων του διαδικτύου, μπορούν να χωριστούν σε μερικά διακριτά πεδία, αυτά είναι :

- Αξιοπιστία πηγής παροχής δεδομένων
- Ενημερωμένα – επικαιροποιημένα δεδομένα
- Ταχύτητα εξυπηρέτησης αιτήματος – ερωτήματος
- Ακρίβεια δεδομένων

Για την αντιμετώπιση λοιπόν των προβλημάτων αυτών έχουν αναπτυχθεί διάφορα συστήματα διαχείρισης των δεδομένων που υπάρχουν και είναι διαθέσιμα μέσω του διαδικτύου. Κάθε τακτική διαχείρισης των δεδομένων αντιμετωπίζει καλύτερα και κάποια ή κάποιες από τις δυσκολίες που υπάρχουν, όμως πάντα πρέπει να γίνεται επιλογή με βάση τις εκάστοτε υπάρχουσες ανάγκες που πρέπει να καλυφθούν. Σε όλες αυτές τις τεχνικές διαχείρισης των δεδομένων του διαδικτύου και κατ'επέκταση του Σημαιολογικού Ιστού, θα αναφερθούμε παρακάτω.

Αρχικά η πλειοψηφία των παρόχων δεδομένων διαθέτουν ένα συνολικό αρχείο RDF (RDF Dump) που περιέχει ολόκληρο το διασυνδεδεμένο γράφο των δεδομένων τους για την διευκόλυνση τοποθέτησης ερωτημάτων. Μέσω μιας υπηρεσίας παρόμοιας με το πρωτόκολλο HTTP μπορεί κανείς να θέσει κάποιο SPARQL ερώτημα προς όλα τα δεδομένα, με βάση το δοσμένο γράφο και τους υπογράφους που δομούνται μέσω των διασυνδέσεων των δεδομένων του, σε ένα νοητό σημείο που καλείται SPARQL endpoint. Οι εφαρμογές λοιπόν χρησιμοποιούν αυτό το σημείο για να έχουν πρόσβαση σε όλα τα δεδομένα στα οποία 'βλέπει' αυτό. Με αυτή την μέθοδο συνεπώς έχουμε πρόσβαση σε μια μεγάλη ίσως ομάδα δεδομένων αλλά δεν εκμεταλλευόμαστε ολόκληρο τον διαθέσιμο όγκο δεδομένων που υπάρχουν ανοικτά με αυτή την τυποποιημένη μορφή. Πρέπει λοιπόν να υπάρξει μια ένωση όλων των σημείων αυτών έτσι ώστε ένα ερώτημα να τίθεται προς τον μέγιστο δυνατό όγκο πληροφορίας, προκειμένου να λάβουμε την πιο ολοκληρωμένη και σωστή απάντηση.

Πολλοί έχουν εξετάσει τη χρήση μεθόδων που χρησιμοποιούνται και στις κλασσικές βάσεις δεδομένων για την ταχύτερη εξυπηρέτηση αιτημάτων σε αυτές. Τέτοιες μέθοδοι τυχαίνει να

είναι πολύ αποδοτικές όμως δεν αξιοποιούν στο έπακρο τις δυνατότητες που μας παρέχουν τα Διασυνδεδεμένα Δεδομένα στο διαδίκτυο. Ένας πίνακας με τις μεθόδους αυτές, καθώς και τα θετικά και τα αρνητικά στοιχεία καθεμιάς από αυτές παρατίθεται παρακάτω. Οι επιμέρους μέθοδοι δεν θα αναλυθούν εκτενώς στα πλαίσια της παρούσας διπλωματικής.

	data warehousing	search engines	query federation	active discovery query federation	link traversal
Universe of Discourse (UoD):	loaded data	Web of Data	known data sources	Web of Data	Web of Data
Required source interfaces:	mainly RDF dumps	arbitrary	SPARQL endpoints	SPARQL endpoints	Linked Data (look-up) interface
Access to original data:	no	no	yes	yes	yes
Supporting data structures:	indices and statistics	crawled index	statistics	(statistics)	-
Response and throughput:	fast / fast	fast / fast	slow / medium	slow / slow	medium / slow
Recall (w.r.t. UoD):	100%	<100%	100%	<100%	<100%
Precision:	100%	<100%	100%	100%	100%
Up-to-dateness:	low	medium	high	high	high
	hybrid approaches				

Εικόνα 12. Μέθοδοι εξυπηρέτησης ερωτημάτων προς Διασυνδεδεμένα Δεδομένα

Έχουν προταθεί όμως και κάποιες μέθοδοι, οι οποίες έχουν αναπτυχθεί ειδικά για εξυπηρέτηση ερωτημάτων σε Διασυνδεδεμένα Δεδομένα στο διαδίκτυο. Κάποιες από αυτές που βρίσκονται ακόμη σε ερευνητικό στάδιο είναι οι παρακάτω [4]:

- **Ενεργή Αναζήτηση Συλλογής Ερωτημάτων (Active Discovery based Query Federation) :** Η ιδέα πίσω από αυτή την προτεινόμενη μέθοδο είναι η εξής, το αρχικό σύνθετο ερώτημα διαιρείται σε επιμέρους απλούστερα ερωτήματα τα οποία αναζητούν απάντηση σε όλα τα πιθανά SPARQL endpoints. Με αυτό τον τρόπο έχουμε αξιοποίηση των διαθέσιμων δεδομένων σε μεγάλο βαθμό, με παράλληλη βελτιστοποίηση του χρόνου εκτέλεσης με την απλοποίηση του αρχικού πολύπλοκου ερωτήματος.
- **Εκτέλεση Ερωτημάτων με Διάσχιση Συνδέσμων (Linked Traversal based Query Execution) :** Αναλύοντας τα URIs του δοσμένου ερωτήματος αρχίζουμε και διατρέχουμε το σύνολο των Διασυνδεδεμένων Δεδομένων του διαδικτύου, ακολουθώντας όλους τους συνδέσμους που συναντάμε με σκοπό την εξεύρεση και αξιοποίηση κάθε πιθανής πληροφορίας. Αυτή η μέθοδος κάνει χρήση όλης της υπάρχουσας πληροφορίας, όμως όπως γίνεται κατανοητό έχει και αντίστοιχα επιβαρυνμένο κόστος στην εκτέλεσή της

Κλείνοντας μπορεί κανείς να κατανοήσει πως απέχουμε αρκετά από την επίτευξη του στόχου της δημιουργίας και αξιοποίησης ενός παγκόσμιου, ανοιχτού δικτύου διασυνδεδεμένων δεδομένων, μιάς και συναντάμε ακόμη πολλά προβλήματα που δεν επιτρέπουν την αποδοτική χρήση του τεράστιου αυτού όγκου πληροφοριών.

3.3 Semantic Web

Το σημασιολογικό διαδίκτυο (semantic web) είναι η υλοποίηση της ιδέας του Tim Berners-Lee για την εξέλιξη του διαδικτύου σε μια μορφή αναγνώσιμη από τους υπολογιστές. Αυτή η μορφή του διαδικτύου μπορεί να υλοποιηθεί με χρήση της τεχνολογίας και των τεχνολογιών των διασυνδεδεμένων δεδομένων, με εφαρμογή τους στα δεδομένα που ήδη υπάρχουν στο διαδίκτυο. Πιο συγκεκριμένα κάθε πληροφορία που “ανεβαίνει” στο διαδίκτυο θα αποτελεί και μια οντότητα που θα ορίζεται από ένα URI και πλέον θα αποτελεί κομμάτι του σημασιολογικού ιστού, συνεπώς θα αποκτά ταυτότητα, ιδιότητες και θα συνδέεται με σχέσεις με τα υπόλοιπα αντικείμενα του σημασιολογικού ιστού. Σε ένα διαδίκτυο με αυτά τα χαρακτηριστικά η πληροφορία και οι μεταπληροφορίες της θα αυξάνουν την αξία τους καθώς θα μπορούν να αξιοποιηθούν και να προσπελασθούν με πολλαπλούς και συνεχώς αυξανόμενους τρόπους που στο σημερινό χαοτικό και άναρχο διαδίκτυο πληροφοριών είναι αδύνατο να συμβεί. Όλος ο όγκος δεδομένων που υπάρχει θα αρχίσει να αποτελεί πραγματική γνώση, αξιοποιήσιμη, επεκτάσιμη, λειτουργική και χρηστική. Προς αυτόν τον κοινό και ενιαίο τρόπο αναπαράστασης της πληροφορίας έχουν γίνει πολλές προσπάθειες που μας παρέχουν την απαιτούμενη προτυποποίηση.

Για παράδειγμα το schema.org περιέχει “σχήματα”, οντολογίες για την εισαγωγή μεταπληροφορίας με κοινό τρόπο ώστε να περιορίζεται η πολυμορφία στον τρόπο με τον οποίο δηλώνουμε την πληροφορία και τα διάφορα χαρακτηριστικά της. Για παράδειγμα, αν θέλουμε να αναφερθούμε σε πληροφορία σχετική με βιβλία, έχει αναπτυχθεί το “σχήμα” book-schema που μας δίνει έτοιμα τα διάφορα πεδία μεταπληροφορίας που μπορεί να χρειαστούμε. Με την χρήση αυτών των έτοιμων και υποστηριζόμενων σχημάτων, η πληροφορία μας αυξάνει την αξία της καθώς γίνεται κατανοητή και αποδεκτή από περισσότερους ανθρώπους αλλά και μηχανές αναζήτησης.

Έπειτα για την περιγραφή με μοναδικό τρόπο κάποιας οντότητας, χρειαζόμαστε και κάποιο αναγνωριστικό το οποίο θα αναφέρεται πάντα σε αυτή τη συγκεκριμένη οντότητα. Για αυτόν το σκοπό αξιοποιήθηκαν τα URIs (Uniform Resource Identifiers) και τα URLs (Uniform Resource Locators) ώστε να “δείχνουν ” σε ένα μοναδικό και συγκεκριμένο αντικείμενο γνώσης. Για τη διατήρηση και την καλύτερη αξιοποίηση των αναγνωριστικών αυτών έχει δημιουργηθεί το PURL (Persistent Uniform Resource Locator). Το PURL είναι μια συλλογή από αναγνωριστικά οντοτήτων URLs/URIs τα οποία παραμένουν και εξακολουθούν να είναι συνδεδεμένα με συγκεκριμένες οντότητες παρά τις όποιες αλλαγές μπορεί να συμβούν στις οντότητες ή στις βάσεις δεδομένων που φιλοξενούν τη σχετική πληροφορία. Ακόμη το PURL έχει και ιστοσελίδα που παρέχει στους χρήστες την υπηρεσία να κάνουν αναζήτηση για την εύρεση του αναγνωριστικού που τους ενδιαφέρει ή να δημιουργήσουν και να δηλώσουν ένα νέο αναγνωριστικό.

3.4 Vocabularies

Για την ευκολότερη, καλύτερη και ενιαία σημασιολογική σήμανση των δεδομένων , έχουν αναπτυχθεί πολλά λεξικά ανά τομέα. Ένα λεξικό παρέχει μια προτυποποίηση για τον τρόπο με τον οποίο ορίζεται κάθε οντότητα ανάλογα με τον τύπο του δεδομένου που αναπαριστά. Ορίζει τις καθορισμένες ιδιότητες, τα χαρακτηριστικά και τις σχέσεις που διέπουν μια οντότητα αυτού του τύπου. Έχουν αναπτυχθεί λεξικά για τον ορισμό πολλών διαφορετικών ειδών δεδομένων ώστε να επιτυγχάνεται η μεγαλύτερη δυνατή ομοιομορφία στον σημασιολογικό ιστό. Πολλές φορές γίνεται αναφορά στα λεξικά με τον όρο οντολογίες, η διαφοροποίηση μεταξύ τους είναι πολύ λεπτή και γι' αυτό οι δύο αυτοί όροι έχουν ταυτιστεί.

Ακόμη έχουν δημιουργηθεί ιστοσελίδες όπως το lov.okfn.org (Linked Open Vocabularies), όπου παρέχουν στους χρήστες τη δυνατότητα αναζήτησης για την εύρεση του λεξικού ή των λεξικών που ταιριάζουν καλύτερα στον τύπο των δεδομένων στον οποίο αναφέρονται. Συνεπώς πλέον είναι εύκολο αλλά και απαραίτητο κάθε χρήστης ή προγραμματιστής που επιθυμεί να εισάγει σημασιολογία στα δεδομένα του, να χρησιμοποιήσει κάποιο από τα ήδη υπάρχοντα και ευρέως αναγνωρίσιμα λεξικά προκειμένου να αυξάνεται η συνοχή των δεδομένων ολοκληρου του Σημασιολογικού Ιστού.

Τα λεξικά λοιπόν αποτελούν βασικό εργαλείο για την επίτευξη του στόχου ενός συνεκτικού, χρηστικού και εύκολα προσπελάσιμου δικτύου διασυνδεδεμένων δεδομένων που θα περιέχει όλη την παγκόσμια γνώση.

