

ΕΥΧΑΡΙΣΤΙΕΣ

Πριν ξεκινήσω να παρουσιάζω και να αναλύω οτιδήποτε για τη πτυχιακή μου εργασία, νιώθω την ανάγκη και την ηθική υποχρέωση να ευχαριστήσω θερμά κάποιους ανθρώπους, που βοήθησαν σημαντικά στην εκπόνησή της.

Θα ήθελα να αποδώσω ευχαριστίες στους καθηγητές μας, οι οποίοι μας εξόπλισαν με τα απαραίτητα εφόδια, ώστε να ολοκληρώσουμε επιτυχώς τις σπουδές μας. Ιδιαίτερα τον καθηγητή μας Χαλκιάπουλο Κωνσταντίνο, ο οποίος με καθοδήγησε σε όλη την πορεία της εκπόνησης της πτυχιακής μου εργασίας, και με τις κατάλληλες συμβουλές, επεξηγήσεις και την απαιτούμενη καθοδήγηση με βοήθησε να συντάξω την παρακάτω μελέτη.

Τέλος, οφείλω ένα μεγάλο ευχαριστώ πραγματικά μέσα από τη καρδιά μου στην οικογένειά μου, η οποία με στήριξε όλα αυτά τα χρόνια των σπουδών μου, συναισθηματικά και οικονομικά, αλλά και σε όλους τους φίλους μου που με την αγάπη τους και τη στήριξή τους μου έδιναν δύναμη και κουράγιο όλο αυτό το χρονικό διάστημα.

Με εκτίμηση
Ζωή Ναζίρη

ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια υπάρχει εκθετική αύξηση του όγκου της πληροφορίας που δημοσιεύεται στο Διαδίκτυο. Καθώς όμως η πληροφορία αυτή δε συνδέεται με τη σημασιολογία της παρατηρείται δυσκολία στη διαχείρισή της και στην πρόσβαση σε αυτήν. Ο Σημασιολογικός Ιστός, λοιπόν, είναι μια ομάδα μεθόδων και τεχνολογιών που σκοπεύουν να δώσουν τη δυνατότητα στις μηχανές να κατανοήσουν τη “σημασιολογία” των πληροφοριών σχετικά με τον Παγκόσμιο Ιστό.

Ο Σημασιολογικός Ιστός (Semantic Web) αποτελεί επέκταση του Παγκοσμίου Ιστού. Στο Σημασιολογικό Ιστό οι πληροφορίες εμπλουτίζονται με μεταδεδομένα, τα οποία υπακούουν σε κοινά πρότυπα και επιτρέπουν την εξαγωγή γνώσεως από την ήδη υπάρχουσα, καθώς επίσης και το συνδυασμό της υπάρχουσας πληροφορίας με στόχο την εξαγωγή συμπερασμάτων. Απώτερος στόχος του Σημασιολογικού Ιστού είναι η βελτιωμένη αναζήτηση, η εκτέλεση σύνθετων διεργασιών και η εξατομίκευση της πληροφορίας σύμφωνα με τις ανάγκες του κάθε χρήστη.

Στην παρούσα διπλωματική εργασία μελετήθηκε η χρήση των τεχνολογιών του Σημασιολογικού Ιστού για τη βελτίωση της πρόσβασης σε δεδομένα. Συγκεκριμένα αρχικά έγινε εμβάθυνση στις τεχνολογίες και στις θεμελιώδεις έννοιες του Σημασιολογικού Ιστού. Παρουσιάστηκαν αναλυτικά οι βασικές γλώσσες σήμανσης: XML που επιτρέπει τη δημιουργία δομημένων εγγράφων με λεξιλόγιο καθορισμένο από το χρήστη, RDF που προσφέρει ένα μοντέλο δεδομένων για την περιγραφή πληροφοριών με τέτοιο τρόπο ώστε να είναι δυνατή η ανάγνωση και η κατανόησή τους από μηχανές.

Αναφέρθηκαν, ακόμη, οι διάφοροι τρόποι σύνταξης της γλώσσας RDF καθώς και πως γίνεται αναζήτηση σε γράφους RDF με το πρωτόκολλο SPARQL. Στη συνέχεια ακολουθεί η περιγραφή της RDFS, που πρόκειται για γλώσσα περιγραφής του RDF λεξιλογίου. Έχοντας παρουσιαστεί σε προηγούμενο κεφάλαιο η έννοια της οντολογίας, γίνεται αναφορά στη σημασιολογική γλώσσα σήμανσης OWL, που χρησιμοποιείται για την έκδοση και διανομή οντολογιών στο Διαδίκτυο. Έπειτα ακολουθεί μια ανασκόπηση από επιλεγμένα έργα που χρησιμοποιούν τις τεχνολογίες του Σημασιολογικού Ιστού.

Λέξεις-Κλειδιά: Οντολογία, Σημασιολογικός Ιστός, Γλώσσες Σημασιολογικού Ιστού: XML, RDF, RDFS, OWL, Γλώσσα Σημασιολογικών Επερωτημάτων: SPARQL, SPARQL Update, προγραμματιστική διεπαφή Jena.

ABSTRACT

Over the past few years there has been exponential increase in the volume of information published on the Internet. Since information is not connected to its semantics, it is difficult to manipulate and access it. Therefore, the Semantic Web consists of methods and technologies that aim to enable machines to understand information's semantics.

The Semantic Web is an extension of the World Wide Web (WWW). Specifically, information is enriched with metadata, which are subject to common standards and permit knowledge extraction from the existing one and the combination of existing information in order to infer implicit knowledge, as well. Future goals of the Semantic Web are enhanced searching, complicated processes' execution and information personalization according to each user's needs.

This post-graduate diploma thesis researches the usage of Semantic Web technologies for the enhancement of the access to data. More specifically, Semantic Web technologies and essential concepts are studied. Basic markup languages are presented analytically: XML that allows structured documents' creation with user defined vocabulary, RDF that offers a data model for such information description that it is readable and understandable by machines.

Also, various RDF syntaxes and how to search RDF graphs using SPARQL protocol are referred. Below, RDFS description follows, which is a description language of RDF vocabulary. Having introduced the concept of ontology in previous chapter, the semantic markup language OWL is presented, that is used for ontology publishing and distribution on the Internet. A review of selected projects of the last years, which are characterized by the application of technologies of the Semantic Web, is presented.

Keywords: Ontology, Semantic Web, Semantic markup languages: XML, RDF, RDFS, OWL, SPARQL Query, Language, SPARQL Update, Jena Semantic Web Framework.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΥΧΑΡΙΣΤΙΕΣ.....	3
ΠΕΡΙΛΗΨΗ.....	4
ABSTRACT.....	5
ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ	9
ΕΙΣΑΓΩΓΗ.....	10
ΚΕΦΑΛΑΙΟ 1°	11
1.1 Ιστορικά στοιχεία του διαδικτύου.....	11
1.2 Εισαγωγή στο Σημασιολογικό Ιστό	13
1.2.1 World Wide Web	13
1.2.2 Εισαγωγή του διαδικτύου	16
ΚΕΦΑΛΑΙΟ 2°	18
2.1 Περιγραφικές Γλώσσες (HTML...).....	18
2.2 Scripting (php, asp..).....	20
2.3 Μεταδεδομένα	24
2.4 Search engines (Google, yahoo, bing)	28
2.5 Σύνδεση μεταδεδομένων με μηχανές αναζήτησης (metadata-seo).....	33
2.6 Ορισμός-Σημασιολογικός Ιστός	34
2.6.1 Ανάγκη για επέκταση του σημερινού Ιστού	36
2.6.2 Οι Στόχοι του Σημασιολογικού Ιστού.....	38
2.6.3 Αρχιτεκτονική Σημασιολογικού Ιστού	41
2.7 Το όραμα του Σημασιολογικού Ιστού	43
2.7.1 Παραδείγματα χρήσης semantic	43
2.7.2 Χαρακτηριστικά και επεξήγηση λειτουργίας	46
2.8 Το μέλλον του semantic.....	48
2.8.1 Είναι υλοποιήσιμο-εφικτό ένα τέτοιο εγχείρημα.....	51
2.9 ΗΑΚΙΑ.....	54
2.9.1 Πράγματα που κάνουν να αναζητούμε μία σημασιολογική αναζήτηση.....	54
2.10 Συστήματα διαχείρισης Γνώσης βασισμένα σε οντολογίες.....	56
2.10.1 Μετά - Διαδικασία Γνώσης.....	56
2.10.2 Διαδικασία Γνώσης.....	57
ΚΕΦΑΛΑΙΟ 3°	59
3.1 Τι είναι μια οντολογία και γιατί τη χρειαζόμαστε	59
3.1.1 Η μορφή μιας οντολογίας και τα βασικά συστατικά της	61

3.2 Μεθοδολογία ανάπτυξης οντολογιών	63
3.3 Βασικές αρχές ανάπτυξης οντολογιών	68
3.4 Περιγραφή της Οντολογικής Γλώσσας OWL.....	70
3.4.1 Επισκόπηση της OWL	70
3.4.2 Επιρροές της OWL	72
3.4.3 Βασικά στοιχεία μοντελοποίησης της OWL	72
3.4.4 Οι υπο - γλώσσες της OWL.....	77
3.5 Περιγραφικές Λογικές	81
3.6 Πρότυπο Πλαισίων	84
ΚΕΦΑΛΑΙΟ 4 ^ο	85
4.1 XML (Extensible Markup Language).....	85
4.1.1 Η ανάγκη για την XML	85
4.1.2 Η Λύση της XML	87
4.1.3 Δημιουργία εγγράφων XML.....	88
4.1.4 Μέθοδος XML – XML schema	88
4.1.5 Αναπαράσταση Γνώσης με χρήση XML	88
4.2 Resource Description Framework (RDF)	91
4.2.1 Αναπαράσταση Δεδομένων σχετικά με Δεδομένα	95
4.2.2 RDF Schema	95
4.2.3 Αναπαράσταση Γνώσης με χρήση RDF	96
4.3 Γλώσσες Οντολογίας - Προκάτοχοι της OWL	97
4.3.1 XML-based Ontology Exchange Language (XOL).....	97
4.3.2 Simple HTML Ontology Extension (SHOE).....	101
4.3.3 Ontology Markup Language (OML).....	104
4.3.4 Ontology Interchange Language (OIL).....	105
4.3.5 DARPA Agent Markup Language+OIL (DAML+OIL).....	107
4.4 Προβλήματα στην Πορεία	109
4.4.1 Συντακτικά Προβλήματα	109
4.4.2 Σημασιολογικά Προβλήματα	110
4.4.3 Εκφραστική Ισχύς.....	112
4.4.4 Υπολογιστικά Προβλήματα	112
4.5 Λύσεις.....	113
4.5.1 Αναγνωσιμότητα.....	113
4.5.2 Χειρισμός Λανθασμένων Γραφημάτων.....	114
4.5.3 Παροχή μιας βιώσιμης Σημασιολογικής Θεωρίας για την OWL	114
4.5.4 Αποφυγή Παράδοξων	116

4.5.5 Διατήρηση Ικανότητας Λήψης Αποφάσεων.....	117
ΚΕΦΑΛΑΙΟ 5 ^ο	118
5.1 Συστήματα βασισμένα σε οντολογίες.....	118
5.1.1 OntoSeek.....	118
5.1.2 Telltale	118
5.1.3 Ontology-Based Semantic Online Classification of Documents	119
5.2 Οντολογίες αναφοράς - προφίλ.....	120
5.2.1 SUMO — the Suggested Upper Merged Ontology	120
5.2.2 WordNet — Μια Online λεξικογραφική βάση δεδομένων	120
5.3 Το Protégé - 2004	123
5.3.1 Γραφικό Περιβάλλον Εργασίας – Βασικές Λειτουργίες.....	123
ΚΕΦΑΛΑΙΟ 6 ^ο	125
6.1 Επίλογος - Συμπεράσματα	125
ΒΙΒΛΙΟΓΡΑΦΙΑ	127
ΙΣΤΟΣΕΛΙΔΕΣ - ΑΝΑΦΟΡΕΣ ΣΕ ΔΗΜΟΣΙΕΥΣΕΙΣ ΣΤΟΝ ΙΣΤΟΧΩΡΟ.....	136

