

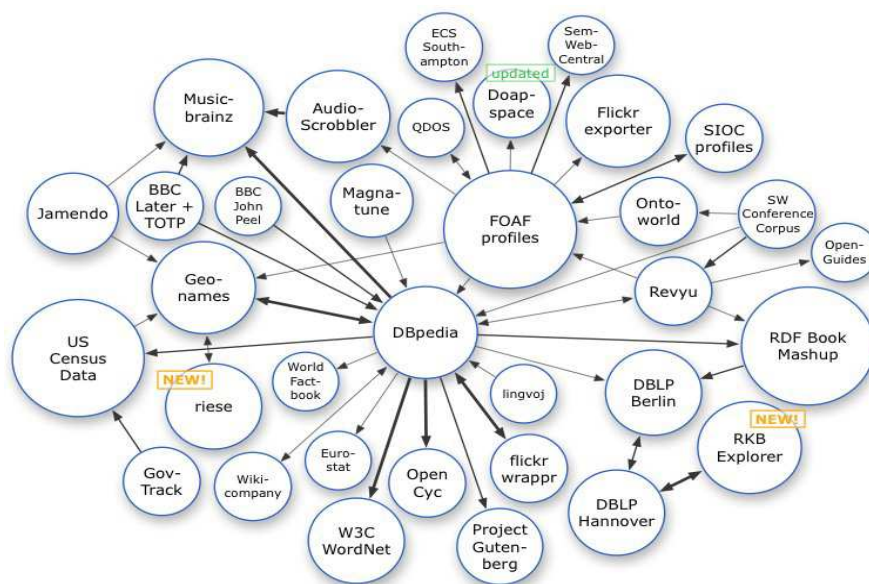


ΑΛΕΞΑΝΔΡΕΙΟ Τ.Ε.Ι. ΘΕΣΣΑΛΟΝΙΚΗΣ  
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



## ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

### Συνδεδεμένα Δεδομένα: Δημοσίευση, Σύνδεση, Προσπέλαση (Linked Data)



Της φοιτήτριας

Μουσταφά Νεντιμέ

Αρ. Μητρώου: 03/2452

Επιβλέπων καθηγητής

Νίτσορ Ηλίας

## ΠΡΟΛΟΓΟΣ

Η παρούσα πτυχιακή πραγματεύεται τη μελέτη των συνδεδεμένων δεδομένων (Linked Data). Ο όρος Συνδεδεμένα Δεδομένα αναφέρεται σε ένα σύνολο βέλτιστων πρακτικών για δημοσίευση και διασύνδεση δομημένων δεδομένων στον Ιστό, οι οποίες, σύμφωνα με τον εμπνευστή του Παγκόσμιου Ιστού και των ίδιων των Συνδεδεμένων Δεδομένων Tim Berners-Lee, αποτελούν τη γέφυρα για τη μετάβαση από τον Ιστό των Εγγράφων στον Ιστό των Δεδομένων, ή αλλιώς Σημασιολογικό Ιστό. Ο Σημασιολογικός Ιστός βασίζεται στα μεταδεδομένα (metadata), ή μεταπληροφορία, τα οποία περιγράφουν τη σημασιολογία του περιεχομένου του Ιστού. Οραματίζεται τον εμπλουτισμό των δεδομένων του Ιστού με σημασιολογία έτσι ώστε να είναι κατανοητά από τους υπολογιστές επιτρέποντας έτσι την εξαγωγή υπονοούμενης (implicit) γνώσης. Το κύριο στοιχείο για την επίτευξη αυτού του στόχου είναι οι οντολογίες. Οι οντολογίες προσφέρουν μια εννοιολογική θεώρηση ενός πεδίου ενδιαφέροντος κάνοντας τη γνώση επαναχρησιμοποιήσιμη και διαμοιραζόμενη. Το μυστικό του Σημασιολογικού Ιστού είναι τα μεταδεδομένα τα οποία κάνουν σαφές το υπονοούμενο ("making explicit the implicit" Tim Berner Lee). Τα μεταδεδομένα θα καταστήσουν τα δεδομένα κατανοητά σε επίπεδο μηχανής ώστε να διευκολύνεται η επεξεργασία και η αναζήτηση τους, θα επιτρέψουν τη δημιουργία κοινής ορολογίας, οργάνωσης και σημασιολογίας στα δεδομένα που διακινούνται στον Ιστό, στο βαθμό που είναι εφικτό, και θα προσδώσουν ευελιξία στη σύνθεση της σημασιολογικής πληροφορίας και στη δημιουργία νέας.

## ΠΕΡΙΛΗΨΗ

Ο Σημασιολογικός Ιστός (Semantic Web) υπόσχεται έναν διαφορετικό τρόπο διαχείρισης και πρόσβασης στις πληροφορίες του Διαδικτύου. Η κεντρική ιδέα στην οποία βασίζεται το όραμα του Σημασιολογικού Ιστού είναι η κατανόηση των πληροφοριών από τους ίδιους τους υπολογιστές μέσω της περιγραφής των ίδιων των δεδομένων και των σχέσεων μεταξύ τους. Ο Σημασιολογικός Ιστός υπόσχεται ένα νέο τρόπο χρήσης του διαδικτύου, τόσο στο επίπεδο αναζήτησης πληροφορίας όσο και στο επίπεδο δημιουργίας και χρήσης εφαρμογών. Η κατανόηση της σημασίας δεδομένων από τους ίδιους τους υπολογιστές μπορεί να οδηγήσει σε ταχύτατες αναζητήσεις πληροφοριών με ακρίβεια που ποτέ μέχρι σήμερα δεν ήταν δυνατή.

Η παρούσα πτυχιική εργασία επικεντρώνεται στην παρουσίαση των συνδεδεμένων δεδομένων. Ειδικότερα μας παρουσιάζει τι είναι αυτά τα δεδομένα, πώς προέκυψαν και πώς χρησιμοποιούνται σήμερα. Επίσης αναλύει ορισμένα από τα διαθέσιμα εργαλεία για τη δημοσίευσή τους, για την προσπέλαση και την αξιοποίησή τους.

## ABSTRACT

The Semantic Web promises a different way of managing and accessing information on the Internet. The central idea behind the Semantic Web vision is to understand the information from their own computers by describing the data itself and relationships between them. The Semantic Web promises a new way of using the Internet, both in terms of information search and in terms of creating and using applications. Understanding the importance of data from the computers themselves can lead to rapid information searches with precision never until now was not possible.

This thesis focuses on the presentation of linked data. In particular shows what these data are, how they arose and how they are used today. It also analyzes some of the tools available for publication, access and use.

## **ΕΥΧΑΡΙΣΤΙΕΣ**

Πριν την παρουσίαση των αποτελεσμάτων της παρούσας εργασίας, αισθάνομαι την ανάγκη να ευχαριστήσω ορισμένους από τους ανθρώπους που γνώρισα, συνεργάστηκα μαζί τους και έπαιξαν πολύ σημαντικό ρόλο στην παρουσίασή της. Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της πτυχιακής μου κ. Ηλία Νίτσο για την εμπιστοσύνη και την συνεχή καθοδήγησή του. Τέλος, πολλά ευχαριστώ στην οικογένειά μου για την υποστήριξη που μου προσέφεραν σε όλα τα χρόνια των σπουδών μου.

«Έχω ένα όνειρο για τον Παγκόσμιο Ιστό, οι υπολογιστές να καταστούν ικανοί να αναλύουν όλα τα δεδομένα που υπάρχουν στον Ιστό– περιεχόμενο, συνδέσεις, και συναλλαγές μεταξύ ανθρώπων και υπολογιστών. Ένας «Σημασιολογικός Ιστός», ο οποίος θα το κάνει αυτό δυνατόν , θα αργήσει να εμφανιστεί, αλλά όταν το κάνει οι καθημερινοί μηχανισμοί του εμπορίου, της γραφειοκρατίας και η καθημερινότητα μας θα διαχειρίζονται από μηχανές που θα μιλάνε σε μηχανές. Οι «Ευφυείς πράκτορες» τους οποίους οι άνθρωποι ονειρεύονται για αιώνες θα γίνουν πραγματικότητα.» Tim Berners-Lee, 1999

## ΠΕΡΙΕΧΟΜΕΝΑ

|                                                                                |    |
|--------------------------------------------------------------------------------|----|
| ΠΡΟΛΟΓΟΣ.....                                                                  | 2  |
| ΠΕΡΙΛΗΨΗ.....                                                                  | 3  |
| ABSTRACT.....                                                                  | 4  |
| ΕΥΧΑΡΙΣΤΙΕΣ.....                                                               | 5  |
| ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ.....                                                         | 10 |
| ΕΙΣΑΓΩΓΗ.....                                                                  | 11 |
| <br>                                                                           |    |
| ΚΕΦΑΛΑΙΟ 1.....                                                                | 13 |
| 1.1 ΤΙ ΕΙΝΑΙ ΣΥΝΔΕΔΕΜΕΝΑ ΔΕΔΟΜΕΝΑ.....                                         | 13 |
| 1.2 ΙΣΤΟΡΙΚΗ ΕΞΕΛΙΞΗ.....                                                      | 13 |
| 1.3 ΔΟΜΗ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΙΣΤΟΥ.....                                             | 16 |
| 1.4 Η ΤΕΧΝΟΛΟΓΙΑ ΤΩΝ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ.....                               | 19 |
| 1.4.1 ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΤΟΥ ΜΟΝΤΕΛΟΥ RDF.....                                      | 21 |
| 1.4.2 ΠΛΗΡΟΦΟΡΙΑΚΟΙ ΚΑΙ ΜΗ ΠΛΗΡΟΦΟΡΙΑΚΟΙ ΠΟΡΟΙ.....                            | 22 |
| 1.4.3 DEREFERENCING ΤΩΝ HTTP URIS.....                                         | 22 |
| 1.4.4 RDF SCHEMA.....                                                          | 23 |
| 1.5 ΚΑΝΟΝΕΣ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ.....                                        | 25 |
| <br>                                                                           |    |
| ΚΕΦΑΛΑΙΟ 2.....                                                                | 27 |
| 2.1 ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ.....                                  | 27 |
| 2.2 ΔΗΜΟΣΙΕΥΣΗ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ.....                                     | 27 |
| 2.2.1 ΔΗΜΟΣΙΕΥΣΗ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ<br>ΣΤΑΤΙΣΤΙΚΑ RDF ΑΡΧΕΙΑ.....      | 28 |
| 2.2.2 ΔΗΜΟΣΙΕΥΣΗ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ<br>ΣΧΕΣΙΑΚΕΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ..... | 29 |
| 2.2.3 ΔΗΜΟΣΙΕΥΣΗ ΑΛΛΟΥ ΤΥΠΟΥ ΠΛΗΡΟΦΟΡΙΩΝ ΣΕ ΜΟΡΦΗ<br>ΣΥΝΔΕΔΕΜΕΝΟΥ ΑΡΧΕΙΟΥ..... | 30 |
| 2.3 ΜΕΤΑΔΕΔΟΜΕΝΑ.....                                                          | 31 |
| 2.4 ΕΡΓΑΛΕΙΑ ΓΙΑ ΤΗ ΔΗΜΟΣΙΕΥΣΗ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ.....                     | 32 |
| 2.4.1 ΛΕΞΙΛΟΓΙΑ.....                                                           | 32 |
| 2.4.1.1 FOAF.....                                                              | 32 |

|                                                                                                                          |        |
|--------------------------------------------------------------------------------------------------------------------------|--------|
| 2.4.1.2 DC.....                                                                                                          | 33     |
| 2.4.1.3 SEMANTICALLY INTERLINKED ONLINE COMMUNITIES                                                                      | 33     |
| 2.4.1.4 DESCRIPTION OF A PROJECT.....                                                                                    | 34     |
| 2.4.1.5 SIMPLE KNOWLEDGE ORGANIZATION SYSTEMS.....                                                                       | 34     |
| 2.4.1.6 CREATIVE COMMONS.....                                                                                            | 34     |
| 2.5 LINKING OPEN DATA PROJECT.....                                                                                       | 35     |
| 2.6 ONTOLOGIES.....                                                                                                      | 37     |
| 2.7 ΕΡΓΑΛΕΙΑ ΚΑΤΑΣΚΕΥΗΣ ΟΝΤΟΛΟΓΙΩΝ.....                                                                                  | 38     |
| 2.8 ΓΛΩΣΣΕΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ ΟΝΤΟΛΟΓΙΩΝ.....                                                                                | 39     |
| 2.8.1 KIF.....                                                                                                           | 39     |
| 2.8.2 OWL.....                                                                                                           | 39     |
| <br>ΚΕΦΑΛΑΙΟ 3.....                                                                                                      | <br>43 |
| 3.1 ΕΡΓΑΛΕΙΑ ΓΙΑ ΤΗ ΔΗΜΟΣΙΕΥΣΗ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ                                                                    | 43     |
| 3.1.1 ΔΗΜΟΣΙΕΥΣΗ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ<br>ΣΧΕΣΙΑΚΕΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ.....                                           | 43     |
| 3.1.2 ΔΗΜΟΣΙΕΥΣΗ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΧΡΗΣΗ<br>ΕΦΑΡΜΟΓΗΣ ΠΕΡΙΤΥΛΙΓΜΑΤΟΣ ΑΠΟ ΥΠΑΡΧΟΥΣΕΣ ΕΦΑΡΜΟΓΕΣ<br>Η WEB APIS..... | 47     |
| 3.2 ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΩΝ.....                                                                                          | 48     |
| 3.2.1 TALIS ASPIRE.....                                                                                                  | 48     |
| 3.2.2 FACETES WIKIPEDIA SEARCH.....                                                                                      | 50     |
| 3.2.3 DBPEDIA MOBILE.....                                                                                                | 50     |
| 3.2.4 ΣΗΜΑΣΙΟΛΟΓΙΚΑ WIKIS.....                                                                                           | 52     |
| 3.2.4.1 ΠΩΣ ΕΠΙΤΥΓΧΑΝΟΝΤΑΙ ΤΑ ΣΗΜΑΣΙΟΛΟΓΙΚΑ<br>ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ.....                                                       | 52     |
| 3.2.4.2 ΑΠΟΔΟΣΗ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΣΧΟΛΙΑΣΜΟΥ.....                                                                           | 53     |
| 3.2.5 SEMANTIC MEDIAWIKIS.....                                                                                           | 53     |
| 3.2.6 MULTIMEDIAN E-CULTURE.....                                                                                         | 57     |
| 3.2.7 FREEBASE .....                                                                                                     | 57     |
| 3.2.8 GEONAMES .....                                                                                                     | 58     |
| <br>ΚΕΦΑΛΑΙΟ 4.....                                                                                                      | <br>62 |
| 4.1 ΣΗΜΑΣΙΟΛΟΓΙΚΕΣ ΜΗΧΑΝΕΣ ΑΝΑΖΗΤΗΣΗΣ.....                                                                               | 62     |



|                                               |           |
|-----------------------------------------------|-----------|
| <b>4.2 ΣΗΜΑΣΙΟΛΟΓΙΚΕΣ ΠΥΛΕΣ.....</b>          | <b>67</b> |
| <b>4.3 ΕΛΕΓΧΟΣ ΔΗΜΟΣΙΕΥΣΗΣ ΔΕΔΟΜΕΝΩΝ.....</b> | <b>69</b> |
| <b>ΣΥΜΠΕΡΑΣΜΑΤΑ.....</b>                      | <b>70</b> |
| <b>ΒΙΒΛΙΟΓΡΑΦΙΑ.....</b>                      | <b>72</b> |

## Ευρετήριο εικόνων

|                                                                                                                                                          |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Εικόνα 1. Κλασσικός Ιστός ή Ιστός των εγγράφων.....                                                                                                      | 12 |
| Εικόνα 2. Δομή σημασιολογικού Ιστού.....                                                                                                                 | 14 |
| Εικόνα 3. Οι Εκδοχές της Αρχιτεκτονικής του Semantic Web.....                                                                                            | 17 |
| Εικόνα 4. Επιλογές στην Δημοσίευση Συνδεδεμένων Δεδομένων.....                                                                                           | 28 |
| Εικόνα 5. Διάγραμμα-σύννεφο των Linking Open Data που δίνει μια γενική<br>εικόνα των δημοσιευμένων συνόλων δεδομένων και τις σχέσεις<br>μεταξύ τους..... | 34 |
| Εικόνα 6. Επίπεδα εκφραστικότητας OWL.....                                                                                                               | 41 |
| Εικόνα 7. ODEMapster επεξεργαστή αντιστοιχίσεων.....                                                                                                     | 42 |
| Εικόνα 8. D2R Server.....                                                                                                                                | 43 |
| Εικόνα 9. Δομή D2RQ map.....                                                                                                                             | 44 |
| Εικόνα 10. Η Αρχιτεκτονική του Open-Link Virtuoso.....                                                                                                   | 46 |
| Εικόνα 11. Αντιστοίχιση μ των μοτίβων URL σε σύνολα ερωτημάτων<br>SQL, στη διαμόρφωση του Triplify για το WordPress (κώδικας PHP)                        | 47 |
| Εικόνα 12. Διάγραμμα της λειτουργίας του RDF Book Mashup.....                                                                                            | 48 |
| Εικόνα 13. Η σύνδεση της εφαρμογής Talis Aspire συνδεδεμένα περιοδικά,<br>DBPedia και LCSH.....                                                          | 49 |
| Εικόνα 14. Talis Aspire: Μια ακαδημαϊκή λίστα πόρων.....                                                                                                 | 50 |
| Εικόνα 15. Talis Aspire: Απλοποιημένη Εφαρμογή Αρχιτεκτονικής.....                                                                                       | 50 |
| Εικόνα 16. Η DBPedia στο δίκτυο Διασυνδεδεμένων Δεδομένων.....                                                                                           | 51 |
| Εικόνα 17. Πληροφορίες για το Βερολίνο από την DBpedia Mobile.....                                                                                       | 52 |
| Εικόνα 18. Semantic MediaWiki logo.....                                                                                                                  | 54 |
| Εικόνα 19. Τα τμήματα οργάνωσης του semantic mediawiki.....                                                                                              | 55 |
| Εικόνα 20. Όψεις στα δεδομένα.....                                                                                                                       | 57 |
| Εικόνα 21. Γραφικό εργαλείο DartGrid για την διαδικασία Mapping.....                                                                                     | 62 |
| Εικόνα 22. Η λογική δομή του Wordnet .....                                                                                                               | 62 |
| Εικόνα 23. Η αρχιτεκτονική της μηχανής αναζήτησης WebOWL.....                                                                                            | 67 |

## ΕΙΣΑΓΩΓΗ

Κατά την διάρκεια των τελευταίων ετών, η μορφή του διαδικτύου αλλάζει. Από τον κλασσικό παγκόσμιο ιστό με τους υπερσυνδέσμους μεταβαίνουμε στον σημασιολογικό ιστό. Ένα από τα κεντρικά σημεία όπου επικεντρώνεται η κριτική για τον κλασσικό παγκόσμιο ιστό αποτελεί το γεγονός ότι η πληροφορία που αναρτάται σ' αυτόν βρίσκεται στην συντριπτική της πλειοψηφία σε μη δομημένη μορφή. Έτσι περνάμε στον σημασιολογικό ιστό και μεταβαίνουμε από την απλή δημοσίευση εγγράφων και δεδομένων, στην δημιουργία δομημένων εγγράφων και στην σύνδεσή τους (Bizer et al., 2008). Ο Σημασιολογικός Ιστός σαν οργανωμένη προσπάθεια, άρχισε να παίρνει σάρκα και οστά μετά τη δημοσίευση του Οδικού Χάρτη για το Σημασιολογικό Ιστό (Semantic Web Road Map) το Σεπτέμβριο του 1998 στο World Wide Web Consortium (εν συντομία W3C), της διεθνούς κοινοπραξίας που θέτει τα πρότυπα των τεχνολογιών επάνω στις οποίες βασίζεται ο Παγκόσμιος Ιστός. Παρέχει μέσω των προτύπων του την δυνατότητα περιγραφής της πληροφορίας σε σημασιολογικό και όχι μόνο σε συντακτικό επίπεδο, βάσει λεξιλογίων αναπαράστασης γνώσης ('οντολογιών') και καθίσταται δυνατός ο άμεσος και ακριβής εντοπισμός πληροφοριών καθώς και η ανάπτυξη προηγμένων εφαρμογών.

Το μεγαλύτερο μέρος σήμερα του σήμερα διαθέσιμου περιεχομένου στον Παγκόσμιο Ιστό είναι εξαιρετικά δύσκολο να επεξεργαστεί αυτόματα: "... τα δεδομένα που μεταδίδονται μέσω του Web είναι σε μεγάλο βαθμό αναλώσιμα δεδομένα τα οποία φαίνονται καλά, αλλά έχουν μικρή δομή".

Ο Παγκόσμιος Ιστός έχει εξελιχθεί από έναν παγκόσμιο χώρο πληροφοριών των συνδεδεμένων εγγράφων σε έναν χώρο όπου συνδέονται τα έγγραφα και τα δεδομένα. Η υποστήριξη αυτής της εξέλιξης είναι ένα σύνολο βέλτιστων πρακτικών για τη δημοσίευση και τη σύνδεση δομημένων δεδομένων στον Παγκόσμιο Ιστό γνωστή ως Συνδεδεμένα Δεδομένα ( Linked Data).

Το όραμα του Σημασιολογικού Ιστού είναι να καταστήσει τις διαθέσιμες πληροφορίες διαχειρίσιμες και εκμεταλλεύσιμες από μη-ανθρώπινο νου, δηλαδή τους υπολογιστές. Ο Παγκόσμιος Ιστός παρέχει άπειρο πλούτο σε πληροφορία που όμως είναι αναγνώσιμη και κατανοητή μόνο από τον άνθρωπο, καθώς η έμφαση δίνεται στο τρόπο παρουσίασής της παρά στη σημασιολογία της. Ο

Σημασιολογικός Ιστός φιλοδοξεί στην απονομή δομής και σημασιολογίας σε αδόμητη πληροφορία με στόχο τον διαμοιρασμό, την αποτελεσματική εύρεση και επαναχρησιμοποίηση της (Berners Lee, 1998). Με την μετεξέλιξη του Παγκόσμιου Ιστού σταδιακά στο Σημασιολογικό Παγκόσμιο Ιστό (Semantic Web) νέα μοντέλα και πρότυπα (XML, RDF, OWL) αναπτύσσονται για την προώθηση αυτής της διαδικασίας. Έτσι ανοίγονται νέοι ορίζοντες στη χρήση του Διαδικτύου με την έκφραση, μετάδοση και αναζήτηση πληροφοριών με χρήση αυτών των προτύπων. Το πλαίσιο RDF θεωρείται σημαντική συνεισφορά για την καθιέρωση του Σημασιολογικού Ιστού. Πρόκειται για ένα μοντέλο περιγραφής και επεξεργασίας μεταδεδομένων, δηλαδή δεδομένων για τα δεδομένα. Ο Σημασιολογικός Ιστός βασίζεται στις Οντολογίες. Μια Οντολογία έχει διττό ρόλο, από τη μια εξασφαλίζει ένα κοινό λεξιλόγιο και από την άλλη εξασφαλίζει μια κοινή κατανόηση του τομέα γνώσης τον οποίο μοντελοποιεί. Οι βασικές εφαρμογές των τεχνολογιών του Σημασιολογικού Ιστού είναι η σημασιολογική αναζήτηση σε μηχανές αναζήτησης, η χρήση πρακτόρων (software agents) για την εύρεση πληροφοριών και οι Σημασιολογικές Πύλες.

Η παρούσα εργασία αποτελείται από τέσσερα κεφάλαια. Στο πρώτο κεφάλαιο παρουσιάζονται τα συνδεδεμένα δεδομένα, η ιστορική τους εξέλιξη, η δομή του σημασιολογικού ιστού, γίνεται επίσης αναφορά στην τεχνολογία των συνδεδεμένων δεδομένων και στους κανόνες για τη δημοσίευση στοιχείων. Στο δεύτερο κεφάλαιο γίνεται αναφορά επί των βασικών αρχών συνδεδεμένων δεδομένων και της δημοσίευσής τους. Ο Σημασιολογικός Ιστός βασίζεται στις Οντολογίες. Μια Οντολογία είναι ένα σύνολο από δεδομένα με σημασιολογικό περιεχόμενο. Μια Οντολογία αποτελείται από έννοιες (ή κλάσεις), που είναι οτιδήποτε μπορεί να ειπωθεί κάτι για αυτό, από σχέσεις, που αναπαριστούν τις αλληλεπιδράσεις μεταξύ των εννοιών από αξιώματα, που περιγράφουν προτάσεις οι οποίες είναι πάντα αληθείς και από στιγμιότυπα που αναπαριστούν συγκεκριμένα στοιχεία. Στο τρίτο κεφάλαιο παρουσιάζεται αναλυτικά η δημοσίευση συνδεδεμένων δεδομένων με τη χρήση εργαλείων για τη δημοσίευσή τους. Τέλος, στο τέταρτο κεφάλαιο καταγράφονται οι σημασιολογικές μηχανές αναζήτησης, οι σημασιολογικές πύλες και παρουσιάζονται στοιχεία για τον έλεγχο δημοσίευσης δεδομένων.

## ΚΕΦΑΛΑΙΟ 1

### 1.1 ΤΙ ΕΙΝΑΙ ΣΥΝΔΕΔΕΜΕΝΑ ΔΕΔΟΜΕΝΑ

Τα Συνδεδεμένα Δεδομένα (Linked Data) αποτελούν έναν νέο τρόπο για την οργάνωση του "ιστού των πραγμάτων". Έτσι όλα τα αντικείμενα έχουν ένα URI στο web με τη χρήση του οποίου ο άνθρωπος όσο και τα προγράμματα μπορούν να αποκτήσουν δομημένη (structured) γνώση σχετική με το αντικείμενο. Η γνώση αυτή περιλαμβάνει συνδέσμους (links) προς άλλα αντικείμενα, δημιουργώντας ένα διασυνδεδεμένο σύννεφο δεδομένων (linked data cloud). Μαζί με μια τυποποιημένη γλώσσα ερωτημάτων (SPARQL), τα συνδεδεμένα δεδομένα αποτελούν ένα βήμα προς την ανάπτυξη «έξυπνων» εφαρμογών.