3.5 Μορφές Κωδικοποίησης Δεδομένων (Serialization Formats)

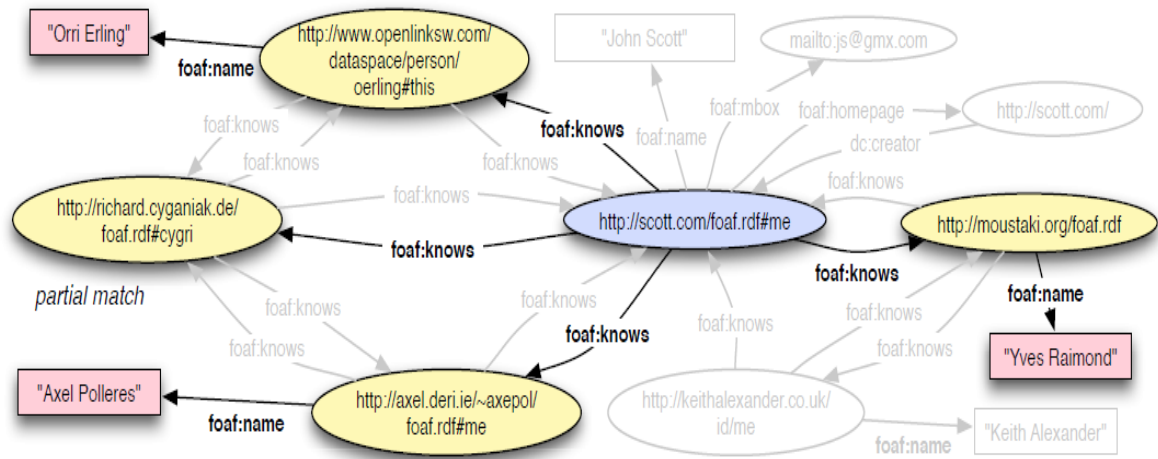
Τα δεδομένα που δημιουργούμε ή εξάγουμε, προκειμένου να αποθηκευτούν σε μία βάση δεδομένων θα πρέπει να έρθουν σε μία κατάλληλη μορφή. Αυτή η μορφή αποτελεί μια κωδικοποίηση των δεδομένων που επιτρέπει την αποθήκευση, μεταφορά και επαναδημιουργία των δεδομένων σε μορφή αναγνωρίσιμη από τον άνθρωπο (αποκωδικοποίηση). Έχουν αναπτυχθεί πολλές διαφορετικές κωδικοποιήσεις για την επίτευξη των παραπάνω σκοπών, καθεμία από τις οποίες έχει και διαφορετικά χαρακτηριστικά, πλεονεκτήματα και μειονεκτήματα. Πιο συγκεκριμένα για την εξυπηρέτηση των αναγκών που δημιουργούνται για την καλύτερη λειτουργία και αξιοποίηση του σημασιολογικού ιστού, αναπτύχθηκαν πολλές κωδικοποιήσεις. Τις σημαντικότερες από αυτές θα αναλύσουμε παρακάτω.

Θα αναφερθούμε στις κωδικοποιήσεις : RDF-Triples, Turtle, JSON και JSON-LD.

3.5.1 RDF- Triples

Το RDF είναι ένας τρόπος αναπαράστασης της μεταπληροφορίας των δεδομένων μας, κάθε χαρακτηριστικό ή σχέση μίας οντότητας περιγράφεται από μια τριάδα, η οποία αποτελείται από Υποκείμενο-Ρήμα-Αντικείμενο. Με τον τρόπο αυτό διευκολύνεται η αναζήτηση και η εύρεση των επιθυμητών οντοτήτων, ένας χρήστης μπορεί να κάνει μια αναζήτηση υποβάλλοντας μια τριάδα ή έναν συνδυασμό από τριάδες, τον οποίο απαιτεί να ικανοποιείται. Το RDF μοντέλο αποτελεί ένα μέσο για την εύκολη και άμεση αναπαράσταση της πληροφορίας μας σε ένα γνωσιακό γράφο με όλα τα χαρακτηριστικά και τις σχέσεις που διέπουν τα δεδομένα μας. Η δυνατότητα αυτή, οπτικοποίησης των δεδομένων, είναι πολύ

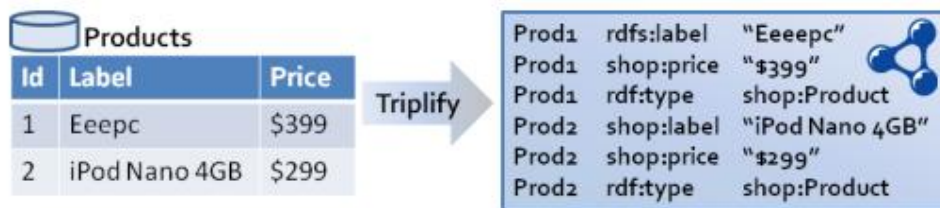
ουσιαστική και χρήσιμη καθώς κάνει εύκολη την παρατήρηση συγκεκριμένων χαρακτηριστικών ενός συνόλου δεδομένων με απλή εποπτεία του γράφου που έχει παραχθεί.



Εικόνα13.Οπτικοποίηση ενός Γράφου RDF δεδομένων

Μια βάση δεδομένων που περιέχει εξολοκλήρου δεδομένα σε RDF μορφή, καλείται RDF Triple-Store δηλαδή μια “αποθήκη τριάδων”.

Το RDF είναι λοιπόν ένας ευρέως διαδεδομένος τρόπος αναπαράστασης μεταπληροφορίας και για το λόγο αυτό έχουν αναπτυχθεί και διάφορα εργαλεία για την αυτοματοποιημένη δημιουργία – μετατροπή δεδομένων σε RDF. Ένα τέτοιο εργαλείο είναι το Triplify και έρχεται να δουλέψει είτε σαν πρόσθετο σε μια υπάρχουσα εφαρμογή είτε σαν αυτόνομο πρόγραμμα. Το Triplify δέχεται τα δεδομένα μια Βάσης και αναλαμβάνει να εξάγει και να μορφοποιήσει τη σημασιολογική πληροφορία των δεδομένων, παράγοντας RDF ή JSON δεδομένα. Συνεπώς το Triplify παρέχει έναν αυτοματοποιημένο τρόπο μετατροπής δεδομένων σε μορφή κατάλληλη για τον σημασιολογικό ιστό.



Εικόνα 14. Η βασική λειτουργία του Triplify

3.5.2 Turtle

Η Turtle είναι μια αναπαράσταση μέσω απλού κειμένου (plain text) του μοντέλου των RDF Triples. Δηλαδή διευκολύνει την αποτύπωση και ενισχύει την αναγνωσιμότητα μεταδεδομένων με μορφή τριάδων RDF. Δίνει τη δυνατότητα αποθήκευσης και χρήσης συντομεύσεων για την αποφυγή της επανάληψης μεγάλων αναγνωριστικών (URIs, URLs), μας επιτρέπει να ορίζουμε πολλαπλές ιδιότητες του ίδιου υποκειμένου χωρίς να χρειάζεται επανορισμός του με τη χρήση semicolon, “ ; ” μας προσφέρει ακόμη τη δυνατότητα να δηλώσουμε ένα ανώνυμο όρισμα σε μια τριάδα. Αυτές είναι μερικές μόνο από τις διευκολύνσεις που μας παρέχει η Turtle για τη συγγραφή και αποθήκευση RDF δεδομένων.

3.5.3 JSON & JSON-LD

Η JSON (JavaScript Object Notation) είναι ένας δομημένος και οργανωμένος τρόπος αποθήκευσης δεδομένων, με τη μορφή ιδιοτήτων – χαρακτηριστικών του οριζόμενου αντικειμένου – οντότητας. Αρχικά κατασκευάστηκε ως εναλλακτική στη χρήση της XML, η JSON λειτουργεί ανεξαρτήτως γλώσσας προγραμματισμού (Language Independent). Έχει στόχο την διευκόλυνση των χρηστών παρέχοντας μια πιο αναγνώσιμη μορφή στον τρόπο με τον οποίο ορίζονται τα δεδομένα από αυτούς. Ένα αντικείμενο σε JSON λοιπόν, ορίζεται από μια μεταβλητή – όνομα και τις ιδιότητές του με τις αντίστοιχες τιμές αυτών. Χρησιμοποιώντας το όνομα του αντικειμένου ακολουθούμενο από μια ιδιότητά του, έχουμε πρόσβαση στην τιμή αυτής. Ακόμη έχουμε τη δυνατότητα αποθήκευσης και προσπέλασης πινάκων με δεδομένα σε JSON, δηλαδή ενός πίνακα αντικειμένων με ιδιότητες.

Αντίστοιχα η JSON-LD (JavaScript Object Notation for Linked Data) είναι μια ακόμη μορφή αναπαράστασης διασυνδεδεμένων δεδομένων η οποία χαρακτηρίζεται για την ταχύτητά της και την αναγνωσιμότητα και την ευχρηστία της για τους ανθρώπους. Έχει βασιστεί πάνω στη JSON που αναφέραμε στην παραπάνω παράγραφο και έχει στόχο την εύκολη μετατροπή των δεδομένων με μορφή JSON σε μία νέα που θα είναι κατάλληλη για χρήση με διασυνδεδεμένα δεδομένα στο διαδίκτυο.

3.6 SPARQL

SPARQL (SPARQL Protocol And RDF Query Language) είναι μια γλώσσα τοποθέτησης ερωτημάτων σε βάσεις δεδομένων με δεδομένα σε RDF μορφή. Η SPARQL ορίστηκε ως πρότυπο από την RDF DAWG (Data Access Working Group) του W3C και θεωρείται μια από τις βασικές τεχνολογίες του Σημασιολογικού Ιστού. Με χρήση SPARQL ένας χρήστης έχει τη δυνατότητα να εκτελέσει όλες τις δυνατές αναζητήσεις- ερωτήματα (queries) πάνω σε μια αποθήκη τριάδων (triple store), όπως θα έκανε κάποιος με χρήση π.χ. SQL πάνω σε μία κλασσική σχεσιακή βάση δεδομένων. Η SPARQL κάνει χρήση της λογικής των τριάδων που χαρακτηρίζει τα RDF δεδομένα και γι' αυτόν τον λόγο τα ερωτήματα (queries) που θέτουμε έχουν τη μορφή τριάδων, μέσω των οποίων προσπαθούμε να βρούμε “ταιριάσματα” μέσα στη Βάση που κάνουμε αναζήτηση. Τα ερωτήματα που θέτουμε, SPARQL queries, τοποθετούνται – υποβάλλονται έναντι μιας διεύθυνσης μιας Βάσης, η διεύθυνση αυτή ονομάζεται SPARQL endpoint και αποτελείται από ένα URI. Μπορούμε ακόμη να θέσουμε ερώτημα με διαμοιρασμένη αναζήτηση (federated search), δηλαδή ένα ερώτημα το οποίο θα τοποθετηθεί έναντι πολλών Βάσεων και θα επιστρέψει ένα σύνολο αποτελεσμάτων – ταιριασμάτων.

Στην παρακάτω εικόνα βλέπουμε ένα παράδειγμα υποβολής ερωτήματος σε SPARQL.

```

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name ?email
WHERE {
    ?person a foaf:Person.
    ?person foaf:name ?name.
    ?person foaf:mbox ?email.
}

```

Εικόνα 14.

Στο συγκεκριμένο παράδειγμα, αναζητούμε δεδομένα που να είναι τύπου Person, έχουν ένα πεδίο name και ένα πεδίο email. Μετά την υποβολή του παραπάνω ερωτήματος θα έχουμε μια σειρά από ταιριάσματα URI στη μεταβλητή person, ονομάτων στη μεταβλητή name και ηλεκτρονικών διευθύνσεων στη μεταβλητή email. Ενώ θα παρουσιαστούν μόνο οι μεταβλητές name και email, καθώς αυτές έχουν οριστεί στο SELECT κομμάτι του query.

Βασικό δομικό στοιχείο της SPARQL είναι το BGP (Basic Graph Pattern), με αρχή αυτό το απλουστευμένο ερώτημα (query), μπορεί κανείς να συνθέσει πιο πολύπλοκα και σύνθετα ερωτήματα για να βρει όποιο δεδομένο αναζητά. Η απλουστευμένη αυτή μορφή τριάδας (υποκείμενο – ρήμα – αντικείμενο) φαίνεται στο επόμενο σχήμα.



Εικόνα 15. Σχηματική απεικόνιση μέσω γράφου ενός ερωτήματος BGP με δύο μεταβλητές.

3.7 Linked Data Fragments

Τα Linked Data Fragments έρχονται να εξισορροπήσουν τα προβλήματα που παρατηρούνται όταν ακολουθούμε μία από τις δύο συμβατικές μεθόδους ικανοποίησης των SPARQL queries. Τα προβλήματα αυτά συνοπτικά είναι στην μία περίπτωση αυξημένη καθυστέρηση και περιορισμένη εξυπηρέτηση του χρήστη όταν το SPARQL endpoint βλέπει ολόκληρο το triple-store, ενώ στην άλλη απώλεια της ενημερωμένης εικόνας των δεδομένων όταν αποφασίσουμε να αποθηκεύσουμε τοπικά το triple-store. Τα Linked Data Fragments, έρχονται να μοιράσουν το φορτίο εργασίας ανάμεσα στον server και το χρήστη, ζητώντας ουσιαστικά από τον χρήστη να τοποθετεί τα ερωτήματά του , queries, με πιο έξυπνο και αποδοτικό τρόπο. Η λογική των Linked Data Fragments ουσιαστικά υποστηρίζει ότι αντί να τοποθετούμε ερωτήματα σε ολόκληρο το triple-store, τα τοποθετούμε έναντι ενός μέρους του triple-store το οποίο ορίζεται από μια άλλη τριάδα υποκείμενο-ρήμα-αντικείμενο ,

π.χ. subject : “”

predicate:”<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>”

object:”<http://dbpedia.org/ontology/Restaurant>”

,με την τριάδα του παραδείγματος ορίζεται το Linked Data Fragment το οποίο βλέπει μόνο όλα τα αντικείμενα τύπου “Restaurant” με συνέπεια να αυξάνει κατά πολύ την αποκρισμότητα του server και περιορίζει σε έναν προκαθορισμένο βαθμό τα αντικείμενα έναντι των οποίων θα υποβάλλουμε τα ερωτήματά μας.