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Αρχιτεκτονική Semantic Web.....	41
Εικόνα 2: Τα δύο πρώτα επίπεδα της οντολογίας που χρησιμοποιείται στον πράκτορα MELISA.....	44
Εικόνα 3: Το σχεδιάγραμμα της δομής του πράκτορα MELISA.....	44
Εικόνα 4: Η διεπαφή χρήστη του πράκτορα MELISA.....	46
Εικόνα 5: Semantic Web.....	48
Εικόνα 6: Επίπεδα ανάπτυξης μιας οντολογίας.....	61
Εικόνα 7: Σχήμα οντολογίας για κινηματογραφικές ταινίες.....	62
Εικόνα 8: Σχήμα 2.....	62
Εικόνα 9: Enterprise Ontology.....	64
Εικόνα 10: ON-TO-KNOWLEDGE.....	67
Εικόνα 11: Παράδειγμα αναπαράστασης οντολογιών.....	71
Εικόνα 12: Selected classes and properties of the Profile.....	74
Εικόνα 13: Παράδειγμα σχεδιασμού της OWL.....	83
Εικόνα 14: Μοντέλο DTD.....	89
Εικόνα 15: OML Framework.....	105
Εικόνα 16: Δομή της OWL.....	115
Εικόνα 17: Διάγραμμα Venn.....	117
Εικόνα 18: Παράδειγμα WordNet.....	121
Εικόνα 19: Overall system architecture.....	124
Εικόνα 20: An ontology as a filter.....	125

ΕΙΣΑΓΩΓΗ

Ο Παγκόσμιος Ιστός χρειάζεται να εξελιχθεί σε κάτι πολύ παραπάνω από μια απλή πλατφόρμα που επιτρέπει σε χρήστες και οργανισμούς να φιλοξενούν ιστοσελίδες. Είναι ανάγκη να αναπτυχθεί ένας μηχανισμός, ο οποίος θα επιτρέπει στις μηχανές να συλλέγουν περιεχόμενο από διαφορετικές πηγές, να επεξεργάζονται την πληροφορία και να ανταλλάσσουν τα αποτελέσματα με άλλα προγράμματα ηλεκτρονικών υπολογιστών (πράκτορες) και χρήστες.

Προς αυτήν την κατεύθυνση, κινείται το όραμα της ανάπτυξης του *Σημασιολογικού Ιστού* (Semantic Web), ενώ σημαντική μπορεί να είναι η συνεισφορά των οντολογιών στην κατεύθυνση αυτή. Οι οντολογίες διευκολύνουν το διαμοιρασμό και την επαναχρησιμοποίηση της γνώσης, ώστε να υπάρχει η δυνατότητα επικοινωνίας μεταξύ ετερογενών και διανεμημένων συστημάτων εφαρμογών. Αποτελούν, λοιπόν, ένα από τα βασικότερα εργαλεία *σημασιολογικής διαλειτουργικότητας* (semantic interoperability), γιατί προσφέρουν λεξιλόγια για την περιγραφή των εννοιών και των σχέσεών τους και λειτουργούν ως εργαλεία αυτόματης “διερμηνείας” και συσχέτισης των μεταδεδομένων που χρησιμοποιούνται από διάφορες εφαρμογές και υπηρεσίες πληροφόρησης.

ΚΕΦΑΛΑΙΟ 1^ο

1.1 Ιστορικά στοιχεία του διαδικτύου

Ο παγκόσμιος ιστός είναι μια μεγάλη επιτυχημένη επίτευξη αφού διακινεί την μεγαλύτερη ποσότητα διαθέσιμης πληροφορίας και είναι ανυπολόγιστος ο ρυθμός αύξησης των ανθρώπων που τον χρησιμοποιούν.

Έχει αρχίσει να διαπερνά τις περισσότερες πλευρές της καθημερινής μας ζωής αλλά και της ζωής των επιχειρήσεων και οργανισμών.

1960: Μια έρευνα ανατέθηκε από την κυβέρνηση των Ηνωμένων Πολιτειών για την κατασκευή ισχυρών, χωρίς σφάλματα, κατανοητών δικτύων υπολογιστών σε συνεργασία με ιδιωτικά εμπορικά συμφέροντα.

1980: Το Εθνικό Ίδρυμα Επιστημών των ΗΠΑ χρηματοδότησε ένα νέο δίκτυο κατά τη δεκαετία του 1980. Παράλληλα η ιδιωτική χρηματοδότηση για άλλα εμπορικά δίκτυα, γέννησε τη διασύνδεση πολλών δικτύων και ανέπτυξε νέες τεχνολογίες δικτύωσης.

1990: Η δεκαετία του 1990 καθρέπτιζε ένα διεθνές δίκτυο που οφειλόταν στην εμπορευματοποίηση. Το δίκτυο αυτό ενσωματώθηκε σε κάθε σημείο της σύγχρονης ανθρώπινης ζωής. Εκείνη την εποχή μια τυπική ιστοσελίδα ήταν αποθηκευμένη σε έναν διακομιστή σε γλώσσα HTML, έτοιμη να σταλεί στον περιηγητή ενός χρήστη ως απάντηση σε ένα αίτημα.

2000: Οι ιστοσελίδες πλέον, δημιουργούνται με χρήση λογισμικού διαχείρισης περιεχομένου ή λογισμικού wiki, οπότε η διαδικασία δημιουργίας και εξυπηρέτησης των σελίδων έχει γίνει πιο αυτοματοποιημένη και δυναμική, αρχικά με πολύ λίγο περιεχόμενο.

2010: Οι κατασκευαστές ιστοσελίδων στήνουν αρχικά μια βάση δεδομένων με το επιθυμητό περιεχόμενο, χρησιμοποιώντας για τη δουλειά αυτή ειδικές σελίδες επεξεργασίας. Οι επισκέπτες των ιστοσελίδων αυτών μπορούν να βλέπουν και να διαβάζουν το περιεχόμενο στην τελική αναγνώσιμη μορφή τους.

Ένα μεγάλο βήμα στην τεχνολογία της πληροφορίας αποτελεί η εμφάνιση του παγκόσμιου ιστού. Η ύπαρξή του αλλάζει εντελώς τον τρόπο με τον οποίο ο άνθρωπος εκτιμούσε τη χρησιμότητα των ηλεκτρονικών υπολογιστών στην καθημερινότητά του.

Πριν από το World Wide Web στο Internet βλέπαμε πραγματικά μόνο οθόνες με πλήρες κείμενο (και συνήθως μόνο σε μια γραμματοσειρά). Έτσι, αν και ήταν αρκετά καλό για την ανταλλαγή πληροφοριών, αλλά και για την πρόσβαση σε πληροφορίες, ήταν όμως οπτικά πολύ βαρετό.

Σε μια προσπάθεια να κάνουν αυτό το περιβάλλον πιο αισθητικό, εταιρείες όπως η CompuServe και η AOL άρχισαν να αναπτύσσουν αυτό που συνηθίζαμε να αποκαλούμε GUIs (ή γραφικές διεπαφές χρήστη). Το GUIs πρόσθεσε λίγο χρώμα και ένα κομμάτι στη διάταξη, αλλά ήταν ακόμα αρκετά βαρετό.

Πράγματι οι προσωπικοί υπολογιστές IBM, μόλις τώρα αρχίζουν να υιοθετούν τα Windows - αφού το MS-DOS ήταν αρκετά πρωτόγονο. Έτσι, το Διαδίκτυο μπορεί να είναι χρήσιμο, αλλά και όμορφο. Η εμφάνισή του άλλαξε τόσο που το κατέστησε δυνατό να εμφανίζει και να ανταλλάσσει φωτογραφίες και ήχο.

Η ιστοσελίδα είχε μερικούς σημαντικούς προκατόχους, ίσως ο πιο σημαντικός από αυτούς είναι το Xanadu έργο του Τεντ Νέλσον, ο οποίος ασχολήθηκε με την έννοια του Hypertext - όπου μπορείτε να κάνετε κλικ σε μια λέξη και θα σας μεταφέρει κάπου αλλού. Ο Ted Nelson προέβλεπε με το Xanadu σε μια τεράστια βιβλιοθήκη πληροφοριών από όλων των κόσμων. Για να κάνουμε κλικ σε συνδέσμους, όπως ονομάζονταν, ο Ντάγκλας Ένγκελμπαρτ εφηύρε το ποντίκι, το οποίο έμελλε να γίνει αργότερα ένα πολύ σημαντικό μέρος των προσωπικών υπολογιστών. Έτσι, η ιδέα να κάνουμε κλικ σε μια λέξη ή μια εικόνα

για να μας μεταφέρει κάπου αλλού ήταν ένα βασικό θεμέλιο του web.

Ένα άλλο σημαντικό δομικό στοιχείο ήταν η διεύθυνση URL ή Uniform Resource Locator. Αυτό θα επιτρέψει μια ακόμη επιλογή για να βρούμε το δρόμο μας γύρω από την ονομασία ενός ιστότοπου. Κάθε τοποθεσία στον παγκόσμιο ιστό έχει μια μοναδική διεύθυνση URL (όπως www.nethistory.info).

Το άλλο χαρακτηριστικό ήταν η Hypertext Markup Language (HTML), η γλώσσα που επέτρεψε στις σελίδες να εμφανίσουν διαφορετικές γραμματοσειρές και μεγέθη, εικόνες, χρώματα κ.λπ. Πριν από την HTML, δεν υπήρχε τέτοιο πρότυπο, και οι «GUIs» που μιλήσαμε πριν, ανήκαν μόνο σε διαφορετικούς υπολογιστές ή διαφορετικό λογισμικό.

Ήταν ο Tim Berners Lee, που δημιούργησε το World Wide Web. Οι πρώτες δοκιμές του World Wide Web ήταν στα εργαστήρια του CERN (ένα από τα μεγαλύτερα ερευνητικά εργαστήρια της Ευρώπης), στην Ελβετία, τον Δεκέμβριο του 1990. Μέχρι το 1991, πρόγραμμα περιήγησης και web server λογισμικό ήταν διαθέσιμα, και το 1992 υπήρχαν μερικές προκαταρκτικές ιστοσελίδες σε μέρη όπως το Πανεπιστήμιο του Ιλινόις. Μέχρι το τέλος του 1992, υπήρχαν περίπου 26 τοποθεσίες.

Το πρώτο πρόγραμμα περιήγησης που έγινε ευρέως διαθέσιμο ήταν το Mosaic, το 1993. Το Mosaic ήταν τόσο αργό που πραγματικά δεν χειριζόταν λήψη φωτογραφιών και επίσης τα εγχώρια μόντεμ λειτουργούσαν στο ένα έκτο της τρέχουσας ταχύτητας μόντεμ στην καλύτερη περίπτωση. Το συγκεκριμένο μέσο δεν είχε πολλές πιθανότητες εξέλιξης αφού ήταν ήδη πολύ χάλια.

Την 30η Απριλίου του 1993 οι διευθυντές του CERN έκαναν μια δήλωση που ήταν ένα πραγματικό ορόσημο στην ιστορία του Διαδικτύου. Την ημέρα αυτή, δήλωσαν ότι η τεχνολογία WWW θα μπορεί ελεύθερα να χρησιμοποιηθεί από οποιονδήποτε, χωρίς αμοιβή η οποία θα είναι πληρωμένη από το CERN. Η απόφαση αυτή - σε μεγάλο βαθμό σύμφωνα με τις αποφάσεις των προηγούμενων πρωτοπόρων του Internet, για να κάνουν τα προϊόντα τους να διατίθενται ελεύθερα - ήταν σημαντική.

Το πρόγραμμα περιήγησης πραγματικά άρχισε να αλλάζει τα πάντα. Μέχρι το τέλος του 1994 υπήρχαν ένα εκατομμύριο αντίτυπα του προγράμματος περιήγησης κατά τη χρήση των οποίων υπήρχε πράγματι ταχεία ανάπτυξη !!