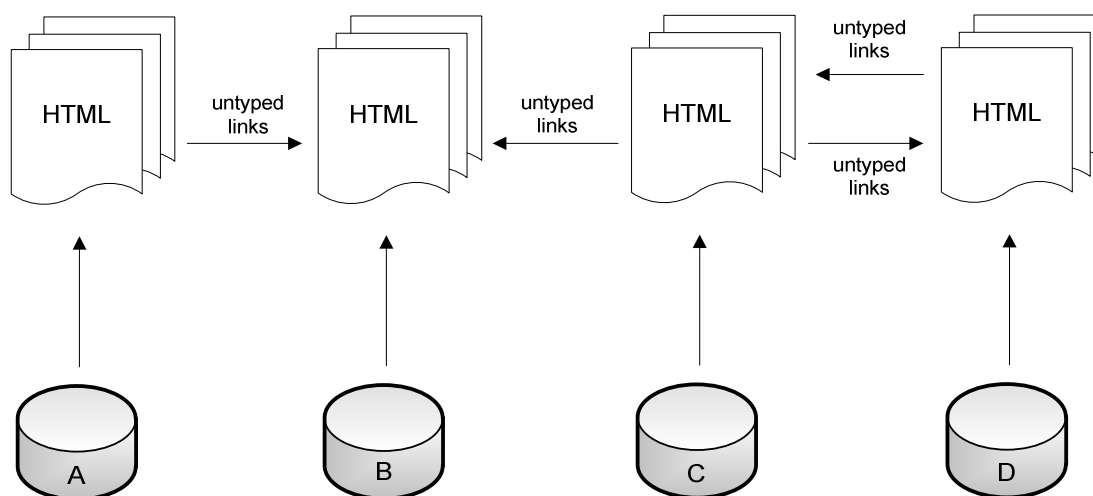
Ο όρος Συνδεδεμένα Δεδομένα αναφέρεται σε ένα σύνολο πρακτικών με στόχο την δημοσίευση και διασύνδεση δομημένων δεδομένων στον Ιστό. Οι πρακτικές αυτές σύμφωνα με τον Berners-Lee (2006), αποτελούν τη γέφυρα για τη μετάβαση από τον Ιστό των Εγγράφων στον Ιστό των Δεδομένων, ή Σημασιολογικό Ιστό.

Τεχνικά, τα συνδεδεμένα δεδομένα (Linked Data) αναφέρονται σε στοιχεία που δημοσιεύονται στο Διαδίκτυο με τέτοιο τρόπο ώστε να είναι αναγνώσιμα από μηχανήματα, η σημασία τους να ορίζεται ρητά και να συνδέονται με άλλα εξωτερικά σύνολα δεδομένων.

### 1.2 ΙΣΤΟΡΙΚΗ ΕΞΕΛΙΞΗ

Ο Berners-Lee κατά το 1989 εργαζόμενος στα εργαστήρια του CERN της Γενεύης, σκέφτηκε να δημιουργήσει ένα σύστημα το οποίο θα επέτρεπε σε ερευνητές που βρίσκονταν σε απομακρυσμένα μεταξύ τους μέρη σε όλο τον κόσμο να οργανώσουν και να συνδέσουν τον όγκο των εγγράφων που διέθεταν. Στο στάδιο αυτό πρότεινε την τοποθέτηση συνδέσμων (links) μέσα στο περιεχόμενο (κείμενο) του κάθε εγγράφου ώστε ο κάθε ένας να μπορεί να έχει πρόσβαση σε οποιοδήποτε έγγραφο. Στα πρώτα χρόνια (περίπου από τις αρχές του 90, έως το 2000), οι περισσότεροι άνθρωποι ήταν απασχολημένοι με την ανάπτυξη λογισμικού όπως οι πράκτορες χρηστών και server, καθώς και περιεχομένου για τον Ιστό των Εγγράφων (Web of Documents).

Στο παρακάτω σχήμα (Εικόνα 1) παρουσιάζεται εποπτικά η μορφή του κλασσικού Ιστού ή αλλιώς Ιστού των Εγγράφων:



Εικόνα 1. Κλασσικός Ιστός ή Ιστός των εγγράφων

Έτσι ο κλασσικός παγκόσμιος ιστός έχει αναπτυχθεί πάνω στην ιδέα της δημιουργίας υπερ-συνδέσεων μεταξύ εγγράφων(που αποκαλούνται Web documents ή Web pages ή HTML pages) του ιστού. Αυτές οι υπερσυνδέσεις αποτελούν τη βάση για την πλοήγηση στον ιστό. Ενοποιούν όλα τα έγγραφα σε ένα μοναδικό παγκόσμιο χώρο. Με απλά λόγια, ο παγκόσμιος ιστός έχει αυξηθεί, επειδή είναι εύκολο να γράψει κανείς μια ιστοσελίδα και εύκολα να την συνδέσεις με άλλες σελίδες. Η ιστορία της ανάπτυξης του Παγκόσμιου Ιστού μπορεί να μετρηθεί με τον αριθμό των ιστοσελίδων που έχουν δημοσιευτεί και τον αριθμό των συνδέσεων μεταξύ των ιστοσελίδων που υπάρχουν.

Σε μια τόσο ξαφνική και απότομη ανάπτυξη του δικτύου και αύξηση του όγκου των πληροφοριών που μεταδίδονται, τέθηκαν ζητήματα αναζήτησης, διαχωρισμού και επεξεργασίας των επιθυμητών πληροφοριών για τους χρήστες. Στον τομέα της επεξεργασίας και ερμηνείας των επιστρεφόμενων πληροφοριών η αυτόματη μηχανική υποστήριξη είναι περιορισμένη και επομένως τη λειτουργία αυτή πρέπει να επιτελέσει ο άνθρωπος-χρήστης. Οι μηχανές αναζήτησης είναι αδύνατο να συνδέσουν μεταξύ τους συσχετισμένα γεγονότα τα οποία εμφανίζονται σε διαφορετικές ιστοσελίδες. Αν και υπάρχει ένας τεράστιος όγκος διαθέσιμων δεδομένων και στοιχείων, εν τούτοις δε διαθέτουν καμία γνώση σχετικά με το περιεχόμενο μιας ιστοσελίδας με αποτέλεσμα να μην μπορούν να

μετατρέψουν τα δεδομένα αυτά σε πληροφορία χρήσιμη για το χρήστη. Αυτό έχει ως επακόλουθο κόστος σε χρόνο και κόπο, δυσχρηστία και περιορισμό του εξεταζόμενου μεγέθους εισερχομένων πληροφοριών κατά τις απαιτήσεις του κάθε χρήστη.

Έτσι ο Tim Berners-Lee έθεσε την ιδέα και τα θεμέλια για την δημιουργία του Σημασιολογικού Ιστού μέσω του οποίου τα περιεχόμενα του WWW να προσπελαύνονται και να ερμηνεύονται από μηχανές. Όπως ισχυρίζεται ο ίδιος, ο Σημασιολογικός Ιστός αποτελεί μία επέκταση του σημερινού Ιστού, στον οποίο η πληροφορία βρίσκεται σε κατανοητή από τις μηχανές μορφή, διευκολύνοντας τη συνεργασία των υπολογιστών, αλλά και των ανθρώπων. Αντίθετα με τον παραδοσιακό Ιστό, ο Σημασιολογικός Ιστός δίνει τη δυνατότητα τοποθέτησης κατάλληλων «ετικετών» στα ίδια τα δεδομένα και στη συνέχεια την περιγραφή τους καθώς και τη δημιουργία συνδέσεων μεταξύ τους. Έτσι ο Ιστός μετατρέπεται σε μία παγκόσμια βάση δεδομένων, ένα ενιαίο δίκτυο με δομικά στοιχεία τα ίδια τα δεδομένα και όχι τα έγγραφα που τα παρουσιάζουν. Οι μηχανές αναζήτησης θα γίνουν πιο αποτελεσματικές απ' ότι είναι τώρα και οι χρήστες θα μπορούν να βρουν τις ακριβείς πληροφορίες που ψάχνουν. Η προσθήκη σημασιολογικής πληροφορίας στα δεδομένα, η οποία είναι κατανοητή από τις μηχανές, ανοίγει το δρόμο για την ανάπτυξη «έξυπνων» εφαρμογών που εκμεταλλεύονται κάποιο κομμάτι όλης αυτής της σημασιολογικής πληροφορίας για την εκτέλεση πολύπλοκων λειτουργιών. Οι πρακτικές οι οποίες προτάθηκαν από τον Tim Berners-Lee για τη δημοσίευση και τη διασύνδεση δομημένων δεδομένων στον Παγκόσμιο Ιστό αποτελούν έκφραση του όρου Συνδεδεμένα Δεδομένα (Linked Data). Μέσω των Συνδεδεμένων Δεδομένων (Linked Data) το διαδίκτυο των Ιστοσελίδων μεταλλάσσεται σε ένα διαδίκτυο δεδομένων που επικοινωνούν σημασιολογικά.

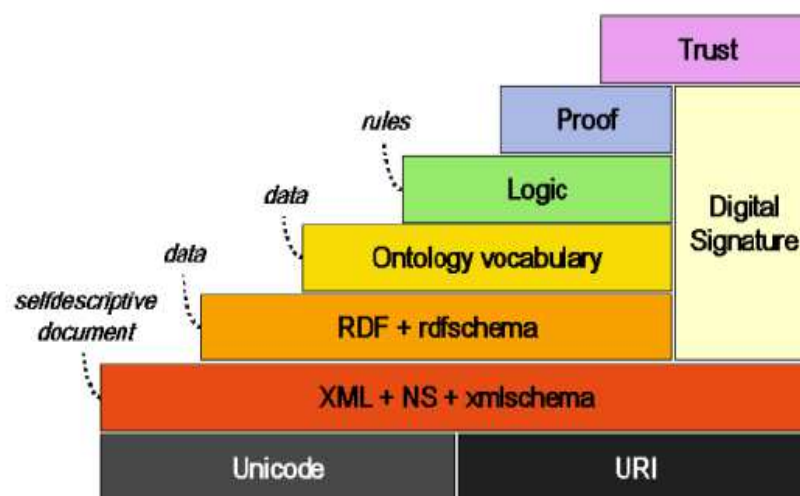
Ο Σημασιολογικός Ιστός είναι μια τεχνολογία για την ανταλλαγή δεδομένων, όπως το υπερκείμενο Web είναι για κοινή χρήση εγγράφων. Ο παγκόσμιος ιστός των Hypertext Markup Language (HTML) εγγράφων εξαπλώνεται πολύ γρήγορα, εν μέρει επειδή ο συντάκτης ενός εγγράφου HTML ή ενός προγράμματος, που παράγει HTML έγγραφα, μπορεί να δει αμέσως τα αποτελέσματα της εργασίας του. Περαιτέρω, με την σύνδεση σε άλλους επιλεγόμενους πόρους, θα μπορούσε να προσθέσει άμεση αξία στη σελίδα του. Αυτό δεν συμβαίνει στον ίδιο βαθμό με τα RDF δεδομένα στον Παγκόσμιο Ιστό.

Ο Ιστός των δεδομένων μπορεί να προσπελαστεί με μηχανές περιήγησης συνδεδεμένων δεδομένων, ακριβώς όπως ο παραδοσιακός ιστός εγγράφων είναι προσβάσιμος με τη χρήση HTML browsers. Ωστόσο, αντί να ακολουθούν δεσμούς μεταξύ HTML σελίδες, οι μηχανές περιήγησης συνδεδεμένων δεδομένων επιτρέπουν στους χρήστες να πλοηγούνται μεταξύ διαφορετικών πηγών δεδομένων ακολουθώντας RDF συνδέσμους. Αυτό επιτρέπει τον χρήστη να ξεκινήσει με μία πηγή δεδομένων και στη συνέχεια να προχωρήσει μέσα από δυνητικά ατελείωτες πηγών δεδομένων που συνδέονται με RDF συνδέσμους. Ενώ οι πρωτογενείς μονάδες του υπερκειμένου Web είναι τα HTML (HyperText Markup Language) έγγραφα που συνδέονται με untyped hyperlinks, τα συνδεδεμένα δεδομένα βασίζονται σε έγγραφα που περιέχουν δεδομένα σε RDF (Resource Description Framework) format (Klyne και Carroll, 2004).

Ένας σύνδεσμος RDF δηλώνει απλά ότι ένα κομμάτι δεδομένων έχει κάποια σχέση με ένα άλλο κομμάτι των δεδομένων. Αυτές οι σχέσεις μπορούν να έχουν διαφορετικούς τύπους. Για παράδειγμα, ένας σύνδεσμος RDF που συνδέει δεδομένα σχετικά με άτομα μπορεί να δηλώσει ότι δύο άτομα γνωρίζουν ο ένας τον άλλον. Επίσης ένας RDF σύνδεσμος που συνδέει πληροφορίες σχετικά με ένα πρόσωπο με πληροφορίες για δημοσιεύσεις σε μια βιβλιογραφική βάση δεδομένων, μπορεί να ισχυριστεί ότι ένα άτομο είναι ο συντάκτης ενός ειδικού εγγράφου.

### 1.3 ΔΟΜΗ ΤΟΥ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΙΣΤΟΥ

Η δομή του σημασιολογικού ιστού όπως την πρότεινε ο Tim Berners-Lee φαίνεται στην Εικόνα 2





## Εικόνα 2. Δομή σημασιολογικού Ιστού

Στην βάση της δομής βρίσκονται τα Unicode και uri μέσω των οποίων τονίζεται η σημασία ενός κοινού συντακτικού. Το Unicode αποτελεί ένα πρότυπο για την ανταλλαγή συμβόλων ενώ τα URIs σχετίζονται με την αναφορά σε οντότητες και αποτελούν το πρότυπο για την αναγνώριση και τον εντοπισμό ενός ονόματος ή μιας πηγής (π.χ. μιας ιστοσελίδας) σε ένα δίκτυο, με τη χρήση μιας προκαθορισμένης μορφής απόδοσης ονομάτων. Στο επόμενο επίπεδο βρίσκεται η XML (eXtensible Markup Language), η οποία είναι μια γλώσσα για τη δόμηση δεδομένων. Το XML έγγραφο είναι πολύ πιο εύκολα προσβάσιμο από υπολογιστές, μιας και κάθε κομμάτι πληροφορίας που περιέχει περιγράφεται κατάλληλα από μεταδεδομένα. Το XMLSchema αποτελεί ένα πλαίσιο για τον καθορισμό γραμματικών κανόνων με σκοπό την δημιουργία «έγκυρων» XML αρχείων. Αποτελεί μία σημαντικά πιο πλούσια γλώσσα σε σχέση με το Document Type Definition, για τον προσδιορισμό της δομής XML εγγράφων. Ένα XML Schema επιτρέπει τη δημιουργία νέων τύπων με την επέκταση ή τον περιορισμό υφιστάμενων σχημάτων. Σε συνδυασμό με τη βασισμένη σε XML σύνταξη, αυτό το χαρακτηριστικό καθιστά δυνατή τη δημιουργία XML σχημάτων από άλλα σχήματα που έχουν δημιουργηθεί ήδη, με αποτέλεσμα τη μείωση του συνολικού χρόνου και κόστους. Τέλος, το XML Schema παρέχει ένα πλούσιο σετ από τύπους δεδομένων που μπορούν να χρησιμοποιηθούν σε έγγραφα XML, ενώ τα DTDs επιτρέπουν μόνο τη χρήση στοιχειοσειρών (strings). Οι τεχνολογίες-πρότυπα των δύο αυτών βασικών επιπέδων είναι ευρύτερα αποδεκτές με αποτέλεσμα ο αριθμός των XML αρχείων να αυξάνεται διαρκώς. Στο τρίτο κατά σειρά επίπεδο βρίσκεται το RDF (Resource Description Framework) το οποίο αποτελεί την αρχή του Σημασιολογικού Ιστού. Σύμφωνα με το W3C, το RDF ένα θεμέλιο για την επεξεργασία μεταδεδομένων, το οποίο παρέχει την δυνατότητα διαλειτουργικότητας (interoperability) μεταξύ των εφαρμογών οι οποίες ανταλλάσσουν πληροφορία στο Παγκόσμιο Ιστό. Το RDF βασίζεται στην ιδέα ότι τα πράγματα που περιγράφονται έχουν κάποιες ιδιότητες, οι οποίες με τη σειρά τους έχουν κάποιες τιμές και ότι οι πηγές δεδομένων ή πόροι (resources), μπορούν να περιγραφούν με προτάσεις οι οποίες καθορίζουν αυτές τις ιδιότητες και τιμές.

Το RDF Schema καθορίζει μια απλή γλώσσα επάνω στο RDF. Παρέχει έναν τρόπο για την περιγραφή πηγών πληροφοριών, με τη χρήση απλών προτάσεων. Το RDF και το RDF Schema είναι κωδικοποιημένα με XML σύνταξη αλλά δεν εφαρμόζουν υποχρεωτικά τη δενδρική δομή της XML.

Στο παραπάνω επίπεδο βρίσκονται οι Οντολογίες. Συγκεκριμένα, οι οντολογίες χρησιμοποιούνται για να επεξηγήσουν τους όρους και τις έννοιες που περιγράφουν και αναπαριστούν τη γνώση πάνω σε ένα συγκεκριμένο τομέα π.χ. στην πληροφορική. Οι γλώσσες οντολογιών, δίνουν τη δυνατότητα στους χρήστες να αναπαριστούν με ξεκάθαρο τρόπο domain models. Οι κύριες απαιτήσεις από αυτές είναι :

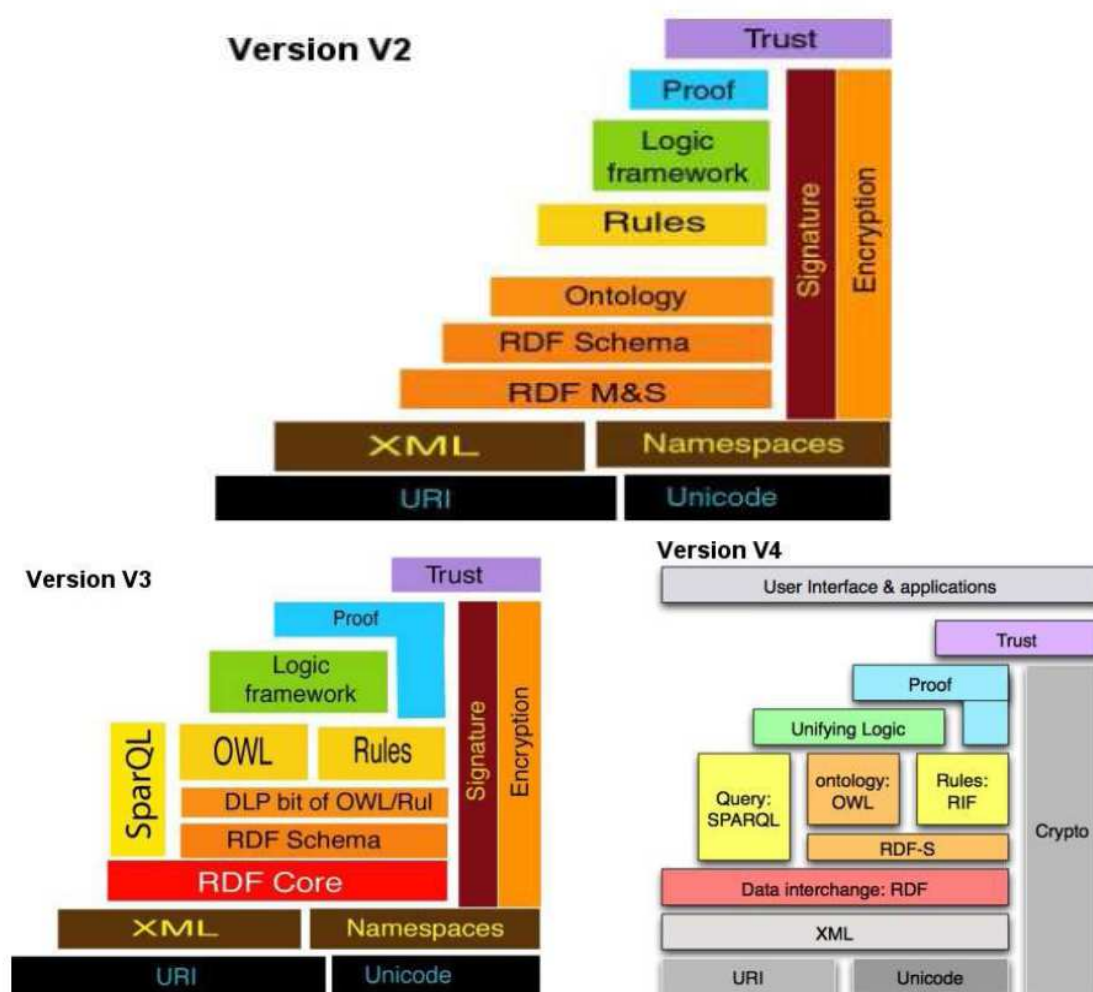
- Καλά-ορισμένη σύνταξη (well-defined syntax),
- Τυπική σημασιολογία (formal semantics).
- Αποτελεσματική υποστήριξη συλλογισμών (efficient reasoning support).
- Επαρκή εκφραστική δύναμη (efficient expressive power).
- Ευκολία στην έκφραση (convenience of expression).

Η λογική τοποθετείται στο επόμενο επίπεδο όπου με την εφαρμογή της θα μπορούσε να εξαχθεί νέα γνώση από την πληροφορία που δηλώνεται ρητά. Γενικά, η λογική προσφέρει, πρώτα απ' όλα, τυπικές γλώσσες (formal languages) για την αναπαράσταση γνώσης. Δεύτερο, η λογική παρέχει αυτό που αναφέραμε και νωρίτερα ως τυπική σημασιολογία (formal semantics), δηλαδή το νόημα των προτάσεων ορίζεται ρητά χωρίς περιθώρια παρερμηνείας. Τρίτο, παρέχει τη δυνατότητα αυτόματης εξαγωγής συμπερασμάτων από δεδομένη γνώση, μετατρέποντας έτσι την άρρητη (implicit) γνώση σε ρητή (explicit).

Στα δύο ανώτερα επίπεδα τοποθετούνται η τεκμηρίωση και η εμπιστοσύνη. Καλύπτουν την ανάγκη για έλεγχο της εγκυρότητας των δηλώσεων (statements) που γίνονται στο Σημασιολογικό ιστό. Το επίπεδο της τεκμηρίωσης (proof), το οποίο βοηθά ώστε τα αποτελέσματα που συμπεραίνονται από δεδομένα στο Semantic Web να μπορούν να οδηγήσουν πίσω στις υποθέσεις που τα προκάλεσαν. Για παράδειγμα, αν κάποιος στείλει σε μια σελίδα A την απόδειξη ότι μπορεί να τη χρησιμοποιήσει τότε η σελίδα A θα πρέπει να μπορεί να ελέγξει και να επαληθεύσει την ύπαρξη της απόδειξης. Το επίπεδο της εμπιστοσύνης (trust), όπου σε συνδυασμό με την τεχνολογία των ψηφιακών

υπογραφών (digital signatures), θα εξασφαλίζει, με αυτοματοποιημένο τρόπο, το βαθμό στον οποίο οι πληροφορίες που διακινούνται, επεξεργάζονται και συμπεραίνονται στο Semantic Web είναι αξιόπιστες.

Οι άλλες διαφορετικές εκδοχές της διαστρωμάτωσης των τεχνολογιών στις οποίες βασίζεται η εξέλιξη του Semantic Web (Gerber et al., 2008), όπως προκύπτουν με την εξέλιξη του όρου του Semantic Web από την πρώτη του εμφάνιση, είναι οι παρακάτω:



Εικόνα 3 : Οι Εκδοχές της Αρχιτεκτονικής του Semantic Web

#### 1.4 Η ΤΕΧΝΟΛΟΓΙΑ ΤΩΝ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

Τα συνδεδεμένα δεδομένα στηρίζονται σε δύο τεχνολογίες που έχουν θεμελιώδη σημασία για το Web: Uniform Resource identifiers (URIs) (Berners-Lee et al., 2005) και το πρωτόκολλο Hypertext Transfer Protocol (HTTP) (Fielding et al., 1999). Ενώ τα URLs (Uniform Resource Locators), έχουν γίνει γνωστά ως

διευθύνσεις για έγγραφα και άλλες οντότητες που μπορούν να βρίσκονται στο Web, τα URIs ( Uniform Resource Identifiers) παρέχουν ένα πιο γενικό μέσο για τον εντοπισμό κάθε οντότητας που υπάρχει στην κόσμο.

Τα URIs και HTTP συμπληρώνονται από μια τεχνολογία που είναι κρίσιμη για τον Ιστό των Δεδομένων, την RDF. Ενώ οι πρωτογενείς μονάδες του υπερκειμένου Web είναι τα HTML (HyperText Markup Language) έγγραφα που συνδέονται με untyped hyperlinks, τα συνδεδεμένα δεδομένα βασίζονται σε έγγραφα που περιέχουν δεδομένα σε RDF (Resource Description Framework) format (Klyne και Carroll, 2004).

Η RDF είναι μία γλώσσα, ή ένα μοντέλο δεδομένων, για την αναπαράσταση πληροφορίας σχετικής με πόρους στον Παγκόσμιο Ιστό. Χρησιμοποιείται κυρίως για την αναπαράσταση μεταδεδομένων σχετικά με πόρους του Ιστού, όπως είναι για παράδειγμα ο τίτλος και ο συγγραφέας μίας ιστοσελίδας, ή τα πνευματικά δικαιώματα ενός εγγράφου στον Ιστό. Όπως ακριβώς το παραδοσιακό Διαδίκτυο των HTML εγγράφων μπορεί να σαρωθεί από crawlers ακολουθώντας τους υπερσύνδεσμους ανάμεσα στα έγγραφα, έτσι γίνεται και με τον Ιστό των Δεδομένων ακολουθώντας όμως RDF συνδέσμους. Από τα δεδομένα που προκύπτουν, οι μηχανές αναζήτησης μπορούν να παρέχουν εκλεπτυσμένες δυνατότητες αναζήτησης, παρόμοιες με αυτές που παρέχονται από τις συμβατικές σχεσιακές βάσεις δεδομένων.

Η RDF μπορεί να χρησιμοποιηθεί και για την περιγραφή πραγμάτων, τα οποία μπορούν να προσδιοριστούν στον Ιστό, αλλά δεν μπορούν να ανακτηθούν άμεσα μέσα σε αυτόν. Τέτοιες περιπτώσεις μεταδεδομένων είναι οι πληροφορίες για τα προϊόντα ενός on-line καταστήματος

Η διαφοροποίηση της RDF από την HTML (και ο βασικός λόγος χρήσης της) είναι ότι η RDF είναι κατανοητή από μηχανές, ενώ η HTML όχι.

Οι βασικές έννοιες της RDF είναι τρεις : οι πόροι, οι ιδιότητες και οι προτάσεις (Αντωνίου και Harmelen, 2009).

Οι πόροι είναι όλα εκείνα τα «πράγματα» για τα οποία θέλουμε να μιλήσουμε, είτε έχουν κάποια αναπαράσταση μέσα στον Ιστό, είτε όχι. Σε κάθε πόρο αντιστοιχεί ένα αναγνωριστικό, το URI του, το οποίο μπορεί να είναι μία διεύθυνση ιστού (URL) ή κάποιο άλλο είδος μοναδικού αναγνωριστικού. Οι ιδιότητες είναι και αυτές μια κατηγορία πόρων που περιγράφουν τις σχέσεις μεταξύ πόρων. Τέτοιες σχέσεις είναι οι : «τίτλος», «δημιουργός», «ημερομηνία γέννησης», κ.α.