Τα Linked Data Fragments είναι μια πιο πρόσφατη ιδέα και σχετίζεται με τον τρόπο με τον οποίο υποβάλλεται ένα ερώτημα (SPARQL Query) σε κάποια βάση διασυνδεδεμένων δεδομένων. Ένα SPARQL Query είναι ένα απλό ή σύνθετο query που υποβάλλει ένας χρήστης σε μια βάση που περιέχει δεδομένα σε RDF μορφή, δηλαδή σε μορφή τριάδων (Subject – Predicate – Object). Ουσιαστικά με τα data fragments ο χρήστης είναι υποχρεωμένος να περιορίσει την αναζήτηση που χρειαστεί να εκτελεστεί στη βάση και αυτό το πετυχαίνει παρέχοντας κάθε φορά τουλάχιστον ένα από τα τρία μέρη της κάθε τριάδας του ερωτήματος που υποβάλλει στη βάση.

3.8 RDF Databases

Για την καλύτερη διαχείριση των δεδομένων που είναι κωδικοποιημένα σε RDF με μορφή τριάδων, απαιτείται η αποθήκευσή τους με τον κατάλληλο τρόπο. Για αυτό το σκοπό οι κλασσικές σχεσιακές βάσεις δεδομένων δεν μας εξυπηρετούν, για αυτό λοιπόν το λόγο αναπτύχθηκαν οι Βάσεις δεδομένων RDF. Δηλαδή βάσεις δεδομένων οι οποίες έχουν σχεδιαστεί για να αποθηκεύουν τριάδες RDF (υποκείμενο – ρήμα – αντικείμενο) και όλες οι λειτουργίες πάνω στα δεδομένα τους, γίνονται με χρήση της SPARQL. Έτσι λοιπόν έχουμε μια νέα μορφή βάσεων δεδομένων η οποία συμμορφώνεται με τις προδιαγραφές που θέτει η διεθνής κοινοπραξία W3C και ανήκει στις μη σχεσιακές βάσεις, NoSQL. Οι βάσεις του τύπου που μόλις περιγράφηκε, διέπονται από κάποια κοινά χαρακτηριστικά που τις κάνουν ιδιαίτερα λειτουργικές, αυτά είναι συνοπτικά τα εξής :

- Φορητότητα δεδομένων : Ευκολία στη μεταφορά δεδομένων από τη μία βάση σε μια άλλη καθώς και το συνδυασμό αποτελεσμάτων από διαφορετικές βάσεις. Αυτό οφείλεται στο γεγονός ότι γίνεται χρήση προτυποποιημένων μεθόδων για όλες τις ενέργειες των βάσεων αυτών.
- Διαλειτουργικότητα προγραμματιστικών εργαλείων : Έχουν αναπτυχθεί προγραμματιστικά εργαλεία σε πολλές διαφορετικές προγραμματιστικές γλώσσες όπως Python, Java και Ruby. Όλα όμως τα εργαλεία αυτά δημιουργούν και ενεργούν πάνω σε βάσεις δεδομένων οι οποίες είναι εν τέλει συμβατές μεταξύ τους. Συνεπώς δεν είναι απαραίτητη η γνώση παραπάνω του ενός εργαλείου σε οποιαδήποτε γλώσσα προγραμματισμού προκειμένου να κανείς να μπορεί να αναπτύξει και να διαχειριστεί μια τέτοια RDF βάση δεδομένων.
- Ευκολία μεταπήδησης από μία βάση σε μία καινούρια : Δεν δημιουργείται μεγάλο πρόβλημα, αν προκύψει η ανάγκη αλλαγής της βάσης και μεταπήδησης σε μια βάση άλλης εταιρείας. Επιπλέον, έχουν αναπτυχθεί ‘οικουστήματα’ χρηστών που μπορούν να βοηθήσουν σε αυτή την αλλαγή και να επιλύσουν τα όποια προβλήματα πιθανόν να παρουσιαστούν.
- Μελλοντική εγκυρότητα των δεδομένων : Το γεγονός ότι οι βάσεις αυτές έχουν φτιαχτεί πάνω στις αρχές του RDF, που είναι το πρότυπο για τη διαχείριση των Διασυνδεδεμένων Δεδομένων, και στηρίζονται σε δομικά στοιχεία όπως τα URIs τις κάνουν εξαιρετικά αξιόπιστες ως προς το γεγονός ότι τα δεδομένα τους θα είναι συμβατά και στο μέλλον. Δίνοντας με αυτόν τον τρόπο μεγαλύτερη αξία στα δεδομένα μας και μεγαλύτερη διαχρονικότητα στην ίδια τη βάση δεδομένων.

4

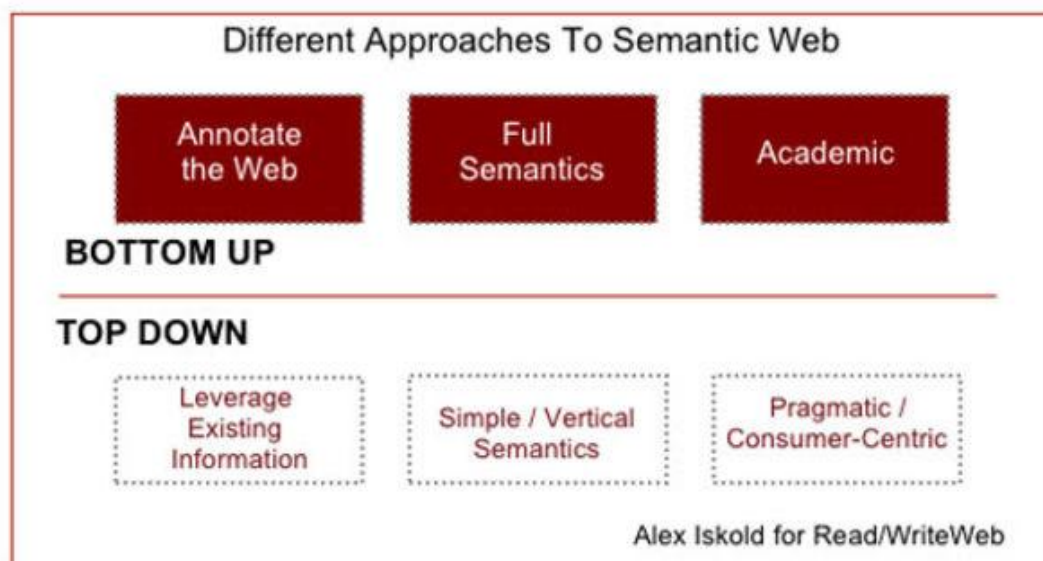
Εφαρμογές με χρήση του Σημασιολογικού Ιστού

Το βασικό χαρακτηριστικό κάθε εφαρμογής που έχει αναπτυχθεί για χρήση στον Σημασιολογικό Ιστό, είναι πως κάνουν χρήση δεδομένων που περιέχονται σε αυτόν και είναι συνεπώς διαθέσιμα σε μορφή RDF. Τα δεδομένα αυτά, τα χρησιμοποιούν οι εφαρμογές με τρόπο τέτοιο που να αξιοποιεί τη σημασιολογική διάσταση που διαθέτουν και κατά συνέπεια προσφέρουν στους χρήστες τους δυνατότητες οι οποίες στο δεδομένο διαδίκτυο θα ήταν αδύνατες.

Η ανάπτυξη εφαρμογών για το Σημασιολογικό Ιστό μπορεί να γίνει ακολουθώντας μία από τις δύο ευδιάκριτες στρατηγικές [5] που εμφανίζονται κατά την υλοποίηση οποιασδήποτε εφαρμογής, αυτές είναι οι Bottom-up και Top-down τεχνικές. Ξεκινώντας κανένας μπορεί να κατανοήσει πως ο Σημασιολογικός Ιστός έχει σχεδιαστεί σαν ένα επίπεδο πάνω από το υπάρχον διαδίκτυο, με αυτή τη λογική η στρατηγική του bottom-up είναι αυτή που προκύπτει ως η πλέον ταιριαστή.

Bottom-up σχεδίαση σημαίνει ότι βασιζόμενοι στην υπάρχουσα πληροφορία του διαδικτύου προσπαθούμε να εισάγουμε τα νέα στοιχεία που χρειάζονται, δηλαδή την μεταπληροφορία σε RDF. Όμως με τη bottom-up σχεδίαση δημιουργείται μια σειρά προβλημάτων, όπως η πολυπλοκότητα των δημιουργούμενων γράφων, η δυσκολία ανάγνωσης και κατανόησης των RDF και OWL από τους ανθρώπους, η ανάγκη εισαγωγής νέας δομημένης μεταπληροφορίας και μετατροπής της υπάρχουσας μας οδηγεί στη σκέψη δημιουργίας ενός αλγορίθμου που θα το κάνει αυτό με αυτόματο τρόπο κάτι που ίσως όμως να μην είναι τελικά εφικτό, ενώ συνεχίζει να υπάρχει το πρόβλημα της παρουσίας πλεονάζουσας πληροφορίας.

Top-down σχεδίαση, συνεπάγεται την προσπάθεια δημιουργίας λογισμικού που θα εξάγει την απαιτούμενη δομημένη πληροφορία ακριβώς όπως κάνει και ένας άνθρωπος, δηλαδή διαβάζοντας το κείμενο. Ήδη έχουν αναπτυχθεί μια σειρά από εφαρμογές που αναλύουν και εμπλουτίζουν τα δεδομένα που χρησιμοποιούν με αυτόν τον τρόπο.



Εικόνα 16.Διάγραμμα τεχνικών σχεδίασης εφαρμογών στο Σημασιολογικό Ιστό[5]

4.1 Αναζήτηση στον Σημασιολογικό Ιστό

Μία μεγάλη κατηγορία εφαρμογών που μπορούν να κάνουν χρήση του Σημασιολογικού Ιστού είναι οι μηχανές αναζήτησης. Εκμεταλλευόμενες τα πλεονεκτήματα που μπορεί να προσφέρει ο Σημασιολογικός Ιστός και τα Διασυνδεδεμένα Δεδομένα, έχουν αναπτυχθεί μηχανές αναζήτησης για το Σημασιολογικό Ιστό.

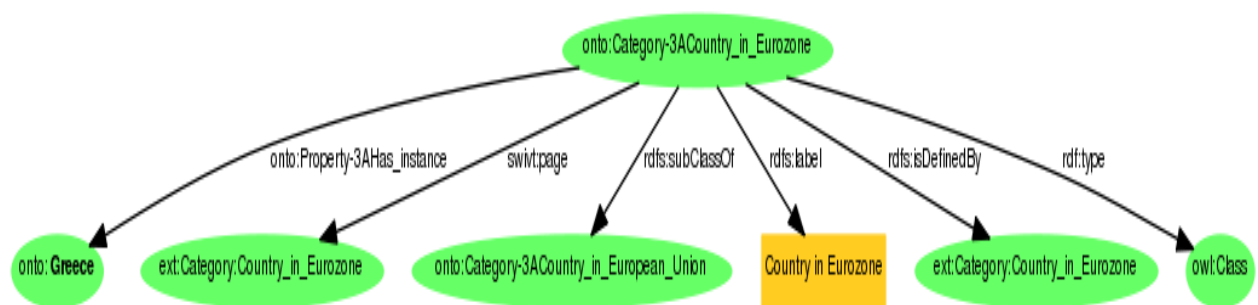
Αρχικά καλό είναι να κατανοήσει κανείς τι συνιστά τη σημασιολογική αναζήτηση. Για το σκοπό αυτό θα αναφερθούμε στον αλγόριθμο της Google με την ονομασία Google Hummingbird. Η Google με τον αλγόριθμο αυτό στην προσπάθειά της να επιστρέψει τα

πλέον ταιριαστά αποτελέσματα στην αναζήτηση του κάθε χρήστη, επιδιώκει να δώσει απάντηση όχι μόνο στο ερώτημα « Τι ψάχνει ο χρήστης ; » αλλά και στο ερώτημα «Γιατί ο χρήστης κάνει αυτή την αναζήτηση ; ». Με αυτόν τον τρόπο η διατύπωση του ερωτήματος παίζει ρόλο στην απάντηση που θα επιστραφεί και συνεπώς η Google αναμένει να πετύχει ταχύτερες και πιο στοχευμένες αναζητήσεις στο διαδίκτυο μέσω του νέου αλγορίθμου της, ο οποίος ξεκίνησε να λειτουργεί του Σεπτεμβρίου του 2013. Εν γένει με τη σημασιολογική αναζήτηση ο χρήστης θα μπορεί να αλληλεπιδράσει με μια μηχανή αναζήτησης κάνοντας ‘ομιλητική αναζήτηση’ (conversational search), αναμένοντας καλύτερα και πιο στοχευμένα αποτελέσματα. Μπορεί κανείς να καταλάβει τη σημασία της ομιλητικής αναζήτησης μέσω του παρακάτω παραδείγματος, π.χ. αντί να θέσει κανείς το ερώτημα « γήπεδα ποδοσφαίρου 5x5», θα μπορεί να ρωτήσει «Ποιά γήπεδα 5x5 βρίσκονται κοντά στο σπίτι μου ; » αναμένοντας η μηχανή αναζήτησης να κατανόησει τη λέξη ‘ ποιά ’, τι είναι ένα γήπεδο 5x5, να βρει τη διεύθυνση του σπιτιού του χρήστη (εφόσον υπάρχει στο λογαριασμό του) και τελικά να εμφανίσει με τα στοχευμένα πλέον αποτελέσματα. Η σημασιολογική αναζήτηση ξεπερνά τα όρια που θέτει η αναζήτηση με βάση το ταίριασμα λέξεων και ανοίγει πολλά μονοπάτια για την καλύτερη και αποδοτικότερη αξιοποίηση της γνώσης.