Την ίδια χρονιά ο Marc Andreessen ίδρυσε το Netscape Corporation, και το World Wide Web Consortium, τα οποία διαχειρίζονται την ανάπτυξη του World Wide Web, που ιδρύθηκε από τον Tim Berners Lee.

Τότε θα αρχίσει πραγματικά η ανάπτυξη. Κάθε χρόνο 1994-2000, το Διαδίκτυο βλέπει μαζική ανάπτυξη, που όμοιά της δεν είχαμε δει με οποιαδήποτε προηγούμενη τεχνολογία. Η εποχή του Διαδικτύου είχε αρχίσει.

Οι πρώτες μηχανές αναζήτησης άρχισαν να εμφανίζονται στα μέσα της δεκαετίας του 1990, και δεν χρειάστηκε πολύς χρόνος για το Google ώστε να έρθει στο προσκήνιο, και να δημιουργήσει μια δεσπόζουσα θέση στην αγορά.

Αρχικά, το διαδίκτυο χρησιμοποιείται κυρίως για την εμφάνιση πληροφοριών. Οι On line αγορές, και η αγορά εμπορευμάτων ήρθε λίγο αργότερα. Η πρώτη μεγάλη εμπορική περιοχή ήταν του Amazon, μια εταιρεία η οποία αρχικά επικεντρώνεται αποκλειστικά στις αγορές του βιβλίου. Η έννοια του Amazon αναπτύχθηκε το 1994. Το Amazon είδε ότι οι on line αγορές ήταν ο εμπορικός τρόπος του μέλλοντος, και επέλεξε την αγορά του βιβλίου ως ένα πεδίο όπου θα μπορούσε να επιτευχθεί άμεσα.

1.2 Εισαγωγή στο Σημασιολογικό Ιστό

Ο Παγκόσμιος Ιστός έχει αλλάξει κατά πολύ τον τρόπο με τον οποίο διαρθρώνεται πλέον η επικοινωνία μεταξύ των ανθρώπων και ειδικά τον τρόπο με τον οποίο η πληροφορία που υπάρχει διαθέσιμη ανά τον κόσμο, διαδίδεται και ανακτάται. Από την άλλη πλευρά, η αποθήκευση όγκου πληροφοριών σε βάσεις δεδομένων οδήγησε στην εμφάνιση του εξής προβλήματος: την καθιέρωση και διατήρηση της σημασιολογίας των δεδομένων που είναι αποθηκευμένα στις βάσεις δεδομένων. Αυτό το πρόβλημα σημασιολογίας δεδομένων παρέμενε ελέγξιμο εφόσον ελέγξιμες ήταν και οι αλλαγές που θα μπορούσαν να προκληθούν στη κάθε βάση, αφού οι πληροφορίες ήταν διαθέσιμες και μπορούσαν να αλλοιωθούν μόνο από συγκεκριμένο αριθμό ατόμων και συγκεκριμένο αριθμό εφαρμογών.

Με την εμφάνιση λοιπόν του Παγκόσμιου Ιστού το τοπίο αυτό άλλαξε. Αμέτρητοι πλέον χρήστες και εφαρμογές μπορούν και έχουν πρόσβαση στις βάσεις δεδομένων που είναι διαθέσιμες στον Ιστό. Υπό αυτές τις συνθήκες, η σημασιολογία κάθε πληροφορίας πρέπει να είναι διαθέσιμη στον κάθε χρήστη μαζί με την ίδια την πληροφορία. Όταν ως χρήστης εννοείται κάποιο φυσικό πρόσωπο, αυτό μπορεί να επιτευχθεί με την κατάλληλη επιλογή κάποιας σχηματικής παρουσίασης για semantic data. Όταν όμως πρόκειται για κάποια εφαρμογή που θα αποκτήσει πρόσβαση στη βάση, η σημασιολογία πρέπει να είναι δομημένη σε μορφή που θα είναι προσπελάσιμη και κατανοητή από την μηχανή που θα αναλάβει την επεξεργασία της. Ως εκ τούτου, κρίθηκε απαραίτητη η επέκταση του σημερινού Ιστού, η οποία κατέληξε στην δημιουργία του Σημασιολογικού Ιστού, γνωστό και ως Semantic Web.

Ο Σημασιολογικός Ιστός λοιπόν αποτελεί μια επέκταση του σημερινού Ιστού η οποία έχει ως σκοπό την αυτοματοποίηση των λειτουργιών και των εφαρμογών του διαδικτύου. Η αυτοματοποίηση αυτή μπορεί να επιτευχθεί εφόσον η γνώση και η πληροφορία που υπάρχει αποθηκευμένη και δημοσιευμένη αυτή τη στιγμή στο σημερινό Παγκόσμιο Ιστό αποκτήσει τυπικό νόημα και σημασιολογία και δομηθεί με ένα τέτοιο τρόπο ώστε να γίνεται κατανοητή από τις μηχανές που την επεξεργάζονται. Υπάρχει πλέον παγκοσμίως η διάθεση συνεργασίας μεταξύ των, ανά τον κόσμο, χρηστών του ιστού με σκοπό να διασπείρουν πληροφορίες εμπλουτισμένες από τη γνώση όχι μόνο ενός ατόμου, αλλά πολλών. Αυτό οδήγησε στην δημιουργία των wikis, ιστοσελίδων οι οποίες μπορούν να εμπλουτιστούν και να επανεκδοθούν από τον καθένα. Συνδυάζοντας τις τεχνολογίες του semantic web και των wikis, έχουμε οδηγηθεί στην δημιουργία των Semantic wikis.

Σήμερα επιπλέον υπάρχει μια τάση από τον μεμονωμένο χρήστη να μετατοπίσει τη γνώση από συστήματα που βασίζονται στην επιφάνεια εργασίας σε συστήματα βασισμένα στον Σημασιολογικό Ιστό. Καθένας λοιπόν μπορεί να αποθηκεύσει κάποια πληροφορία ως έναν πόρο (resource) του Σημασιολογικού Ιστού που χαρακτηρίζεται από μια μοναδική σειρά χαρακτήρων, το URI του. Η συσκευή στην οποία κάποιος αποθηκεύει τις πληροφορίες αυτές ονομάζεται Σημασιολογική Επιφάνεια Εργασίας (Semantic Desktop).

1.2.1 World Wide Web

Πολλοί χρησιμοποιούν τους όρους Διαδίκτυο και Παγκόσμιος Ιστός εναλλακτικά, αλλά οι δύο όροι δεν είναι ταυτόσημοι. Ο Παγκόσμιος Ιστός (World Wide Web) είναι ένα παγκόσμιο σύνολο εγγράφων, εικόνων και άλλων πόρων, συνδεδεμένων λογικά μεταξύ τους μέσω υπερσυνδέσμων και αναφέρονται με διευθύνσεις URIs (Uniform Resource Identifiers). Τα URIs προσδιορίζουν συμβολικά υπηρεσίες, διακομιστές, βάσεις δεδομένων, καθώς και τα έγγραφα και τους πόρους που μπορούν να προσφέρουν. Το HTTP (Hypertext Transfer

Protocol) είναι το κύριο πρωτόκολλο πρόσβασης στον Παγκόσμιο Ιστό, αλλά είναι μόνο ένα από τα εκατοντάδες πρωτόκολλα επικοινωνίας που χρησιμοποιούνται στο Διαδίκτυο. Οι υπηρεσίες Web επίσης χρησιμοποιούν το πρωτόκολλο HTTP για να επιτρέψουν στα συστήματα λογισμικού την επικοινωνία, προκειμένου να μοιραστούν και να ανταλλάξουν επιχειρηματική λογική και δεδομένα.

Η ιστορία του Παγκοσμίου Ιστού ξεκινά το Νοέμβριο του 1990, όταν ο Tim Berners-Lee με τη βοήθεια του Robert Cailliau δημοσίευσε μια πρόταση σχετικά με την κατασκευή ενός «έργου Υπερκειμένου» με την ονομασία «WorldWideWeb» ("W3") ως "ιστού εγγράφων υπερκειμένου" τα οποία οι χρήστες θα μπορούσαν να επισκέπτονται από "περιηγητές" χρησιμοποιώντας αρχιτεκτονική client-server (πελάτη-διακομιστή). Η πρόταση αυτή υπολόγιζε ότι μέσα σε τρεις μήνες θα αναπτυσσόταν ένας ιστός μόνο για ανάγνωση και ότι θα χρειαζόνταν έξι μήνες για την επίτευξη της δημιουργίας νέων συνδέσμων και υλικού από τους αναγνώστες, με σκοπό την συγγραφή από όλον τον κόσμο, καθώς επίσης και την αυτόματη ειδοποίηση των αναγνωστών όταν δημοσιεύεται νέο υλικό σχετικό με τα ενδιαφέροντά τους. Ενώ ο στόχος του ιστού μόνο για ανάγνωση επιτεύχθηκε, η πρόσβαση στην ελεύθερη συγγραφή περιεχομένου στον παγκόσμιο ιστό από τους απλούς χρήστες χρειάστηκε περισσότερο χρόνο για να ωριμάσει, με την εμφάνιση των wikis, blogs, του Παγκόσμιου Ιστού 2.0 και RSS / Atom.

Ένας υπολογιστής NeXT χρησιμοποιήθηκε από τον Berners-Lee, ως ο πρώτος διακομιστής ιστού στον κόσμο, αλλά και για την συγγραφή του πρώτου προγράμματος περιήγησης στον Ιστό, το WorldWideWeb, το 1990. Μέχρι τα Χριστούγεννα του 1990, ο Berners-Lee είχε κατασκευάσει όλα τα απαραίτητα εργαλεία για έναν λειτουργικό ιστό: το πρώτο πρόγραμμα περιήγησης, τον πρώτο διακομιστή και τις πρώτες ιστοσελίδες, οι οποίες περιέγραφαν το ίδιο το έργο. Στις 6 Αυγούστου του 1991, δημοσιεύτηκε μια σύντομη περίληψη¹ του έργου του Παγκόσμιου Ιστού. Η ημερομηνία αυτή σηματοδότησε το ντεμπούτο του Παγκόσμιου Ιστού ως μια υπηρεσία διαθέσιμη στο κοινό μέσω του Διαδικτύου. Η πρώτη φωτογραφία στο διαδίκτυο μεταφορτώθηκε από τον Berners-Lee το 1992, μια εικόνα του οικοδομικού συγκροτήματος του CERN.

Ο Παγκόσμιος Ιστός είχε ορισμένες διαφορές από τα άλλα συστήματα υπερκειμένου που ήταν τότε διαθέσιμα. Ο Ιστός απαιτούσε μόνο συνδέσμους μονής κατεύθυνσης και όχι αμφίδρομης. Αυτό καθιστούσε δυνατή την σύνδεση με άλλο πόρο χωρίς δράση από τον ιδιοκτήτη του. Επίσης, μείωσε σημαντικά τη δυσκολία της υλοποίησης διακομιστών Ιστού και προγραμμάτων περιήγησης (σε σύγκριση με τα προγενέστερα συστήματα). Στις 30 Απριλίου 1993, το CERN ανακοίνωσε ότι ο Παγκόσμιος Ιστός θα είναι ελεύθερος για όλους, χωρίς οφειλόμενα τέλη. Η ανακοίνωση αυτή δύο μήνες μετά την ανακοίνωση ότι η χρήση του διακομιστή του πρωτοκόλλου Gopher δεν ήταν πλέον δωρεάν, παρήγαγε μια ταχεία στροφή από το Gopher προς τον Ιστό.