Καθορίζονται και αυτές επίσης από URIs, γεγονός το οποίο προσφέρει μία μέθοδο παγκόσμιας, μοναδικής ονομασίας μειώνοντας δραστικά το πρόβλημα της ομοιότητας των ονομάτων για την αναπαράσταση των κατανεμημένων δεδομένων μέχρι σήμερα. Πρόταση είναι μία τριάδα (triple) υποκειμένου - κατηγορήματος - αντικείμενου (subject - predicate - object). Η τριάδα αυτή αντικατοπτρίζει τη βασική δομή μιας απλής πρότασης όπως φαίνεται παρακάτω:

|              |                              |                       |
|--------------|------------------------------|-----------------------|
| <b>Chris</b> | has the <b>email address</b> | <b>chris@bizer.de</b> |
| (υποκείμενο) | (κατηγορημα)                 | (αντικείμενο)         |

Υποκείμενο είναι ένας πόρος, κατηγορημα είναι μία ιδιότητα και αντικείμενο είναι είτε ένας πόρος είτε ένα λεκτικό (literal). Τα λεκτικά είναι αλφαριθμητικά (strings). Μία πρόταση δηλώνει ότι ο πόρος-υποκείμενο συνδέεται με τη σχέση-κατηγορημα με κάποιον άλλο πόρο-αντικείμενο ή ότι ο πόρος-υποκείμενο συνδέεται με τη σχέση-κατηγορημα με κάποιο αλφαριθμητικό-αντικείμενο, το οποίο είναι η λεκτική αναπαράσταση της τιμής που έχει κάποια ιδιότητά του (το κατηγορημα).

Οι σύνδεσμοι RDF (Bizer & Cyganiak & Heath, 2007) παίρνουν τη μορφή της τριάδας όπου το θέμα της είναι μια αναφορά URI στο χώρο ονομάτων ενός συνόλου δεδομένων, ενώ το αντικείμενο της, είναι μια αναφορά URI σε άλλο σύνολο δεδομένων.

Το σύνολο των τριάδων RDF αντιμετωπίζεται από κάποιους ως ένας γράφος RDF. Το URIs που εμφανίζονται ως υποκείμενο και αντικείμενο URIs είναι οι κόμβοι στον γράφο, και κάθε τριάδα είναι ένα κατευθυνόμενο τόξο (βέλος) που συνδέει το υποκείμενο με το αντικείμενο.

#### 1.4.1 ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΤΟΥ ΜΟΝΤΕΛΟΥ RDF

Η χρησιμοποίηση του μοντέλου RDF παρουσιάζει τα παρακάτω πλεονεκτήματα :

- Κάθε χρήστης μπορεί να επισκεφθεί την URI ενός RDF γράφου στον Ιστό για να ανακτήσει επιπλέον πληροφορίες.
- Οι πληροφορίες από διαφορετικές μέσα μπορούν να συγχωνευθούν με φυσικό τρόπο.
- Μέσω του μοντέλου RDF μπορεί ο καθένας να θέσει συνδέσμους (links) μεταξύ δεδομένων που προέρχονται από διαφορετικές πηγές.

- Συνδυάζοντας το μοντέλο με άλλες σχηματικές γλώσσες όπως η RDF-S και η OWL, μπορούμε να αναπαραστήσουμε αυστηρά δομημένα δεδομένα καθώς και ημι-δομημένα δεδομένα (Bizer et al, 2007).

#### 1.4.2 ΠΛΗΡΟΦΟΡΙΑΚΟΙ ΚΑΙ ΜΗ ΠΛΗΡΟΦΟΡΙΑΚΟΙ ΠΟΡΟΙ

Στην ορολογία της Αρχιτεκτονικής του Ιστού, όλα τα αντικείμενα του ενδιαφέροντός μας καλούνται πόροι. Στον ιστό των δεδομένων υπάρχουν πόροι οι οποίοι διακρίνονται σε πληροφοριακούς (informational resources) και μη-πληροφοριακούς (non-informational resources) ή αλλιώς «άλλους» πόρους. Ο διαχωρισμός αυτός είναι πολύ σημαντικός για τα Συνδεδεμένα Δεδομένα.

Πληροφοριακοί είναι οι πόροι των οποίων, όλα τα ουσιώδη χαρακτηριστικά, μπορούν να μεταδοθούν μέσα σε ένα μήνυμα. Όλοι οι πόροι που βρίσκουμε στον παραδοσιακό Ιστό των Εγγράφων, όπως εικόνες, έγγραφα και άλλα αρχεία (πολυ)μέσων, είναι πληροφοριακοί πόροι. Όλοι οι άλλοι πόροι εκτός των πληροφοριακών είναι οι μη-πληροφοριακοί πόροι. Σε αυτή την κατηγορία ανήκουν οι επιστημονικές ιδέες, τοποθεσίες και γενικά όλα τα φυσικά αντικείμενα του κόσμου μας και γενικά οτιδήποτε δεν περιλαμβάνεται στο χώρο πληροφορίας του Ιστού.

Οι πληροφοριακοί πόροι μπορούν να έχουν αναπαραστάσεις, οι οποίες μπορούν να προσπελαστούν χρησιμοποιώντας το πρωτόκολλο HTTP. Αυτές οι αναπαραστάσεις μπορούν να μεταδοθούν μέσα σε κάποιο μήνυμα. (και έτσι ικανοποιείται ο ορισμός των πληροφοριακών πόρων).

#### 1.4.3 DEREFERENCING ΤΩΝ HTTP URIS

Στην περίπτωση που κάποιος χρήστης ανατρέξει στο URI κάποιου πόρου, θα πρέπει να πάρει σαν απάντηση πληροφορίες για τον πόρο αυτόν. Υπάρχει διαφοροποίηση όσον αφορά τους πληροφοριακούς και τους μη-πληροφοριακούς πόρους.

**Πληροφοριακοί Πόροι:** Όταν κάποιος πελάτης επισκεφτεί το URI κάποιου πληροφοριακού πόρου, ο εξυπηρετητής του ιδιοκτήτη του URI αυτού παράγει μία καινούρια αναπαράσταση της τρέχουσας κατάστασης του πόρου και την στέλνει πίσω στον πελάτη χρησιμοποιώντας τον HTTP κωδικό 200 OK.

Έτσι, το dereferencing του URI ενός πληροφοριακού πόρου μπορεί να γίνει είτε άμεσα (επιστροφή εγγράφου που αντιστοιχεί στο ζητούμενο URI), είτε έμμεσα (επιστροφή άλλου εγγράφου ή με την ανακατεύθυνση σε άλλο έγγραφο το οποίο αντιστοιχεί ακριβότερα στις απαιτήσεις του).

**Μη Πληροφοριακοί Πόροι:** Δεν μπορούμε να επισκεφθούμε τους μη-πληροφοριακούς πόρους απ' ευθείας. Στην περίπτωση αυτή μπορούμε να χρησιμοποιήσουμε τον ίδιο τον Ιστό, ένα εύρωστο και κλιμακούμενο σύστημα δημοσίευσης πληροφοριών. Ο εξυπηρετητής στέλνει στον πελάτη το URI ενός πληροφοριακού πόρου που περιγράφει τον μη-πληροφοριακό πόρο, χρησιμοποιώντας τον HTTP κωδικό απόκρισης 303 *See Other*. Αυτό ονομάζεται 303 ανακατεύθυνση. Σε ένα δεύτερο βήμα, ο πελάτης επισκέπτεται αυτό το νέο URI και παίρνει μια αναπαράσταση που περιγράφει τον αρχικό μη-πληροφοριακό πόρο.

Επίσης μια άλλη προσέγγιση είναι η παρακάτω: τα URIs με δέση (#), ή αλλιώς hash URIs (Sauermann & Cyganiak,2008). Τα URIs αυτά περιέχουν το σύμβολο της δέσης '#', το οποίο τα χωρίζει σε δύο μέρη, το μέρος πριν το '#' και το μέρος μετά. Όταν ένας πελάτης αναζητά κάποιο hash URI, το πρωτόκολλο HTTP επιβάλλει την αφαίρεση του τμήματος μετά το '#' πριν την αποστολή της αίτησης στον εξυπηρετητή. Με αυτό τον τρόπο, η αναζήτηση του URI ενός μη-πληροφοριακού πόρου μεταφράζεται στην αναζήτηση του URI ενός πληροφοριακού πόρου που περιγράφει τον αρχικό πόρο .

#### **1.4.4 RDF Schema**

Η RDF Schema, ή αλλιώς RDFS, είναι μία γλώσσα για την περιγραφή των RDF λεξιλογίων, ή αλλιώς των οντολογιών. Η RDF είναι ανεξάρτητη πεδίου, με την έννοια ότι δεν έχει γίνει καμία υπόθεση σχετικά με κάποιο συγκεκριμένο πεδίο χρήσης της. Οι χρήστες είναι υπεύθυνοι για τον ορισμό της δικής τους ορολογίας, μέσω της RDFS. Συγκεκριμένα, η RDFS προσφέρει μηχανισμούς για την περιγραφή ομάδων σχετικών πόρων, που ονομάζονται κλάσεις, και για την περιγραφή των σχέσεων μεταξύ των πόρων. Οι RDFS περιγραφές των RDF λεξιλογίων γράφονται σε RDF και ένα έγγραφο RDFS (ένα σχήμα RDF) δεν είναι παρά ένα έγγραφο RDF. Στη συνέχεια παραθέτουμε τα βασικά στοιχεία μοντελοποίησης της RDFS.

### **Βασικές κλάσεις**

- `rdfs:Resource`, η κλάση όλων των πόρων.
- `rdfs:Class`, η κλάση όλων των κλάσεων (ακόμα και του εαυτού της).
- `rdfs:Literal`, η κλάση όλων των λεκτικών.
- `rdf:Property`, η κλάση όλων των ιδιοτήτων.
- `rdf:Statement`, η κλάση όλων των υποστασιοποιημένων προτάσεων

Ένα βασικό βήμα σε κάθε είδους διαδικασία περιγραφής, είναι ο προσδιορισμός των διαφόρων ειδών πραγμάτων που πρόκειται να περιγραφούν. Το RDF Schema αναφέρεται σε αυτά τα «είδη πραγμάτων» ως *κλάσεις*. Μία κλάση στο RDF Schema αντιστοιχεί στη γενική έννοια ενός *Τύπου* ή μίας *Κατηγορίας*, όπως περίπου η έννοια της κλάσης στις αντικειμενοστραφείς γλώσσες προγραμματισμού. Οι κλάσεις του RDF μπορούν να χρησιμοποιηθούν για να αντιπροσωπεύσουν σχεδόν οποιαδήποτε κατηγορία πράγματος, όπως ιστοσελίδες, ανθρώπους, τύπους εγγράφων, βάσεις δεδομένων ή αφηρημένες έννοιες.

### **Βασικές ιδιότητες για τον ορισμό σχέσεων**

- `rdf:type`, η ιδιότητα που συνδέει έναν πόρο με κάποια κλάση στην οποία ανήκει, ή ισοδύναμα συνδέει έναν πόρο με κάποια κλάση της οποίας είναι στιγμιότυπο.
- `rdfs:subClassOf`, η ιδιότητα που συσχετίζει μία κλάση με τις υπερκλάσεις της. Σημειώνουμε ότι στην RDFS επιτρέπεται η πολλαπλή κληρονομικότητα (multiple inheritance).
- `rdfs:subPropertyOf`, η ιδιότητα που συσχετίζει μία ιδιότητα με τις υπεριδιότητές της.

Τόσο το υποκείμενο όσο και το αντικείμενο των RDF τριπλετών αποτελούν σημασιολογικούς πόρους (resources) που περιγράφονται μοναδικά μέσω ενός URI. Το RDF είναι απλά ένα πλαίσιο αναπαράστασης της γνώσεως και RDFS είναι το πρότυπο που περιγράφει πώς μπορεί να χρησιμοποιηθεί το RDF προκειμένου να περιγραφεί περιεχόμενο του Διαδικτύου. Τα πάντα στο RDFS



είναι URIs, μια αντίληψη που διατηρήθηκε σε όλη την πορεία ανάπτυξης του Σημασιολογικού Ιστού.

### **1.5 ΚΑΝΟΝΕΣ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ**

Ο Berners-Lee (2006) παρουσίασε μια σειρά από «κανόνες» για τη δημοσίευση στοιχείων σχετικά στο Web με τέτοιο τρόπο ώστε όλα τα δημοσιευμένα στοιχεία να γίνουν μέρος ενός ενιαίου παγκόσμιου χώρου δεδομένων. Το 2009 ο ίδιος τροποποίησε ελαφρώς τους κανόνες, προσθέτοντας ρητά το Resource Description Framework (RDF) στον τρίτο κανόνα, ενώ οι συζητήσεις για το εάν το RDF είναι πράγματι ουσιώδες μέρος των Συνδεδεμένων Δεδομένων συνεχίζονται. Οι κανόνες αυτοί είναι:

1. Χρήση URIs για την αναγνώριση όλων των αντικειμένων.
2. Χρήση HTTP URIs έτσι ώστε οι άνθρωποι να μπορούν να αναζητήσουν αυτά τα αντικείμενα που έχουν αναγνωριστεί.
3. Όταν κάποιος αναζητά ένα URI, να παρέχονται χρήσιμες πληροφορίες, χρησιμοποιώντας τα πρότυπα (RDF, SPARQL).
4. Συμπερίληψη και άλλων συνδέσμων σε άλλα URIs έτσι ώστε να μπορούν να ανακαλύψουν περισσότερα πράγματα.

Σύμφωνα με τον πρώτο κανόνα οτιδήποτε υπάρχει στον κόσμο μας, ένα έγγραφο (π.χ. μία HTML σελίδα), κάποιο υπαρκτό πρόσωπο ή αντικείμενο, μία ιδέα, ή γενικά οτιδήποτε έχει αξία σαν πόρος, μπορεί να ταυτοποιηθεί από ένα μοναδικό URI (Uniform Resource Identifier – Καθολικό Αναγνωριστικό Πόρου) και έτσι να εισέλθει στον Ιστό των Δεδομένων.

Σύμφωνα με τον δεύτερο κανόνα πρέπει να χρησιμοποιούνται HTTP URIs, ώστε τα URIs να είναι dereferenceable μέσω του πρωτοκόλλου HTTP. Dereferenceable URI ονομάζεται ένα URI το οποίο αν αναζητηθεί από κάποιο χρήστη (π.χ. μέσω ενός browser), επιστρέφει στο χρήστη πληροφορίες που αναφέρονται στον πόρο που ταυτοποιεί. Η διαδικασία αυτή λέγεται dereferencing. Στην πράξη πολλοί πάροχοι συνδεδεμένων δεδομένων δεν προσφέρουν dereferenceable URIs.

Σύμφωνα με τον τρίτο κανόνα οι πάροχοι συνδεδεμένων δεδομένων θα πρέπει να προσφέρουν τις πληροφορίες που σχετίζονται με τα δεδομένα τους με τη χρήση της RDF και της SPARQL.

Σύμφωνα με τον τέταρτο κανόνα στον ιστό των Συνδεδεμένων Δεδομένων ένας πόρος πρέπει να συνδέεται με άλλους πόρους του Ιστού ώστε να έχουμε έναν ενιαίο, συνεκτικό Ιστό στον οποίο κάποιος μπορεί να περιηγηθεί και να βρει οτιδήποτε αναζητήσει.

### 2.1 ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

Οι βασικές αρχές των συνδεδεμένων δεδομένων είναι οι εξής:

1. Χρήση του RDF μοντέλου δεδομένων για τη δημοσίευση δομημένων δεδομένων στον Παγκόσμιο Ιστό.
2. Χρήση RDF συνδέσεων για την διασύνδεση δεδομένων από διαφορετικές πηγές.

Εφαρμόζοντας τις δύο αρχές που οδηγούν στη δημιουργία δεδομένων στο διαδίκτυο, οι άνθρωποι και οι οργανώσεις μπορούν να δημοσιεύουν και να καταναλώνουν δεδομένα για κάθε θέμα. Αυτά τα δεδομένα συχνά αποκαλούνται ως Ιστός Δεδομένων ή Σημασιολογικός Ιστός.

### 2.2 ΔΗΜΟΣΙΕΥΣΗ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

Με τη δημοσίευση δεδομένων στον Ιστό σύμφωνα με τους κανόνες των Συνδεδεμένων Δεδομένων, οι πάροχοι δεδομένων προσθέτουν τα δεδομένα τους σε ένα παγκόσμιο χώρο δεδομένων, ο οποίος επιτρέπει στα δεδομένα να ανακαλυφθούν και να χρησιμοποιηθούν από διάφορες εφαρμογές. Η δημοσίευση ενός συνόλου δεδομένων ως Συνδεδεμένα Δεδομένα στον Ιστό συνίσταται στα επόμενα τρία βασικά βήματα:

1. Ανάθεση URIs στις οντότητες οι οποίες περιγράφονται από το σύνολο δεδομένων και παροχή δυνατότητας για dereferencing αυτών των URIs πάνω από το πρωτόκολλο HTTP με RDF αναπαραστάσεις.
2. Τοποθέτηση συνδέσμων RDF προς άλλες πηγές δεδομένων στον Ιστό, ούτως ώστε οι πελάτες να μπορούν να περιηγηθούν συνολικά στον Ιστό των Δεδομένων, ακολουθώντας RDF συνδέσμους.
3. Παροχή μετα-δεδομένων σχετικά με τα δημοσιευμένα δεδομένα, έτσι ώστε οι πελάτες να μπορούν να εκτιμήσουν την ποιότητα των δημοσιευμένων δεδομένων και να επιλέξουν μεταξύ διαφορετικών μέσων πρόσβασης.

Η απλούστερη μέθοδος για δημοσίευση Συνδεδεμένων Δεδομένων είναι η δημιουργία στατικών RDF αρχείων και η φόρτωσή τους σε κάποιον εξυπηρετητή. Η πρακτική αυτή χρησιμοποιείται σε περιπτώσεις όπου τα RDF αρχεία δημιουργούνται με το χέρι, π.χ. κατά τη δημοσίευση FOAF αρχείων ή RDF

λεξιλογίων, ή σε περιπτώσεις που τα RDF αρχεία είναι έξοδος εφαρμογών οι οποίες δεν μπορούν να δώσουν έξοδο σε άλλη μορφή, παρά μόνο σε αρχείο.

Οι πάροχοι των δεδομένων μπορούν να επιλέξουν ανάμεσα σε δύο HTTP πρότυπα χρήσης URI για τον εντοπισμό οντοτήτων: το 303 URIs και hash URIs. Και τα δύο μοντέλα διασφαλίζουν ότι οι πελάτες μπορούν να διακρίνουν μεταξύ URIs που αναγνωρίζουν πραγματικές οντότητες και URIs που προσδιορίζουν τα έγγραφα του Ιστού περιγράφοντας αυτές τις πραγματικές οντότητες (Sauermann & Cyganiak, 2008).

Έχουν αναπτυχθεί διάφορα εργαλεία δημοσίευσης Συνδεδεμένων Δεδομένων. Τα εργαλεία είτε επιδίδουν το περιεχόμενο RDF αποθηκών ως Συνδεδεμένα Δεδομένα στον Ιστό είτε παρέχουν όψεις Συνδεδεμένων Δεδομένων σε μη-RDF πηγές δεδομένων. Τα εργαλεία «προστατεύουν» τους εκδότες από την αντιμετώπιση τεχνικών λεπτομερειών όπως η διαπραγμάτευση περιεχομένου και διασφαλίζουν ότι τα δεδομένα δημοσιεύονται σύμφωνα με τις καλύτερες πρακτικές της κοινότητας των Συνδεδεμένων Δεδομένων. Όλα τα εργαλεία υποστηρίζουν το dereferencing των URIs σε RDF περιγραφές. Επιπλέον, κάποια από τα εργαλεία παρέχουν και πρόσβαση σε ερωτήματα SPARQL προς τα σύνολα δεδομένων και υποστηρίζουν τη δημοσίευση RDF αποθηκών.

### **2.2.1 Δημοσίευση Συνδεδεμένων Δεδομένων από στατικά RDF αρχεία**

Η απλούστερη μέθοδος για δημοσίευση Συνδεδεμένων Δεδομένων είναι η δημιουργία στατικών RDF αρχείων και η φόρτωσή τους σε κάποιον εξυπηρετητή. Η πρακτική αυτή χρησιμοποιείται σε περιπτώσεις όπου τα RDF αρχεία δημιουργούνται με το χέρι, π.χ. κατά τη δημοσίευση FOAF αρχείων ή RDF λεξιλογίων, ή σε περιπτώσεις που τα RDF αρχεία είναι έξοδος εφαρμογών οι οποίες δεν μπορούν να δώσουν έξοδο σε άλλη μορφή, παρά μόνο σε αρχείο. Παρακάτω παρουσιάζεται η αναπαράσταση της RDF περιγραφής σχήμα σε XML σειριακή διάταξη.

#### Αναπαράσταση σε RDF/XML :

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:extterms="http://www.myexample.org/terms/">
  <rdf:Description rdf:about="http://www.myexample.org/index.html">
    <extterms:creation-date>August 17, 1987</extterms:creation-date>
    <dc:language>en</dc:language>
    <dc:creator
      rdf:resource="http://www.myexample.org/people/alice" />
  </rdf:Description>
</rdf:RDF>
```

#### Αναπαράσταση σε Notation 3 ή Turtle :

```
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix extterms: < http://www.myexample.org/terms/>.

<http://www.myexample.org/index.html>
  extterms:creation-date "August 17, 1987";
  dc:language "en";
  dc:creator <http://www.myexample.org/people/alice>.
```

#### Αναπαράσταση σε N-Triples :

```
<http://www.myexample.org/index.html>
<http://www.myexample.org/terms/creation-date> "August 17, 1987" .
<http://www.myexample.org/index.html>
<http://purl.org/dc/elements/1.1/language> "en" .
<http://www.myexample.org/index.html>
<http://purl.org/dc/elements/1.1/creator>
<http://www.myexample.org/people/alice> .
```

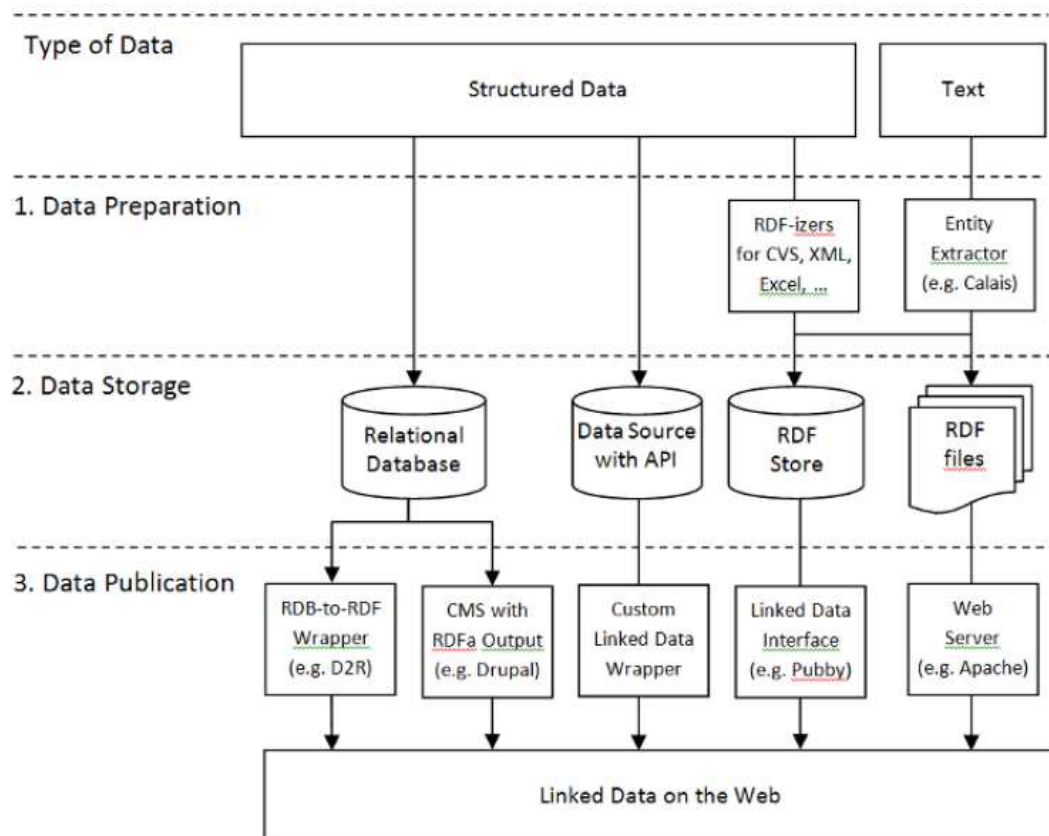
### **2.2.2 Δημοσίευση Συνδεδεμένων Δεδομένων από σχεσιακές βάσεις δεδομένων**

Σε περιπτώσεις όπου τα δεδομένα μας είναι ήδη αποθηκευμένα σε κάποια βάση δεδομένων, είναι συνήθως καλή ιδέα να τα αφήσουμε εκεί και να δημοσιεύσουμε σαν Συνδεδεμένα Δεδομένα μία όψη (view) της βάσης μας. Ένα εργαλείο για δημοσίευση Συνδεδεμένων Δεδομένων με αυτό τον τρόπο είναι ο D2R Server. Ο D2R Server είναι ένας εξυπηρετητής ο οποίος, βασιζόμενος σε μία αντιστοίχιση μεταξύ των σχημάτων της σχεσιακής βάσης δεδομένων και κάποιων RDF όρων, δημοσιεύει μία όψη της βάσης σε μορφή Συνδεδεμένων Δεδομένων και προσφέρει έναν SPARQL ακροδέκτη.

### 2.2.3 Δημοσίευση άλλου τύπου πληροφορίας σε μορφή Συνδεδεμένων Δεδομένων

Όταν θέλουμε να δημοσιεύσουμε Συνδεδεμένα Δεδομένα τα οποία όμως αναπαριστούνται σε άλλες μορφές, όπως CSV ή Microsoft Excel, είναι βέλτιστο να ακολουθηθεί η ακόλουθη μεθοδολογία :

- Μετατροπή των δεδομένων σε RDF με κάποιο πρόγραμμα που γράφουμε οι ίδιοι ή χρησιμοποιώντας κάποιο εργαλείο που εξειδικεύεται σε αυτή τη λειτουργία.
- Μετά τη μετατροπή αποθήκευση των δεδομένων σε κάποια αποθήκη τριάδων<sup>1</sup>.
- Μια αποθήκη τριάδων περιλαμβάνει και μία διαπροσωπεία Συνδεδεμένων Δεδομένων έτσι ώστε τα δεδομένα που είναι αποθηκευμένα σε αυτή να είναι προσπελάσιμα μέσα από τον Ιστό.



Εικόνα 4 : Επιλογές στην Δημοσίευση Συνδεδεμένων Δεδομένων

<sup>1</sup> Μπορούμε να βρούμε μία λίστα από αποθήκες τριάδων στο <http://esw.w3.org/SemanticWebTools>

## 2.3 ΜΕΤΑΔΕΔΟΜΕΝΑ

Τα μεταδεδομένα είναι δομημένα δεδομένα τα οποία περιγράφουν τα χαρακτηριστικά ενός αντικειμένου. Είναι δεδομένα που σχετίζονται με αντικείμενα προς διευκόλυνση των πιθανών χρηστών τους παρέχοντας εκ των προτέρων πλήρη γνώση της ύπαρξης και των χαρακτηριστικών τους. Το πρόθεμα «μετα» υποδηλώνει περιγραφή σε κάποιο ανώτερο επίπεδο ή πιο βασικό. Τα χαρακτηριστικά του κάθε σχήματος μεταδεδομένων αποτελούν έναν περιορισμένο αριθμό στοιχείων, τα ονόματά τους και τη σημασία τους.

Συνήθως, η σημασιολογική περιγραφή αναφέρεται στα περιεχόμενα, το χώρο ύπαρξης, τις φυσικές ιδιότητες, τον τύπο (π.χ. κείμενο ή εικόνα, χάρτης ή μοντέλο) και τη μορφή (π.χ. εκτυπωμένο αντίγραφο, ηλεκτρονικό αρχείο). Χαρακτηριστικά στοιχεία μεταδεδομένων που υποστηρίζουν την πρόσβαση σε έγγραφα περιλαμβάνουν τον συγγραφέα, το χρόνο και τόπο που εκδόθηκε καθώς και τις θεματικές ενότητες που καλύπτονται. Εάν το υλικό των πληροφοριών βρίσκεται σε μη ψηφιακή μορφή (π.χ. εκτυπωμένη έκδοση) παρέχονται επιπρόσθετα μεταδεδομένα που βοηθούν στον εντοπισμό του υλικού αυτού, όπως για παράδειγμα τηλέφωνα βιβλιοθηκών.

Τα μεταδεδομένα αποτελούν τον συνδετικό κρίκο μεταξύ των δημιουργών πληροφοριών και των χρηστών των πληροφοριών αυτών. Εκτός από τη βελτίωση της εύρεσης δεδομένων τα μεταδεδομένα αναπτύσσονται και για άλλους λόγους περιλαμβανομένων της ασφάλειας, της προσωποποίησης των πληροφοριών, της αξιολόγησης του περιεχομένου, τη διαχείριση πνευματικών δικαιωμάτων, τη διατήρηση πολύτιμης γνώσης σε βάθος χρόνου.

Η κοινότητα που παρέχει τις πληροφορίες μπορεί επίσης να καθορίσει και μια λογική ομαδοποίηση των στοιχείων ή να αφήσει τη διαδικασία αυτή στο εκάστοτε σχήμα μεταδεδομένων. Για παράδειγμα, το σχήμα Dublin Core μπορεί να παρέχει τον πυρήνα των διαδικασιών με τις οποίες μπορούν να γίνουν προσθήκες νέων στοιχείων.

Μερικά δημοφιλή σχήματα μεταδεδομένων είναι τα Dublin Core, AACR2 (Anglo-American Cataloging Rules), GILS (Government Information Locator Service), EAD (Encoded Archives Description), IMS (IMS Global Learning Consortium), AGLS (Australian Government Locator Service).

## **2.4 ΕΡΓΑΛΕΙΑ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ**

### **2.4.1 ΛΕΞΙΛΟΓΙΑ**

Διαφορετικές κοινότητες έχουν συγκεκριμένες προτιμήσεις όσον αφορά τα λεξιλόγια που χρησιμοποιούν για τη δημοσίευση δεδομένων στον Ιστό. Ο Ιστός των Δεδομένων είναι επομένως ανοιχτός σε αυθαίρετα λεξιλόγια που χρησιμοποιούνται παράλληλα. Παρά το γεγονός αυτό, θεωρείται καλή πρακτική η επαναχρησιμοποίηση όρων από γνωστά RDF λεξιλόγια όπως:

- Friend-of-a-Friend (FOAF), λεξιλόγιο για την περιγραφή ανθρώπων.
- Dublin Core (DC) περιγράφει χαρακτηριστικά των μετα-δεδομένων.
- Semantically-Interlinked Online Communities ( SIOC), λεξιλόγιο για την αναπαράσταση online κοινοτήτων.
- Description of a Project ( DOAP), λεξιλόγιο για την περιγραφή projects.
- Simple Knowledge Organization System ( SKOS), λεξιλόγιο για την αναπαράσταση ταξινομιών και αόριστα δομημένης γνώσης.
- Music Ontology παρέχει όρους για την περιγραφή καλλιτεχνών, άλμπουμ και τραγουδιών.
- Review Vocabulary, λεξιλόγιο για την αναπαράσταση σχολίων.
- Creative Commons (CC), λεξιλόγιο για την περιγραφή όρων αδειών (Licence Terms)

#### **2.4.1.1 FOAF**

Το λεξιλόγιο αυτό έχει σχεδιαστεί με τέτοιο τρόπο ώστε να καταστήσει κατανοητή από τον υπολογιστή την πληροφορία που υπάρχει σε προσωπικές ιστοσελίδες ή επίσης ιστοσελίδες ομάδων, εταιρειών, οργανισμών κ.ο.κ. Η βασική ιδέα του έργου είναι ότι η πληροφορία των σελίδων που αφορούν ανθρώπους είναι συνήθως της ίδιας σημασίας. Αν επομένως, πολλοί άνθρωποι δημιουργούν τις ιστοσελίδες τους σε FOAF, θα μπορεί να δημιουργηθεί μια Βάση Γνώσεως με αυτούς τους ανθρώπους. Η φιλοδοξία της FOAF είναι ότι αν πολύς κόσμος ακολουθήσει την γλώσσα αυτή, θα είναι δυνατή η εξαγωγή πληροφορίας μέσω προγραμμάτων των οποίων θα μπορούν να διασχίζουν τις σελίδες αυτές και να διατηρούν λίστες σε άλλες σελίδες.