Οι μηχανές αναζήτησης που κάνουν χρήση του σημασιολογικού ιστού, βρίσκονται σήμερα σε ένα αρκετά πρώιμο στάδιο. Αυτή τη στιγμή έχουν αναπτυχθεί τέτοιες μηχανές αναζήτησης οι οποίες ψάχνουν το Σημασιολογικό Ιστό αναζητώντας οντολογίες και διασυνδεδεμένα δεδομένα που να ικανοποιούν το ερώτημα του χρήστη. Όταν ο Σημασιολογικός Ιστός έχει φτάσει σε αρκετά ώριμο στάδιο και μπορέσει να συνδυαστεί με τις τεχνικές της σημασιολογικής αναζήτησης, όπως περιγράφηκαν παραπάνω, θα έχουμε πλέον στα χέρια μας μια σειρά πολύ δυνατών εργαλείων αλληλεπίδρασης των χρηστών με τον παγκόσμιο γράφο γνώσης. Θα αναφέρουμε παρακάτω δύο μηχανές αναζήτησης του Σημασιολογικού Ιστού :

4.1.1 FALCONS

Μέσω του FALCONS μπορεί ο χρήστης να κάνει αναζήτηση στο Σημασιολογικό Ιστό για κάποιον όρο, επιθυμώντας ένα έγγραφο, μια οντολογία, έναν ορισμό-διατύπωση ή ένα αντικείμενο. Μάλιστα στην αναζήτηση οντολογίας επιστρέφει το γράφο δείχνει συνοπτικά τις κοντινότερες διασυνδέσεις που έχει ο όρος που αναζητήσαμε. Για παράδειγμα μια αναζήτηση για την οντολογία Greece, μας επέστρεψε τον παρακάτω γράφο.

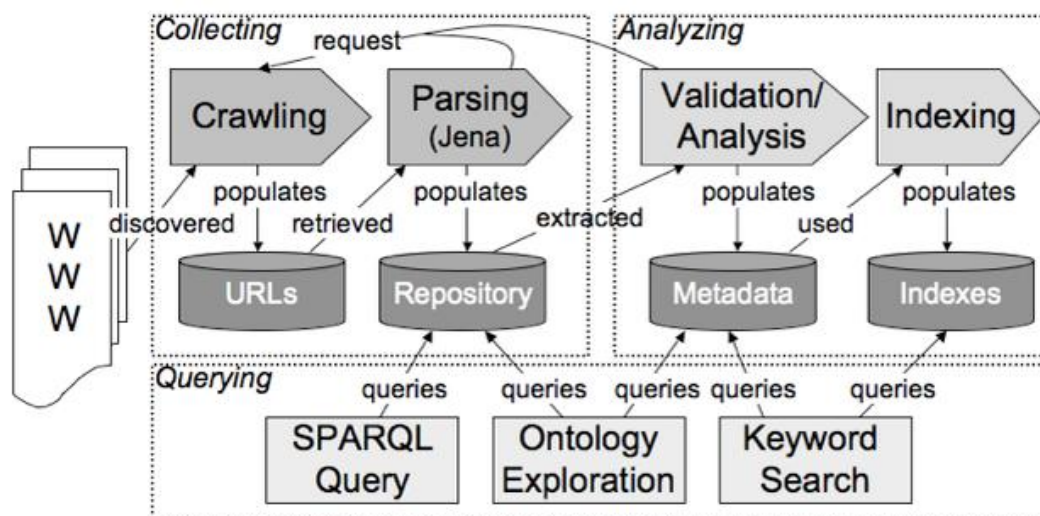


Εικόνα17. Γράφος αναζήτησης της οντολογίας Greece, στο Falcons.

4.1.2 WATSON

Η μηχανή αναζήτησης WATSON, αποτελεί μια πύλη προς τα δεδομένα του Σημασιολογικού Ιστού. Οι λειτουργίες που εκτελεί είναι συνοπτικά οι εξής τρεις :

- Συλλογή διαθέσιμων σημασιολογικών δεδομένων του Ιστού.
- Ανάλυση και εξαγωγή των χρήσιμων μεταδεδομένων και αναγνωριστικών που περιέχει.
- Παροχή αποδοτικών μεθόδων τοποθέτησης ερωτημάτων προς τα δεδομένα που έχει συλλέξει κα εξάγει.



Εικόνα18. Διάγραμμα λειτουργιών και υπηρεσιών του WATSON.

4.2 Περιήγηση στον Σημασιολογικό Ιστό (Semantic Web browsers)

Όπως έχει αναφερθεί κάθε αντικείμενο του Σημασιολογικού Ιστού αναγνωρίζεται από ένα URI, το οποίο οδηγεί και σε μια ιστοσελίδα με όλα τα σχετικά με αυτό δεδομένα εκφρασμένα σε μια από τις αποδεκτές μορφές όπως το RDF/XML ή κάποια μικροδομή. Αυτό που κάνουν λοιπόν οι περιηγητές του Σημασιολογικού Ιστού, είναι ότι ξεκινάνε από ένα URI το οποίο παρέχει ο χρήστης και με αρχή αυτό το αντικείμενο χτίζεται ένας ολόκληρος γράφος από όλους τους συνδέσμους που περιέχει το αρχικό URI. Ο γράφος αυτός συνήθως αποθηκεύεται με τη μορφή μιας session cache, δηλαδή προσωρινή μνήμη για την περιήγηση. Όσο λοιπόν ο χρήστης προχωράει σε μεγαλύτερο βάθος και ακολουθεί νέους συνδέσμους, διευρύνει ακόμη περισσότερο το γράφο και εξερευνά νέες περιοχές του Σημασιολογικού Ιστού. Παρακάτω θα αναφερθούμε εν συντομία σε δύο γνωστούς περιηγητές του Σημασιολογικού Ιστού και τα βασικά χαρακτηριστικά τους.

4.2.1 *Tabulator*

Ο Tabulator είναι ένας περιηγητής γενικού σκοπού για διασυνδεδεμένα δεδομένα και βρίσκεται σε ερευνητικό και πειραματικό στάδιο. Οι βασικοί στόχοι που έχει είναι [6] :

- Δημιουργία μιας πλατφόρμας λογισμικού εντός περιηγητή του διαδικτύου για νέες και υπάρχουσες εφαρμογές βασιζόμενες σε RDF δεδομένα.
- Η εξερεύνηση και ανάδειξη των περιορισμών, δυνατοτήτων και πιθανών βελτιώσεων της αρχιτεκτονικής του Διαδικτύου και του Σημασιολογικού Ιστού.
- Εντοπισμός των προβλημάτων υλοποίησης και χρήσης που απαιτούν διόρθωση.

Για τη λειτουργία του Tabulator μεγαλύτερη σημασία έχει ο λογικός γράφος των διασυνδέσεων των δεδομένων και μικρότερη το δίκτυο των πραγματικών αρχείων. Ο χρήστης μπορεί να προσπελαίνει το γράφο με όλα τα δεδομένα που έχουν διαβαστεί και ανά πάσα στιγμή έχει τη δυνατότητα να εξετάσει την πηγή και την αξιοπιστία των δεδομένων που του παρουσιάζονται. Κατά την εξέταση κάποιου URI μπορεί να ακολουθήσει κάποιον από τους εξερχόμενους συνδέσμους του, γνωρίζοντας πάντα μέσω ενδείξεων, εάν έχει ξαναεπισκεφθεί το συγκεκριμένο αντικείμενο ή αν υπάρχει κάποιο πρόβλημα κατά τη φόρτωσή του.

Μία έντονη διαμάχη σχετικά με τη σχεδίαση της διεπαφής με το χρήστη, εμφανίζεται συχνά ανάμεσα εξερεύνηση και την ανάλυση του ιστού. Για το διαδίκτυο είναι πιο συνηθισμένος ο όρος εξερεύνηση μιας και ο χρήστης ακολουθεί αλληπάλληλους συνδέσμους προκειμένου να φτάσει όσο πιο κοντά μπορεί σε αυτό που αναζητά. Στο Σημασιολογικό Ιστό όμως αυτή η διαδικασία της εξερεύνησης, απαιτεί τη βηματική πρόοδο από κόμβο σε κόμβο , αναζήτηση νέων συνδέσμων και έλεγχο της θέσης μας. Ο Tabulator λοιπόν παρέχει δύο διαφορετικά μοντέλα λειτουργίας, ένα για ανάλυση και ένα για εξερεύνηση. Κατά τη λειτουργία εξερεύνησης, ο χρήστης εξερευνά μια δενδρική δομή χωρίς να γνωρίζει τι άλλα δεδομένα είναι διαθέσιμα. Κατά τη λειτουργία ανάλυσης, ο χρήστης ορίζει ένα πρότυπο το οποίο ζητά από τον περιηγητή να ικανοποιήσει και εν συνεχεία του παρουσιάζει τα αποτελέσματα σε μία από τις διαθέσιμες μορφές. Αυτές είναι οι εξής :

1. Πίνακας
2. Χρονικό Διάγραμμα & Ημερολόγιο
3. Χάρτης

Τέλος ο Tabulator επιτρέπει στο χρήστη να τοποθετήσει κάποιο ερώτημα (query) σε SPARQL. Τα ερωτήματα που μπορεί να θέσει κανείς μπορεί να είναι από πολύ απλά και χωρίς απαίτηση εξειδικευμένων γνώσεων, μέχρι πολύ σύνθετα και συνδυαστικά.

4.2.2 Disco

Ο περιηγητής Disco, λειτουργεί κάνοντας χρήση της Semantic Web Client Library η οποία αναπαριστά όλο το Σημασιολογικό Ιστό σαν έναν ενιαίο RDF γράφο και επιτρέπει σε εφαρμογές να τοποθετούν ερωτήματα σε αυτόν διατυπωμένα σε SPARQL.

Αρχικά, ο περιηγητής βρίσκει την αναπαράσταση του δοσμένου URI και την προσθέτει στο γράφο, στη συνέχεια αυτόματα αναζητά τα URI τα οποία περιέχονται ως σχετικά και τα προσθέτει και αυτά στο γράφο. Έπειτα εκτελεί ταίριασμα προτύπων, θέτοντας το αρχικό URI ως υποκείμενο και ως αντικείμενο για όλες τις πιθανές προτάσεις, δηλαδή (x any any) και (any any x), όπου any είναι μια μεταβλητή για το ταίριασμα προτύπων και x είναι το δοσμένο URI. Τέλος για όλες τις τριάδες που ικανοποίησαν το ταίριασμα προτύπων, επαναλαμβάνεται η αρχική διαδικασία διευρύνοντας το γράφο περιήγησης. Η διαδικασία αυτή επαναλαμβάνεται δίνοντας στο χρήστη τη δυνατότητα να εξερευνήσει το Σημασιολογικό Ιστό σε όποιο βάθος επιθυμεί ορίζοντας ο ίδιος το σημείο έναρξης της αναζήτησής του.