Οι μελετητές συμφωνούν γενικά ότι η κρίσιμη καμπή για τον Παγκόσμιο Ιστό ξεκίνησε με την εισαγωγή του περιηγητή Mosaic το 1993, ένα πρόγραμμα περιήγησης με γραφικά που αναπτύχθηκε από μια ομάδα στο Εθνικό Κέντρο για Εφαρμογές Υπερυπολογιστών στο Πανεπιστήμιο του Ιλινόις (NCSA-UIUC), με επικεφαλής τον Marc Andreessen. Πριν από την έκδοση του Mosaic, τα γραφικά δεν αναμιγνύονταν συνήθως με το κείμενο στις ιστοσελίδες και η δημοτικότητα του Ιστού ήταν μικρότερη με τη χρήση παλαιότερων πρωτοκόλλων στο Διαδίκτυο, όπως το Gopher και το Wide Servers Information Servers (WAIS). Η γραφική διεπαφή χρήστη του Mosaic κατέστησε τον Παγκόσμιο Ιστό το πιο δημοφιλές, μέχρι στιγμής, πρωτόκολλο Internet.

Το λογισμικό περιήγησης στον Παγκόσμιο Ιστό, όπως ο Internet Explorer, Mozilla Firefox, Opera, Apple's Safari και Google Chrome, επιτρέπει στους χρήστες να περιηγηθούν

¹ "Short summary of the World Wide Web project". Groups.google.com. August 6, 1991. Retrieved July 27, 2009.

από τη μία ιστοσελίδα στην άλλη, μέσω υπερσυνδέσμων, που εμπεριέχονται στα έγγραφα. Τα έγγραφα αυτά μπορούν επίσης να περιέχουν οποιονδήποτε συνδυασμό ηλεκτρονικών δεδομένων, συμπεριλαμβανομένων των γραφικών, ήχων, κειμένου, βίντεο, πολυμέσων και διαδραστικού περιεχομένου που εκτελούνται ενώ ο χρήστης βρίσκεται σε αλληλεπίδραση με την σελίδα. Client-side λογισμικό μπορεί να περιλαμβάνει κινούμενα σχέδια, παιχνίδια, εφαρμογές γραφείου και επιστημονικές επιδείξεις. Μέσα από διαδικτυακή έρευνα με χρήση λέξεων-κλειδιών σε μηχανές αναζήτησης όπως το Yahoo! και Google, οι χρήστες παγκοσμίως έχουν εύκολη και άμεση πρόσβαση σε έναν τεράστιο αριθμό σχετικών πληροφοριών. Σε σύγκριση με τα έντυπα μέσα, βιβλία, εγκυκλοπαίδειες και τις παραδοσιακές βιβλιοθήκες, ο Παγκόσμιος Ιστός έχει επιτρέψει την αποκέντρωση των πληροφοριών σε μεγάλη κλίμακα.

Ο Παγκόσμιος Ιστός δίνει επίσης τη δυνατότητα σε άτομα και οργανισμούς να δημοσιεύουν ιδέες και πληροφορίες σε δυναμικά μεγάλο κοινό διαδικτυακά με σημαντικά μειωμένα έξοδα και χρονική καθυστέρηση. Η δημοσίευση μιας ιστοσελίδας, ενός blog, ή η κατασκευή ενός δικτυακού τόπου έχουν μικρό αρχικό κόστος και πολλές δωρεάν υπηρεσίες είναι διαθέσιμες. Ωστόσο, η δημοσίευση και διατήρηση μεγάλων, επαγγελματικών ιστοσελίδων με ελκυστικές, ποικίλες και ενημερωμένες πληροφορίες εξακολουθούν να είναι μια δύσκολη και ακριβή πρόταση. Πολλά άτομα και ορισμένες εταιρείες και ομάδες χρησιμοποιούν τα αρχεία καταγραφής ιστού ή τα blogs, τα οποία χρησιμοποιούνται ευρέως ως εύκολα τροποποιήσιμα ημερολόγια. Ορισμένοι εμπορικοί οργανισμοί ενθαρρύνουν τα μέλη του προσωπικού τους να επικοινωνούν διαδικτυακά σχετικά με τους τομείς εξειδίκευσής τους με την ελπίδα ότι οι επισκέπτες θα εντυπωσιαστούν από την εξειδικευμένη γνώση και τις δωρεάν πληροφορίες, και ως αποτέλεσμα θα τους προσελκύσει η εταιρεία. Συλλογές προσωπικών ιστοσελίδων δημοσιευμένων από μεγάλους παρόχους υπηρεσιών παραμένουν δημοφιλείς, και ολοένα και εξελίσσονται.

Όταν ξεκίνησε ο Παγκόσμιος Ιστός στη δεκαετία του 1990, μια τυπική ιστοσελίδα ήταν αποθηκευμένη σε έναν διακομιστή, μορφοποιημένη σε γλώσσα HTML, έτοιμη να σταλεί στον περιηγητή ενός χρήστη ως απάντηση σε ένα αίτημα. Με την πάροδο του χρόνου, η διαδικασία της δημιουργίας και εξυπηρέτησης των σελίδων έχει γίνει πιο αυτοματοποιημένη και δυναμική. Οι ιστοσελίδες συχνά δημιουργούνται με χρήση λογισμικού διαχείρισης περιεχομένου ή λογισμικού wiki, έχοντας, αρχικά, πολύ λίγο περιεχόμενο. Οι συμμετέχοντες στη δημιουργία των δικτυακών τόπων συμπληρώνουν μια υποκείμενη βάση δεδομένων με περιεχόμενο, χρησιμοποιώντας τις σελίδες επεξεργασίας που είναι σχεδιασμένες για τον σκοπό αυτό, ενώ οι περιστασιακοί επισκέπτες βλέπουν και διαβάζουν αυτό το περιεχόμενο στην τελική αναγνώσιμη μορφή.

Η εμφάνιση του Παγκόσμιου Ιστού αποτέλεσε ένα τεράστιο βήμα στην τεχνολογία της πληροφορίας και άλλαξε εντελώς τον τρόπο με τον οποίο ο άνθρωπος εκτιμούσε τη χρησιμότητα των ηλεκτρονικών υπολογιστών στην καθημερινότητά του. Ενώ παλαιότερα οι ηλεκτρονικοί υπολογιστές αποτελούσαν απλά υπολογιστικές μηχανές, με τη δημιουργία του Ιστού αποτελούν ένα μέσο πρόσβασης στη διάχυτη παγκόσμια γνώση. Ο καθένας λοιπόν έχει πρόσβαση σε βάσεις δεδομένων που βρίσκονται στον Ιστό και είναι δυνατόν να αντλήσει από αυτές οποιαδήποτε πληροφορία. Οι πληροφορίες αυτές παρουσιάζονται με τέτοια μορφή ώστε να γίνονται κατανοητές από τον χρήστη. Ακόμη και πληροφορίες που έχουν αντληθεί κατευθείαν από μια βάση δεδομένων, δεν παρουσιάζονται αυτούσιες, όπως είναι δομημένες στη βάση, αλλά με μια απλουστευμένη μορφή, αντιληπτή από τον εκάστοτε αναγνώστη. Κατανοούμε, λοιπόν, πως η γνώση που είναι αποθηκευμένη στον παγκόσμιο ιστό, είναι γνώση μεταφρασμένη σε μια γλώσσα αντιληπτή από τον άνθρωπο, αφού αυτός είναι και ο χειριστής αυτού του εργαλείου.

Κάθε πληροφορία αντιστοιχεί, επομένως, σε μια λέξη νοηματικά κατανοητή από τον άνθρωπο. Για το λόγο αυτό, οι μηχανές αναζήτησης πληροφοριών του Ιστού είναι έτσι

κατασκευασμένες ώστε να αποδίδουν τα αποτελέσματά τους με βάση τις λέξεις που θα τους δοθούν από τον χρήστη και την αντιστοιχία των λέξεων αυτών με εκείνες που είναι αποθηκευμένες στον ιστό. Όσο ο όγκος των αποθηκευμένων αυτών πληροφοριών παρέμενε σε χαμηλά επίπεδα, ο τρόπος άντλησής τους (βασισμένη σε λέξεις-κλειδιά) ήταν αποτελεσματικός. Με τη ραγδαία όμως αύξησή του, οι αδυναμίες δεν άργησαν να φανούν.

Οι μηχανές αναζήτησης μπορεί να βοηθούν στην εύρεση περιεχομένου που περιέχει συγκεκριμένες λέξεις, αλλά το περιεχόμενο μπορεί να μην είναι ακριβώς αυτό που θέλουμε. Η αναζήτηση βασίζεται στα περιεχόμενα των σελίδων και όχι στο εννοιολογικό νόημα των περιεχομένων της σελίδας ή των πληροφοριών για τη σελίδα. Έτσι, παρόλο που η αναζήτηση μιας λέξης θα μας αποφέρει χιλιάδες άμεσα συσχετιζόμενες πληροφορίες, η ύπαρξη επίσης χιλιάδων πληροφοριών που δεν σχετίζονται άμεσα με το πεδίο της αναζήτησής μας, αλλά τυγχάνει να περιέχουν τη λέξη κλειδί που έχουμε χρησιμοποιήσει, δημιουργούν ένα φόρτο γνώσης, που ο μέσος χρήστης είναι συχνά ανήμπορος να διαχειριστεί αποτελεσματικά. Άλλωστε, πολλές φορές, η ύπαρξη πολλών και άχρηστων πληροφοριών είναι πιο αναποτελεσματική από την ελλιπή πληροφόρηση. Και τελικά, ακόμη και η άμεσα συσχετιζόμενη με την αναζήτηση του χρήστη γνώση είναι διασκορπισμένη σε σελίδες του Ιστού που δεν έχουν καμία αλληλεξάρτηση, κάτι το οποίο δυσχεραίνει την πλοήγηση του μέσου χρήστη.

1.2.2 Εισαγωγή του διαδικτύου

Η αρχική ιδέα του Ιστού όπως το είχε οραματιστεί ο Tim Berners-Lee περιελάμβανε τόσο έγγραφα όσο και δεδομένα (καθώς και πιο αφηρημένες έννοιες). Υποθέτουμε ότι ένας από τους κύριους λόγους για τους οποίους το κομμάτι των δεδομένων δεν τα κατάφερε σε πρώτη φάση να εξαπλωθεί, ίσως να ήταν ότι οι άνθρωποι και τα ιδρύματα επικεντρώθηκαν στο κομμάτι της παρουσίας (όπως η διάταξη, η διάδραση και η δομή) παρά στα πρωτογενή δεδομένα. Παρά τα αδιαμφισβήτητα οφέλη που παρέχει ο Ιστός, μέχρι πρόσφατα οι ίδιοι κανόνες που έδωσαν στον Ιστό των Εγγράφων τη δυνατότητα να ανθίσει δεν εφαρμόζονταν στα δεδομένα με αποτέλεσμα ο Ιστός να είναι κάπως ατελής. Τα πρώτα χρόνια (μιλώντας κατά προσέγγιση από τις αρχές του '90 έως το 2000), οι περισσότεροι άνθρωποι ήταν απασχολημένοι με την ανάπτυξη λογισμικού όπως πράκτορες και εξυπηρετητές, αλλά και περιεχομένου για τον Ιστό των Εγγράφων.

Σύμφωνα με τον Tim Berners-Lee, ο Ιστός σήμερα είναι αρκετά αποτελεσματικός στο να μας βοηθά να δημοσιεύουμε και να ανακαλύπτουμε έγγραφα, αλλά τα μεμονωμένα στοιχεία πληροφοριών μέσα σε αυτά τα έγγραφα (είτε είναι η ημερομηνία ενός γεγονότος, η τιμή ενός αντικειμένου σε ένα κατάλογο ή ένας μαθηματικός τύπος) δεν μπορούν να υποστούν επεξεργασία απ' ευθείας ως δεδομένα. Σήμερα μπορούμε να δούμε τα δεδομένα με ένα φυλλομετρητή (browser), αλλά δεν μπορούμε να χρησιμοποιήσουμε άλλα προγράμματα για το χειρισμό ή την ανάλυσή τους χωρίς πρώτα να τα έχουμε επεξεργαστεί εμείς οι ίδιοι χειρωνακτικά. Αν λυθεί το πρόβλημα αυτό, ο Ιστός θα αναμένεται να μοιάζει περισσότερο με μεγάλη βάση δεδομένων ή φύλλο εργασίας (spreadsheet), παρά ως ένα απλό σύνολο συνδεδεμένων εγγράφων. Οι ψηφιακές πληροφορίες που αφορούν σχεδόν κάθε πλευρά της ζωής μας δημιουργούνται με ραγδαίους ρυθμούς. Μέσα σε όλα αυτά τα δεδομένα βρίσκεται το κλειδί της γνώσης για τη θεραπεία ασθενειών, τη δημιουργία επιχειρηματικής αξίας και της πιο αποτελεσματικής διακυβέρνησης του κόσμου μας.