```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <foaf:Person>
    <foaf:name>Peter Parker</foaf:name>
    <foaf:mbox rdf:resource="mailto:peter.parker@dailybugle.com"/>
  </foaf:Person>
```

#### 2.4.1.2 DC

Το Σύνολο μεταδεδομένων DC (Dublin Core) είναι ένα πρότυπο λεξιλόγιο για crossdomain πληροφορία. Χρησιμοποιείται ευρέως για την περιγραφή ψηφιακού περιεχομένου όπως video, ήχος, εικόνα, κείμενο αλλά και σύνθετα πολυμέσα όπως οι ιστοσελίδες. Οι υλοποιήσεις του είναι τυπικά βασισμένες σε XML και RDF. Το Dublin Core είναι ορισμένο από τον ISO, στο ISO Standard 15836 και το NISO Standard Z.39.85-2007.

Το πρότυπο Dublin Core περιλαμβάνει δύο επίπεδα: το Απλό και το Qualified. Το απλό Dublin Core αποτελείται από δεκαπέντε στοιχεία ενώ το Qualified Dublin Core περιέχει τρία στοιχεία επιπλέον: τα στοιχεία Audience, Provenance και RightsHolder, καθώς και ένα σύνολο από εκλεπτύνσεις στοιχείων που λέγονται Qualifiers και που εκλεπτύνουν τη σημασιολογία των στοιχείων με χρήσιμους για την ανακάλυψη των πόρων τρόπους.

#### 2.4.1.3 Semantically-Interlinked Online Communities ( SIOC)

Το SIOC project έχει σαν στόχο την δημιουργία μιας διασύνδεσης ανάμεσα σε δεδομένα που παρατίθενται στο Διαδίκτυο online (π.χ. από forums, message boards κτλ.). Με βάση ένα δικό του είδος «οντολογίας» (ontology) που έχει δημιουργήσει, παρέχει την δυνατότητα μετατροπής των δεδομένων αυτών σε μορφή RDF. Το αποτέλεσμα της χρησιμοποίησης του SIOC project σε εφαρμογές και ιστοσελίδες, είναι η πραγμάτωση μιας «γέφυρας» ανάμεσα στις παραπάνω εφαρμογές, η οποία θα συνδέει δεδομένα τα οποία σχετίζονται μεταξύ τους, αλλά παρουσιάζονται σε διαφορετικές ιστοσελίδες παράθεσης δεδομένων .

#### **2.4.1.4 Description of a Project ( DOAP)**

Το λεξιλόγιο αυτό έχει σχεδιαστεί για να περιγράφει έργα ανοιχτού κώδικα.

#### **2.4.1.5 Simple Knowledge Organization System ( SKOS)**

Το SKOS ή Simple Knowledge Organization Systems αποτελεί μια κατηγορία των τυπικών γλωσσών οι οποίες σχεδιάζονται για την αναπαράσταση θησαυρών, σχεδίων ταξινόμησης, ταξονομιών, συστημάτων προσυνδυασμένης ευρετηρίασης, ή οποιουδήποτε άλλου τύπου δομημένου ελεγχόμενου λεξιλογίου. Το SKOS δομείται επάνω σε RDF και RDFS, και ο κύριος στόχος τους είναι να επιτρέψει τη δημοσίευση ελεγχόμενων δομημένων λεξιλογίων στο σημασιολογικό ιστό με ένα τρόπο ευκολότερο από τις γλώσσες οντολογιών. Το SKOS αναπτύσσεται αυτήν την περίοδο στο πλαίσιο του W3C.

#### **2.4.1.6 Creative Commons (CC)**

Η πρωτοβουλία Creative Commons ανέπτυξε μια σειρά από «άδειες» τις οποίες οι συγγραφείς μπορούν να επιλέξουν για να περιγράψουν τους όρους σύμφωνα με τους οποίους μπορεί να χρησιμοποιηθεί η δουλειά τους. Οι άδειες αυτές επιτρέπουν την παρουσίαση, τη διάθεση, την αναπαραγωγή και διανομή ενός έργου, ενώ παρέχουν στο δημιουργό και τη δυνατότητα να θέσει τέσσερις επιπλέον περιορισμούς: απόδοσης, μη εμπορικής χρήσης, καμίας παράγωγης εργασίας ή αδειοδότησης για παράγωγη εργασία με κάποιους όρους (αυτό σημαίνει πως τα τεκμήρια αυτά πρέπει να παρέχονται στο κοινό με τους ίδιους όρους που ισχύουν και στα πρωτότυπα). Οι δημιουργοί επιλέγουν τους περιορισμούς που επιθυμούν να ισχύουν για τα έργα τους.

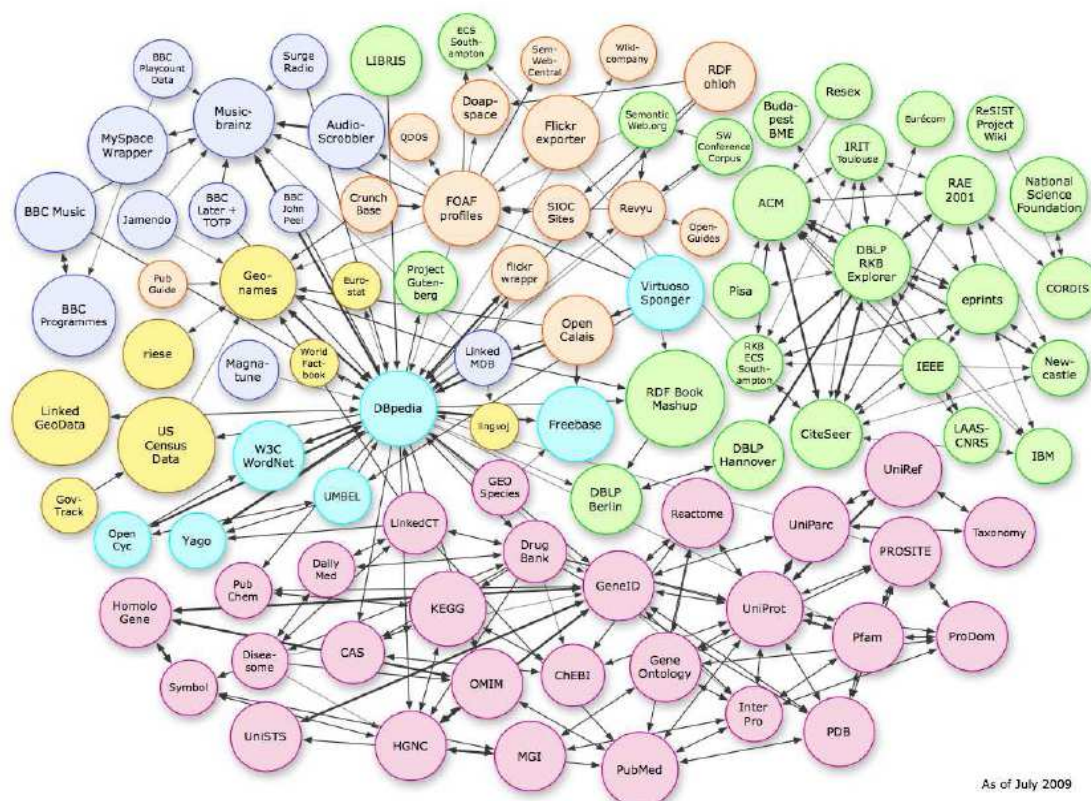
Στην περίπτωση που δεν μπορούν να βρεθούν καλά υπάρχοντα λεξιλόγια που να καλύπτουν όλες τις απαιτούμενες κλάσεις και ιδιότητες, τότε απαιτείται ο προσδιορισμός νέων όρων. Ο προσδιορισμός νέων όρων δεν είναι δύσκολος. Οι RDF κλάσεις και ιδιότητες είναι και οι ίδιες πόροι που ταυτοποιούνται με URIs και δημοσιεύονται στον Ιστό, επομένως ό,τι έχει ειπωθεί για τη δημοσίευση Συνδεδεμένων Δεδομένων εφαρμόζεται και σε αυτά επίσης. Ο ορισμός λεξιλογίων μπορεί να γίνει χρησιμοποιώντας την RDF Vocabulary Description Language 1.0: RDF Schema ή την Web Ontology Language (OWL).

## 2.5 Linking Open Data project

Το πιο εμφανές παράδειγμα υιοθέτησης και εφαρμογής των κανόνων των Συνδεδεμένων Δεδομένων είναι το project «Σύνδεση Ανοιχτών Δεδομένων» (Linking Open Data - LOD), μια κοινοτική προσπάθεια που θεμελιώθηκε τον Ιανουάριο του 2007 από τους Chris Bizer και Richard Cyganiak και υποστηρίχθηκε από το W3C Semantic Web Education και το Outreach Group. Το εγχείρημα αυτό ενθαρρύνει τους ανθρώπους να αναφέρονται σε συγκεκριμένα αντικείμενα (όπως είναι μια ταινία, ένα πρόσωπο ή ένα εστιατόριο) με ένα πρότυπο Κοινό Αναγνωριστικό Πόρου (URI), έτσι ώστε ο καθένας να ξέρει πότε δύο άνθρωποι μιλάνε για το ίδιο αντικείμενο. Ο αρχικός και τεκταινόμενος σκοπός του project είναι να συνυφάνει τον Ιστό των Δεδομένων αναγνωρίζοντας τα υπάρχοντα σύνολα δεδομένων, τα οποία διατίθενται ανοιχτά, μετατρέποντάς τα σε RDF σύμφωνα με τους κανόνες των Συνδεδεμένων Δεδομένων και δημοσιεύοντάς τα στον Ιστό.

Τα μεγαλύτερα διασυνδεδεμένα σύνολα δεδομένων του Ιστού ήταν αρχικά η DBpedia<sup>5</sup> (91 εκατομμύρια τριάδες), το Geonames<sup>6</sup> (60 εκατομμύρια τριάδες), το Musicbrainz<sup>7</sup> (50 εκατομμύρια τριάδες), ο DBtune music server<sup>8</sup> (4 εκατομμύρια τριάδες), η βιβλιογραφία DBLP<sup>9</sup> (15 εκατομμύρια τριάδες), οι αξιολογήσεις Renvu<sup>10</sup> (15 χιλιάδες τριάδες), το σύνολο δεδομένων απογραφής των ΗΠΑ (700 εκατομμύρια τριάδες) και το RDF Book Mashup<sup>11</sup> (αρκετά δισεκατομμύρια τριάδες).

Καθοριστικό ρόλο στην προσπάθεια μετατροπής του σημερινού παγκόσμιου ιστού από Ιστό Εγγράφων (*Web of Documents*) σε Ιστό Δεδομένων (*Web of Data*) κατέχει το *Linked Open Data Project (LOD)*. Στο LDP, εκατοντάδες σύνολα δεδομένων (datasets) από διάφορους οργανισμούς και υπηρεσίες ανά τον κόσμο δημοσιεύονται και συνδέονται μεταξύ τους ακολουθώντας το μοντέλο *RDF (Resource Description Framework)*. Ενδεικτικό του εύρους και της κλίμακας του Ιστού Δεδομένων που προήλθε από το Linking Open Data project φαίνεται στην Εικόνα 5. Κάθε κόμβος στο διάγραμμα - σύννεφο αναπαριστά ένα ξεχωριστό σύνολο δεδομένων που δημοσιεύεται ως Συνδεδεμένα Δεδομένα.



Εικόνα 5. Διάγραμμα-σύννεφο των Linking Open Data που δίνει μια γενική εικόνα των δημοσιευμένων συνόλων δεδομένων και τις σχέσεις μεταξύ τους,

Όπως φαίνεται στο σύννεφο των Συνδεδεμένων Ανοιχτών Δεδομένων υπάρχουν κάποιοι κεντρικοί κόμβοι, με τους οποίους συνδέονται αρκετά από τα υπόλοιπα σύνολα δεδομένων. Ο κεντρικότερος από όλους αυτούς τους κόμβους είναι η DBpedia, ένα από τα πρώτα σύνολα δεδομένων που εμφανίστηκαν στον Ιστό των Συνδεδεμένων Ανοιχτών Δεδομένων. Η DBpedia εξάγει πληροφορίες από τα infoboxes της Wikipedia και τις δημοσιεύει ως Συνδεδεμένα Δεδομένα. Infoboxes είναι τα πλαίσια που βρίσκονται στην πάνω δεξιά πλευρά των σελίδων της Wikipedia και περιέχουν μερικές δομημένες πληροφορίες σχετικά με το αντικείμενο του άρθρου. Επίσης ένα άλλο βασικό σύνολο δεδομένων στο σύννεφο των Συνδεδεμένων Ανοιχτών Δεδομένων είναι το Geonames, το οποίο δημοσιεύει περιγραφές εκατομμυρίων γεωγραφικών τοποθεσιών από όλο τον κόσμο σε μορφή Συνδεδεμένων Ανοιχτών Δεδομένων. Οι δύο αυτοί κόμβοι δηλ. η DBpedia και το Geonames προσφέρουν URIs και πληροφορίες που αναφέρονται σε πολλά από τα αντικείμενα του κόσμου μας στα οποία συχνά θέλουμε να αναφερθούμε. Αυτός είναι και ο λόγος που εξελίχθηκαν σε δύο κεντρικούς κόμβους στο σύννεφο των Συνδεδεμένων Ανοιχτών Δεδομένων.

Η δημοσίευση σε μορφή Συνδεδεμένων Δεδομένων των δεδομένων που περιέχονται σε ένα σύνολο δεδομένων συνδέεται άμεσα με το dereferencing των URIs των πόρων που περιέχονται στο σύνολο αυτό. Εκτός από τη διαδικασία του dereferencing, ένα σύνολο δεδομένων προσφέρει στους πελάτες τη δυνατότητα να εξάγουν πληροφορίες για τους πόρους που τους ενδιαφέρουν και με άλλους τρόπους. Σε αυτούς περιλαμβάνεται η προσφορά ολόκληρης της αποθήκης (dump) και η δυνατότητα σύνδεσης με κάποιο SPARQL ακροδέκτη (SPARQL endpoint), ο οποίος δέχεται από τους πελάτες SPARQL ερωτήματα σχετικά με τα δεδομένα του συγκεκριμένου συνόλου δεδομένων, τα εκτελεί και επιστρέφει σε αυτούς τις απαντήσεις. Ένα σύνολο δεδομένων μπορεί να προσφέρει τα δεδομένα του με έναν ή περισσότερους από τους προαναφερθέντες τρόπους.

Εκτός από τη δημοσίευση και τη διασύνδεση συνόλων δεδομένων, γίνονται και άλλες εργασίες μέσα στα πλαίσια των Συνδεδεμένων Ανοιχτών Δεδομένων, όπως είναι η ανάπτυξη φυλλομετρητών (browsers), ανιχνευτών (crawlers), μηχανών αναζήτησης (search engines) και άλλων εφαρμογών που δρουν πάνω σε Συνδεδεμένα Δεδομένα.

## **2.6 ΟΝΤΟΛΟΓΙΕΣ**

Με την ανάπτυξη του Παγκόσμιου Ιστού δημιουργήθηκε η ανάγκη για επαναχρησιμοποίηση και διαμοιρασμό της γνώσης. Οι οντολογίες στοχεύουν στην εκπλήρωση αυτής της ανάγκης μετατοπίζοντας το σημείο εστίασης στο σχεδιασμό συστημάτων μακριά από τις τεχνολογίες υλοποίησης αλλά προς λύσεις οι οποίες λειτουργούν τόσο στο επίπεδο της τεχνολογίας όσο και στο επίπεδο της γνώσης.

Την τελευταία δεκαετία το ενδιαφέρον για οντολογίες μεταβλήθηκε από αναδυόμενο επιστημονικό θέμα έρευνας σε κυρίαρχο ενδιαφέρον της επιστήμης των υπολογιστών και της ανάπτυξης λογισμικού. Η μεταβολή αυτή οφείλεται κυρίως στην έξαρση της ανάπτυξης του Παγκόσμιου Ιστού και της ανάγκης για μεγαλύτερη συνεργασία και αλληλεπιδραστικότητα μεταξύ υπολογιστικών μηχανών και προγραμμάτων, γεγονός που τονίζεται και στο όραμα του Σημασιολογικού Ιστού από τον Tim B. Lee

Το κίνητρο για την ανάπτυξη, εφαρμογή και διάδοση οντολογιών είναι η καταγραφή γνώσης με έναν ξεκάθαρο και διακριτό τρόπο, δυνατό για επεξεργασία και ανάγνωση τόσο από ανθρώπους όσο και από μηχανές,

απελευθερωμένης από τις λεπτομέρειες χιλιάδων γραμμών κώδικα συστήματος. Επιπλέον, οι τεχνολογίες που χρησιμοποιούνται για την κατασκευή υπολογιστικών συστημάτων αλλάζουν συχνά, σε αντίθεση με την αναπαράσταση μιας θεματικής ενότητας με τη μορφή μιας οντολογίας η οποία είναι αρκετά πιο ανεπηρέαστη στις τεχνολογικές εξελίξεις και πιο ελεύθερη από περιορισμούς υλοποίησης των διάφορων εφαρμογών. Συνεπώς, η χρήση οντολογίας αποσκοπεί στη διαχρονικότητα των πληροφοριών, την ανεξαρτησία των δεδομένων από την εκάστοτε υλοποίηση και την επαναχρησιμοποίησή της σε πολλαπλές ξεχωριστές εφαρμογές.

Απαραίτητο βήμα για κάποιον που σχεδιάζει να χρησιμοποιήσει μια οντολογία σε κάποια εφαρμογή είναι πρώτα ο έλεγχος μεταξύ των υπαρχόντων οντολογιών στο Παγκόσμιο Ιστό, εάν κάποια από αυτές καλύπτει τις απαιτήσεις του. Εάν βρεθεί η κατάλληλη τότε αποφεύγεται ο κόπος σχεδιασμού μιας οντολογίας από την αρχή και επιπλέον, ίσως να είναι επιτεύξιμο κάποιο όφελος από την ανταλλαγή γνώσεων μεταξύ των διάφορων εφαρμογών που ήδη χρησιμοποιούν την οντολογία αυτή και της νέας υπό ανάπτυξη εφαρμογής.

## **2.7 ΕΡΓΑΛΕΙΑ ΚΑΤΑΣΚΕΥΗΣ ΟΝΤΟΛΟΓΙΩΝ**

Ένας καλός επεξεργαστής οντολογιών μπορεί να συμβάλει στην εξοικονόμηση σημαντικού χρόνου κατά την ανάπτυξη οντολογιών, βοηθώντας τους μηχανικούς οντολογιών να επικεντρώσουν τις προσπάθειές τους στα σημασιολογικά χωρίς να ανησυχούν για την συντακτική οργάνωση. Το εργαλείο Protege (Gennari et al.,2003) παρέχει ένα αυτόνομο περιβάλλον ανάπτυξης οντολογιών. Χαρακτηρίζεται από το ανεξάρτητο ως προς τη γραμματική συντακτικού περιβάλλον διεπαφής με το χρήστη και την εσωτερική δομή που επιτρέπει την εύκολη ενσωμάτωσή του σε εφαρμογές. Είναι βολικό για ανεξάρτητη ανάπτυξη οντολογιών και χρησιμοποιείται από μεγάλη κοινότητα χρηστών.

Το εργαλείο SWOOP (Kalyanpur,2005) αξιοποιεί τόσο το Protege όσο και το Ontolingua,(ένα περιβάλλον για την επεξεργασία, δημοσίευση και διαμοιρασμό οντολογιών ανεπτυγμένων πριν από την εμφάνιση του Σημασιολογικού Ιστού) και παρέχει μια βασισμένη στο διαδίκτυο διεπαφή χρήστη για την φυλλομέτρηση, επεξεργασία, αποσφαλμάτωση (Kalyanpur:2006) και δημοσίευση οντολογιών.