5

Περιγραφή Συστήματος

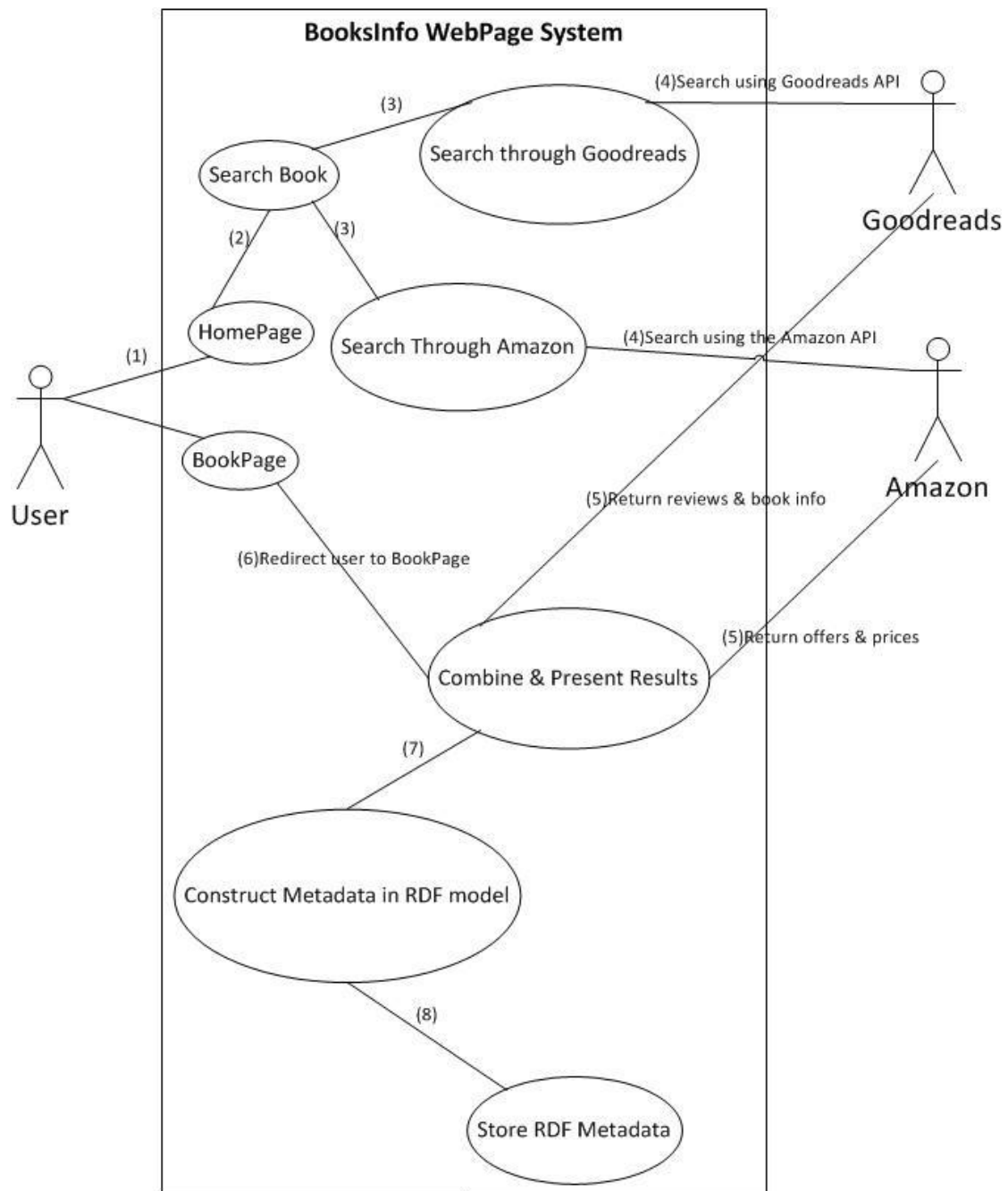
Το σύστημα που υλοποιήθηκε περιλαμβάνει τον χρήστη, τη διεπαφή χρήστη – ιστοσελίδα αναζήτησης τίτλων βιβλίων (www.booksinfo.gr), τις διεπαφές - APIs με τα διάφορα site άντλησης δεδομένων ([goodreads](http://goodreads.com), [amazon](http://amazon.com)), τα ίδια τα site αυτά και τέλος τη βάση στην οποία αποθηκεύονται όλες οι πληροφορίες που αντλήθηκαν και μετατράπηκαν σε RDF δεδομένα.

5.1 Διαγράμματα χρήσης – Use case diagrams

Τα βασικά σενάρια χρήσης του συστήματος που θα εξετάσουμε παρακάτω, μέσα από διαγράμματα – use case diagrams, είναι τα εξής :

- i. Επιτυχής αναζήτηση τίτλου και επιστροφή δεδομένων από όλα τα site άντλησης, δημιουργία πληροφορίας σε RDF μορφή και αποθήκευσή της στη βάση.
- ii. Ανεπιτυχής αναζήτηση τίτλου.

i)

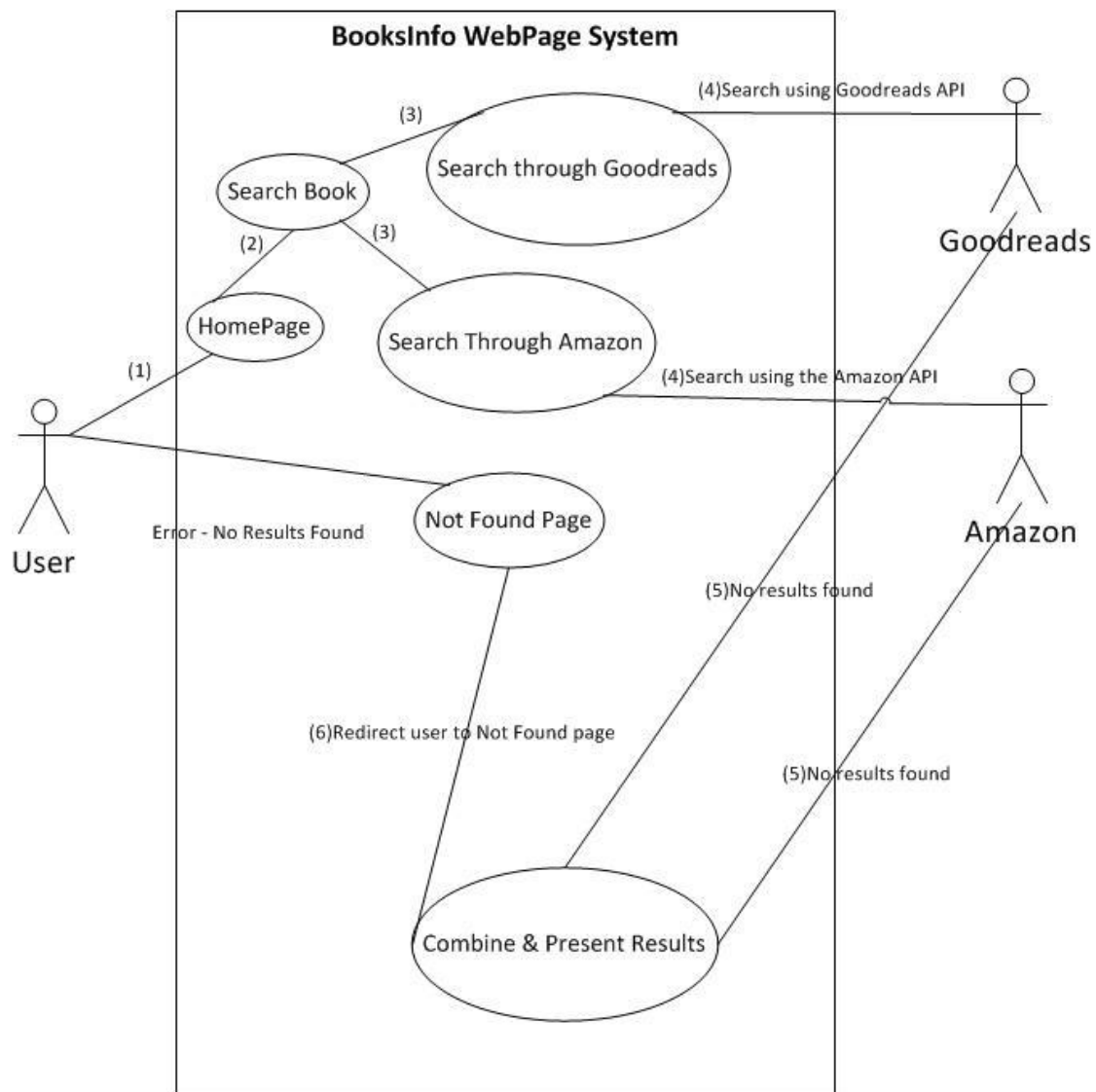


Σχήμα 1

- (1) Ο χρήστης οδηγείται στην αρχική σελίδα του BooksInfo.
- (2) Ο χρήστης εισάγει τις λέξεις – κλειδιά με τις οποίες θα γίνει η αναζήτηση τίτλου.
- (3)-(4) Το σύστημα προωθεί την αναζήτηση στις σελίδες των Amazon και Goodreads, μέσω των APIs τους.

- (5) Επιστρέφονται τα αποτελέσματα των επιμέρους αναζητήσεων.
- (6) Ο χρήστης οδηγείται σε νέα σελίδα με όλα τα δεδομένα που αντλήθηκαν με βάση την αναζήτηση που έκανε.
- (7) Δημιουργία μεταπληροφορίας για τα δεδομένα που αντλήθηκαν και δόμηση αυτής σε μορφή τριάδων RDF (RDF triples).
- (8) Αποθήκευση των RDF δεδομένων σε τοπική Βάση με τη βοήθεια της RDFLib.

ii)



Σχήμα 2

- (1) Ο χρήστης οδηγείται στην αρχική σελίδα του BooksInfo.
- (2) Ο χρήστης εισάγει τις λέξεις – κλειδιά με τις οποίες θα γίνει η αναζήτηση τίτλου.
- (3)-(4) Το σύστημα προωθεί την αναζήτηση στις σελίδες των Amazon και Goodreads, μέσω των APIs τους.
- (5) Οι επιμέρους αναζητήσεις επιστρέφουν κενά αρχεία αποτελεσμάτων.
- (6) Ο χρήστης οδηγείται σε νέα σελίδα με το σχετικό μήνυμα αδυναμίας εύρεσης αποτελεσμάτων για τη συγκεκριμένη αναζήτηση που επιχείρησε.

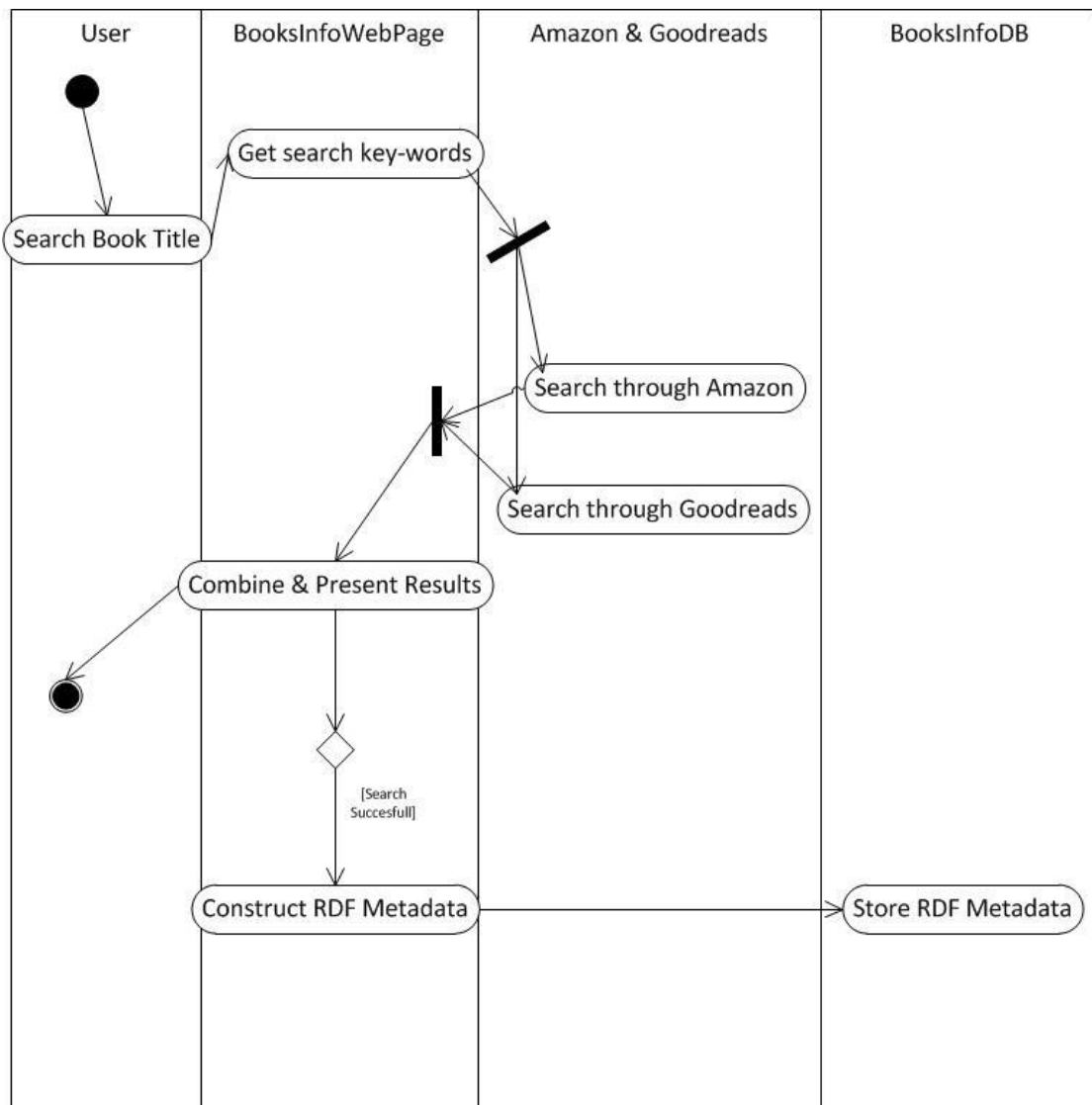
5.2 Σχεδίαση Συστήματος – UML diagrams

Σε αυτήν την ενότητα θα παρατεθούν μερικά διαγράμματα που θα κάνουν πιο κατανοητό τον τρόπο λειτουργίας του συστήματός μας, τη σειρά με την οποία λαμβάνουν χώρα οι διάφορες διεργασίες και με αυτόν τον τρόπο θα γίνουν ευδιάκριτα όλα τα επιμέρους υποσυστήματα και ο ρόλος του καθενός.