Αν και ο Παγκόσμιος Ιστός είναι η μεγαλύτερη αποθήκη πληροφοριών που δημιουργήθηκε ποτέ, με τα περιεχόμενά του να επεκτείνονται σε διάφορες γλώσσες και πεδία γνώσεων, μακροπρόθεσμα, είναι εξαιρετικά δύσκολο να βγει νόημα από το περιεχόμενό του. Οι μηχανές αναζήτησης μπορεί να βοηθούν στην εύρεση περιεχομένου που περιέχει

συγκεκριμένες λέξεις, αλλά το περιεχόμενο μπορεί να μην είναι ακριβώς αυτό που θέλουμε. Η αναζήτηση βασίζεται στα περιεχόμενα των σελίδων και όχι στο εννοιολογικό νόημα των περιεχομένων της σελίδας ή των πληροφοριών για τη σελίδα. Από το 2000 έως το 2006 (χονδρικά) ένας αριθμός τεχνικών καινοτομιών, το RDF το οποίο είναι για τα δεδομένα ό,τι η HTML για τα έγγραφα και η Web Ontology Language (OWL) η οποία μας επιτρέπει να εκφράσουμε το πώς συνδέονται μεταξύ τους οι πηγές δεδομένων, σε συνδυασμό με τις πιο ανοιχτές πρακτικές διαμοιρασμού πληροφοριών, μετακινούν τον Παγκόσμιο Ιστό προς αυτό που ονομάζουμε Σημασιολογικό Ιστό.

Η ύπαρξη του Σημασιολογικού Ιστού δίνει τη δυνατότητα να τοποθετηθούν ετικέτες σε όλο το περιεχόμενο του Ιστού, να περιγράψουμε τι είναι κάθε κομμάτι πληροφορίας και να δώσουμε σημασιολογικό νόημα στο αντικείμενο του περιεχομένου. Έτσι, οι μηχανές αναζήτησης θα γίνουν πιο αποτελεσματικές απ' ό,τι είναι τώρα και οι χρήστες θα μπορούν να βρουν τις ακριβείς πληροφορίες που ψάχνουν. Οι οργανισμοί που παρέχουν διάφορες υπηρεσίες μπορούν να βάλουν ετικέτες με νόημα σε αυτές τις υπηρεσίες. Χρησιμοποιώντας πράκτορες λογισμικού βασισμένους στον Ιστό, μπορούμε να βρούμε δυναμικά αυτές τις υπηρεσίες και να τις χρησιμοποιήσουμε προς όφελός μας ή σε συνεργασία με άλλες υπηρεσίες. Η πρόοδος προς την καλύτερη ολοκλήρωση των δεδομένων θα συντελεστεί μέσω της χρήσης της τεχνολογίας-κλειδί που έκανε τον Παγκόσμιο Ιστό τόσο επιτυχημένο.

Η δύναμη του Ιστού σήμερα, συμπεριλαμβανομένης της δυνατότητας να βρούμε τη σελίδα για την οποία ψάχνουμε, έγκειται στο γεγονός ότι τα έγγραφα τοποθετούνται στον Ιστό σε πρότυπη μορφή και έπειτα συνδέονται μεταξύ τους. Ο Σημασιολογικός Ιστός θα δώσει τη δυνατότητα καλύτερης ολοκλήρωσης των δεδομένων, επιτρέποντας στον καθένα ο οποίος τοποθετεί μεμονωμένα στοιχεία δεδομένων στον Ιστό, να τα συνδέσει με άλλα κομμάτια δεδομένων, χρησιμοποιώντας πρότυπες μορφοποιήσεις.

Το αποτέλεσμα της χρήσης των τεχνολογιών του Σημασιολογικού Ιστού για να:

1. δημοσιεύσουμε δομημένα δεδομένα στον Ιστό,
2. θέσουμε συνδέσμους μεταξύ δεδομένων από μια πηγή δεδομένων προς δεδομένα μέσα σε άλλες πηγές δεδομένων είναι τα Συνδεδεμένα Δεδομένα (Linked Data).

Ως αποτέλεσμα θα προκύψει ο Ιστός των Δεδομένων (Web of Data). Με τη διασύνδεση συνόλων δεδομένων, οι μηχανές θα μπορούν να διασχίζουν έναν ανεξάρτητο ιστό δομημένων πληροφοριών έτσι ώστε να συλλέξουν σημασιολογική γνώση από αυθαίρετες οντολογίες και τομείς. Το αποτέλεσμα θα είναι μια μαζική βάση γνώσης με ελεύθερη πρόσβαση η οποία θα αποτελέσει τα θεμέλια μιας νέας γενιάς εφαρμογών και υπηρεσιών.

ΚΕΦΑΛΑΙΟ 2^ο

2.1 Περιγραφικές Γλώσσες (HTML...)

Η **HTML(HyperText Markup Language)** είναι μια περιγραφική γλώσσα μορφοποίησης δεδομένων και κλήσης άλλων αρχείων. Τα βασικά δομικά στοιχεία της html είναι οι ετικέτες(tags) που αναγνωρίζονται και μεταφράζονται σε εντολές από τους browsers και έτσι παρουσιάζεται η ιστοσελίδα.

Οι ετικέτες περικλείονται από τα σύμβολα <> και ομαδοποιούνται σε ζεύγη για άνοιγμα και κλείσιμο π.χ. <h3>...</h3>

Κάθε σελίδα HTML περικλείεται από τα tags <html>...</html> και χωρίζεται σε δύο τμήματα το <head>...</head> και στο <body>...</body>

Στο **τμήμα <head>** περιέχονται:

μεταπληροφορίες (metatags) για την ιστοσελίδα οι οποίες μπορούν να αξιοποιηθούν από τις μηχανές αναζήτησης βελτιώνοντας σημαντικά την κατάταξη της. Ο τίτλος της ιστοσελίδας ανάμεσα στα tags <title> ...</title> ο οποίος εμφανίζεται στο πάνω μέρος του browser.

Ορίζεται η χρήση του CSS και JavaScript.

Στην περιοχή **body** ορίζεται το κύριο μέρος της ιστοσελίδας και μπορεί να περιέχει ετικέτες μορφοποίησης, όπως <p> ορίζει έναρξη παραγράφου.

<h1><h2>...<h6> Το κείμενο ανάμεσα σε αυτές τις ετικέτες έχει την μορφή επικεφαλίδας της οποίας το μέγεθος μειώνεται όσο μεταβαίνουμε από την <h1> στην <h6>.

<hr> εμφανίζει μια οριζόντια γραμμή στην σελίδα μας.

<i> εμφανίζει το κείμενο ανάμεσα στο εν λόγω tag με πλάγιους χαρακτήρες.

 εμφανίζει το κείμενο με έντονη μορφή.

<u> εμφανίζει το κείμενο υπογραμμισμένο.

 αλλαγή γραμμής.

Λίστες Ταξινομημένες ή αταξινομητες με τα tags και αντίστοιχα π.χ. Red Green Blue

Πλαίσια (frames) ορίζονται με το tag <frame>

Εικόνες φορτώνονται με το tag img όπου ορίζουμε το όνομα και την διαδρομή του αρχείου εικόνας(src), τις διαστάσεις της(width-height) και την επεξήγηση(title) π.χ.

Υπερσυνδέσμοι σε άλλες σελίδες με το tag <A>... όπου ορίζουμε το όνομα και την διαδρομή της σελίδας(HREF), την επεξήγηση(title) π.χ.

HERE

Πίνακες δεδομένων ορίζονται με τα tags <table>...</table> Το <table> έχει διάφορες ιδιότητες όπως width border, caption π.χ.

<table width="100%" border="1" <caption> Πελάτες</caption>>

Οι γραμμές του πίνακα ορίζονται με το tag <tr>...</tr> ενώ οι στήλες με το tag <th>...</th> για επικεφαλίδες στηλών και <td>...</td> για δεδομένα στηλών. Π.χ

<table><tr><th>Name</th><th>address</th><th>phone</th></tr>

<tr><th rowspan="2">J.papas</th><th>Heracleion</th></tr>

<tr><th>123-444</th></tr></table>

Η rowspan ορίζει σε πόσες γραμμές απλώνεται ένα κελί παρόμοια είναι και το colspan

Φόρμες εισαγωγής δεδομένων από τον client και αποστολής τους στον server. Ορίζονται με τα tags <form>...</form> Η <form> έχει 2 ιδιότητες την **Action** που ορίζει σε

ποια σελίδα θα στείλουμε τα δεδομένα και την **Method** που ορίζει τη μέθοδο του πρωτοκόλλου HTTP θα χρησιμοποιήσουμε για την αποστολή των δεδομένων **GET ή POST** π.χ. <form action="page1.php"

method="get">

Με την μέθοδο POST τα δεδομένα στέλνονται μέσα στα HTTP μηνύματα που στέλνει ο browser και δεν τα βλέπει ο χρήστης ενώ με την μέθοδο GET τα δεδομένα προστίθενται στο url π.χ.

<http://www.teicrete.gr/page1.php?a1=12>

Μία φόρμα μπορεί να περιέχει τα παρακάτω στοιχεία:

1) Στοιχείο **Text** για την εισαγωγή κειμένου π.χ.

<input type="text" id="c1" name="code" value="0"> Πριν από το στοιχείο text μπορούμε να εμφανίσουμε μία ετικέτα για να γνωρίζει ο χρήστης τι είναι αυτό που εισάγει π.χ. <label

for="c1">κωδικός; </label>

2) Στοιχείο **Textarea** για την εισαγωγή κειμένου περισσότερων από μία γραμμή π.χ.

<textarea name="area1">περιγραφή</textarea>

Στοιχείο για την εισαγωγή συνθηματικού(**password**) π.χ.

<input type="password" name="myPasswd">

3) Στοιχείο **checkbox** για να τσεκάρει ο χρήστης μία ή περισσότερες επιλογές π.χ.

<input type="checkbox" name="lang1" value="English"> English

<input type="checkbox" name="lang2" value="German"> German

<input type="checkbox" name="lang3" value="France"> France

4) Στοιχείο **radio button** για να τσεκάρει ο χρήστης μία επιλογή π.χ.

<input type="radio" name="thesex" value="male"> male

<input type="radio" name="thesex" value="female">female

5) Στοιχείο **select** ορίζει μία pop up list για να επιλέξει ο χρήστης μία επιλογή π.χ.

<select name="combo">

<option>Ag.Nikolaos</option>

<option>heraklion</option>

<option>rethymno</option>

<option>hania</option>

</select>

6) Στοιχείο **button** ορίζουμε κουμπιά διαφόρων τύπων όπως submit για αποστολή της φόρμας στον server , reset για επαναφορά της φόρμας π.χ.

<input type="submit" name="button1" value="Αποστολή">

<input type="reset" name="reset" value="reset">

2.2 Scripting (php, asp..)

Η **PHP (Hypertext Preprocessor)** είναι γλώσσα προγραμματισμού για την δημιουργία δυναμικών ιστοσελίδων.

Είναι **Server-side scripting** δηλαδή το πρόγραμμα εκτελείται στον web-server και μπορεί να κάνει ό,τι κάνει ένα CGI script.

CGI (Common Gateway Interface- Κοινή Διασύνδεση Πύλης): είναι η μέθοδος ή ο τρόπος με τον οποίο ο webserver και ο browser χειρίζονται τα δεδομένα. Πιο συγκεκριμένα προσδιορίζει:

α) Το πέρασμα δεδομένων από web browser στον server β) Τον τρόπο μεταφοράς των δεδομένων του server στα δικά του προγράμματα και γ) τον τρόπο μεταφοράς των δεδομένων από τον server στον Web browser.