Το εργαλείο SMORE είναι σχεδιασμένο ώστε να επιτρέπει τη συγγραφή OWL εγγράφων ταυτόχρονα με τη σχεδίαση ιστοσελίδων καθώς διαθέτει ενσωματωμένο και τον Ekit HTML Editor. Μπορεί ο χρήστης να σχεδιάσει έναν δικτυακό τόπο, να σχεδιάσει παράλληλα και τα semantics (οντολογία) και να δημιουργήσει μια ολοκληρωμένη, σημασιολογικά τεκμηριωμένη σελίδα. Είναι, επομένως, ένα ισχυρό εργαλείο το οποίο αποτελεί στην ουσία επέκταση του SWOOP. Ιδιαίτερη βαρύτητα δίνει το SMORE στην υποστήριξη της OWL. Ενώ οι υπόλοιποι editors υποστηρίζουν την RDF κατά κύριο λόγο, ο SMORE είναι προσανατολισμένος στην επεξεργασία OWL εγγράφων.

## **2.8 ΓΛΩΣΣΕΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ ΟΝΤΟΛΟΓΙΩΝ**

### **2.8.1 KIF**

Η KIF (Knowledge Interchange Format) είναι από τις πρώτες γλώσσες αναπαράστασης γνώσης. Αποτελεί μια γλώσσα για την ανταλλαγή γνώσης μεταξύ διαφορετικών εφαρμογών, αναπτύχθηκε για να λύσει το πρόβλημα της ανομοιογένειας μεταξύ γλωσσών που χρησιμοποιούνται για αναπαράσταση γνώσης. Περιλαμβάνει επεξηγηματικές δηλώσεις (που σημαίνει ότι είναι δυνατόν να «καταλάβει» το νόημα των εκφράσεων που δηλώνονται χωρίς να χρειάζεται διερμηνέα για να τις χειριστεί). Επιτρέπει τον ορισμό αντικειμένων, συναρτήσεων και σχέσεων. Έχει προταθεί ως πρότυπο για την περιγραφική λογική Η KIF αποτελεί μια συγκεκριμένη γλώσσα λογικής η οποία προτάθηκε σαν πρότυπο για να περιγράψει αντικείμενα σε ευφυείς εφαρμογές. Επιπλέον, έχει ειδικά σχεδιαστεί για να χρησιμοποιηθεί ως interlingua, μια γλώσσα μεσάζουσα στη διαδικασία μεταφράσεων από άλλες γλώσσες.

### **2.8.2 OWL**

Η γλώσσα OWL παρέχει ένα ακόμα πλουσιότερο λεξιλόγιο για την περιγραφή κλάσεων και ιδιοτήτων. Είναι δυνατή η ακόμα εκτενέστερη απόδοση σημασιολογίας στα δεδομένα, ορίζοντας ξένες ή όμοιες κλάσεις, θέτοντας περιορισμούς πολλαπλότητας ή πεδίου τιμών στις ιδιότητες κ.ά.

Η Web Ontology Language (OWL <http://www.w3.org/2004/OWL/> ) συγκεκριμένα, αποτελεί μια αναθεώρηση της DAML+OIL

(<http://www.daml.org/2000/12/reference.html>) που ενσωματώνει χρήσιμες πληροφορίες από το σχεδιασμό και την εφαρμογή της DAML+OIL. Καθώς αποτελεί μέρος του σημασιολογικού ιστού δε θα μπορούσε παρά να στοχεύει στη χρήση και επεξεργασία των πληροφοριών από εφαρμογές παρά στις περιπτώσεις όπου το περιεχόμενο χρειάζεται απλώς να παρουσιαστεί σε ανθρώπους-χρήστες. Μπορεί να χρησιμοποιηθεί για να αναπαραστήσει τη σημασία των όρων που συναντώνται σε λεξικά και των σχέσεων μεταξύ τους. Αυτή η αναπαράσταση των όρων και των σχέσεων που τους διέπουν και τους συνδέουν ονομάζεται οντολογία

Με την OWL γίνεται λοιπόν εφικτός ο ορισμός Οντολογιών. Οντολογία είναι η θεωρία για την ύπαρξη. Οι οντολογίες προσφέρουν μία εννοιολογική θεωρία ενός πεδίου ενδιαφέροντος, περιγράφοντας τις βασικές οντότητες και τις σχέσεις μεταξύ τους. Αυτή η ενιαία περιγραφή του πεδίου είναι αναγκαία για να υπερκεραστούν διαφορές στην ορολογία. Έτσι μπορεί να οριστεί ότι δύο διαδικτυακοί πόροι παρότι έχουν διαφορετικό όνομα έχουν την ίδια σημασιολογία, ή ότι δύο διαδικτυακοί πόροι με ίδιο όνομα έχουν διαφορετική σημασιολογία (W3C, OWL).

Η δημιουργία του σημασιολογικού ιστού στηρίζεται στην επέκταση των υπαρχόντων πλαισίων περιγραφής μετα-δεδομένων και ακριβέστερα στην ύπαρξη σημασιολογικού περιεχομένου το οποίο δύναται να υπόκειται σε αυτόματη επεξεργασία από τον υπολογιστή χωρίς την ανθρώπινη παρεμβολή.

Η OWL παρέχει τρεις επιμέρους γλώσσες σχεδιασμένες για διαφορετική χρήση η κάθε μια. Η σχέση αυτών των γλωσσών φαίνεται στην Εικόνα 6.

**OWL Lite:** Καλύπτει ουσιαστικά την ανάγκη για ταξινόμηση και ιεράρχηση και υποστηρίζει απλούς περιορισμούς. Υποστηρίζει εκείνους τους χρήστες που χρειάζονται κυρίως μια ιεραρχία ταξινόμησης και απλούς περιορισμούς. Χρησιμοποιεί μόνο μερικά από τα γλωσσικά χαρακτηριστικά της OWL και έχει περισσότερους περιορισμούς στη χρήση των χαρακτηριστικών αυτών από την OWL DL ή την OWL Full.

**OWL DL** (Description Logic): Παρέχει τη μέγιστη δυνατή εκφραστικότητα, ενώ παράλληλα εγγυάται υπολογιστική ακεραιότητα (όλοι οι υπολογισμοί είναι υπολογίσιμοι σε πεπερασμένο χρόνο). Αυτό επιτυγχάνεται με τον χρήση μόνο καταληκτικών (decidable) εκφράσεων λογικής πρώτης τάξης (first order logic). Υποστηρίζει εκείνους τους χρήστες που θέλουν τη μέγιστη εκφραστικότητα



διατηρώντας την υπολογιστική πληρότητα (όλα τα συμπεράσματα είναι εγγυημένα να είναι υπολογίσιμα) και decidability (όλοι οι υπολογισμοί θα τελειώσουν σε πεπερασμένο χρόνο).

**OWL Full:** Παρέχει τη μέγιστη δυνατή εκφραστικότητα, την ελευθερία σύνταξης της RDF χωρίς ωστόσο να εγγυάται την ακεραιότητα των υπολογισμών.



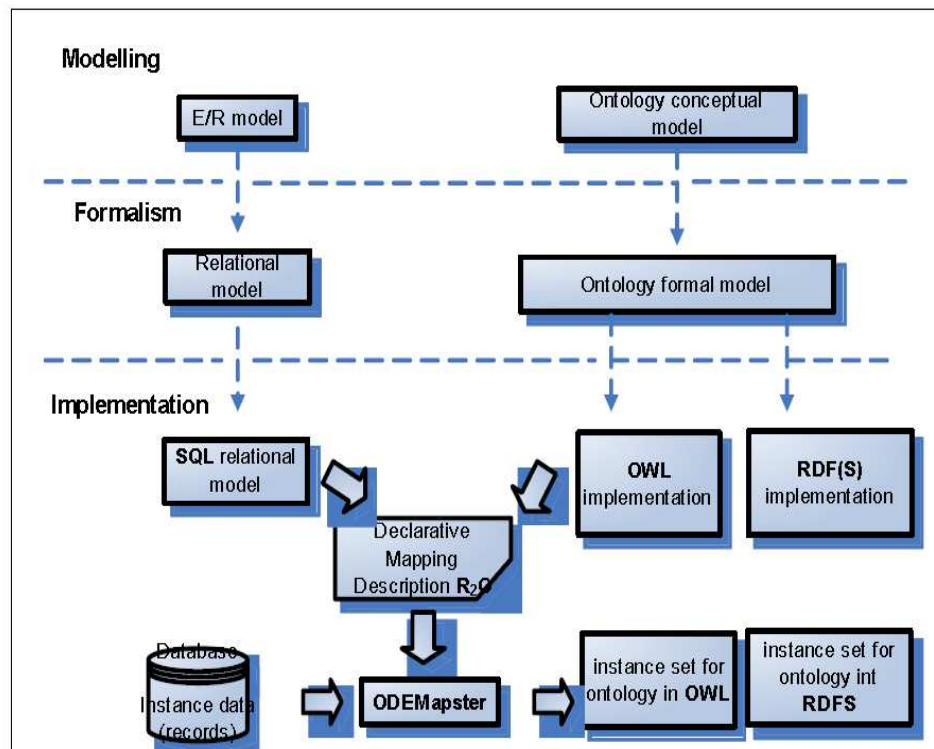
Εικόνα 6. Επίπεδα εκφραστικότητας OWL

Η επιλογή της κατάλληλης υπο-γλώσσας της OWL για την ανάπτυξη της οντολογίας εξαρτάται από τις ανάγκες της. Η επιλογή μεταξύ της OWL Lite και της OWL DL εξαρτάται από το βαθμό στον οποίο οι χρήστες απαιτούν τα περισσότερο-εκφραστικά κατασκευάσματα που παρέχονται από την OWL DL. Η επιλογή μεταξύ της OWL DL και της OWL Full εξαρτάται κυρίως από το βαθμό στον οποίο οι χρήστες απαιτούν τις ευκολίες μετα-διαμόρφωσης του σχήματος RDF (π.χ. καθορισμός κλάσεων άλλων κλάσεων, ή σύνδεση των ιδιοτήτων με τις κλάσεις). Κατά χρησιμοποίηση της OWL Full σε σύγκριση με την OWL DL, η υποστήριξη συλλογισμού είναι λιγότερο προβλέψιμη αφού πλήρεις OWL Full εφαρμογές δεν υπάρχουν.

### **R2O (Relational to Ontology)**

Η R2O είναι μια δηλωτική γλώσσα η οποία περιγράφει αντιστοιχίσεις ανάμεσα σε σχεσιακά σχήματα βάσεων δεδομένων και οντολογίες υλοποιημένες σε RDF(S) ή OWL. Η προσέγγιση η οποία ακολουθείται από τη συγκεκριμένη γλώσσα περιλαμβάνει τον ορισμό των αντιστοιχίσεων ανάμεσα στα στοιχεία του σχεσιακού σχήματος και της οντολογίας, η οποία προϋπάρχει και δεν παράγεται κατά τη διαδικασία της αντιστοίχισης, και την αυτόματη επεξεργασία των

αντιστοιχίσεων αυτών από τον ODEMapster επεξεργαστή αντιστοιχίσεων (εικόνα 7).



Εικόνα 7. ODEMapster επεξεργαστή αντιστοιχίσεων

### Αρχιτεκτονική R2O

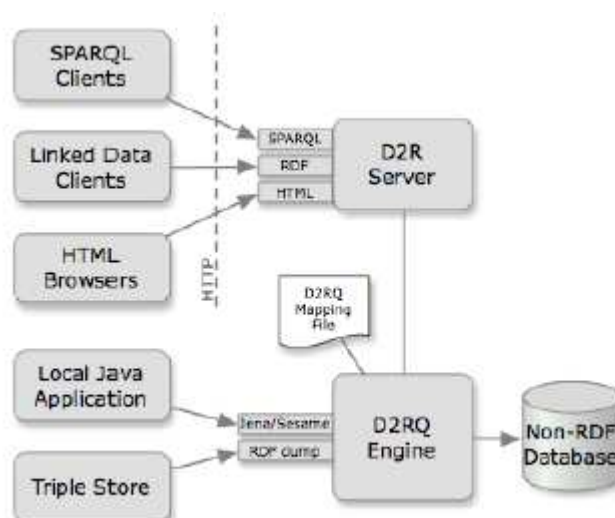
Η R2O είναι μια γλώσσα ανεξάρτητη των Σχεσιακών Συστημάτων Διαχείρισης Βάσεων Δεδομένων και μπορεί να χρησιμοποιηθεί για οποιαδήποτε βάση υλοποιεί το SQL πρότυπο. Είναι επίσης δυνατή η χρήση της από υπάρχοντα εργαλεία, που στόχο έχουν την απεικόνιση ενός σχεσιακού σχήματος σε μια οντολογία, για τη δήλωση αντιστοιχίσεων ανάμεσα στα δύο αυτά σχήματα.

### 3.1 ΕΡΓΑΛΕΙΑ ΓΙΑ ΤΗ ΔΗΜΟΣΙΕΥΣΗ ΣΥΝΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

#### 3.1.1 Δημοσίευση Συνδεδεμένων Δεδομένων από σχεσιακές βάσεις δεδομένων

**D2R Server:** είναι ένα εργαλείο για τη δημοσίευση του περιεχομένου των σχεσιακών βάσεων δεδομένων στο Σημασιολογικό Ιστό, ένα παγκόσμιο χώρο πληροφοριών που αποτελείται από συνδεδεμένα δεδομένα.

- Είναι ένα εργαλείο το οποίο αντιστοιχίζει(mapping) μία Σχεσιακή Βάση Δεδομένων σε μορφή RDF και δημοσιεύει το περιεχόμενο της ως Linked Data
- Επιτρέπει την πλοήγηση και την ανάκτηση των RDF δεδομένων
- Πρόσβαση μέσω SPARQL
- Χρησιμοποιείται για την δημοσίευση επιλεγμένων δεδομένων της βάσης περιεχομένου του LiferayServer



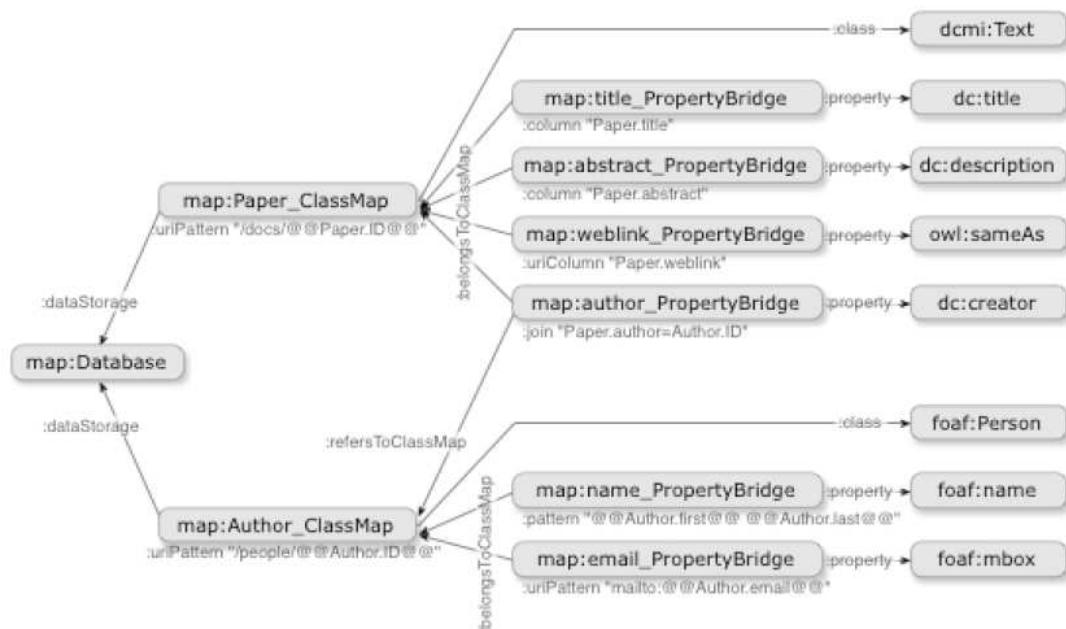
Εικόνα 8. D2R Server

Ο D2R Server χρησιμοποιεί την D2RQ γλώσσα χαρτογράφησης για να συλλάβει αντιστοιχίσεις μεταξύ σχημάτων βάσεων δεδομένων συγκεκριμένων εφαρμογών και σχημάτων RDFS ή OWL οντολογιών. Μια D2RQ χαρτογράφηση καθορίζει το πώς ταυτοποιούνται οι πόροι και πώς παράγονται οι τιμές των ιδιοτήτων από το περιεχόμενο της βάσης δεδομένων. Το κεντρικό αντικείμενο στη D2RQ είναι το ClassMap. Ένα ClassMap αναπαριστά μια χαρτογράφηση από ένα σύνολο οντοτήτων που περιγράφεται μέσα στη βάση δεδομένων, προς μια

κλάση ή μια ομάδα παρόμοιων κλάσεων πόρων. Κάθε ClassMap έχει ένα σύνολο από PropertyBridges, οι οποίες καθορίζουν το πώς δημιουργούνται οι περιγραφές των πόρων. Οι τιμές των ιδιοτήτων μπορούν να δημιουργηθούν απ' ευθείας από τις τιμές της βάσης δεδομένων ή με τη χρήση «καλουπιών» ή πινάκων μετάφρασης. Η D2RQ υποστηρίζει υπό όρους (conditional) χαρτογραφήσεις σε επίπεδο ClassMap και PropertyBridge, τη χαρτογράφηση n:m σχέσεων και το χειρισμό υψηλά κανονικοποιημένων δομών πινάκων όπου οι περιγραφές οντοτήτων εκτείνονται σε αρκετούς πίνακες.

Ο D2R Server περιλαμβάνει ένα εργαλείο που παράγει αυτόματα μια D2RQ χαρτογράφηση από τη δομή του πίνακα μιας βάσης δεδομένων. Το εργαλείο παράγει ένα νέο RDF λεξιλόγιο για κάθε βάση δεδομένων, χρησιμοποιώντας τα ονόματα των πινάκων ως ονόματα κλάσεων και τα ονόματα των στηλών ως ονόματα ιδιοτήτων. Η χαρτογράφηση μπορεί να προσαρμοστεί στη συνέχεια, αντικαθιστώντας τους αυτόματα παραγόμενους όρους με όρους από γνωστά RDF λεξιλόγια.

Το παρακάτω σχήμα δείχνει τη δομή ενός παραδείγματος D2RQ map:

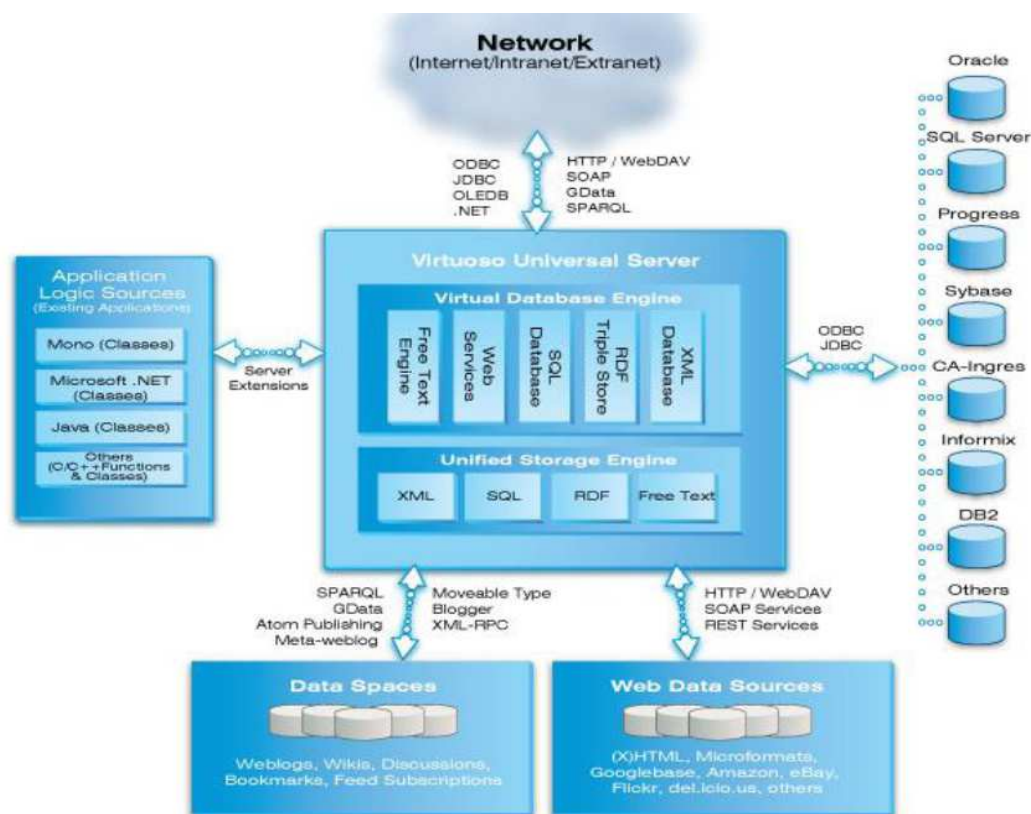


Εικόνα 9. Δομή D2RQ map

Η δημοσίευση μιας σχεσιακής βάσης δεδομένων ως Συνδεδεμένα Δεδομένα περιλαμβάνει τα παρακάτω βήματα:

1. Φόρτωση και εγκατάσταση του λογισμικού του server όπως έχει περιγραφεί στη σελίδα του D2R Server.
2. Αυτόματη παραγωγή από τον D2R Server μιας D2RQ αντιστοίχισης από το σχήμα της βάσης δεδομένων.
3. Προσαρμογή της αντιστοίχισης, αντικαθιστώντας τους παραγόμενους με αυτόματο τρόπο όρους με όρους από γνωστά και δημόσια προσβάσιμα RDF λεξιλόγια.
4. Προσθήκη της νέας πηγής δεδομένων στη λίστα συνόλων δεδομένων του ESW Wiki56 στην κατηγορία Συνδεδεμένα Δεδομένα και στη λίστα των σημείων SPARQL και τοποθέτηση αρκετών RDF συνδέσμων από το FOAF προφίλ προς URIs κεντρικών πόρων μέσα στη νέα πηγή δεδομένων, ώστε οι crawlers να μπορούν να ανακαλύψουν τα δεδομένα μας.

**Virtuoso** : Το πρόγραμμα Virtuoso είναι ένα εμπορικό προϊόν της εταιρείας Open-Link. Υπάρχει και μια open-source έκδοση με περιορισμένες όμως δυνατότητες. Το Virtuoso προσφέρει γενικά την δυνατότητα διασύνδεσης ετερογενών δεδομένων από πολλαπλές πηγές. Από τις βασικές υπηρεσίες του είναι η διασύνδεση σχεσιακών βάσεων σε RDF σχήματα ώστε να προκύψει μια οντολογία.



Εικόνα 10. Η Αρχιτεκτονική του Open-Link Virtuoso

Η αρχιτεκτονική του Virtuoso (εμπορική έκδοση) παρουσιάζεται στην εικόνα 10 και όπως φαίνεται υποστηρίζει οποιαδήποτε σχεσιακή βάση μπορεί να συνδεθεί μέσω του JDBC-ODBC ενώ μπορεί να αντλεί και δεδομένα από δίκτυο ή από άλλες πηγές.

**Triplify:** Το Triplify (<http://www.triplify.org>) λειτουργεί μονάχα σε web περιβάλλοντα και ξεχωρίζει για την απλότητα του. Είναι ένα πολύ ελαφρύ πρόσθετο (plug in) και μπορεί να χρησιμοποιηθεί σε πολλές Web εφαρμογές. Το εργαλείο αυτό βασίζεται στην χαρτογράφηση HTTP-URI αναζητήσεων σε ερωτήματα SQL μαζί με κάποιες προσθήκες. Επιστρέφει τα δεδομένα σε RDF σχήματα και συγκεκριμένα ως Linked Data. Το Triplify είναι ακόμα σε δοκιμαστικό στάδιο και η απλότητα του μπορεί να το κάνει ιδιαίτερα χρήσιμο σε εφαρμογές με χαμηλές απαιτήσεις σημασιολογίας.

Στον πυρήνα του Triplify υπάρχει ο ορισμός μιας διαμόρφωσης, η οποία είναι προσαρμοσμένη για ένα συγκεκριμένο σχήμα βάσης δεδομένων. Η διαμόρφωση αυτή περιέχει όλες τις πληροφορίες που χρειάζονται για την υλοποίηση του Triplify έτσι ώστε να παράγει σύνολα τριάδων RDF, οι οποίες επιστρέφονται σε αιτήσεις URL που δέχεται μέσω HTTP. Η διαμόρφωση Triplify ορίζεται ως μια τριάδα (s,φ,μ) όπου :

- s είναι ένας προεπιλεγμένος χώρος ονομάτων σχημάτων,
- φ είναι μια αντιστοίχιση προθεμάτων του χώρου ονομάτων σε URI
- μ είναι μια αντιστοίχιση μοτίβων URL σε ερωτήματα SQL.

Ο προεπιλεγμένος χώρος ονομάτων σχημάτων s χρησιμοποιείται στη δημιουργία αναγνωριστικών URI για προβολές στηλών βάσεων δεδομένων που δεν έχουν αντιστοιχιστεί σε υπάρχοντα λεξιλόγια. Αφού η ίδια διαμόρφωση μπορεί να χρησιμοποιηθεί για πλήθος διαδικτυακών εφαρμογών που έχουν την ίδια διάταξη βάση δεδομένων, τα δεδομένα αυτών των διαδικτυακών εφαρμογών μπορούν να ενσωματωθούν ακόμα και χωρίς αντιστοίχιση σε υπάρχοντα λεξιλόγια. Η αντιστοίχιση φ των προθεμάτων του χώρου ονομάτων σε URI ορίζει συντομεύσεις για συχνά χρησιμοποιούμενους χώρους ονομάτων. Η αντιστοίχιση μ αντιστοιχίζει αναφερόμενους πόρους RDF σε σύνολα προβολών SQL που

περιγράφουν αυτούς τους πόρους. Τα μοτίβα στο πεδίο της  $\mu$  αξιολογούνται ως προς τις αιτήσεις σε URI που δέχεται το Triplify μέσω HTTP.