Τα διαγράμματα αυτά είναι τα εξής:

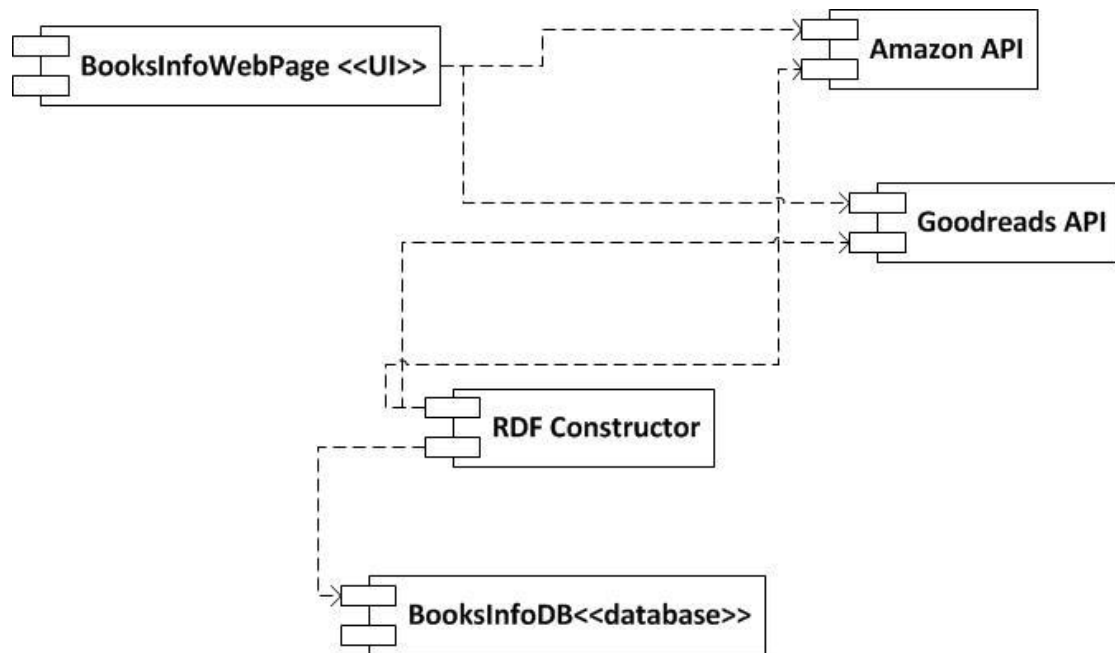
- i. Διάγραμμα Δραστηριότητας - Activity Diagram
- ii. Διάγραμμα Οντοτήτων - Component Diagram
- iii. Συνεργατικό Διάγραμμα - Collaboration Diagram

i)



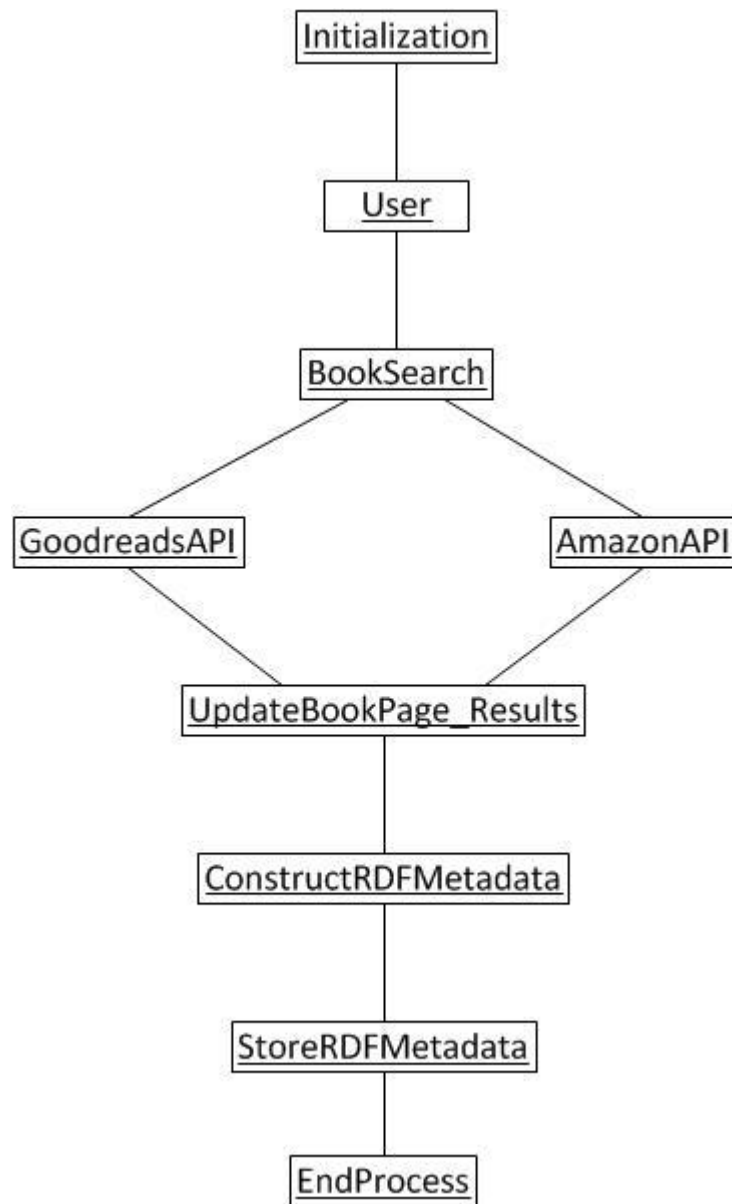
Σχήμα 3

ii)



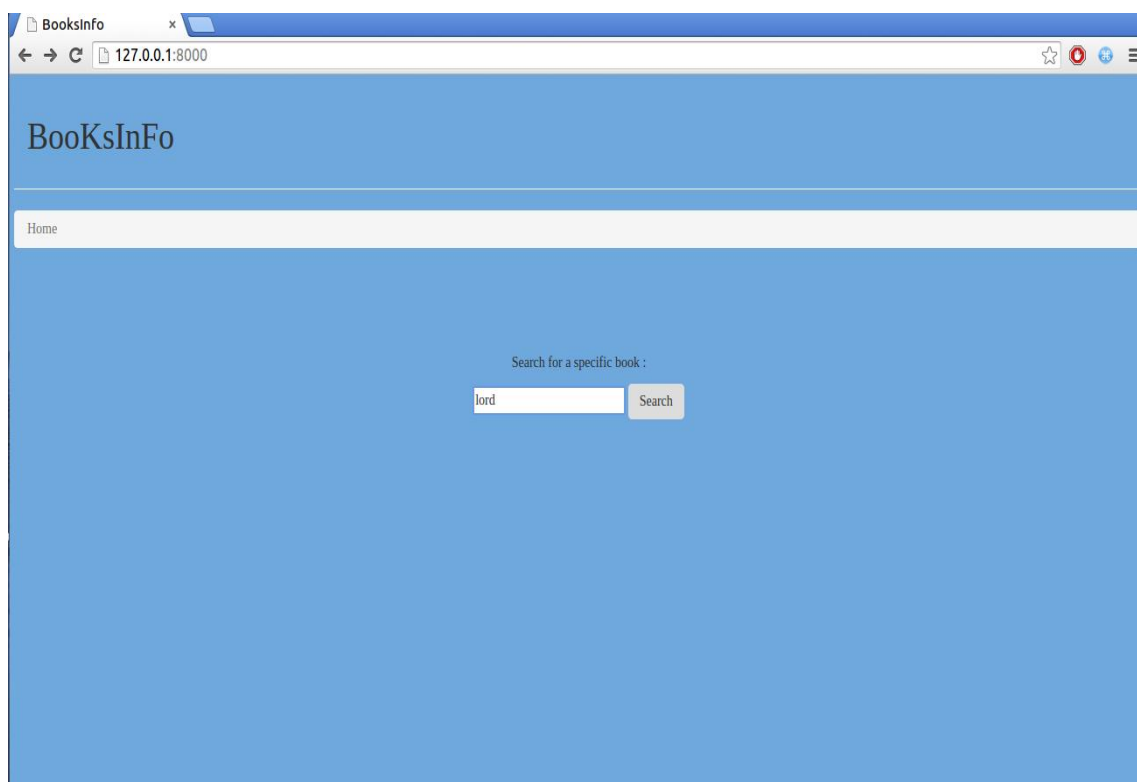
Σχήμα 4

iii)



Σχήμα 5

5.3 Screenshots Χρήσης Ιστοσελίδας



Εικόνα 19. Αρχική σελίδα του BooksInfo

BookKsInFo

Home / Lord of the Flies

Lord of the Flies

ISBN : 0140283331
 Publisher : Penguin Books
 Publication Year : 1999
 Number of Pages : 182
 Average Rating : 3.60
 Ratings Count : 1248994
 Reviews Count : 18618

Description : William Golding's compelling story about a group of very ordinary small boys marooned on a coral island has become a modern classic. At first, it seems as though it's all going to be great fun; but the fun before long becomes furious & life on the island turns into a nightmare of panic & death. As ordinary standards of behavior collapse, the whole world the boys know collapses with them—the world of cricket & homework & adventure stories—and another world is revealed beneath, primitive & terrible. **Lord of the Flies** remains as provocative today as when it was 1st published in 1954, igniting passionate debate with its startling, brutal portrait of human nature. Though critically acclaimed, it was largely ignored upon its initial publication. Yet soon it became a cult favorite among both students and literary critics who compared it to J.D. Salinger's **The Catcher in the Rye** in its influence on modern thought & literature. Labeled a parable, an allegory, a myth, a morality tale, a parody, a political treatise, even a vision of the apocalypse, **Lord of the Flies** has established itself as a classic.

Lord of the Flies Reviews

[Write a review](#) (you'll need to sign in to your Goodreads account or sign up) (showing 1-8 of 18,765)

By **Nora** ★☆☆☆☆ · September 28, 2007

basic premise. So I picked it up again, only to wish I hadn't. There's a reason why they teach this book in middle school—in order to enjoy this book, one's intellectual cognizance must be that of a child... [...more](#)

By **Emily May** (The United Kingdom) ★★★★★ · March 20, 2014

Kids are evil. Don't you know?

I've just finished rereading this book for my book club but, to be honest, I've liked it ever since my class were made to read it in high school. Overall, *Lord of the Flies* doesn't seem to be very popular, but I've always liked the almost Hobbesian look at the state... [...more](#)

By **Andrew** (The United States) ★★★★★ · August 27, 2008

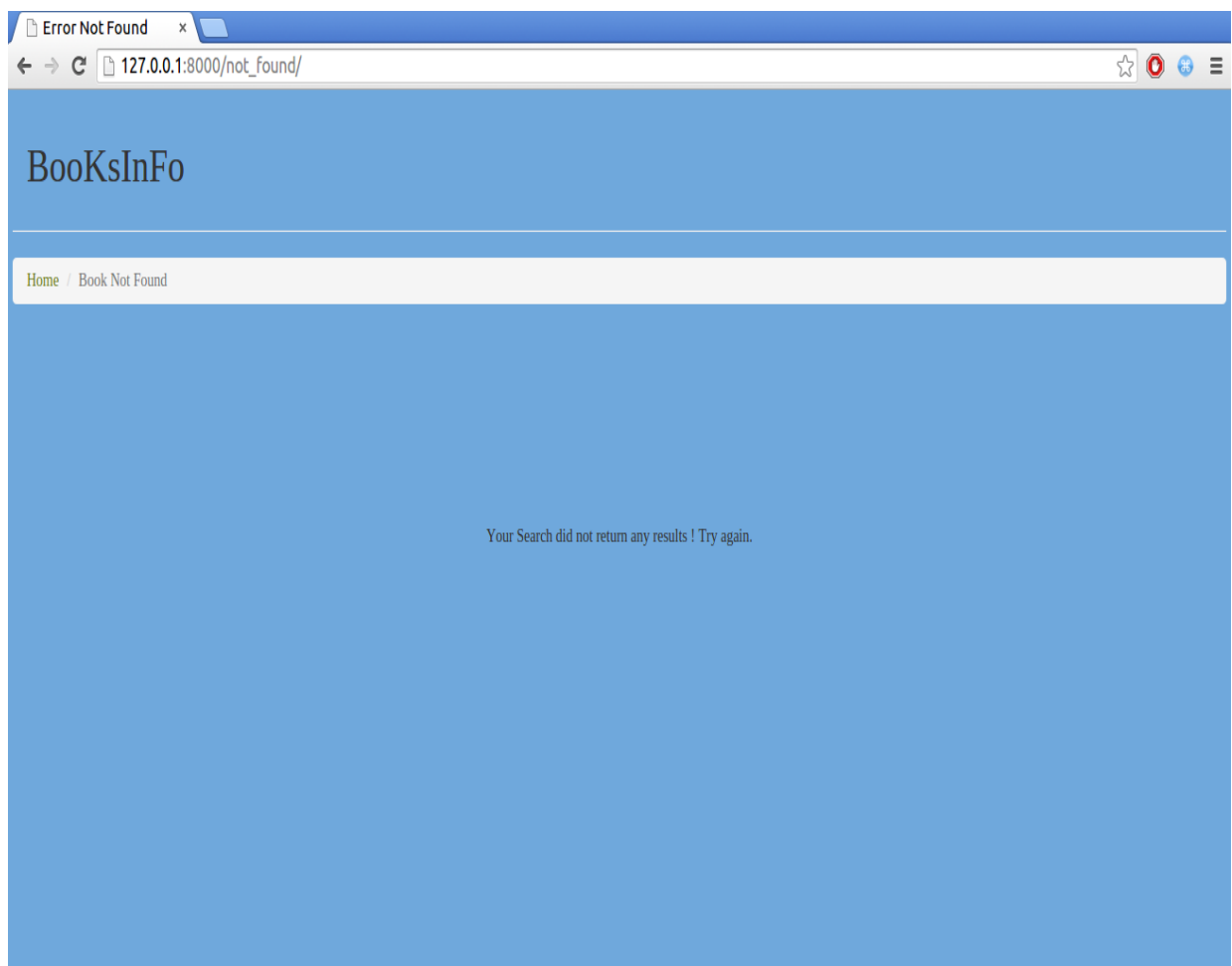
I was tempted to give this five stars, since in so many ways it strikes me as the kind of masterpiece, like *Heart of Darkness*, that I imagine will retain its horror and readability for centuries. The prose veers (or as Golding would say it, "tends") from plain to painterly. The story is well known... [...more](#)