Η εκτέλεση του προγράμματος στον web server γίνεται με την βοήθεια του **μεταγλωττιστή PHP(parser)** που πρέπει να έχει εγκατασταθεί στον web server.

Η PHP είναι HTML ενσωματωμένη scripting γλώσσα. Δηλαδή ο κώδικας PHP συμπεριλαμβάνεται μέσα σε μία σελίδα HTML. Κατά την μεταγλώττιση η PHP αφήνει την σελίδα HTML όπως είναι μέχρι να συναντήσει ένα από τα ειδικά tags έναρξης php “<?” και τότε αρχίζει να μεταγλωττίζει το κείμενο ως κώδικα PHP μέχρι να συναντήσει το επόμενο PHP tag κλεισίματος“ ?>” δηλαδή οτιδήποτε βρίσκεται έξω από τα tags της PHP μένει όπως είναι. Το αποτέλεσμα της μεταγλώττισης μορφοποιείται σε HTML και όλη η ιστοσελίδα αποστέλλεται στον browser. Από τα παραπάνω συμπεραίνεται ότι η **PHP είναι διερμηνέας(interpreter)** δηλαδή δεν παράγει εκτελέσιμο αρχείο.

Είναι command line scripting δηλαδή ένα Php script μπορεί να τρέξει χωρίς Browser και Server.

Χρειαζόμαστε μόνο τον μεταγλωττιστή.

Υποστηρίζει πολλά πρωτόκολλα για επικοινωνία με άλλες υπηρεσίες όπως IMAP,SNMP,NNTP,HTTP,POP3 καθώς επίσης και τα socket.

Υποστηρίζεται από τους περισσότερους web servers.

Πλεονεκτήματα της PHP

Οι ανταγωνιστές της PHP είναι ο Perl, Microsoft Active Server Pages(ASP), Java Server Pages(JSP), Allaire Cold Fusion. Σε σύγκριση με τους ανταγωνιστές έχει τα παρακάτω πλεονεκτήματα.

1) Απόδοση, με τις δοκιμές που έγιναν η PHP έχει καλύτερη απόδοση από τους ανταγωνιστές της. Με ένα φθηνό διακομιστή μπορεί να εξυπηρετήσει εκατομμύρια επισκεπτών ημερησίως.

2) Διασυνδέσεις με πολλά Συστήματα Βάσεων Δεδομένων όπως Mysql, Oracle, Informix, Interbase, Sybase, msql, PostgreSQL, dbm κ.λ.π. Με την χρήση του ODBC μπορεί να συνδεθεί με οποιαδήποτε βάση παρέχει αυτό τον οδηγό. **ODBC(Open DataBase Connectivity)** είναι μια διεπαφή γλώσσας C που επιτρέπει στο εκτελέσιμο πρόγραμμα να λειτουργεί ανεξάρτητα από ΣΔΒΔ.

3) Ενσωματωμένες βιβλιοθήκες για συνηθισμένες web διαδικασίες όπως σύνδεση με υπηρεσίες άλλων δικτύων, αποστολή email, να δουλέψετε με cookies, να δημιουργήσετε έγγραφα pdf και εικόνες gif.

4) Χαμηλό κόστος. Η PHP είναι λογισμικό ανοικτού κώδικα(Open Source software) έτσι διατίθεται δωρεάν στην διεύθυνση <http://www.php.net> με άδεια ανοικτού κώδικα ή με χαμηλό κόστος για με εμπορική άδεια.

5) Διαθεσιμότητα κώδικα προέλευσης. Επειδή είναι Open Source μπορείτε να πάρετε και να τροποποιήσετε τον κώδικα προέλευσης.

6) Μεταφερσιμότητα. Τρέχει σε όλα τα λειτουργικά unix και windows

7) Ευκολία στην εκμάθηση. Η σύνταξη των εντολών της βασίζεται στην C και την Perl.

Σύνταξη και Εντολές της PHP

- Ο κώδικα αρχίζει με <? και τελειώνει με ?>.
- Κάθε εντολή τελειώνει με Ελληνικό ερωτηματικό (;)
- Τα σχόλια αρχίζουν με //
- Οι μεταβλητές ορίζονται με την απόδοση τιμής. Το όνομα τους ξεκινά με \$ και ακολουθεί γράμμα ή κάτω παύλα και στην συνέχεια οτιδήποτε εκτός από χαρακτήρες +, -, &, *. Είναι case-sensitive π.χ. \$total και \$Total είναι διαφορετικές μεταβλητές.

Υπάρχουν οι γνωστοί τύποι μεταβλητών όπως συμβολοσειρές (strings), ακέραιοι (integers), αριθμοί κινητής υποδιαστολής (floats), λογικές τιμές (booleans), πίνακες (arrays) και αντικείμενα (objects) .

Δομές ελέγχου

```
If (συνθήκη){εντολές}else{εντολές}
```

```
Switch(μεταβλητή) case τιμή1: εντολές break;case τιμή2: εντολές; break; ... default: εντολές;
```

Δομές επανάληψης

```
while(συνθήκη){εντολές}
```

```
Do{ εντολές}while(συνθήκη)
```

```
for(Αρχική τιμή; Τελική τιμή; Μεταβολή) {εντολές}
```

Ορισμοί Πινάκων

```
$Pin1= new array(); //ορίζει ένα κενό πίνακα
```

```
$colors = array('red', 'blue', 'green', 'yellow');
```

```
$Pin2[]= "TEI"; $Pin2[]= "Ηράκλειο"; //Με αυτό τον τρόπο ο τελεστής [] ορίζει τον πίνακα
```

```
$Pin2 και του δίνει και περιεχόμενο στις θέσεις $Pin2[0] και $Pin2[1] τις τιμές "TEI" και "Ηράκλειο" αντίστοιχα.
```

Η εμφάνιση των χρωμάτων που περιέχονται στον πίνακα \$colors που ορίσαμε παραπάνω γίνεται με τον βρόγχο που ακολουθεί

```
foreach ($colors as $color) { echo " $color\n";}
```

Μεταφορά Αρχείων

Η PHP μπορεί να λαμβάνει αρχεία από οποιοδήποτε συμβατό browser. Αυτό δίνει τη δυνατότητα στους χρήστες να κάνουν uploads αρχείων κειμένου καθώς και αρχείων binary. Με την φόρμα που ακολουθεί γίνεται upload αρχείου ή εικόνας που επιλέγει ο χρήστης

```
<form enctype=multipart/form-data method="post" action="...">
<input type=hidden name=MAX_FILE_SIZE value=150000>
<input type=hidden name=completed value=1>
<p>Επέλεξε Εικόνα: <input type="file" name="imagefile"></p>
<input type="submit" name="submit" value=" submit ">
</form>
```

Η μεταβλητή `MAX_FILE_SIZE` ορίζει το μέγιστο μέγεθος αρχείου που θέλουμε να κάνουμε αποδεκτό μ' αυτή τη φόρμα. Αυτή η τιμή δεν μπορεί να υπερβεί τη ρύθμιση `upload_max_filesize` στο αρχείο `php.ini`. Τα αρχεία αρχικά αποθηκεύονται σε προκαθορισμένο προσωρινό κατάλογο του server μέχρι να ελεγχθεί αν το μέγεθός του υπερβαίνει το `MAX_FILE_SIZE`. Η μεταβλητή που ορίζεται για τα αρχεία που έχουν γίνει upload είναι η `$_FILES`. Οι πληροφορίες που περιέχει είναι το αρχικό όνομα του αρχείου που έχει γίνει upload, το μέγεθος του, τον τύπο του και τέλος το προσωρινό όνομα με το πλήρη διαδρομή.

Για Παράδειγμα αν με την φόρμα που ορίσαμε λίγο παραπάνω φορτώσουμε μία εικόνα με όνομα `toyota-Yaris.jpg` και μέγεθος 55,3 kb τότε τα περιεχόμενα της `$_FILES` είναι τα ακόλουθα:

```
$_FILES['imagefile']['name'] toyota-Yaris.jpg (αρχικό όνομα)
$_FILES['imagefile']['type'] image/jpeg (τύπος αρχείου)
$_FILES['imagefile']['size'] 56675 (μέγεθος αρχείου)
```

```
$_FILES['imagefile']['tmp_name' ] C: \WINDOWS\Temp\phpA9.tmp (προσωρινό όνομα)
```

Τα παραπάνω εμφανίζονται αν στην `php` χρησιμοποιήσουμε μία βοηθητική μεταβλητή για να πάρουμε το περιεχόμενο και στην συνέχεια την εμφανίσουμε. Για παράδειγμα

```
$tmp1=$_FILES['imagefile']['size']; echo "$tmp1";
```

Η συνάρτηση **`move_uploaded_file()`** αντιγράφει ένα αρχείο από μία θέση σε μία άλλη. Απαιτεί σαν ορίσματα το πηγαίο αρχείο με τη διαδρομή της θέσης που βρίσκεται και το όνομα του αρχείου προορισμού με τη διαδρομή που θα τοποθετηθεί. Για παράδειγμα η `move_uploaded_file($_FILES['imagefile']['tmp_name'], "latest.img");`

Σύμφωνα με τα προηγούμενα αντιγράφει το αρχείο `C: \WINDOWS\Temp\phpA9.tmp` στον τρέχων κατάλογο με όνομα `latest.img`

Προσπέλαση αρχείων.

Η συνάρτηση **`fopen()`** Ανοίγει ένα αρχείο για διάβασμα (`r`) ή για γράψιμο (`w`). Η `fopen()` επιστρέφει ένα ακέραιο αριθμό (file handler) που καθορίζει το αρχείο. Π.χ.

```
$fp = fopen("file.txt", "r");
```

Ανοίγει το αρχείο `file.txt` για διάβασμα και ο file handler τοποθετείται στην μεταβλητή `$fp`. Μετά από την διαχείριση ενός αρχείου πρέπει να απελευθερώσουμε τον file handler με την εντολή `fclose()`.

Η συνάρτηση **`fread()`** διαβάζει δεδομένα από ένα αρχείο π.χ. `$data=fread($fp,10)` διαβάζει 10 bytes από το αρχείο που ορίζεται από την `$fp` και τα αποθηκεύει στη μεταβλητή `$data`.

Η συνάρτηση **`fwrite()`** αποθηκεύει δεδομένα στο αρχείο π.χ. `fwrite($fp, $line)` αποθηκεύει στο αρχείο που ορίζεται από την `$fp` το περιεχόμενο της μεταβλητής `$line`.

Η εικόνα που περιέχεται στο `latest.img` που αναφέραμε παραπάνω δεν είναι κατάλληλη να την αποθηκεύσουμε σε βάση δεδομένων διότι μπορεί να περιέχει χαρακτήρες που για την `php` έχουν κάποια σημασία. Τέτοιοι χαρακτήρες είναι τα μονά και διπλά εισαγωγικά η ανάποδη κάθετος κλπ. Για να μην μεταφράσει την σημασία αυτών των χαρακτήρων η `php` αλλά να τους θεωρήσει κανονικούς χαρακτήρες προσθέτουμε πριν από κάθε τέτοιο χαρακτήρα την ανάποδη κάθετος π.χ. το “ γίνεται \” Μπορούμε να προσθέτουμε την ανάποδη κάθετος όπου χρειάζεται σε ένα ολόκληρο κείμενο αυτόματα χρησιμοποιώντας την συνάρτηση **`addslashes()`** ή να την αφαιρέσουμε αν την έχουμε προσθέσει με την συνάρτηση **`stripslashes()`**.