```
$triplify['queries']=array(
  'post'=>array(
    "SELECT id, post_author AS 'sioc:has_creator->user', post_date AS 'dcterms:created', post_title AS 'dc:title',
    post_content AS 'sioc:content', post_modified AS 'dcterms:modified'
    FROM posts WHERE post_status='publish'",
    "SELECT post_id AS id, tag_id AS 'tag:taggedWithTag->tag' FROM post2tag",
    "SELECT post_id AS id, category_id AS 'belongsToCategory->category' FROM post2cat",
  ),
  'tag'=>"SELECT tag_ID AS id, tag AS 'tag:tagName' FROM tags",
  'category'=>"SELECT cat_ID AS id, cat_name AS 'skos:prefLabel', category_parent AS 'skos:narrower->category'
  FROM categories",
  'user'=>array(
    "SELECT id, user_login AS 'foaf:accountName', user_url AS 'foaf:homepage',
    SHA(CONCAT('mailto:',user_email)) AS 'foaf:mbox_sha1sum', display_name AS 'foaf:name'
    FROM users",
    "SELECT user_id AS id, meta_value AS 'foaf:firstName' FROM usermeta WHERE meta_key='first_name'",
    "SELECT user_id AS id, meta_value AS 'foaf:family_name' FROM usermeta WHERE meta_key='last_name'",
  ),
  'comment'=>"SELECT comment_ID AS id, comment_post_id AS 'sioc:reply_of->user', comment_author AS 'foaf:name',
  comment_author_url AS 'foaf:homepage', SHA(CONCAT('mailto:',comment_author_email)) AS 'foaf:mbox_sha1sum',
  comment_type, comment_date AS 'dcterms:created', comment_content AS 'sioc:content', comment_karma,
  FROM comments WHERE comment_approved='1'",
);
```

Εικόνα 11. Αντιστοίχιση  $\mu$  των μοτίβων URL σε σύνολα ερωτημάτων SQL, στη διαμόρφωση του Triplify για το WordPress (κώδικας PHP)

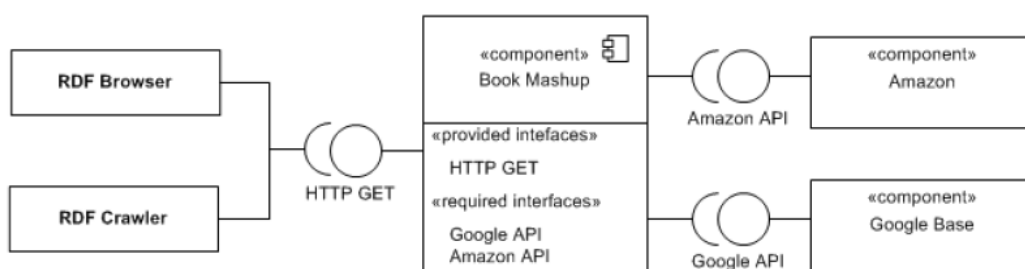
### 3.1.2 Δημοσίευση Συνδεδεμένων Δεδομένων με χρήση εφαρμογών περιτυλιγμάτων γύρω από υπάρχουσες εφαρμογές ή Web APIs

Πλήθος εφαρμογών του Ιστού έχουν ξεκινήσει να κάνουν τα δεδομένα τους διαθέσιμα μέσω Web APIs (Application Programming Interfaces). Παραδείγματα τέτοιων πηγών δεδομένων είναι το eBay, η Amazon, το Yahoo, το Google και το Google Base. Το κάθε API προσφέρει διαφορετικές διεπαφές για τη διατύπωση ερωτημάτων και την ανάκτηση απαντήσεων, κάτι το οποίο αποτελεί σοβαρό εμπόδιο στην καθολικότητα και επεκτασιμότητα εφαρμογών που επωφελούνται από τα δεδομένα που είναι δημοσιευμένα στον Ιστό. Αυτοί οι περιορισμοί μπορούν να ξεπεραστούν με την υλοποίηση εφαρμογών-περιτυλιγμάτων (wrappers). Οι εφαρμογές αυτές βρίσκονται ανάμεσα στον πελάτη και στο API και σε γενικές γραμμές κάνουν τα εξής :

1. Αναθέτουν HTTP URIs στους μη πληροφοριακούς πόρους για τους οποίους το API προσφέρει πληροφορίες.
2. Όταν ένα από αυτά τα URIs γίνεται dereferenced ζητώντας τον τύπο application/rdf+xml, η εφαρμογή μετατρέπει την αίτηση του πελάτη στην ισοδύναμη αίτηση που είναι σύμφωνη με τις προδιαγραφές του API και τη στέλνει στο API.

3. Τα αποτελέσματα της API αίτησης μετατρέπονται σε RDF και στέλνονται πίσω στον πελάτη.

Ένα χαρακτηριστικό παράδειγμα υλοποίησης αυτής της πρακτικής είναι το RDF Book Mashup (<http://www4.wiwiiss.fu-berlin.de/bizer/bookmashup/>). Το RDF Book Mashup αναθέτει ένα HTTP URI σε κάθε βιβλίο που έχει ένα ISBN αριθμό και όταν ένα από αυτά τα URIs γίνεται dereferenced, το RDF Book Mashup ζητά δεδομένα για το αντίστοιχο βιβλίο, το συγγραφέα του, καθώς και για κριτικές και προσφορές πώλησης από το Amazon API και το Google Base API. Τα δεδομένα που συγκεντρώνει, τα μετατρέπει στη συνέχεια σε RDF και τα επιστρέφει στον πελάτη.



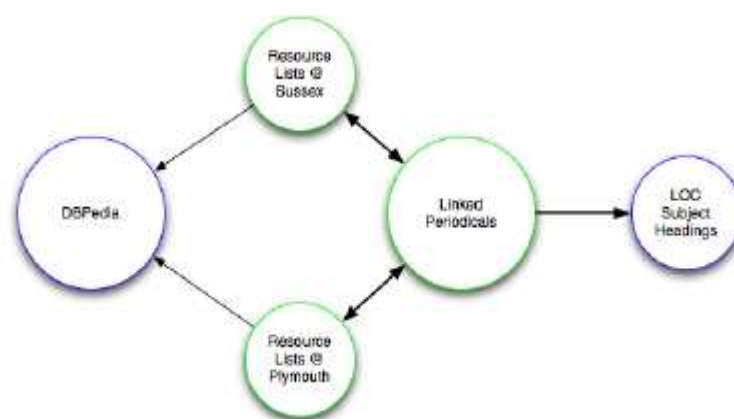
Εικόνα 12. Διάγραμμα της λειτουργίας του RDF Book Mashup

## 3.2 ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΗΣ

### 3.2.1 Talis Aspire

Στο χώρο του εκπαιδευτικού ψηφιακού υλικού, η πιο γνωστή εφαρμογή Διασυνδεδεμένων Δεδομένων είναι το Talis Aspire (<http://www.talisaspire.com/>), ένα σύστημα διαχείρισης καταλόγου ψηφιακών πηγών για μαθήματα που χρησιμοποιείται από τα πανεπιστήμια του Plymouth και του Sussex. Το Talis Aspire ενοποιεί τις περιγραφές των ψηφιακών αντικειμένων χρησιμοποιώντας οντολογίες, επιτρέπει την ενοποίηση και την αντιστοίχισή τους με άλλες σε διαφορετικά βιβλιογραφικά αποθετήρια και οντολογίες οι οποίες έχουν σχεδιαστεί για την περιγραφή των σχέσεων μεταξύ ψηφιακών εκπαιδευτικών αντικειμένων. Επίσης όπως αναφέρει ο Clarke (2009), το Talis Aspire βελτιώνει την διαλειτουργικότητα των δεδομένων και παροτρύνει στους χρήστες να εμπλουτίσουν με σημασιολογικά δεδομένα και την χρησιμοποίηση της δυνατότητας σύστασης ψηφιακού υλικού σε άλλους.





Εικόνα 13. Η σύνδεση της εφαρμογής Talis Aspire συνδεδεμένα περιοδικά, DBPedia και LCSH

Το TALIS Aspire είναι ένα μοντέλο λογισμικού για τη διαχείριση καταλόγων ακαδημαϊκών πόρων. Οι κατάλογοι των πόρων είναι ένα βασικό μέρος του περιεχομένου των μαθημάτων που δίνονται στους φοιτητές προκειμένου να μελετήσουν περισσότερα αποδοτικά. Παρέχουν μια διεπαφή ανάγνωσης καθώς και ένα ισχυρό περιβάλλον για την επεξεργασία των καταλόγων.

**BROADMINSTER UNIVERSITY** Sign out | Feedback

Home My Bookmarks My Lists

### Financial Accounting and Reporting

Used by ABF203

[Table of contents](#) [\[show\]](#)

(94 items)

**Recommended Text** (3 items)

- Financial accounting and reporting** - Elliott, Barry., Elliott, Jamie., 2008. Book  
The 12th edition of Elliott and Elliott is the recommended text to buy for this module.
- Public finance quarterly.** Journal  
It is recommended that you maintain awareness of the current issue, as well as reviewing the last 3years backfile.
- Financial Accounting and Reporting, 12/E - Pearson Education EMA Catalogue** Web Page  
This is the companion website and student resources for the 12th edition

**Supplementary Resources** (11 items)

- Financial reporting** - Association of Certified and Corporate Accountants., 2003. Book
- Financial reporting : International stream. [ACCA study text]** - Association of Certified and Corporate Accountants., 2004. Book
- Accounting Standards Board - Welcome** Web Page
- Accounting theory and practice** - Glautier, M.W.E., Underdown, Brian., 2001. Book
- IASB Home** Web Page  
This is the home page of the International Accounting Standards Board and is where you can find information about international accounting standards (IFRS and IAS) as well as recent information on

**Actions**  
Save Discard changes

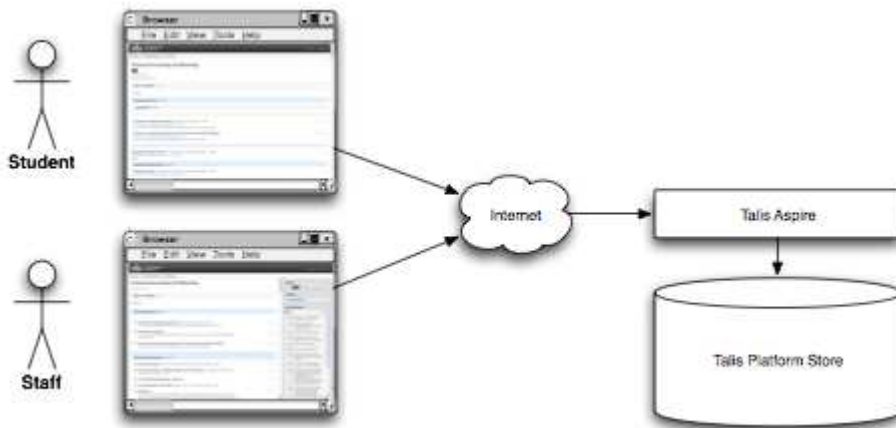
**Sections**  
[New Section](#)

**My Bookmarks**  
Find:

- YouTube - The Language of Business: Finance 101
- No new revolution in economics? Taking Thompson and fine forward
- Applied financial economics
- Broad-basing green stock market indices: a concept note
- Public finance quarterly.
- Private and public finance
- Structural Evolution of the Protein Kinase-Like Superfamily | SciVee
- YouTube - Starting a New Business - Part 1, Introduction
- Annals of Clinical Microbiology and Antimicrobials | Full text | Exogenous re-infection by a novel Streptococcus pneumoniae serotype 14 as a cause of recurrent meningitis in a child from The Gambia
- PLoS ONE: Rapid Regulatory T-Cell Response Prevents Cytokine Storm in CD28 Superagonist Treated Mice

Εικόνα 14. Talis Aspire: Μια ακαδημαϊκή λίστα πόρων

Μια απλοποιημένη εφαρμογή του μοντέλου Talis Aspire παρουσιάζεται παρακάτω:



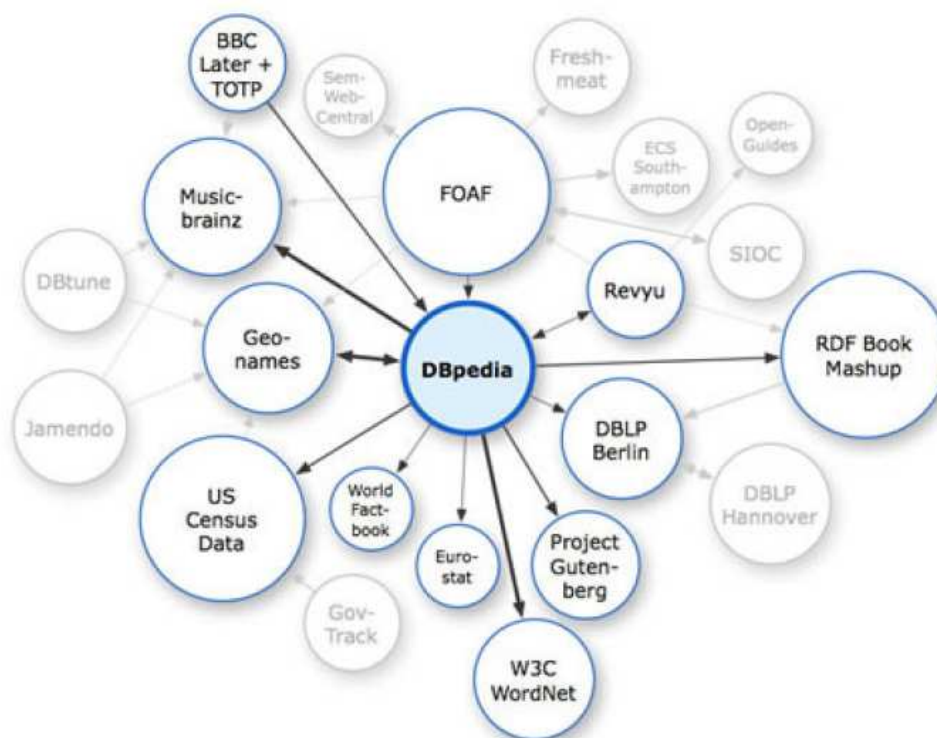
Εικόνα 15. Talis Aspire: Απλοποιημένη Εφαρμογή Αρχιτεκτονικής

### 3.2.2 Faceted Wikipedia Search

Το Faceted Wikipedia Search επιτρέπει την υποβολή σύνθετων ερωτημάτων στη Wikipedia. Οι απαντήσεις που θα παραχθούν για τα ερωτήματα δεν παράγονται από ταίριασμα κάποιων λέξεων κλειδιών, αλλά έχουν ως βάση δομημένα δεδομένα που έχουν εξαχθεί από διαφορετικά άρθρα της Wikipedia.

### 3.2.3 DBpedia Mobile

Είναι ένα λογισμικό που χρησιμοποιεί τοποθεσίες της DBpedia. Η Dbpedia είναι μία προσπάθεια από μία ανοικτή κοινότητα χρηστών για την εξαγωγή δομημένης πληροφορίας από την εγκυκλοπαίδεια Wikipedia και να κάνει την πληροφορία αυτή διαθέσιμη στο διαδίκτυο. Η DBpedia Mobile παρέχει τη δυνατότητα στους χρήστες να ανακαλύψουν, αναζητήσουν και να δημοσιεύουν Συνδεδεμένα Δεδομένα σε σχέση με τη γεωγραφική τους θέση με τη χρήση φορητών μέσων.



Εικόνα 16. Η DBpedia στο δίκτυο Διασυνδεδεμένων Δεδομένων

Με βάση τη θέση της φορητής συσκευής που υποδεικνύεται από το GPS της, η DBpedia Mobile δημιουργεί ένα διαδραστικό χάρτη που δείχνει κοντινές τοποθεσίες από την DBpedia. Οι τοποθεσίες μπορεί να είναι σημειωμένες σε οποιαδήποτε από τις γλώσσες που υποστηρίζονται από την DBpedia και απεικονίζονται με σύμβολα που βασίζονται σε χαρτογράφηση επιλεγμένων κατηγοριών YAGO. Από αυτό το χάρτη, ο χρήστης μπορεί να ανακαλύψει πληροφορίες για ότι τον περιβάλλει, ακολουθώντας συνδέσμους προς άλλα σύνολα δεδομένων. Η DBpedia δεν αντλεί δεδομένα από ένα σταθερό αριθμό πηγών δεδομένων, αλλά μπορεί να χρησιμοποιηθεί για πρόσβαση σε πηγές δεδομένων που είναι συνδεδεμένες με τη DBpedia ή θα συνδεθούν στο μέλλον.



Εικόνα 17. Πληροφορίες για το Βερολίνο από την DBpedia Mobile

### **3.2.4 Σημασιολογικά Wikis**

Τα Σημασιολογικά Wikis δίνουν τη δυνατότητα στους χρήστες να περιγράφουν τυπικά κάποια πληροφορία σημειώνοντας σημασιολογικά σχόλια στις σελίδες του wiki που περιγράφουν αυτή τη πληροφορία. Έτσι ενώ τα κλασικά wikis στηρίζονται σε φυσική γλώσσα, τα σημασιολογικά wikis επιτρέπουν την επιπλέον περιγραφή της με τη χρήση μιας τυπικής γλώσσας σχολιασμού με μεταδεδομένα. Με την προσθήκη μεταδεδομένων στο κλασικό περιεχόμενο των wikis, αυτό αποκτά πρόσθετη αξία μέσω της αποτελεσματικότερης αναζήτησης, της ευκολότερης ανταλλαγής πληροφορίας και της επαναχρησιμοποίησης της εμπειριεχόμενης γνώσης.

Στην αρχιτεκτονική επιπέδων του Semantic Web ακολουθούνται δύο βασικές αρχές:

- Συμβατότητα προς τα κατώτερα επίπεδα. Οι πράκτορες λογισμικού και οι διαδικτυακές εφαρμογές που αντιλαμβάνονται πλήρως ένα επίπεδο θα πρέπει να είναι αυτομάτως σε θέση να χρησιμοποιήσουν και τις πληροφορίες των κατώτερων επιπέδων.
- Αποσπασματική κατανόηση των ανώτερων επιπέδων. Οι πράκτορες λογισμικού που αντιλαμβάνονται πλήρως ένα επίπεδο θα πρέπει να έχουν τη δυνατότητα να χρησιμοποιήσουν κάποιο μέρος από τις πληροφορίες των ανώτερων επιπέδων.

Τα οφέλη της χρησιμοποίησης των semantic έναντι των παραδοσιακών wiki είναι πολλά. Η αναζήτηση παύει να είναι βασισμένη σε δίχως νόημα λέξεις και αποκτά σημασιολογικό υπόβαθρο. Έτσι, η άντληση πληροφοριών δεν απορρέει από μια απλή αντιστοιχία λέξεων, που πολλές φορές μπορεί να είναι και τυχαία, αλλά από την εκτέλεση μιας διαδικασίας ερωταπαντήσεων. Τα αποτελέσματα των queries αυτών βασίζονται στην ισχύ τους στο γεγονός ότι τόσο οι πληροφορίες που είναι αποθηκευμένες στον ιστό όσο και οι ερωτήσεις των χρηστών είναι δομημένες με την ίδια γλώσσα αναπαράστασης γνώσης.

#### **3.2.4.1 Πώς επιτυγχάνονται τα σημασιολογικά χαρακτηριστικά**

Στα κλασικά wikis η απαιτούμενη λειτουργικότητα που οφείλει να καλύπτεται συνοψίζεται στη διαχείριση και εξαγωγή των δεδομένων, στη διαχείριση των διαφορετικών εκδόσεων του περιεχομένου, στον αποτελεσματικό έλεγχο πρόσβασης, εφόσον είναι απαραίτητος. Τα σημασιολογικά wikis έχοντας επιπλέον χαρακτηριστικά, πρέπει να δίνουν λύση σε βασικά θέματα όπως τον

τρόπο με τον οποίο: (α) θα αποδοθεί σημασιολογικός σχολιασμός στα περιεχόμενα, (β) θα περιγραφεί το περιεχόμενο σε κάποια τυπική γλώσσα και (γ) θα γίνεται η πλοήγηση (Tazzoli et al, 2004).

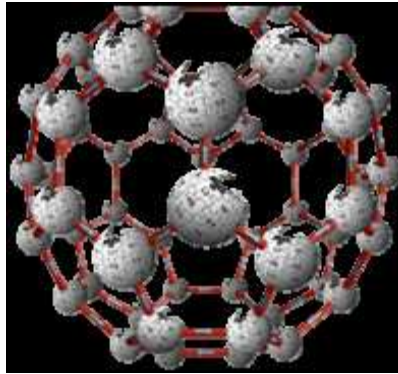
#### **3.2.4.2 Απόδοση Σημασιολογικού Σχολιασμού**

Ο σχολιασμός γενικότερα, αλλά και ο σημασιολογικός σχολιασμός ειδικότερα, για να οριστεί δανείζεται από την γραμματική και συντακτική ορολογία. Έτσι περιγράφεται από μία τριπλέτα <υποκείμενο-κατηγορημα-αντικείμενο>. Το υποκείμενο του σχολιασμού μπορεί είτε να έχει οριστεί τυπικά, είτε όχι. Έτσι για παράδειγμα η σημείωση στο περιθώριο μιας παραγράφου ενός κειμένου είναι μια άτυπη παραδοχή ότι η σημείωση αναφέρεται στην παράγραφο, Αν όμως χρησιμοποιηθεί ένας σαφής δείκτης προς το υποκείμενο του σχολιασμού, όπως το URI, τότε αυτό ορίζεται τυπικά.

Το κατηγορημα σε έναν σημασιολογικό σχολιασμό μπορεί να είναι επίσης τυπικά ορισμένο ή όχι. Στο παράδειγμα της σημείωσης στο περιθώριο μιας παραγράφου, μπορούμε να συμπεράνουμε, αν και δεν ορίζεται σαφώς, πως η σημείωση μπορεί να υποδηλώνει ένα σχόλιο προς τα γραφόμενα, μια επισήμανση για λάθος σύνταξη, κτλ. Αν όμως χρησιμοποιηθεί ένας σύνδεσμος προς έναν καλώς ορισμένο όρο μιας οντολογίας που περιγράφει και ορίζει το είδος της σημείωσης, τότε το κατηγορημα είναι τυπικά ορισμένο.

### **3.5 Semantic MediaWiki**

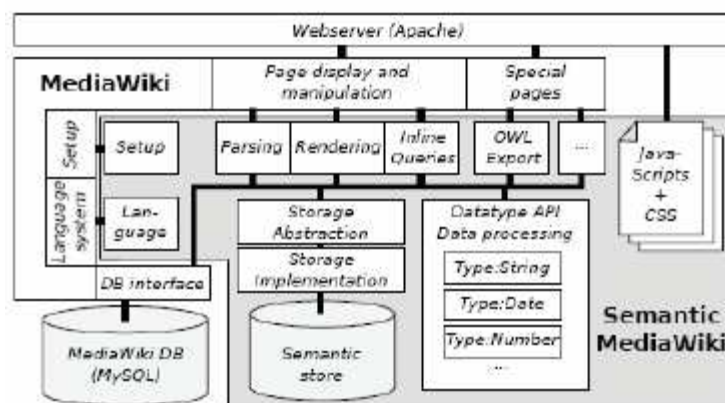
Το Semantic MediaWiki (SMW) ([http://ontoworld.org/wiki/Semantic\\_MediaWiki](http://ontoworld.org/wiki/Semantic_MediaWiki)) αποτελεί τη σημασιολογική επέκταση του MediaWiki. Το MediaWiki είναι ευρέως διαδεδομένο, καθώς χρησιμοποιείται από τη Wikipedia. Συνεπώς ο απώτερος στόχος της ανάπτυξης του Semantic MediaWiki είναι να επιτρέψει τη σημασιολογική απεικόνιση του περιεχομένου της Wikipedia. Αυτό σημαίνει ότι πρέπει να δοθεί ιδιαίτερη έμφαση στην απόδοση του αναπτυσσόμενου συστήματος και τις δυνατότητες επέκτασης του, δεδομένου ότι τα περιεχόμενα της πιο διάσημης ηλεκτρονικής εγκυκλοπαίδειας θα συνεχίσουν να πληθαίνουν (Krotzsch,Vrandecic, Volkel, 2006).



Εικόνα 18. Semantic MediaWiki logo

Το SMW χρησιμοποιεί έναν πολύ απλό τρόπο για σημασιολογικό σχολιασμό. Οι χρήστες χρησιμοποιούν ειδικό συντακτικό το οποίο αντιστοιχίζεται από το SMW σε τυπικές περιγραφές της γλώσσας οντολογιών OWL DL. Έτσι η αναγκαία σημασιολογική πληροφορία προστίθεται από τους χρήστες καθώς συγγράφουν το κείμενο της σελίδας. Κάθε σελίδα-άρθρο αντιστοιχίζεται σε ένα οντολογικό στοιχείο.

Το Semantic MediaWiki είναι λογισμικό ανοικτού κώδικα και υπόκειται στην GNU GPL (<http://www.gnu.org/copyleft/gpl.html>) άδεια. Υλοποιείται σε PHP και εγκαθίσταται ως extension σε μια παρούσα εγκατάσταση MediaWiki, η οποία απαιτεί Apache web server και κάποια βάση δεδομένων (MySQL ή Postgres).



Εικόνα 19. Τα τμήματα οργάνωσης του semantic mediawiki

Στο SMW γίνεται προσπάθεια επαναχρησιμοποίησης ήδη παραδεδεγμένων όρων από τα κλασικά wikis, αντιστοιχίζοντας τους στους κατάλληλους σημασιολογικούς όρους. Οι κατηγορίες (categories) είναι ένας εύκολος τρόπος απόδοσης σημασιολογίας και κατηγοριοποίησης των σελίδων. Η

έννοια των κατηγοριών είναι διαδεδομένη στο MediaWiki -αλλά και σε άλλα wikis- και μεταφράζεται τυπικά σε OWL κλάση. Οι υπάρχουσες συσχετίσεις (relations) μεταξύ σελίδων επιφορτώνονται με σημασιολογικό περιεχόμενο και αποτελούν τις OWL object properties. Οι ιδιότητες (attributes) επιτρέπουν στους χρήστες να ορίσουν συσχετίσεις άρθρων με οντότητες που δεν είναι άρθρα και αντιστοιχίζονται στις OWL data properties. Τέλος, κάθε άρθρο χρησιμοποιεί ένα URI που διαφέρει από το URL της σελίδας, έτσι ώστε να διαχωρίζονται οι έννοιες από τις απλές HTML σελίδες. Παρόλα αυτά το συγκεκριμένο URI εφόσον χρησιμοποιηθεί παραπέμπει στη σελίδα με το αρχικό URL.

### **Platypus**

Το Platypus 5wiki χρησιμοποιεί έναν WYSIWYG (what you see is what you get) editor για τη συγγραφή περιεχομένου σε XHTML, ενώ Cascading Style Sheets (CSS) υποστηρίζουν τη μορφοποίηση. Επιπλέον τηρούνται τα απαραίτητα ώστε η εφαρμογή να είναι συμβατή με τις επιταγές του Web Accessibility Initiative (WAI) σε επίπεδο AAA, για ιστοσελίδες φιλικές προς άτομα με αναπηρία. Οι σελίδες του Platypus wiki αποθηκεύονται σε XHTML με μεταδεδομένα σε RDF, ενώ το URL της σελίδας ταυτίζεται με το URI της. Για να αναπαρασταθεί ένας σύνδεσμος προς μια σελίδα χρησιμοποιείται η σύμβαση namespace:pagename, που μεταφράζεται στο URL: `http://hostname/namespace/pagename`, ενώ το URL `http://hostname/namespace/pagename/index.rdf` επιστρέφει τα RDF μεταδεδομένα της σελίδας. Όπως στα περισσότερα wikis, αν η σελίδα namespace:pagename δεν υπάρχει, τότε μπορεί να δημιουργηθεί από τον χρήστη.

### **OntoWiki**

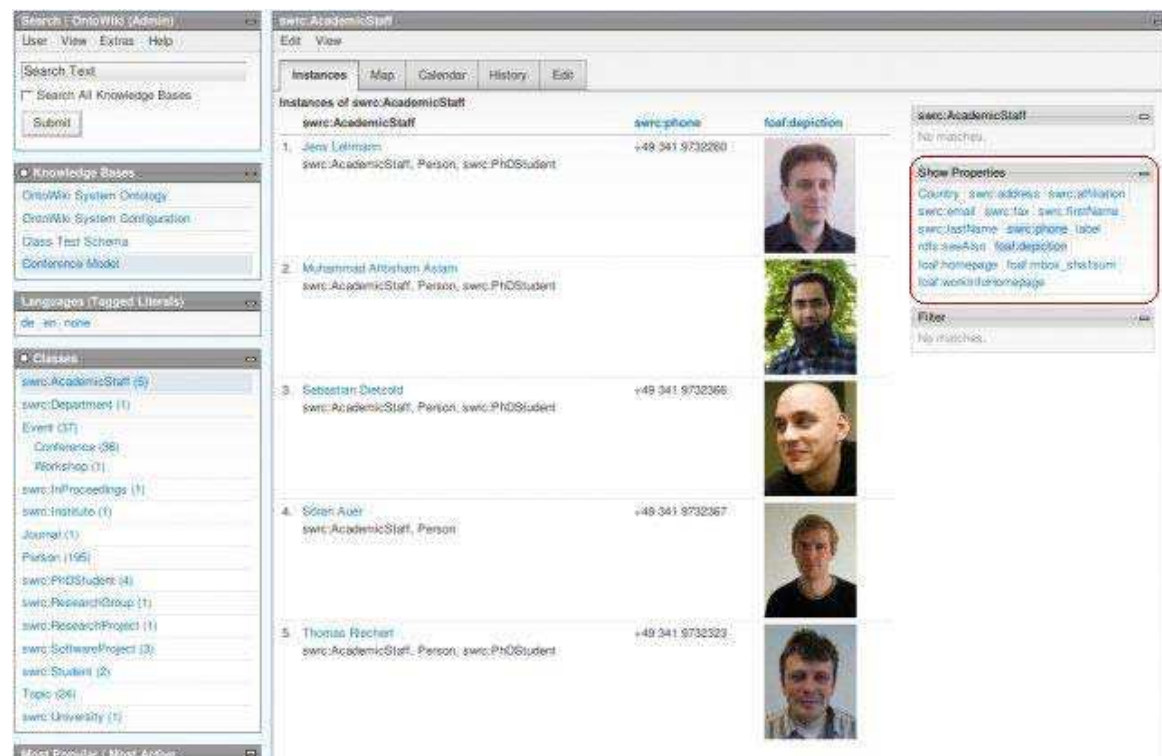
Το OntoWiki (<http://aksw.org/Projects/OntoWiki>) χρησιμοποιείται ως εργαλείο διαχείρισης γνώσης σε περιπτώσεις όπου η γνωσιακή βάση τροφοδοτείται από πολλούς διαφορετικούς χρήστες και σε βάθος χρόνου. Το OntoWiki προσφέρει μια οπτική αναπαράσταση των δεδομένων, σαν χάρτη πληροφοριών με ξεχωριστές όψεις στα μεμονωμένα στιγμιότυπα (Auer, Dietzold, Riechert, 2006). Αν και το όνομα παραπέμπει σε σημασιολογικό wiki όπως τα υπόλοιπα που περιγράφονται σε αυτό το κεφάλαιο, εντούτοις το OntoWiki παρεκκλίνει σχεδιαστικά και μπορεί να θεωρηθεί ως συμπληρωματικό εργαλείο για ένα συμβατικό wiki. Η βασική διαφορά με τα σημασιολογικά wikis που στοχεύουν στην ενσωμάτωση των RDF/OWL σχολιασμών στο κείμενο του wiki,



είναι ότι το OntoWiki διαχωρίζει το κείμενο και την επεξεργασία του από τη διαχείριση του σχετιζόμενου σημασιολογικού περιεχομένου.