Reviews from Goodreads.com

Matching Amazon Product
 Lord of the Flies -- Book
[Product Details](#) [Customer Reviews](#)

Find it as an E-book at [Gutenberg Project](#) Read reviews at [GoodReads](#) Or buy it online : [Amazon](#) [Book Depository](#) - Free WorldWide Shipping

Εικόνα 20. Επιτυχής αναζήτηση και εμφάνιση αποτελεσμάτων



Εικόνα 21. Ανεπιτυχής αναζήτηση και εμφάνιση αντίστοιχου μηνύματος

6

Υλοποίηση Συστήματος

Στο πρώτο υποκεφάλαιο θα αναλυθούν τα επιμέρους κομμάτια που υλοποιήθηκαν προκειμένου να ολοκληρωθεί το σύστημα, καθώς και τα αντίστοιχα κομμάτια κώδικα. Έπειτα στο δεύτερο θα αναφερθούμε στα διάφορα προγραμματιστικά εργαλεία, τις πλατφόρμες και τις διεπαφές που χρησιμοποιήθηκαν.

6.1 Λεπτομέρειες Υλοποίησης

Για την υλοποίηση του συστήματος, χρειάστηκε να κατασκευαστεί αρχικά η ιστοσελίδα – διεπαφή χρήστη για την αναζήτηση τίτλων βιβλίων, έπειτα δημιουργήθηκαν οι απαραίτητες διασυνδέσεις με τα site άντλησης πληροφοριών και τέλος θα δημιουργούσαμε τη βάση στην οποία θα αποθηκευόταν όλη η καινούρια πληροφορία που δημιουργείται και θα μπορούσε να αποτελέσει κομμάτι του σημασιολογικού ιστού των διασυνδεδεμένων δεδομένων.

Για την επεξεργασία των επιστρεφόμενων xml αρχείων, φτιάξαμε ακόμη μια συνάρτηση η οποία αναζητά την τιμή ενός πεδίου μέσα στο δοσμένο xml αρχείο και την επιστρέφει προκειμένου να δομηθεί η πληροφορία γύρω από το συγκεκριμένο βιβλίο.

```
def book_info(book_xml, string):
    # print book_xml
    element = book_xml.getElementsByTagName(string)
    # print element
    if element!=[]:
        if element[0].firstChild:
            element = element[0].firstChild.toprettyxml()
            element = element.lstrip('<![CDATA[')
            element = element.rstrip(']>')
            # print element
            return element
    else:
        return None
```

Εικόνα 23. Συνάρτηση επεξεργασίας xml αρχείων και εξαγωγής επιθυμητής πληροφορίας

Τέλος υλοποιήσαμε ακόμη μια συναρτήση για το Goodreads API, η οποία μας επιστρέφει όλες τις κριτικές, τα σχόλια και τις βαθμολογίες των χρηστών για τον συγκεκριμένο τίτλο βιβλίου.

```
#####New get reviews method #####  
def get_reviews(book_title):  
    key = "[REDACTED]"  
    secret = "[REDACTED]"  
  
    goodreads_client = GoodReadsClient(key, secret)  
    return goodreads_client.get_reviews(book_title)  
#####
```

Εικόνα 24. Συνάρτηση για την άντληση σχολίων χρηστών από το Goodreads μέσω του API

Αντίστοιχα για την επικοινωνία μας με το amazon.com κάναμε χρήση του παρεχόμενου API, και εδώ δεν υπήρξε κάποια ιδιαίτερη δυσκολία. Από το amazon θέλαμε να κάνουμε αναζήτηση για προϊόντα που είναι σχετικά με το βιβλίο που μας επέστρεψε η αναζήτηση στο goodreads και τα αντίστοιχα σχόλια χρηστών.

```
#####New amazon product search method#####  
def get_amazon_prod(bookIdentifier):  
    config = {  
        'access_key': '[REDACTED]',  
        'secret_key': '[REDACTED]',  
        'associate_tag': 'booksinfo0e-20',  
        'locale': 'uk'  
    }  
    api = amazonproduct.API(cfg=config)  
    items = api.item_search('Books', Keywords=bookIdentifier)  
    return items  
#####
```

Εικόνα 25. Συνάρτηση για την αναζήτηση προϊόντων στο amazon μέσω του API

Εδώ χρειάστηκε να εξερευνήσουμε το αρχείο που παίρναμε σαν αποτέλεσμα για να κατανοήσουμε πλήρως τη δομή του και τελικά χρειάστηκε να το μετατρέψουμε εμείς σε xml, για να διευκολύνουμε την επεξεργασία του και την αναζήτηση πεδίων μέσα σε αυτό.

6.1.2 Υλοποίηση Οντολογιών – Διασυνδεδεμένων Δεδομένων και Καταχώρηση σε Βάση Δεδομένων

Με βάση το schema.org/Book, γεμίσαμε τα διάφορα πεδία ιδιοτήτων των βιβλίων που παρουσιάστηκαν στην ιστοσελίδα μας σαν αποτελέσματα αναζητήσεων και δημιουργήσαμε με αυτόν τον τρόπο δομημένες οντότητες.

Παρακάτω φαίνεται πως μέσα στο HTML αρχείο της σελίδας παρουσίασης του αποτελέσματος της αναζήτησης, γίνεται ταυτόχρονα η μορφοποίηση της σελίδας και το γέμισμα των πεδίων με τη λέξη `itemprop`.

```
<div class="row" id = "Book Information" itemscope itemtype = "http://www.schema.org/Book">
  <h2><a class="link" itemprop="name" href="{BookURL | safe}"> {{ BookTitle | safe }}</a></h2>
  <table>
    <tr>
      <td>
        <table>
          <tr>
            <td>
              <i>ISBN</i> : <span itemprop="isbn">{{BookISBN | safe}}</span><br>
              <i>Publisher</i> : <span itemprop="publisher">{{BookPublisher | safe}}</span><br>
              <i>Publication Year</i> : <span itemprop="datePublished">{{BookPubYear | safe}}</span><br>
              <i>Number of Pages</i> : <span itemprop="numberOfPages">{{BookPages | safe}}</span><br>
              <i>Average Rating</i> : <span itemprop="aggregateRating">{{BookAvgRating | safe}}</span><br>
              <i>Ratings Count</i> : <span itemprop="interactionCount">{{BookRatingsCount | safe}}</span><br>
              <i>Reviews Count</i> : <span itemprop="interactionCount">{{BookReviewsCount | safe}}</span><br>
            </td>
          </tr>
          <tr>
            <td>
              
            </td>
          </tr>
        </table>
      </td>
      <td>
        <i>Description</i> : <span itemprop="description">{{BookDescription | safe}}</span><br>
      </td>
    </tr>
  </table>
</div>
```

Εικόνα 26. Κώδικας που εισάγει σημασιολογική μεταπληροφορία στα δεδομένα που παρουσιάζονται στο χρήστη

Εν συνεχεία αποθηκεύουμε αυτές τις οντότητες σε μια βάση δεδομένων προκειμένου να διευκολυνθούν μελλοντικές αναζητήσεις. Με αυτό τον τρόπο θα μπορούσε να ξεκινήσει η δημιουργία μιας διαρκώς ανανεώμενης βάσης με δομημένη και σημασιολογική πληροφορία για τα βιβλία. Η βιβλιοθήκη που χρησιμοποιήσαμε ήταν η RDFLib η οποία επιτρέπει την εύκολη δημιουργία ενός RDF triplestore και θα αναλυθεί περισσότερο παρακάτω.

6.2 Πλατφόρμες και Προγραμματιστικά εργαλεία

6.2.1 Django

Το Django είναι ένα Web Framework δηλαδή μια πλατφόρμα, γραμμένη σε python, πάνω στην οποία κατασκευάσαμε το site μας. Το Django παρέχει έτοιμες πολλές από τις υπηρεσίες που περιλαμβάνει μια μέση ιστοσελίδα, όπως σελίδα διαχειριστή με πολλαπλά δικαιώματα, δυνατότητα εύκολης σύνδεσης με βάση δεδομένων, δυνατότητα εγγραφής και εισόδου χρηστών και πολλά άλλα που διευκολύνουν σε μεγάλο βαθμό την κατασκευή μιας λειτουργικής ιστοσελίδας ή μιας διαδικτυακής εφαρμογής.

6.2.2 PyCharm

Το PyCharm είναι ένα περιβάλλον ανάπτυξης κώδικα για project σε Python. Μέσα σε αυτό το περιβάλλον έγινε και η ανάπτυξη ολόκληρου του project για την υλοποίηση ολόκληρου του συστήματός μας.

6.2.3 Amazon WebServices API

Η διεπαφή μας με τα δεδομένα της Amazon έγινε μέσω των Amazon Web Services, όπου αποτελούν μία συλλογή απομακρυσμένων υπηρεσιών που παρέχονται με τυποποιημένο τρόπο. Για τις ανάγκες της ιστοσελίδας μας και την αναζήτηση προϊόντων και σχολίων χρηστών γύρω από αυτά, κάναμε χρήση της μεθόδου ItemSearch του Amazon API. Η συγκεκριμένη μέθοδος δέχεται αρχικά μία λέξη κλειδί με βάση την οποία κάνει αναζήτηση προϊόντων, επειδή όμως τα δεδομένα που θα επιστρέφονταν θα ήταν πάρα πολλά δίνουμε ακόμη μία παράμετρο ώστε να περιορίσουμε την αναζήτησή μας. Ακόμη μπορούμε να προσδιορίσουμε τον αριθμό των αποτελεσμάτων που θέλουμε να μας επιστραφούν. Για τις δικές μας αναζητήσεις δίνουμε σαν λέξη κλειδί τον κωδικό ISBN του βιβλίου που εξάγαμε από το Goodreads, μιας και η αναζήτηση στο Goodreads γίνεται πρώτη, και ως προσδιοριστική παράμετρο δίνουμε το “Keywords = BookIdentifier” .

6.2.4 Goodreads API

Η αντίστοιχη διεπαφή για τα δεδομένα που παρέχει το Goodreads έγινε πάλι με το αντίστοιχο API που παρέχετε δωρεάν. Το Goodreads API παρέχει στο χρήστη πληθώρα λειτουργιών με τη μορφή μεθόδων, όπως μπορεί κανείς εύκολα να διαπιστώσει διαβάζοντας τον εκτενή οδηγό χρηστών στο www.goodreads.com/api .