Η μεταβλητή `$image` για να είναι κατάλληλη να αποθηκευθεί σε βάση δεδομένων πρέπει να κάνουμε τις παρακάτω εντολές: `$instr = fopen("latest.img", "rb");`


```
$image = addslashes(fread($instr,filesize("latest.img")));
```

Πρόσβαση σε MYSQL βάση

Για να συνδεθούμε με την βάση χρησιμοποιούμε την συνάρτηση:

```
mysql_connect(hostname , username ,password)
```

Hostname είναι η διεύθυνση του υπολογιστή που τρέχει ο mysql server

Η παραπάνω συνάρτηση σε περίπτωση επιτυχίας επιστρέφει έναν αναγνωριστή συνδέσμου τον οποίο πρέπει να αποθηκεύουμε σε μία μεταβλητή για να τον χρησιμοποιήσουμε σε επόμενες εντολές

```
π.χ. $conn=mysql_connect(hostname , username ,password)
```

Σε περίπτωση αποτυχίας η συνάρτηση mysql_connect επιστρέφει false.

mysql_select_db(όνομα βάσης, αναγνωριστής συνδέσμου") Χρησιμοποιείται για την επιλογή βάσης δεδομένων π.χ. mysql_select_db("test1",\$conn)

mysql_create_db(όνομα βάσης, αναγνωριστής συνδέσμου") Χρησιμοποιείται για την δημιουργία βάσης δεδομένων π.χ. mysql_create_db("dimopras",\$conn)

mysql_drop_db(όνομα βάσης, αναγνωριστής συνδέσμου") Χρησιμοποιείται για την διαγραφή βάσης δεδομένων π.χ. mysql_drop_db("dimopras",\$conn)

mysql_query(ερώτημα, αναγνωριστής συνδέσμου") Χρησιμοποιείται για την εκτέλεση ερωτημάτων.

Το αποτέλεσμα του ερωτήματος το αποθηκεύουμε σε μεταβλητή ειδικά αν έχουμε ερώτημα ανάκλησης δεδομένων από πίνακα π.χ.

```
$query1="select * from users";
```

```
$result=mysql_query($quer1,$conn);
```

Για να επεξεργαστούμε το περιεχόμενο του αποτελέσματος σε γραμμές χρησιμοποιούμε τη συνάρτηση mysql_fetch_array() η οποία κάθε φορά επιστρέφει μία γραμμή από το αποτέλεσμα για παράδειγμα έστω ότι θέλω να πάρω τα στοιχεία των χρηστών που είναι καταχωρημένοι στον πίνακα users. Τα στοιχεία των χρηστών είναι ο κωδικός το όνομα και το τηλέφωνο δηλαδή τα πεδία code,

fullname και phone αντίστοιχα.

```
$sql="select code,fullname,phone from users ";
```

```
$result=mysql_query($sql,$conn);
```

```
while($row = mysql_fetch_array($result))
```

```
{ $code=$row['code'];
```

```
$fullname=$row['fullname'];
```

```
$phone=$row['phone'];
```

```
}
```

Αν κάτι δεν πάει καλά κατά την εκτέλεση των εντολών μπορούμε να χρησιμοποιήσουμε την συνάρτηση die() για να εμφανίσουμε κάποιο μήνυμα δικό μας ή το αντίστοιχο μήνυμα της mysql με την χρήση της συνάρτησης mysql_error() ή και τα δύο όπως στο παράδειγμα που ακολουθεί:

```
$conn=mysql_connect("127.0.0.1","user1","niki") or die("μή εφικτή η σύνδεση")  
.mysql_error ());
```

2.3 Μεταδεδομένα

Τα μεταδεδομένα είναι δομημένες πληροφορίες οι οποίες περιγράφουν, εξηγούν, εντοπίζουν, ή αλλιώς καθιστούν ευκολότερη την ανάκτηση, τη χρήση, ή τη διαχείριση μιας πηγής πληροφοριών. Τα μεταδεδομένα ονομάζονται συχνά δεδομένα σχετικά με τα δεδομένα ή πληροφορίες σχετικά με τις πληροφορίες.

Τα μεταδεδομένα μπορούν να περιγράψουν πόρους σε οποιοδήποτε επίπεδο ομαδοποίησης. Μπορούν να περιγράψουν μια συλλογή, ένα μόνο πόρο, ή ένα συνθετικό μέρος μιας μεγαλύτερης πηγής (για παράδειγμα, μια φωτογραφία σε ένα άρθρο).

Τα μεταδεδομένα μπορούν να ενσωματωθούν σε ένα ψηφιακό αντικείμενο ή να αποθηκευτούν χωριστά. Τα μεταδεδομένα είναι συχνά ενσωματωμένα σε HTML έγγραφα και στις κεφαλίδες των αρχείων εικόνας. Η αποθήκευση των μεταδεδομένων μαζί με το αντικείμενο που περιγράφουν διασφαλίζει ότι τα μεταδεδομένα δε θα χαθούν, ώστε να αποφεύγονται προβλήματα σύνδεσης μεταξύ των δεδομένων και των μεταδεδομένων, και βοηθά στο να διασφαλιστεί ότι τα μεταδεδομένα και το αντικείμενο θα ενημερώνονται ταυτόχρονα. Ωστόσο, είναι αδύνατο να ενσωματωθούν μεταδεδομένα σε ορισμένους τύπους αντικειμένων (για παράδειγμα, εκθέματα). Επίσης, η αποθήκευση μεταδεδομένων χωριστά μπορεί να απλοποιήσει τη διαχείριση των μεταδεδομένων και να διευκολύνει την έρευνα και την ανάκτηση. Ως εκ τούτου, τα μεταδεδομένα συνήθως αποθηκεύονται σε ένα σύστημα βάσης δεδομένων και συνδέονται με τα αντικείμενα που περιγράφουν.

Ένας σημαντικός λόγος για τη δημιουργία περιγραφικών μεταδεδομένων είναι η διευκόλυνση στην ανακάλυψη των σχετικών πληροφοριών. Πέρα, όμως, από την ανακάλυψη των πόρων, τα μεταδεδομένα μπορούν να βοηθήσουν στην οργάνωση ηλεκτρονικών πηγών πληροφόρησης, στη διευκόλυνση της διαλειτουργικότητας, στην παροχή ψηφιακής ταυτοποίησης και στην υποστήριξη της αρχειοθέτησης και συντήρησης.

Τα μεταδεδομένα εξυπηρετούν τις ίδιες λειτουργίες στην ανακάλυψη των πόρων, όπως η καλή αρχειοθέτηση:

- επιτρέποντας την εύρεση πόρων με τα σχετικά κριτήρια,
- προσδιορίζοντας τους πόρους,
- ομαδοποιώντας παρόμοιους πόρους,
- διακρίνοντας ανόμοιους πόρους, και
- δίνοντας πληροφορίες τοποθεσίας.

Καθώς ο αριθμός των δικτυοκεντρικών πόρων αυξάνεται εκθετικά, το σύνολο των σελίδων ή πυλών είναι όλο και πιο χρήσιμο για την οργάνωση συνδέσεων με τους πόρους με βάση το κοινό ή το θέμα. Οι κατάλογοι αυτοί μπορούν να κατασκευαστούν ως στατικές σελίδες, με τα ονόματα και τις τοποθεσίες των πόρων "ενσωματωμένα" στην HTML. Ωστόσο, είναι πιο αποτελεσματικό και όλο και πιο σύνηθες να κατασκευάζονται αυτές οι σελίδες δυναμικά από μεταδεδομένα που αποθηκεύονται σε βάσεις δεδομένων. Διάφορα εργαλεία λογισμικού μπορούν να χρησιμοποιηθούν για την αυτόματη εξαγωγή και διαμόρφωση των πληροφοριών για τις εφαρμογές Ιστού.

Η περιγραφή ενός πόρου με μεταδεδομένα επιτρέπει την κατανόηση από τους ανθρώπους και από τις μηχανές με τρόπους που προωθούν τη διαλειτουργικότητα. Η διαλειτουργικότητα είναι η ικανότητα ανταλλαγής δεδομένων με ελάχιστη απώλεια περιεχομένου και λειτουργικότητας μεταξύ πολλαπλών συστημάτων με διαφορετικό υλικό και πλατφόρμες λογισμικού, δομές δεδομένων και διεπαφές. Με τη χρήση προκαθορισμένων συστημάτων μεταδεδομένων, κοινών πρωτοκόλλων μεταφοράς, και διαβάσεις μεταξύ των συστημάτων, η αναζήτηση των πόρων σε όλο το διαδίκτυο μπορεί να γίνει πιο ομαλά.

Τα περισσότερα σχήματα μεταδεδομένων περιλαμβάνουν στοιχεία, όπως αριθμούς για το μοναδικό προσδιορισμό του έργου ή του αντικειμένου στο οποίο αναφέρεται. Η τοποθεσία

ενός ψηφιακού αντικειμένου μπορεί επίσης να δοθεί με τη χρήση ενός ονόματος αρχείου, μια διεύθυνση URL (Uniform Resource Locator), ή ένα περισσότερο μόνιμο αναγνωριστικό όπως PURL (Persistent URL) ή DOI (Digital Object Identifier). Τα μόνιμα αναγνωριστικά προτιμώνται διότι οι τοποθεσίες των αντικειμένων συχνά αλλάζουν, κάνοντας το πρότυπο URL (και επομένως και την εγγραφή των μεταδεδομένων) μη έγκυρη. Εκτός από τα πραγματικά στοιχεία που επισημαίνουν το αντικείμενο, τα μεταδεδομένα μπορούν να συνδυαστούν για να ενεργήσουν ως ένα σύνολο στοιχείων ταυτότητας, διαφοροποιώντας ένα αντικείμενο από άλλο για σκοπούς επικύρωσης.

Οι πιο πρόσφατες προσπάθειες μεταδεδομένων επικεντρώνονται γύρω από την ανακάλυψη πρόσφατα δημιουργημένων πόρων. Ωστόσο, υπάρχει αυξανόμενη ανησυχία ότι οι ψηφιακοί πόροι δε θα επιβιώσουν σε εύχρηστη μορφή στο μέλλον. Η ψηφιακή πληροφορία είναι εύθραυστη. Μπορεί να καταστραφεί ή να τροποποιηθεί, ηθελημένα ή αθέλητα. Μπορούν να αχρηστευτούν καθώς τα μέσα αποθήκευσης και οι τεχνολογίες υλικού και λογισμικού αλλάζουν.

Τα μεταδεδομένα είναι το κλειδί για τη διασφάλιση της επιβίωσης και της προσβασιμότητας στους πόρους στο μέλλον. Η αρχειοθέτηση και η διατήρηση απαιτούν ειδικά στοιχεία για την παρακολούθηση της καταγωγής ενός ψηφιακού αντικειμένου (από πού προήλθε και πώς έχει αλλάξει κατά την πάροδο του χρόνου), τις λεπτομέρειες των φυσικών χαρακτηριστικών, και την τεκμηρίωση της συμπεριφοράς, προκειμένου να τα εξομοιώσει σε τεχνολογίες του μέλλοντος.

Τα συστήματα μεταδεδομένων (ονομάζεται και σχήμα) είναι σύνολα στοιχείων μεταδεδομένων που έχουν σχεδιαστεί για έναν συγκεκριμένο σκοπό, όπως η περιγραφή ενός συγκεκριμένου είδους πόρων πληροφοριών. Ο ορισμός ή η σημασία αυτών των στοιχείων είναι γνωστός ως η σημασιολογία του συστήματος. Οι τιμές που δίνονται στα στοιχεία μεταδεδομένων είναι το περιεχόμενο. Τα συστήματα μεταδεδομένων γενικά καθορίζουν τα ονόματα των στοιχείων και την σημασιολογία τους. Προαιρετικά, μπορούν να προσδιορίζουν τους κανόνες για τη διατύπωση του περιεχομένου (για παράδειγμα, πώς να αναγνωρίζεται ο βασικός τίτλος), κανόνες για την αναπαράσταση του περιεχομένου (για παράδειγμα, κεφαλαιοποίηση), και για τις επιτρεπόμενες τιμές περιεχομένου (για παράδειγμα, οι όροι πρέπει να χρησιμοποιούνται από ένα συγκεκριμένο ελεγχόμενο λεξιλόγιο).