### Όψεις πάνω στα δεδομένα

Τα διάφορα στιγμιότυπα παρατίθενται σε μορφή λίστας με παραμετροποιήσιμη δομή, καθώς ο χρήστης επιλέγει ποιες από τις διαθέσιμες στήλες να συμπεριλάβει. Οι λίστες στιγμιότυπων γεμίζουν με τα αποτελέσματα ενός ερωτήματος βάσει κάποιας λέξεως κλειδί ή με το αποτέλεσμα της πλοήγησης μέσω facet ή αποτελούν στιγμιότυπα μιας συγκεκριμένης προεπιλεγμένης κλάσης.



Εικόνα 20. Όψεις στα δεδομένα

Οι σελίδες που παρουσιάζουν τα στιγμιότυπα συνδυάζουν όλες τις ιδιότητες που συνδέονται με το στιγμιότυπο και βεβαίως με το τύπο της κλάσης του. Επιπλέον προσφέρονται δύο εξειδικευμένες όψεις πάνω στα στιγμιότυπα. (α) Επισκόπηση χάρτη: Αν το στιγμιότυπο έχει κάποια ιδιότητα της οποίας το πεδίο τιμών είναι γεωγραφική πληροφορία (γεωγραφικό μήκος και πλάτος), τότε εμφανίζεται ένας χάρτης όπου τοποθετείται το στιγμιότυπο. (β) Επισκόπηση ημερολογίου: Αν το στιγμιότυπο έχει κάποια ιδιότητα της οποίας το πεδίο τιμών είναι ημερομηνία (xsd:date), τότε εμφανίζεται ένα ημερολόγιο με την συγκεκριμένη ημερομηνία.



### 3.2.6 MULTIMEDIAN E-CULTURE

Ο κύριος στόχος του Mutimedian e-culture έργου είναι να επιδείξει πως οι καινοτόμες τεχνολογίες του Σημασιολογικού Ιστού και παρουσίασης δεδομένων μπορούν να χρησιμοποιηθούν για να παρέχουν καλύτερη ευρετηρίαση και αναζήτηση σε μία τεράστια συλλογή εικονικών συλλογών αντικειμένων πολιτισμικής κληρονομιάς. Η αρχιτεκτονική του βασίζεται σε ανοικτά πρότυπα (XML,RDF,OWL,SPARQL) και έχει ως βασική υπόθεση ότι η χρήση γνώσης αυστηρά υπό τη μορφή οντολογιών, θησαυρών, λεξιλογίων είναι ιδιαίτερα χρήσιμη για την ανάκτηση πληροφορίας σε γνωστικά πεδία πλούσια σε πληροφορίες.

### 3.2.7 FREEBASE

Η Freebase είναι μία πρακτική, κλιμακωτή βάση δεδομένων με σκοπό τη δόμηση γενικής ανθρώπινης γνώσης. Η έμπνευσή της προέρχεται από ευρέως διαδεδομένες κοινότητες όπως τη Wikipedia και το Σημασιολογικό Ιστό. Τα δεδομένα στη Freebase δημιουργούνται, δομούνται και συντηρούνται συνεργατικά. Τη στιγμή αυτή περιέχει παραπάνω από 125 εκατομμύρια εγγραφές, περισσότερους από 4000 τύπους και περισσότερες από 7000 ιδιότητες. Η δημόσια πρόσβαση για ανάγνωση και εγγραφή στη Freebase γίνεται μέσω ενός API επερώτησης γράφων βασισμένο στο HTTP πρωτόκολλο που χρησιμοποιεί την Metaweb Query Language (MQL) ως γλώσσα επερώτησης και χειρισμού δεδομένων. Η MQL παρέχει ένα εύχρηστο αντικειμενοστραφές περιβάλλον για τα δεδομένα στη Freebase και είναι σχεδιασμένο για να υποστηρίξει συνεργατικές και διαδικτυακές εφαρμογές προσανατολισμένες στα δεδομένα.

Τα κύρια συστατικά μέρη της Freebase είναι τα εξής:

- Ένα κλιμακώσιμο αποθετήριο πλειάδων το οποίο έχει υποστήριξη για ακύρωση αλλαγών (undo) μέσα από ένα ενσωματωμένο σύστημα εκδόσεων. Αυτό το χαρακτηριστικό δίνει τη δυνατότητα για ευρεία δημιουργία και συντήρηση συνεργατικών δεδομένων
- Ένα API βασισμένο σε HTTP πρωτόκολλο και τη μορφοποίηση JSON με το οποίο μπορούν να συνταχθούν ερωτήματα ανάγνωσης και εγγραφής χρησιμοποιώντας την γλώσσα Metaweb Query Language. Η MQL είναι μία εύχρηστη, αντικειμενοστρεφής

γλώσσα επερωτήσεων με μία δενδρική μορφή των αποτελεσμάτων.

- Ένα ελαφρύ, συνεργατικό σύστημα τύπων. Είναι περισσότερο μία γενική συλλογή δομικών μηχανισμών και συμβάσεων παρά από ένα σύνολο οντολογιών και αναπαραστάσεων. Υπάρχει η δυνατότητα ύπαρξης συγκρουόμενων τύπων και ιδιοτήτων για να αντανακλώνται οι διάφορες απόψεις και συλλήψεις των χρηστών.
- Ένα τεράστιο σύνολο δεδομένων. Περιλαμβάνει πρόσωπα, πολυμέσα, τοποθεσίες και πολλά ακόμα. Τα δεδομένα παρέχονται μέσα από μία άδεια Creative Commons για σχεδόν κάθε χρήση.
- Φιλοσοφία πλήρης κανονικοποίησης: κάθε οντότητα, θέμα ή έννοια στη Freebase έχει το δικό του μοναδικό GUID.

### 3.2.8 GEONAMES

Η GeoNames είναι μία γεωγραφική βάση δεδομένων διαθέσιμη και προσβάσιμη μέσα από διάφορες υπηρεσίες διαδικτύου και με την άδεια Creative Commons

Η βάση δεδομένων περιέχει περισσότερα από 8 εκατομμύρια γεωγραφικά ονόματα που αντιστοιχούν σε παραπάνω από 6.5 εκατομμύρια μοναδικά χαρακτηριστικά. Όλα τα χαρακτηριστικά κατηγοριοποιούνται σε μία από εννέα κατηγορίες χαρακτηριστικών και έπειτα υποκατηγοριοποιούνται με έναν από 645 κωδικούς χαρακτηριστικών. Πέρα από τα ονόματα των περιοχών σε μία ποικιλία από γλώσσες, τα δεδομένα περιλαμβάνουν και τις συντεταγμένες (γεωγραφικό πλάτος, γεωγραφικό μήκος), το υψόμετρο, το πληθυσμό και τη διοικητική διαίρεση και τους ταχυδρομικούς κώδικες.

#### Drupal

Το Drupal είναι ένα ανοιχτού κώδικα Σύστημα διαχείρισης περιεχόμενου. Χρησιμοποιείται για την δημιουργία και την διαχείριση πολλών και διαφορετικών ιστότοπων. Η σταθερότητα, ευελιξία και η ασφάλεια του είναι μερικά από τα σημεία αναφοράς του. Το Drupal έχει ένα βασικό στρώμα, ή πυρήνα, που παρέχει τα βασικά χαρακτηριστικά του Drupal και υποστηρίζει αρθρωτές μονάδες που προσθέτουν επιπλέον λειτουργικότητα ή χαρακτηριστικά. Οι μονάδες που περιέχονται στον πυρήνα του Drupal παρέχουν στους χρήστες τη δυνατότητα να υποβάλλουν, αναθεωρούν, κατηγοριοποιούν ύλη, να εκτελούν αναζητήσεις, να υποβάλλουν σχόλια, να λαμβάνουν μέρος σε φόρουμ συζητήσεων, να ψηφίζουν

σε ψηφοφορίες και να δουλεύουν σε συνεργατικά εγγραφικά έργα, χωρίς την απαίτηση να γνωρίζουν HTML. Οι μονάδες του πυρήνα επιτρέπουν επίσης στους χρήστες να υποβάλλουν και να βλέπουν προσωπικά προφίλ, να επικοινωνούν μεταξύ τους ή τους διαχειριστές του ιστοτόπου. Η διαχείριση εκδόσεων του Drupal, επίσης ένα χαρακτηριστικό του πυρήνα, παρακολουθεί τις αλλαγές του περιεχομένου της ύλης, το ποιος άλλαξε κάτι, τι άλλαξε, την ημερομηνία και ώρα της αλλαγής κ.ο.κ. Παρέχεται από το σύστημα ένα ημερολόγιο με σχόλια αλλαγών και δίνει τη δυνατότητα για μετάβαση του περιεχομένου σε προηγούμενη έκδοση.

Το σύστημα διαχείρισης περιεχομένου Drupal 7 πρωτοπορεί στην ανάπτυξη του RDFa για ευρεία υιοθεσία του Semantic Web (σημασιολογικού διαδικτύου). Συνεχώς η Google και η Bing θα βασίζονται σε δεδομένα δομημένα και αναγνώσιμα από μηχανές από τις ιστοσελίδες τις οποίες σκανάρουν. Ο σχεδιασμός του Drupal 7 ενσωματώνει σημασιολογικά δεδομένα που κάνουν την από μηχανή – προς μηχανή αναζήτηση έμφυτη για μια ιστοσελίδα βασισμένη στο Drupal 7. Προσθετη αξία μπορεί να δοθεί με το RDFa, δίνοντας στις μηχανές αναζήτησης περισσότερα στοιχεία, μη ορατά στους ανθρώπους, όπως το γεωγραφικό μήκος και πλάτος ενός χώρου για την εμφάνισή του σε ένα χάρτη, ή να παρέχει την ημερομηνία σε μορφή ISO για σωστή απεικόνιση στα αποτελέσματα αναζήτησης για διαφορετικές χώρες.

### **GridVine**

Το GridVine αποτελεί ένα σημασιολογικό δίκτυο με δυνατότητα κλιμάκωσης που υποστηρίζει την δημιουργία τοπικών σχημάτων ενώ ταυτόχρονα παρέχει καθολική σημασιολογική διαλειτουργικότητα. Ακολουθεί την αρχή της ανεξαρτησίας δεδομένων και διαχωρίζει το λογικό από το φυσικό στρώμα. Το λογικό στρώμα αποτελείται από το σημασιολογικό στρώμα για την διαχείριση και την αντιστοίχιση δεδομένων και μεταδεδομένων σχημάτων, ενώ το φυσικό στρώμα αποτελείται από ένα δομημένο δίκτυο ομότιμων κόμβων που δρομολογεί μηνύματα με υψηλή απόδοση. Το δίκτυο αυτό χρησιμοποιείται για να υλοποιήσει διάφορες λειτουργίες στο λογικό στρώμα όπως αναζήτηση βάσει γνωρίσματος, διαχείριση σχημάτων και αντιστοίχιση σχημάτων.

Οι κόμβοι στο GridVine έχουν την δυνατότητα να δημοσιεύουν τους διαθέσιμους πόρους δημιουργώντας RDF τριπλέτες (μεταδεδομένα). Μια RDF τριπλέτα αποθηκεύεται τρεις φορές στο δίκτυο χρησιμοποιώντας τρία διαφορετικά

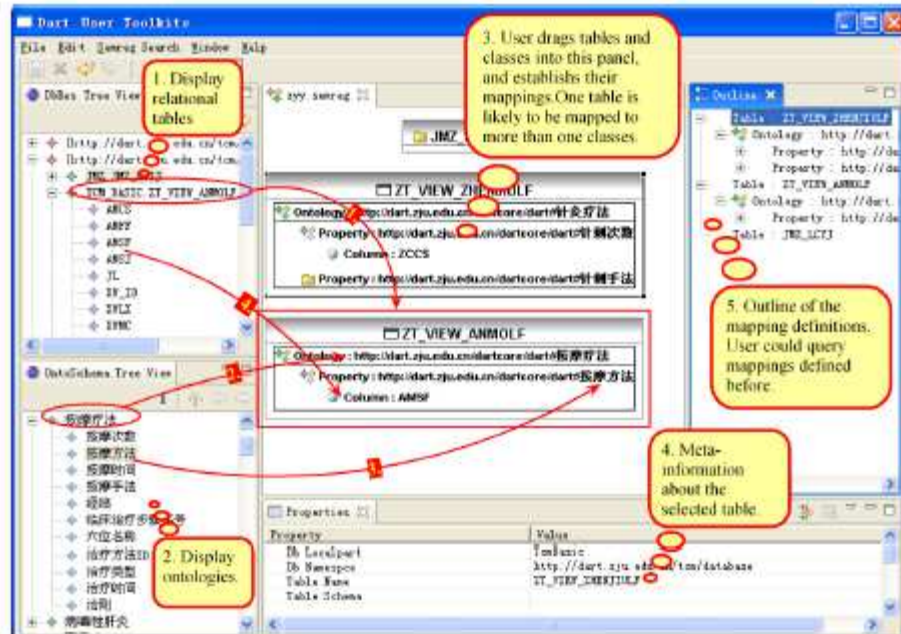
κλειδιά, βάσει των τιμών του υποκειμένου, του κατηγορήματος και του αντικειμένου. Επομένως, τα είδη επερωτήσεων υποστηρίζονται επίσης από το σύστημα GridVine. Επιπλέον, υποστηρίζονται οι αναζητήσεις χρησιμοποιώντας το πρόθεμα μιας συμβολοσειράς π.χ. χρησιμοποιώντας την αρχή μιας συμβολοσειράς που αναπαριστά μια τιμή αντικειμένου. Αυτές οι αναζητήσεις υποστηρίζονται εύκολα χρησιμοποιώντας τους μηχανισμούς δρομολόγησης του PGrid.

Το σύστημα GridVine επιτρέπει στους κόμβους να βρίσκουν νέα σχήματα από γνωστά βασικά σχήματα (χρησιμοποιώντας RDFS) παρέχοντας έτσι κληρονομικότητα σχήματος. Κάθε κόμβος έχει επίσης την δυνατότητα να δημιουργεί μια αντιστοίχιση μεταξύ δύο σχημάτων και έτσι δημιουργούνται οι σύνδεσμοι μετάφρασης μεταξύ των κόμβων του δικτύου (χρησιμοποιώντας OWL). Με αυτόν τον τρόπο οι επερωτήσεις στέλνονται από ένα σημασιολογικό πεδίο σε ένα άλλο. Υπάρχουν δύο τεχνικές που χρησιμοποιούνται για την επίλυση μεταφραστικών συνδέσμων, η επαναληπτική και η αναδρομική επίλυση. Με την επαναληπτική επίλυση ο κόμβος που θέτει μια επερώτηση RDF προσπαθεί να βρει και να επεξεργαστεί όλους τους συνδέσμους της μετάφρασης μόνος του, ενώ με την αναδρομική επίλυση περισσότεροι του ενός κόμβοι χρησιμοποιούνται για την αποστολή της επερώτησης και την μετάφραση της.

### **DartGrid**

Το Dartgrid αναπτύχθηκε από Κινέζους ερευνητές και χρησιμοποιήθηκε για την διασύνδεση πολλών σχεσιακών βάσεων σχετικά με την Κινέζικη παραδοσιακή ιατρική και την σημασιολογική αναζήτηση σε αυτές.

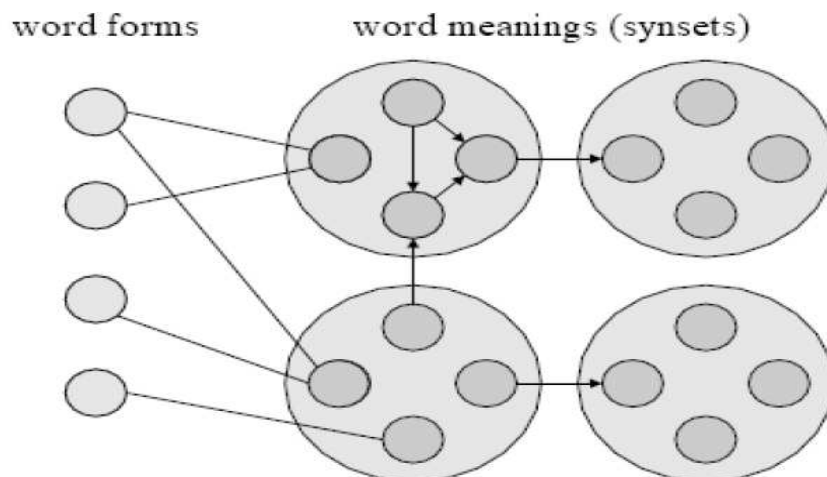
Και αυτό το εργαλείο βασίζεται στην μετατροπή των σημασιολογικών ερωτημάτων σε SQL ερωτήματα με βάση κάποια χαρτογράφηση που παρέχει ο χρήστης μεταξύ οντολογίας και βάσης. Αυτό που κάνει το DartGrid να διαφέρει είναι η παρουσία γραφικών εργαλείων τόσο για την διαδικασία χαρτογράφησης (Εικόνα 21) όσο και για την υποβολή σημασιολογικών ερωτημάτων. Τα εργαλεία αυτά στηρίζονται στο D2RQ. Επιπλέον η εφαρμογή προσφέρει και αναζήτηση κειμένου στις διασυνδεδεμένες βάσεις.



Εικόνα 21. Γραφικό εργαλείο DartGrid για την διαδικασία Mapping

### WordNet— Μια Online λεξικογραφική βάση δεδομένων

Το WordNet είναι μια ελεύθερα διαθέσιμη on-line λεξικογραφική βάση δεδομένων. Το τμήμα γλωσσολογίας του Πανεπιστημίου Πρίνσετον τη δημιούργησε σαν αποτέλεσμα της ψυχολογολογικής έρευνας. Παρόλα αυτά την τελευταία δεκαετία το WordNet αποδείχτηκε πολύ χρήσιμη πηγή για αυτοματοποιημένη επεξεργασία της φυσικής γλώσσας. Τεχνικά, το WordNet είναι ένας ηλεκτρονικός θησαυρός, ορίζοντας μεγάλα σύνολα εννοιών λέξεων, διασυνδεδεμένα με σημασιολογικούς δείκτες. Η λογική δομή του WordNet φαίνεται στην Εικόνα 22.



Εικόνα 22. Η λογική δομή του WordNet

**4.1 ΣΗΜΑΣΙΟΛΟΓΙΚΕΣ ΜΗΧΑΝΕΣ ΑΝΑΖΗΤΗΣΗΣ**

Τα σημασιολογικά συστήματα αναζήτησης έχουν κάνει την εμφάνισή τους εδώ και λίγα χρόνια. Παρά το γεγονός ότι όλα περιγράφονται με τον όρο «σημασιολογικά» υπάρχουν σημαντικές διαφορές μεταξύ τους, τόσο σε επίπεδο σχεδιασμού όσο και σε επίπεδο λειτουργικότητας. Από τη μια μεριά υπάρχουν συστήματα που χρησιμοποιούν λεξικά και θησαυρούς συγκεκριμένων όρων για να αναλύσουν και να κατανοήσουν το νόημα και τη σημασία ενός ερωτήματος ή κάποιων λέξεων-κλειδιών. Μια τεχνική που ονομάζεται Latent Semantic Analysis και είναι βασισμένη σε αυτή την αρχή χρησιμοποιείται κατά κόρον από τις σημερινές μεγάλες μηχανές αναζήτησης όπου συχνά αναφέρεται ως Latent Semantic Indexing ή LSI. Στο άλλο άκρο υπάρχουν οι πιο αμιγείς προσεγγίσεις με συστήματα που επεξεργάζονται αποκλειστικά σημασιολογικές περιγραφές δεδομένων καθώς και οντολογίες, σε RDF ή OWL. Αυτή είναι και η κατηγορία σημασιολογικών μηχανών αναζήτησης στην οποία ανήκει η WebOWL.

Οι Esmaili et al. προτείνουν μια μέθοδο κατηγοριοποίησης των μηχανών αναζήτησης σημασιολογικού ιστού. Σύμφωνα με αυτή υπάρχουν δύο κατηγορίες μηχανών αναζήτησης σημασιολογικού ιστού: μηχανές προσανατολισμένες σε Οντολογίες (Ontology Search Engines) και Σημασιολογικές μηχανές, οι οποίες προσπαθούν να βελτιώσουν τα αποτελέσματα αναζήτησης χρησιμοποιώντας πρότυπα του Σημασιολογικού Ιστού και οντολογικές γλώσσες (Semantic Search Engines) για την ανίχνευση σχέσεων μεταξύ πόρων (resources) του Διαδικτύου.

Ο σκοπός μιας Σημασιολογικής Μηχανής Αναζήτησης είναι να μπορεί να απαντά ερωτήματα για σημασιολογικά δεδομένα. Κατά συνέπεια, τα ερωτήματα αυτά είναι λογικό να γίνονται σε επίπεδο σημασιολογικών κόμβων, επιτρέποντας στον χρήστη να καθορίζει σημασιολογικά αντικείμενα και τις επιθυμητές σχέσεις μεταξύ αυτών των αντικειμένων. Το εάν το σύστημα διασπά κάθε σημασιολογικό κόμβο σε ένα σύνολο εγγραφών ή εάν μπορεί να χειριστεί τον κάθε κόμβο ως ξεχωριστό στοιχείο δεν έχει σημασία καθώς η εσωτερική αναπαράσταση δεν θα πρέπει να γίνεται αντιληπτή από τον χρήστη. Το βασικό στοιχείο των σημασιολογικών ερωτημάτων είναι ότι επιτρέπουν στον χρήστη να χρησιμοποιήσει σημασιολογικά δεδομένα για να προσδιορίσει το ερώτημα με

πολύ μεγαλύτερη ακρίβεια από ότι στις κλασσικές γλώσσες ερωτημάτων όπως η SQL.

### **Swoogle**

Η Swoogle είναι μια μηχανή αναζήτησης η οποία ανήκει στην πρώτη κατηγορία. Η Swoogle ( Ding *et al*,2004) είναι μια σημασιολογική μηχανή αναζήτησης που χρησιμοποιεί έναν crawler βασισμένο στη Jena για να ανακαλύψει αρχεία RDF και OWL. Το Swoogle επικεντρώνεται σε ολόκληρο το αρχείο με τα σημασιολογικά δεδομένα και όχι στα ίδια τα δεδομένα. Έτσι, ένα ερώτημα προς το σύστημα δεν επιστρέφει την ίδια την απάντηση αλλά μια λίστα με αρχεία οντολογιών. Το σύστημα χρησιμοποιεί μια βάση MySQL για να αποθηκεύει οντολογίες και εφαρμόζει έναν αλγόριθμο κατάταξης παρόμοιο με τον PageRank (Brin and Page,1998), ο οποίος βασίζεται σε σημασιολογικές αναφορές ανάμεσα στις οντολογίες. Τα ερωτήματα γίνονται με βάση λέξεις-κλειδιά στο κλασσικό πεδίο κειμένου που υπάρχει σε όλες τις μηχανές αναζήτησης. Καμία σημασιολογική περιγραφή των ερωτημάτων δεν υποστηρίζεται, όμως παρέχεται και μια πιο πολύπλοκη λειτουργία ερωτημάτων όπου ο χρήστης, μέσω μιας φόρμας, συμπληρώνει περιορισμούς σε ένα γενικό ερώτημα SQL ειδικά σχεδιασμένο για την βάση του Swoogle.

### **Semantic Web Search Engine (SWSE)**

Η δεύτερη κατηγορία μηχανών αναζήτησης αφορά τις μηχανές όπου η αναζήτηση δε στηρίζεται στην ευρετηριοποίηση των πόρων του Σημασιολογικού Ιστού αλλά στην ενοποίηση του συνόλου των πόρων που έχουν συλλεχθεί και στη διαχείρισή τους ως μια ενότητα. Η μηχανή αναζήτησης Semantic Web Search Engine (SWSE) αποτελεί μια μηχανή αναζήτησης η οποία ανήκει σε αυτή την κατηγορία. Βασίζεται σε ένα μοντέλο το οποίο είναι προσανατολισμένο στις οντότητες και συνεπώς επιτρέπει την περιγραφή οντοτήτων, όπως είναι οι άνθρωποι, οι περιοχές και γενικά οι πόροι. Κατά συνέπεια, τα αποτελέσματα της αναζήτησης είναι δομημένες περιγραφές των οντοτήτων. Το σύστημα λειτουργεί σε ένα σύνολο από RDF εγγραφές και έχει ως στόχο την παροχή δυνατότητας αναζήτησης και διενέργειας ερωτημάτων σε δεδομένα τα οποία προέρχονται από την ενοποίηση ενός μεγάλου αριθμού πηγών.

Η Semantic Web Search Engine ή SWSE χρησιμοποιεί το API της Jena για την αποθήκευση και την επεξεργασία δεδομένων RDF. Αντί να χρησιμοποιεί κάποιο σύστημα βάσης δεδομένων, η μηχανή SWSE αποθηκεύει όλα τα

δεδομένα στη μνήμη του συστήματος πράγμα που φυσικά περιορίζει το σύστημα μόνο για πειραματική και ερευνητική χρήση. Η υποβολή ερωτημάτων γίνεται μέσω του γνωστού πεδίου κειμένου όπου οι χρήστες πληκτρολογούν λέξεις-κλειδιά. Η SWSE επιστρέφει σημασιολογικές οντότητες που περιέχουν αυτές τις λέξεις-κλειδιά στα πεδία `rdfs:label`, `rdfs:comment` και `rdfs:literal` των ιδιοτήτων τους.

## **HAKIA**

Η μηχανή αναζήτησης Hakia είναι μια σημασιολογική μηχανή (semantic), δηλ. χρησιμοποιεί μια σημασιολογική προσέγγιση για να καθορίσει εάν οι λέξεις κλειδιά τις αναζήτησης συσχετίζονται με ένα webpage ακόμα κι αν οι λέξεις κλειδιά δεν υπάρχουν στο webpage.

Η Hakia χρησιμοποιεί τρεις τεχνολογίες για την εύρεση της σημασίας των δεδομένων:

OntoSem - sense repository( γλωσσική βάση δεδομένων όπου οι λέξεις είναι ταξινομημένες σε διάφορες κατηγορίες).

QDEX - Query indexing technique (αντικατάσταση για το δείκτη που οι περισσότερες μηχανές χρησιμοποιούν για να σώσουν το περιεχόμενο με δείκτη για τις ερωτήσεις στις οποίες απαντά το περιεχόμενο, μειώνοντας έτσι τον όγκο πληροφορίας που πρέπει να ψάχνουν).

SemanticRank algorithm (ταξινομεί το περιεχόμενο βάσει περισσότερης ανάλυσης της πρότασης. Η αξιοπιστία και η ηλικία του περιεχομένου χρησιμοποιούνται επίσης για να καθορίσουν τη σχετικότητα).

## **XSearch**

Το σύστημα XSearch είναι μια σημασιολογική μηχανή αναζήτησης για δεδομένα XML χωρίς υποστήριξη για RDF/OWL. Ο λόγος που αναφέρεται εδώ είναι επειδή το XSearch αναλύει σημασιολογικά αρχεία και ερωτήματα XML και επιστρέφει μόνο το κομμάτι του αρχείου που απαντά στο ερώτημα αντί για ολόκληρο το αρχείο. Το κάθε αρχείο XML χρησιμοποιείται σαν ένα δέντρο και επιστρέφονται μόνο οι κόμβοι των οποίων η σημασία συνδέεται με το ερώτημα. Αυτό επιτυγχάνεται με τη χρήση σημασιολογικής πληροφορίας στα ερωτήματα τα οποία όμως διατηρούν μια αρκετά απλή σύνταξη έτσι ώστε να μπορεί να τα σχηματίσει και ένας μη ειδικός.