Για τους σκοπούς της διπλωματικής εργασίας κάναμε χρήση δύο μεθόδων που σχεδιάσαμε με βάση τις οδηγίες του Goodreads προκειμένου να παίρνουμε τα επιθυμητά αποτελέσματα. Αυτές οι δύο μέθοδοι είναι οι get_reviews και get_book_info, μαζί με αυτές χρειάστηκε να υλοποιήσουμε ακόμη 2 μεθόδους για τη σωστή λειτουργία τους. Οι 2 αυτές μέθοδοι είναι οι parse_reviews και parse_book_info και επιτυγχάνουν να δώσουν την απαραίτητη μορφοποίηση στα επιστρεφόμενα αποτελέσματα από τις get_reviews και get_book_info.

```
##### New get book info method
def get_book_info(self, title, author=None):
    url = "%sbook/title.xml" % self.BASE_URL
    query_params = {
        "title": title,
        "author": author,
    }
    url_handler = self.unauthorized_request(url, query_params)
    return self.parser.parse_book_info(url_handler)

def parse_book_info(self, url_handler):
    goodreads_dom = self.parse_result(url_handler)
    info = goodreads_dom
    return info
```

Εικόνα 27. Συναρτήσεις για τη μορφοποίηση των αποτελεσμάτων που επιστρέφει η αναζήτηση πληροφοριών μέσω του Goodreads API

```
#####New get book reviews method
def get_reviews(self, title, author=None):
    url = "%sbook/title.xml" % self.BASE_URL
    query_params = {
        "title": title,
        "author": author,
    }

    url_handler = self.unauthorized_request(url, query_params)
    return self.parser.parse_reviews(url_handler)

###New review parsing method#####
def parse_reviews(self, url_handler):
    goodreads_dom = self.parse_result(url_handler)
    reviews = goodreads_dom.getElementsByTagName("reviews_widget")
    return reviews
```

Εικόν 28. Συναρτήσεις για τη μορφοποίηση των αποτελεσμάτων που επιστρέφει η αναζήτηση σχολίων χρηστών μέσω του Goodreads API

6.2.5 *Bootstrap*

Το Bootstrap είναι μια βιβλιοθήκη που περιέχει μια σειρά από πολλά εργαλεία σε HTML, CSS και JavaScript. Τα εργαλεία αυτά επιτρέπουν στους χρήστες να σχεδιάζουν και να υλοποιούν εύκολα και γρήγορα διάφορα κομμάτια μιας ιστοσελίδας ή μιας web εφαρμογής. Αυτές οι δυνατότητες περιλαμβάνουν κουμπιά, λίστες, φόρμες και πολλά widgets όπως ημερολόγια και navigation bars. Ακόμη ο χρήστης μπορεί να οργανώσει το περιεχόμενο μιας σελίδας με το σύστημα πλέγματος – grid system, το οποίο χωρίζει κάθε σελίδα σε στήλες και γραμμές ώστε να προσδιορίζεται ευκολότερα η θέση κάθε στοιχείου στην κάθε σελίδα.

Εν κατακλείδι, το βασικό πλεονέκτημα του Bootstrap είναι ότι επιτρέπει σε κάποιον προγραμματιστή να σχεδιάσει μια όμορφη και φιλική προς το χρήστη ιστοσελίδα ή web εφαρμογή χωρίς να χρειάζεται να έχει γνώσεις γραφιστικής και design.

6.2.6 *RDFLib*

Για τις ανάγκες της διπλωματικής εργασίας και της ιστοσελίδας αναζήτησης τίτλων βιβλίων, χρησιμοποιήσαμε τη βιβλιοθήκη RDFLib στην οποία θα αναφερθούμε σε επόμενο κεφάλαιο. η οποία είναι γραμμένη στη γλώσσα προγραμματισμού Python και επιτρέπει την εύκολη αλληλεπίδραση με μία RDF βάση δεδομένων. Η βιβλιοθήκη RDFLib παρέχει όλες τις δυνατότητες που θα ήθελε ένας χρήστης πάνω σε μια βάση τριάδων RDF, με πολύ απλές και κατανοητές εντολές. Αρχικό βήμα για τη δημιουργία μιας βάσης με το RDFLib, είναι η δημιουργία ενός γράφου, έστω g, όπου κάθε εντολή που αναφέρεται στα δεδομένα αυτού θα πρέπει να ξεκινάει με το αναγνωριστικό του, π.χ. g.add().

Οι βασικές εντολές και λειτουργίες είναι οι εξής :

- Add() : Άμεση εισαγωγή της δοσμένης τριάδας, που αποτελεί το όρισμα της εντολής.
- Parses() : Παρέχει δυνατότητα εισαγωγής μιας ή πολλαπλών τριάδων μέσω έτοιμου αρχείου.
- Set() : Επιτρέπει την μεταβολή κάποιου υπάρχοντος πεδίου κάποιας τριάδας της βάσης.

- **Remove()** : Διαγραφή της δοσμένης ως όρισμα τριάδας. Σε περίπτωση που κάποιο από τα τρία πεδία οριστεί ως None, τότε γίνεται ταίριασμα προτύπων και όποια τριάδα το ικανοποιεί διαγράφεται αυτόματα. Μπορούμε δηλαδή να διαγράψουμε αυτόματα μια ολόκληρη ομάδα τριάδων. Για παράδειγμα με την εντολή `g.remove(ntua,None,None)`, διαγράφονται όλες οι τριάδες που περιέχουν ως υποκείμενο τον όρο `ntua`.
- **Επαναληπτικοί βρόχοι** : Μπορούμε μέσα από μια επαναληπτική διαδικασία να διατρέψουμε όλα ή ορισμένα από τα στοιχεία του συνολικού γράφου, όπως θα κάναμε σε μια συνδεδεμένη λίστα ή σε έναν πίνακα. `For sub,pred,obj in graph:` .
- **Έλεγχος ύπαρξης τριάδας** : Μπορούμε να δώσουμε μία ορισμένη ή ημιορισμένη τριάδα και να ελέγξουμε εάν υπάρχει κάποια αντιστοιχία με τις τριάδες του γράφου μας. Για παράδειγμα `if (ece, RDF.type, school) in graph:` .
- **Πράξεις μεταξύ γράφων** : Επιτρέπονται οι πράξεις μεταξύ υπαρχόντων γράφων, δηλαδή έχουμε την ένωση δύο γράφων (+), τη διαφορά τους (-), την τομή τους (&).
- **Triples()** : Επιτρέπει την παραγωγή τριάδων βάσει ταιριασμάτων με την τριάδα του ορίσματος. Αποτελεί έναν εύκολο τρόπο δημιουργίας μιας ομάδας τριάδων για το γέμισμα της βάσης μας.
- **Query()** : Το βασικό εργαλείο για οποιαδήποτε αναζήτηση στη βάση μας, μπορούμε να θέσουμε ένα πολύ απλοϊκό μέχρι και ένα πολυσύνθετο ερώτημα προς τα δεδομένα της βάσης.

7

Επίλογος

Σε αυτή την ενότητα θα συνοψίσουμε τα συμπεράσματα και τα αποτελέσματα στα οποία καταλήξαμε μέσα από τη διπλωματική εργασία και την μελέτη όλων των σχετικών τεχνολογιών. Κλείνοντας θα κάνουμε κάποιες επισημάνσεις και θα παραθέσουμε μερικές ιδέες για μελλοντική χρήση στον τομέα του Σημασιολογικού Ιστού και πιο συγκεκριμένα για την επέκταση της ιδέας που αναπτύχθηκε για τη διπλωματική εργασία.

7.1 Σύνοψη και συμπεράσματα

Διαβάζοντας και κατανοώντας σε αρκετό βάθος την πλειοψηφία των τεχνολογιών που αναπτύσσονται και αξιοποιούνται γύρω από το Σημασιολογικό Ιστό και τα Διασυνδεδεμένα Δεδομένα, καθώς και χρησιμοποιώντας κάποιες από αυτές για την ανάπτυξη της ιστοσέλιδας BooksInfo και των λειτουργιών της, υπάρχουν πολλά συμπεράσματα τα οποία προέκυψαν.

Αρχικά η αξία του Σημασιολογικού Ιστού και των Διασυνδεδεμένων Δεδομένων γινόταν όλο και πιο ξεκάθαρη καθώς προχωρούσε η εργασία, οι δυνατότητες και οι προοπτικές εξέλιξης και ανάπτυξης νέων εργαλείων γύρω από τις τεχνολογίες αυτές δείχνουν να είναι πραγματικά τεράστιες. Η δυνατότητα αξιοποίησης όλου του όγκου της παγκόσμιας γνώσης που βρίσκεται στο διαδίκτυο, προκαλεί δέος και δημιουργεί εικόνες επιστημονικής φαντασίας ακόμη και σε κάποιον που δεν έχει εξειδικευμένες γνώσεις. Η αυτοματοποιημένη σημασιολογική

αναζήτηση μέσα στο διαδίκτυο θα επιτρέψει στους ανθρώπους να απαντούν γρήγορα ερωτήματα με την ήδη υπάρχουσα γνώση, αξιοποιώντας την έτσι πλήρως και με τη μέγιστη αποδοτικότητα.

Από επιχειρηματικής άποψης, με την προσθήκη σημασιολογίας και τη διασύνδεση των δεδομένων μας, έχουμε ως άμεσο αποτέλεσμα την αύξηση της αξίας της ιστοσελίδας μας μιάς και πλέον περισσότεροι σύνδεσμοι οδηγούν σε αυτήν. Επιπρόσθετα, νέες εφαρμογές που κάνουν χρήση της σημασιολογίας της πληροφορίας, μπορούν να δώσουν εύκολα και γρήγορα λύση σε πολυσύνθετα προβλήματα που προηγουμένως η επίλυσή τους ήταν μη αποδοτική για μια επιχείρηση.

Βέβαια είναι γεγονός πως προκειμένου να εισαχθεί σημασιολογία και να δημιουργηθούν όλες οι απαιτούμενες διασυνδέσεις στην υπάρχουσα πληροφορία απαιτείται μεγάλη προσπάθεια. Αυτή η προσπάθεια όμως έχει ήδη ξεκινήσει και συνεχώς εξαπλώνεται και σε άλλα γνωστικά πεδία με νέα εργαλεία να έρχονται στο προσκήνιο για να μας διευκολύνουν.

Η παρούσα διπλωματική συνεπώς κατάφερε να καταδείξει την αξία του Σημασιολογικού Ιστού επισημαίνοντας τις δυνατότητες και τις υπηρεσίες που θα παρέχει καθώς και την συνολική αξία που θα προκύψει από την αποδοτική αξιοποίηση της πληροφορίας. Επίσης παρουσιάσαμε μια σειρά από σημαντικές εφαρμογές που έχουν ήδη αναπτυχθεί ή αναπτύσσονται αυτή τη στιγμή και κάνουν χρήση του Σημασιολογικού Ιστού. Επισημάναμε τα σημαντικότερα επιτεύγματα των τεχνολογιών αυτών καθώς και το σκοπούς που εξυπηρετούν. Ακόμη, η επιδίωξη ανάδειξης του τομέα των ψηφιακών βιβλίων στο Σημασιολογικό Ιστό καθώς και της απουσίας συντονισμένης προσπάθειας ένταξής του σε αυτόν επετεύχθει. Τέλος, επισημάνθηκε η συνεισφορά των κοινωνικών δικτύων και των διεπαφών τους στην προώθηση των ψηφιακών βιβλίων και την εξάπλωση του Σημασιολογικού Ιστού.

7.2 Μελλοντικές επεκτάσεις

Μία από τις επεκτάσεις που θα μπορούσαν να γίνουν πάνω στην εφαρμογή της διπλωματικής εργασίας και αναφέρθηκε παραπάνω είναι η βελτίωση της λειτουργίας της βάσης δεδομένων και η αξιοποίηση των δεδομένων της για την επίτευξη μεγαλύτερης ταχύτητας και αποδοτικότητας στα αποτελέσματα που παρέχονται στους χρήστες. Ακόμη η δημοσιοποίηση και η κοινή χρήση των δομημένων δεδομένων που δημιουργούνται, για την προώθηση της χρήσης και των αρχών του Σημασιολογικού Ιστού.

Θα μπορούσαν ακόμη να προστεθούν περισσότερες ιστοσελίδες για την άντληση δεδομένων, προκειμένου η αναζήτηση να είναι πιο πλήρης και η γνώση που θα αποθηκεύουμε να είναι η μέγιστη δυνατή και όσο πιο διασταυρωμένη γίνεται. Επίσης το γραφιστικό κομμάτι της ιστοσελίδας μας και η συνολική διεπαφή με το χρήστη μπορεί να βελτιωθεί και να προσφέρει συνολικά πιο εύκολη και πρακτική χρήση για όποιον την επισκέπτεται.

Ενώ τέλος θα μπορούσε να σχεδιαστεί και μια εφαρμογή για συσκευές smartphones και tablets, για την αύξηση των χρηστών της και την επιτάχυνση της συλλογής και της δόμησης δεδομένων.

8

Βιβλιογραφία

1. Ricardo Baeza-Yates : From Capturing Semantics to Semantic Search: A Virtuous Cycle (ESWC 2008)
2. C. Pedrinaci, J. Domingue, and A.K.A. de Medeiros : A Core Ontology for Business Process Analysis (ESWC 2008)
3. Tom Heath and Christian Bizer : Linked Data: Evolving theWeb into a Global Data Space
4. Olaf Hartig and Andreas Langeegger : A Database Perspective on Consuming Linked Data on the Web
5. http://readwrite.com/2007/09/20/the_top-down_semantic_web
6. Tim Berners Lee, Yuhsin Chen, Lydia Chilton, Dan Connolly, Ruth Dhanaraj, James Hollenbach, Adam Lerer, and David Sheets : Tabulator: Exploring and Analyzing linked data on the Semantic Web
7. Aurona Gerber, Alta van der Merwe, and Andries Barnard : A functional semantic web architecture
8. <http://en.wikipedia.org>
9. <http://semanticweb.org>
10. <http://linkeddata.org>

11. <https://open-data.europa.eu/en/data>
12. <http://www.wikimedia.org>
13. <http://linkedgeodata.org/>
14. <https://purl.org/docs/index.html>
15. <http://lod2.eu>
16. <http://linkedgeodata.org>
17. <http://linkeddatafragments.org>
18. <http://triplify.org>
19. <http://musicbrainz.org>
20. <http://dbpedia.org>