## **Sindice**

Η μηχανή Sindice (Tummarello et al, 2007) λειτουργεί μέσω ενός API που προσφέρει δυνατότητες αναζήτησης, αποθήκευσης και ανανέωσης δεδομένων. Επικεντρώνεται σε δεδομένα RDF με στόχο δυνατότητας κλιμάκωσης την αποθήκευση 300 εκατομμυρίων οντοτήτων RDF από περίπου 3 εκατομμύρια πηγές (αρχεία). Χρησιμοποιεί μια βάση Berkeley για να αποθηκεύει έναν πίνακα κατακερματισμού (hash table) από πόρους RDF. Το σύστημα δεν υποστηρίζει ερωτήματα σε RDF και η αναζήτηση γίνεται απλά με αντιστοίχιση λέξεων-κλειδιών ή URIs με τους identifiers των resources. Η κατάταξη των αποτελεσμάτων γίνεται βάσει κάποιων μετα-δεδομένων για το κάθε στοιχείο, όπως το hostname της πηγής, το μέγεθος της ή και εξωτερικούς παράγοντες (π.χ. η τιμή του PageRank που δίνει το Google για το site στο οποίο ανήκει η πηγή).

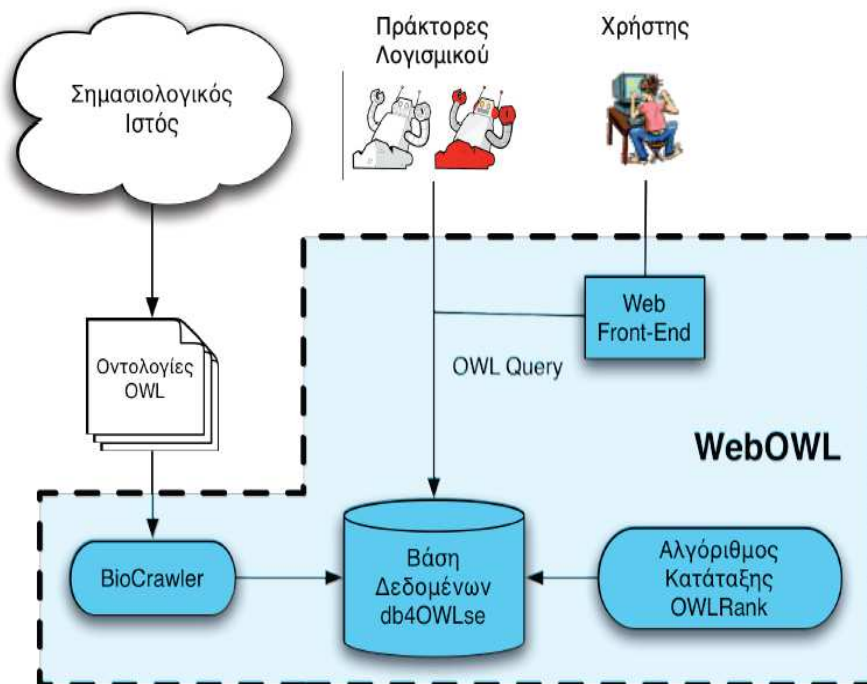
## **SemSearch**

Το σύστημα SemSearch (Lei et al, 2006) χρησιμοποιεί το Sesame για να παρέχει σημασιολογική αναζήτηση. Ο κύριος στόχος του συστήματος είναι η απόκρυψη της σημασιολογικής πολυπλοκότητας από τους τελικούς χρήστες. Για αυτό το λόγο, το σύστημα επικεντρώνεται στη διάδραση με το χρήστη και στους τρόπους με τους οποίους ο χρήστης μπορεί να σχηματίσει σημασιολογικά ερωτήματα. Για το σκοπό αυτό, χρησιμοποιούνται μια σειρά από διαφορετικά επίπεδα επεξεργασίας του κάθε ερωτήματος το οποίο εισάγεται από τον χρήστη μέσω του κλασσικού πεδίου κειμένου που συναντάμε σε όλες τις σημερινές μηχανές αναζήτησης. Στη συνέχεια το ερώτημα αναλύεται και τελικά μετασχηματίζεται σε ερώτημα SeRQL. Το σύστημα επιστρέφει τα αποτελέσματα της αναζήτησης μαζί με μια μικρή επεξήγηση του λόγου για τον οποίο έγινε η επιλογή του κάθε αποτελέσματος.

## **WebOWL**

Η WebOWL είναι μια σημασιολογική μηχανή αναζήτησης για δεδομένα σε γλώσσα OWL. Είναι βασισμένη στις αρχές των σημερινών μεγάλων εμπορικών μηχανών αναζήτησης αλλά εστιάζεται στα ίδια τα σημασιολογικά δεδομένα και όχι στις ιστοσελίδες (αρχεία) που τα περιέχουν. Με βάση αυτή τη λογική, η WebOWL χρησιμοποιεί έναν αλγόριθμο κατάταξης ο οποίος λειτουργεί σε επίπεδο σημασιολογικών οντοτήτων, δηλαδή κλάσεων και individuals. Το σύστημα περιλαμβάνει όλα τα κύρια συστατικά μιας μηχανής αναζήτησης τα οποία είναι: α)

ένα σύστημα web crawling, β) μια κεντρική βάση δεδομένων και γ) ένας αλγόριθμος κατάταξης αποτελεσμάτων. Με μόνη εξαίρεση το σύστημα crawling, λειτουργεί στο επίπεδο ολόκληρων domains και οντολογιών, όλα τα άλλα μέρη λογισμικού της WebOWL, συμπεριλαμβανομένου και του αλγορίθμου κατάταξης, λειτουργούν βάσει διακριτών σημασιολογικών οντοτήτων της OWL και όχι βάσει ολόκληρων σελίδων ή οντολογιών.



Εικόνα 23. Η αρχιτεκτονική της μηχανής αναζήτησης WebOWL

Η Σημασιολογική Μηχανή Αναζήτησης WebOWL χρησιμοποιεί ένα Web front-end μέσω του οποίου επιτρέπει στους χρήστες να υποβάλλουν ερωτήματα και να πλοηγούνται στα αποτελέσματα. Αυτή η διεπαφή χρηστών αναπτύχθηκε κυρίως για να επιδείξει τις δυνατότητες της μηχανής αναζήτησης και όχι ως προϊόν έρευνας στους τρόπους αλληλεπίδρασης των χρηστών με σημασιολογικά δεδομένα. Παρόλα αυτά, η ανάπτυξή της έφερε στην επιφάνεια πολλά προβλήματα και ερωτήματα για το πως θα μπορέσουν τελικά οι χρήστες τέτοιων συστημάτων να χρησιμοποιήσουν εύκολα και αποτελεσματικά τις τόσο πολύπλοκες δομές πληροφοριών που σχηματίζουν τα σημασιολογικά δεδομένα.

### **Μηχανές αναζήτησης προσανατολισμένες στον άνθρωπο**

Μηχανές αναζήτησης όπως η Falcons και η SWSE παρέχουν υπηρεσίες αναζήτησης βασισμένες σε λέξεις-κλειδιά και προσανατολισμένες προς χρήστες-

ανθρώπους και ακολουθούν ένα παρόμοιο πρότυπο αλληλεπίδρασης με τους υπάρχοντες ηγέτες της αγοράς όπως η Google και η Yahoo. Ο χρήστης βλέπει μια μπάρα αναζήτησης στην οποία μπορεί να βάλει λέξεις-κλειδιά σχετικά με το αντικείμενο ή το θέμα για το οποίο ενδιαφέρεται και η εφαρμογή επιστρέφει μια λίστα αποτελεσμάτων που μπορεί να είναι σχετικά με το ερώτημα. Ωστόσο, πέρα από το να παρέχουν απλά συνδέσμους αποτελεσμάτων της αναζήτησης προς τα έγγραφα πηγές στα οποία αναφέρονται οι ζητούμενες λέξεις-κλειδιά, και η SWSE και η Falcons παρέχουν μια πιο λεπτομερή διεπιφάνεια στον χρήστη που αξιοποιεί την υποκείμενη δομή των δεδομένων. Και οι δύο παρέχουν μια περίληψη της οντότητας την οποία επιλέγει ο χρήστης από τη λίστα αποτελεσμάτων, σε παραβολή με επιπλέον δομημένα δεδομένα που έχουν «περπατηθεί» (crawled) από τον Ιστό και συνδέσμους προς σχετιζόμενες οντότητες.

#### **4.2 ΣΗΜΑΣΙΟΛΟΓΙΚΕΣ ΠΥΛΕΣ**

Ο Παγκόσμιος Ιστός έχει καταστήσει ηλεκτρονικά διαθέσιμο έναν τεράστιο όγκο πληροφορίας και έχει φοβερή επιτυχία τόσο σε όρους διαθέσιμης πληροφορίας όσο και σε ρυθμό ανάπτυξης του αριθμού των χρηστών του. Το Διαδίκτυο έχει αποκτήσει πλέον περισσότερους από ένα δισεκατομμύριο χρήστες και πάνω από ένα δισεκατομμύριο έγγραφα. Η επιτυχία αυτή υπήρξε σε μεγάλο βαθμό αποτέλεσμα της απλότητάς του, η οποία έδινε στους προγραμματιστές, στους παρόχους πληροφορίας και στους χρήστες εύκολη πρόσβαση σε νέο περιεχόμενο. Παρ' όλα αυτά, η ίδια η απλότητά του και η εντυπωσιακή επέκτασή του είναι που δημιουργούν σοβαρά έως και απρόσκοπτα εμπόδια στην περαιτέρω ανάπτυξη του Παγκόσμιου Ιστού. Το γενικό πρόβλημα στο να βρει κανείς πληροφορίες στον Παγκόσμιο Ιστό έγκειται στο ότι οι αναζητήσεις είναι ανακριβείς και συχνά δίνουν πολλές χιλιάδες αποτελέσματα. Επιπλέον οι χρήστες αντιμετωπίζουν το πρόβλημα ότι πρέπει να διαβάσουν το έγγραφο που επιστρέφεται από την αναζήτηση προκειμένου να εντοπίσουν την πληροφορία που αναζητούσαν. Αυτοί οι περιορισμοί εμφανίζονται φυσικά στις υπάρχουσες Δικτυακές Πύλες που βασίζονται σε αυτήν την τεχνολογία, κάνοντας την αναζήτηση, την πρόσβαση, την εξαγωγή, τη μετάφραση και την επεξεργασία της πληροφορίας μια δύσκολη και χρονοβόρα διαδικασία. Για αυτούς τους λόγους οι τεχνολογίες του σημασιολογικού ιστού μπορούν να βελτιώσουν αισθητά τη

διαδικασία διανομής της πληροφορίας ξεπερνώντας τα προβλήματα που παρουσιάζουν οι σημερινές Δικτυακές Πύλες. Με αυτή την έννοια, οι Δικτυακές Πύλες που βασίζονται στις τεχνολογίες του Σημασιολογικού Ιστού αποτελούν την επόμενη γενιά Δικτυακών Πυλών και ονομάζονται Σημασιολογικές Πύλες (ΣΠ). Σημασιολογικές πύλες είναι οι παρακάτω:

### **Mindswap**

Οι δημιουργοί αυτής της Πύλης ισχυρίζονται ότι αυτή είναι η πρώτη ιστοσελίδα που χρησιμοποίησε το Σημασιολογικό Ιστό και υλοποιείται με τεχνολογίες συμβατές με OWL. Η ιστοσελίδα αυτή προσπαθεί να κάνει επίδειξη των αυξημένων δυνατοτήτων που προσφέρουν σε μια ιστοσελίδα οι τεχνολογίες του Σημασιολογικού Ιστού. Παρ' όλα αυτά η ιστοσελίδα αυτή δεν δείχνει τις τεχνολογικές δυνατότητες που προσφέρει η εφαρμογή τεχνολογιών του Σημασιολογικού Ιστού: για παράδειγμα ούτε η αναζήτηση ούτε το μενού εφαρμόζουν τεχνολογίες του Σημασιολογικού Ιστού.

### **Οι Πύλες της Karlsruhe**

Το AIFB στο Πανεπιστήμιο της Καρλσρούης έχει αναπτύξει μια από τις πρώτες Σημασιολογικές Πύλες. Σχεδιάζοταν να αποτελέσει την πλατφόρμα για ανταλλαγή πληροφορίας και συνεργασία για την κοινότητα KA2 (Knowledge Annotation initiative of the Knowledge Acquisition community). Για αυτό το λόγο δημιουργήθηκε μια Οντολογία ως βάση για να χαρακτηριστούν δικτυακά έγγραφα για την κοινότητα των χρηστών με ενδιαφέρον για την ανάκτηση γνώσης ώστε να γίνει δυνατή η «έξυπνη» πρόσβαση σε αυτά. Η Πύλη αυτή όμως έχει σταματήσει να υποστηρίζεται και να συντηρείται. Μολονότι η Πύλη βασίζεται συνολικά σε Οντολογία, οι λειτουργικότητες που προσφέρει είναι πολύ απλές και σε επίπεδο πρόσβασης στην πληροφορία από τον τελικό χρήστη είναι σε πολύ χαμηλότερο επίπεδο από αντίστοιχες συμβατικές Πύλες.

### **OntoWeb Portal**

Η OntoWeb είναι μια Δικτυακή Πύλη για μια κοινότητα ανθρώπων, από τα ακαδημαϊκά ιδρύματα και τη βιομηχανία, που δείχνουν ενδιαφέρον για το Σημασιολογικό Ιστό. Η Πύλη αυτή είχε δημιουργηθεί σαν μέρος του προγράμματος OntoWeb της Ευρωπαϊκής Ένωσης.

### 4.3 ΕΛΕΓΧΟΣ ΔΗΜΟΣΙΕΥΣΗΣ ΔΕΔΟΜΕΝΩΝ

Η πιο κύρια αρχή των Συνδεδεμένων Δεδομένων είναι τα δεδομένα να προέρχονται μέσω dereferenceable URIS. Δεν αρκεί όμως μόνο το θέμα αυτό. Θα πρέπει τα δημοσιευμένα δεδομένα να συμβαδίζουν με τις συνήθεις πρακτικές που διασφαλίζουν ότι τα δεδομένα μπορούν να ανακαλυφθούν εύκολα από τις εφαρμογές των Συνδεδεμένων δεδομένων, να γίνουν δηλαδή μέρος του ιστού των δεδομένων. Καλό είναι πριν από τη δημοσίευση δεδομένων, ο εκδότης να συμβουλευεται τα παρακάτω στοιχεία:

1. Έχουν τα δεδομένα συνδέσμους προς άλλα δεδομένα; Οι σύνδεσμοι RDF συνδέουν δεδομένα από άλλες διαφορετικές πηγές σε ένα παγκόσμιο δίκτυο και επιτρέπουν στις εφαρμογές Συνδεδεμένων Δεδομένων να πλοηγούνται ανάμεσα στις πηγές αυτές. Είναι απαραίτητο τα νέα δεδομένα να έχουν συνδέσμους προς άλλες πηγές δεδομένων.

2. Παρέχονται μεταδεδομένα προέλευσης; Για να μπορούν να αναγνωρίζουν οι εφαρμογές την προέλευση των δημοσιευμένων δεδομένων και άρα να κρίνουν και την ποιότητά τους, είναι απαραίτητο να παρέχονται και μεταδεδομένα προέλευσης μαζί με τα κύρια δεδομένα.

3. Παρέχονται μεταδεδομένα σχετικά με την αδειοδότηση; Τα δεδομένα θα πρέπει να παρέχουν μεταδεδομένα που περιγράφουν τους όρους χρήσης τους.

4. Έχουν χρησιμοποιηθεί όροι από ευρέως χρησιμοποιούμενα λεξιλόγια; Για να είναι πιο εύκολα κατανοητά στις εφαρμογές τα Συνδεδεμένα Δεδομένα, συνίστανται η χρήση διαδεδομένων λεξιλογίων για την αναπαράσταση των δεδομένων.

5. Χαρτογραφούνται οι όροι που ορίσαμε σε όρους άλλων λεξιλογίων;

6. Παρέχονται μεταδεδομένα για το επίπεδο του συνόλου των δεδομένων; Εκτός από τα δεδομένα που πρέπει να είναι αυτο-περιγραφικά, είναι επιθυμητό οι εκδότες των δεδομένων να παρέχουν μεταδεδομένα που να περιγράφουν τα χαρακτηριστικά ολόκληρου του συνόλου δεδομένων.

## ΣΥΜΠΕΡΑΣΜΑΤΑ

Ο Παγκόσμιος Ιστός αρχικά δημιουργήθηκε προκειμένου οι πληροφορίες που βρίσκονται σε αυτόν να «καταναλώνονται» από ανθρώπους, και παρά το γεγονός ότι όλα όσα υπάρχουν σε αυτόν μπορούν να γραφούν από μηχανές, κανένα από αυτά τα δεδομένα δεν είναι κατανοητό για τις μηχανές. Είναι πολύ δύσκολο να αυτοματοποιηθούν τα πάντα στον Παγκόσμιο Ιστό, και εξαιτίας του όγκου των πληροφοριών που περιέχονται σε αυτόν, είναι αδύνατη η χειρονακτική διαχείριση τους. Ο Σημασιολογικός Ιστός (Semantic Web) αποτελεί μια επέκταση του παγκόσμιου ιστού όπου οι πληροφορίες έχουν καλά ορισμένες έννοιες, διευκολύνοντας τους ανθρώπους και τους υπολογιστές να συνεργαστούν καλύτερα. Περνά από την κλασσική αναπαράσταση πληροφοριών με τη μορφή εγγράφων σε φυσική γλώσσα τα οποία είναι κατανοητά μόνο από ανθρώπους, στη σημασιολογική αναπαράσταση πληροφοριών η οποία μπορεί να γίνει κατανοητή από υπολογιστές. Τα σημασιολογικά δεδομένα είναι μετα-δεδομένα που έχουν ως στόχο την περιγραφή πληροφοριών ή πόρων (resources). Είναι δομημένα σε διαφορετικά επίπεδα (layers) με το κάθε επίπεδο να είναι πιο εξειδικευμένο και περίπλοκο από αυτό που βρίσκεται ακριβώς από κάτω του.

Καθώς ο Παγκόσμιος Ιστός σταδιακά μετεξελίσσεται στο Σημασιολογικό Παγκόσμιο Ιστό (Semantic Web) νέα μοντέλα και πρότυπα (XML, RDF, OWL) αναπτύσσονται για την προώθηση αυτής της διαδικασίας. Η έκφραση, μετάδοση και αναζήτηση πληροφοριών με χρήση αυτών των προτύπων ανοίγει νέους ορίζοντες στη χρήση του Διαδικτύου.

Οι βασικές τεχνολογίες του Σημασιολογικού ιστού αντιστοιχούν σε ένα σύνολο διαβαθμισμένων προδιαγραφών. Τα στοιχεία κυρίως που χρησιμοποιούνται είναι το Resource Description Framework (RDF), το RDF Schema και η Web Ontology language (OWL). Βασισμένη σε αυτά τα βασικά πρότυπα είναι η SPARQL, μια προτυποποιημένη γλώσσα επερωτήσεων για το RDF η οποία επιτρέπει τη σύνδεση αποκεντροποιημένων συλλογών RDF δεδομένων. Όλες αυτές οι γλώσσες «χτίστηκαν» στα θεμέλια των URIs, XML, και XML namespaces.

Ένα έγγραφο του Σημασιολογικού ιστού μπορεί να αποτελείται από γεγονότα, κλάσεις, και ορισμούς ιδιοτήτων, λογικών περιορισμών και

μεταδεδομένων. Η πλήρης κατανόηση του περιεχομένου του απαιτεί λοιπόν αρκετή εργασία. Οι προγραμματιστές πρέπει να αντιμετωπίσουν το σχεδιαστικό ζήτημα του πότε και σε πόσο βάθος πρέπει οι μηχανές αναζήτησης να εξάγουν συμπεράσματα. Αυτή η εργασία εξαγωγής συμπερασμάτων παράγει επιπλέον δεδομένα, περιορισμούς και μεταδεδομένα που επίσης χρειάζεται να ταξινομηθούν, πιθανότατα μαζί με την αντίστοιχη δικαιολόγηση ή την προέλευσή τους. Οι συμβατικές μηχανές αναζήτησης δεν επιχειρούν να κατανοήσουν το περιεχόμενο λόγω της δυσκολίας του έργου και της έρευνας που απαιτεί στην κατανόηση κειμένου. Τα εργαλεία που υπάρχουν διαθέσιμα αυτή τη στιγμή για τη δημοσίευση, ανακάλυψη και κατανάλωση των συνδεδεμένων δεδομένων είναι πάρα πολλά και καλύπτουν συνήθως διαφορετικές ανάγκες. Χρειάζεται επομένως προσοχή στην επιλογή των κατάλληλων εργαλείων για την κάθε εργασία.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

Berners-Lee, T., Fielding, R., & Masinter, L. (1998). RFC 2396: *Uniform Resource Identifiers (URI): Generic Syntax*

Berners-Lee, T., et al. (2005). *Uniform Resource Identifier (URI): Generic Syntax. Request for Comments: 3986*. <http://tools.ietf.org/html/rfc3986>, ανακτήθηκε στις 2/3/2012

Berners-Lee Tim, (2006). *Linked Data*, [www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html) ανακτήθηκε 1/3/2012

Bizer C., Heath T., Idehen K., Berners-Lee T. (2008). *Linked Data on the Web (LDOW2008)*. In *Proceeding of the 17th international conference on World Wide Web, Beijing, China*. pp. 1265-1266

Brin, S., Page, L. (1998). *The anatomy of a large-scale hypertextual web search engine*. *WWW7/Computer Netw.* 1998;30:107-117

Ding, L., Kolari, P., Ganjugunte, S.; Finin, T.; and Joshi, A. (2004). Modeling and evaluating trust network inference. In *Seventh International Workshop on Trust in Agent Societies at AAMAS*

Esmaili, K.S., Abolhassani, H. (2006). *A categorization scheme for Semantic Web Search Engines*. In *IEEE International Conference on Computer Systems and Applications*.

Gerber, S., M. Comellas-Bigler, B.A. Goetz, and K.P. Locher. (2008). Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science* 321: 246-250

Klyne, G., Carroll, J. (2004). *Resource Description Framework (RDF): Concepts and Abstract Syntax - W3C Recommendation*.

Fielding, R., J. Gettys, J. Mogul, L. Masinter, P. Leach and T. Berners-Lee (1999). *Hypertext Transfer Protocol HTTP/1.1*. RFC2616, June

Antoniou G. & Van Harmelen F. (2009): *Εισαγωγή στο Σημασιολογικό Ιστό (μεταφρασμένο)*, Εκδόσεις Κλειδάριθμος.

C. Bizer, R. Cyganiak & T. Heath, (2007). *How to publish Linked Data on the Web*. Retrieved from Free University of Berlin: <http://www4.wiwi.fu-berlin.de/bizer/pub/LinkedDataTutorial/>

Sauermann, L., Cyganiak, R. (2008): *Cool URIs for the Semantic Web*. W3C Interest Group Note, <http://www.w3.org/TR/cooluris/> , ανακτήθηκε στις 1/3/2012

Gennari, J. H., Musen, M. A., Ferguson, R., Grosso, W. E., Crubzy, M., Eriksson, H., Noy, N. F., Tu, S. W., 2003. The evolution of Protégé: An environment for knowledge-based systems development. *International Journal of Human- Computer Studies* 58 (1), 89–123.



Kalyanpur, A., Parsia, B., Hendler, J. (2005). *A Tool for Working with Web Ontologies*, To Appear in the International Journal on Semantic Web and Information Systems

Kalyanpur, A. et al. (2006). *SMORE-semantic markup, ontology, and RDF editor*. (Defense Technical Information Center:

Lei, Y., Uren, V., & Motta, E. (2006). SemSearch: A search engine for the semantic web. In S. Staab & V. Svatek (Eds.), *Managing Knowledge in A World of Networks* (pp. 238-245). Berlin: Springer. doi: 10.1007/11891451

Tazzoli, R. et al (2004). Towards a Semantic Wiki Wiki Web. *Proceedings of Second International Semantic Web Conference*. Hiroshima, Japan, in CD-ROM

Tummarello, G., Delbru, R., and Oren, E. (2007). Sindice.com: Weaving the open linked data. In Aberer,

K., Choi, K.-S., Noy, N. F., Allemang, D., Lee, K.-I., Nixon, L. J. B., Golbeck, J., Mika, P., Maynard, D.,

Mizoguchi, R., Schreiber, G., and Cudr\_e-Mauroux, P., editors, ISWC/ASWC, volume 4825 of Lecture Notes in Computer Science, pages 552{565. Springe

OWL, <http://www.w3.org/2004/OWL>

DAML+OIL , <http://www.daml.org/2000/12/reference.html>.

Clarke, B.H. (2009), Early adolescents' use of social networking sites to maintain friendship and explore identity: implications for policy, *Policy & Internet*, 1(1), art. 3, DOI: 10.2202/1944-2866.1018

RDF Book Mashup, <http://www4.wiwiw.fu-berlin.de/bizer/bookmashup/>.

Talis Aspire,<http://www.talisaspire.com/>

Semantic MediaWiki (SMW),[http://ontoworld.org/wiki/Semantic\\_MediaWiki](http://ontoworld.org/wiki/Semantic_MediaWiki)

GNU GPL, <http://www.gnu.org/copyleft/gpl.html>

OntoWiki, <http://aksw.org/Projects/OntoWiki>

<http://d2rqupdate.cs.technion.ac.il/tutorial>