Μπορεί επίσης να υπάρχουν κανόνες σύνταξης για την κωδικοποίηση των στοιχείων και του περιεχομένου τους. Ένα σχήμα μεταδεδομένων, χωρίς προβλεπόμενους συντακτικούς κανόνες, καλείται ανεξάρτητη σύνταξη. Τα μεταδεδομένα μπορούν να κωδικοποιηθούν σε οποιαδήποτε προσδιορίσιμη σύνταξη. Πολλά σημερινά συστήματα μεταδεδομένων χρησιμοποιούν SGML (Standard Generalized Mark-up Language) ή XML (Extensible Mark-up Language). Η XML, που αναπτύχθηκε από το World Wide Web Consortium (W3C), είναι μία ανεπτυγμένη μορφή της HTML που επιτρέπει τοπικά καθορισμένα σύνολα ετικετών και εύκολη ανταλλαγή δομημένων πληροφοριών. Η SGML είναι ένα υπερσύνολο τόσο της HTML όσο και της XML και επιτρέπει την πλουσιότερη σήμανση ενός εγγράφου. Χρήσιμα εργαλεία XML γίνονται ευρέως διαθέσιμα καθώς η XML διαδραματίζει ολοένα και σημαντικότερο ρόλο στην ανταλλαγή μιας ποικιλίας δεδομένων στον Ιστό.

Το ποιος δημιουργεί τα μεταδεδομένα εξαρτάται από τον πόρο που περιγράφεται, τα διαθέσιμα εργαλεία και το αναμενόμενο αποτέλεσμα. Ωστόσο, είναι συνήθως συλλογική προσπάθεια.

Πολλά βασικά διαρθρωτικά και διοικητικά μεταδεδομένα παρέχονται από το τεχνικό προσωπικό που αρχικά ψηφιοποιεί ή αλλιώς δημιουργεί το ψηφιακό αντικείμενο, ή παράγονται μέσω μιας αυτοματοποιημένης διαδικασίας. Για τα περιγραφικά μεταδεδομένα, είναι καλύτερο σε ορισμένες περιπτώσεις, εάν ο δημιουργός του πόρου παρέχει τις πληροφορίες. Αυτό ισχύει πολύ στην τεκμηρίωση επιστημονικών σετ δεδομένων στα οποία ο δημιουργός έχει καλύτερη κατανόηση του σκεπτικού για τα δεδομένα και τις χρήσεις στις

οποίες θα μπορούσαν να τεθούν, και για τα οποία υπάρχουν ελάχιστες ή και καθόλου πληροφορίες με τις οποίες θα μπορούσε να λειτουργήσει ένα αυτοματοποιημένο εργαλείο.

Ωστόσο, πολλά έργα έχει διαπιστωθεί ότι είναι πιο αποτελεσματικό να έχουν αρχειοθέτες ή άλλους επαγγελματίες πληροφοριών να δημιουργούν τα περιγραφικά μεταδεδομένα, επειδή οι συντάκτες ή οι δημιουργοί των δεδομένων δεν έχουν το χρόνο ή τις δεξιότητες. Σε άλλες περιπτώσεις, χρησιμοποιείται ο συνδυασμός του ερευνητή και του επαγγελματία πληροφοριών. Ο ερευνητής μπορεί να δημιουργήσει έναν σκελετό, συμπληρώνοντας τα στοιχεία που μπορούν να προσκομισθούν πιο εύκολα. Στην συνέχεια, τα αποτελέσματα μπορούν να συμπληρωθούν ή να αξιολογηθούν από τον ειδικό πληροφοριών για την συνοχή και την συμμόρφωση με την σύνταξη και τις τοπικές οδηγίες.

Πολλές πρωτοβουλίες έργων μεταδεδομένων έχουν αναπτύξει εργαλεία και τα έχουν κοινοποιήσει σε άλλους, μερικές φορές δωρεάν. Ένας αυξανόμενος αριθμός εμπορικών λογισμικών εργαλείων γίνεται επίσης διαθέσιμος. Τα εργαλεία δημιουργίας εμπίπτουν σε διάφορες κατηγορίες:

- Τα *πρότυπα* επιτρέπουν στο χρήστη να εισάγει τις τιμές των μεταδεδομένων σε προκαθορισμένα πεδία που ταιριάζουν με το σύνολο των στοιχείων που χρησιμοποιούνται. Το πρότυπο, στην συνέχεια, θα δημιουργήσει μια διαμορφωμένη σειρά των χαρακτηριστικών των στοιχείων και τις αντίστοιχες τιμές.

- Τα *εργαλεία σήμανσης* διαρθρώνουν τα χαρακτηριστικά των μεταδεδομένων και τις τιμές στην καθορισμένη γλώσσα σχήματος. Τα περισσότερα από αυτά τα εργαλεία παράγουν XML ή SGML Document Type Definitions (DTD). Μερικά πρότυπα περιλαμβάνουν τέτοια σήμανση ως μέρος της τελικής μετάφρασης των μεταδεδομένων.

- Τα *εργαλεία εξόρυξης* δημιουργούν αυτόματα τα μεταδεδομένα από την ανάλυση των ψηφιακών πόρων. Τα εργαλεία αυτά γενικά περιορίζονται σε πόρους σε μορφή κειμένου. Η ποιότητα των εξορυγμένων μεταδεδομένων μπορούν να ποικίλλουν σημαντικά ανάλογα με τους αλγόριθμους του εργαλείου καθώς και το περιεχόμενο και τη δομή της πηγής κειμένου. Τα εργαλεία αυτά θα πρέπει να χρησιμοποιούνται ως ενίσχυση για τη δημιουργία μεταδεδομένων. Τα μεταδεδομένα που προκύπτουν θα πρέπει να αξιολογούνται και να επεξεργάζονται χειροκίνητα.

- Τα *εργαλεία μετατροπής* μεταφράζουν από ένα σχήμα μεταδεδομένων σε άλλο. Η ομοιότητα των στοιχείων στις μορφές της πηγής και του στόχου επηρεάζουν το ποσό της επιπλέον επεξεργασίας και χειροκίνητης εισαγωγής μεταδεδομένων που μπορεί να απαιτηθεί.

Τα εργαλεία μεταδεδομένων γενικά αναπτύσσονται για την υποστήριξη συγκεκριμένων σχημάτων μεταδεδομένων ή ομάδων στοιχείων. Οι ιστοσελίδες για το κάθε σχήμα συχνά έχουν συνδέσμους για τα σχετικά εργαλεία.

Η αυτόματη δημιουργία μεταδεδομένων ή από συντάκτες πληροφοριών οι οποίοι δεν είναι εξοικειωμένοι με την καταλογοποίηση, ευρετηρίασης, ή τον έλεγχο λεξιλογίου μπορεί να δημιουργήσει προβλήματα στην ποιότητα. Υποχρεωτικά στοιχεία μπορεί να λείπουν ή να χρησιμοποιούνται λανθασμένα. Η σύνταξη του σχήματος μπορεί να έχει λάθη που εμποδίζουν την σωστή επεξεργασία των μεταδεδομένων. Η ορολογία του περιεχομένου των μεταδεδομένων μπορεί να είναι ασυνεπής, γεγονός που θα καθιστά δύσκολο να βρεθούν οι σχετικές πληροφορίες. Το Framework of Guidance for Building Good Digital Collections, που διατίθεται στην ιστοσελίδα του NISO, διατυπώνει έξι αρχές που ισχύουν για σωστά μεταδεδομένα:

- Τα σωστά μεταδεδομένα πρέπει να είναι ανάλογα με το υλικό της συλλογής, τους χρήστες της συλλογής και την τρέχουσα και πιθανή χρήση του ψηφιακού αντικειμένου.

- Τα σωστά μεταδεδομένα υποστηρίζουν τη διαλειτουργικότητα.

- Τα σωστά μεταδεδομένα χρησιμοποιούν ελεγχόμενο λεξιλόγιο για να αντικατοπτρίζουν το τι, πού, πότε και ποιος του περιεχομένου.

- Τα σωστά μεταδεδομένα περιλαμβάνουν μια σαφή δήλωση σχετικά με τις προϋποθέσεις και τους όρους χρήσης για το ψηφιακό αντικείμενο.

- Τα σωστά μεταδεδομένα είναι αντικείμενα κι αυτά και ως εκ τούτου θα πρέπει να έχουν ιδιότητες αρχειοθέτησης, ανθεκτικότητας, μοναδική ταυτοποίηση, κλπ. Θα πρέπει να είναι αυθεντικά και επαληθεύσιμα.

- Τα σωστά μεταδεδομένα υποστηρίζουν τη μακροπρόθεσμη διαχείριση των αντικειμένων των συλλογών.

Υπάρχει σε εξέλιξη ένας αριθμός προσπαθειών για την αντιμετώπιση της πρόκλησης για την ποιότητα των μεταδεδομένων:

- Τα εργαλεία δημιουργίας μεταδεδομένων βελτιώνονται με χαρακτηριστικά όπως πρότυπα, λίστες επιλογής που περιορίζουν την επιλογή σε ένα συγκεκριμένο πεδίο, και βελτιωμένοι κανόνες επικύρωσης.

- Τα προγράμματα διαλειτουργικότητας λογισμικού που αυτοματοποιούν τη «διάβαση» μεταξύ των διαφορετικών σχημάτων συνεχώς αναπτύσσονται και τελειοποιούνται.

- Οι δημιουργοί περιεχομένου εκπαιδεύονται επισήμως στην κατανόηση των εννοιών των μεταδεδομένων και του ελεγχόμενου λεξιλογίου, και στη χρήση των εργαλείων που σχετίζονται με τα μεταδεδομένα.

- Υπάρχοντα ελεγχόμενα λεξιλόγια που μπορεί να έχουν αρχικά σχεδιαστεί για συγκεκριμένη χρήση ή για περιορισμένο κοινό αποκτούν ευρύτερη χρήση και γνώση.

- Κοινότητες χρηστών αναπτύσσουν και τελειοποιούν τα σχήματα μεταδεδομένων που αφορούν ειδικό κοινό, τα προφίλ εφαρμογών, τα ελεγχόμενα λεξιλόγια, και τις κατευθυντήριες γραμμές χρήσης.

Διαφορετικά συστήματα εξυπηρετούν διαφορετικές ανάγκες και κοινό. Συμπληρωματικά συστήματα μπορούν να χρησιμοποιηθούν για να περιγράψουν την ίδια πηγή για πολλαπλούς σκοπούς και για την εξυπηρέτηση μιας σειράς ομάδων χρηστών. Το Resource Description Framework (RDF), που αναπτύχθηκε από το World Wide Web Consortium (W3C), είναι ένα μοντέλο δεδομένων για την περιγραφή των πόρων για τον Ιστό που παρέχει έναν μηχανισμό για την ενσωμάτωση πολλαπλών συστημάτων μεταδεδομένων. Στο RDF ένα namespace ορίζεται από μια διεύθυνση URL που παραπέμπει σε έναν πόρο Ιστού που περιγράφει το σχήμα μεταδεδομένων που χρησιμοποιείται στην περιγραφή. Πολλά namespaces μπορούν να οριστούν, επιτρέποντας σε στοιχεία από διαφορετικά συστήματα να συνδυαστούν σε μια ενιαία περιγραφή των πόρων. Πολλές περιγραφές, δημιουργημένες σε διαφορετικές χρονικές στιγμές για διαφορετικούς σκοπούς, μπορούν επίσης να συνδέονται μεταξύ τους. Το RDF εκφράζεται γενικά σε XML.

Η διαλειτουργικότητα και η ανταλλαγή μεταδεδομένων διευκολύνεται περαιτέρω από διαβάσεις μεταδεδομένων. Η διάβαση είναι μια χαρτογράφηση των στοιχείων, της σημασιολογίας, και της σύνταξης από το ένα σχήμα μεταδεδομένων σε αυτά ενός άλλου.

Η διάβαση επιτρέπει στα μεταδεδομένα που δημιουργούνται από μια κοινότητα να χρησιμοποιηθούν από άλλη ομάδα που χρησιμοποιεί ένα διαφορετικό πρότυπο μεταδεδομένων. Ο βαθμός στον οποίο οι διαβάσεις είναι επιτυχημένες εξαρτάται από την ομοιότητα των δύο συστημάτων, την διακριτικότητα των στοιχείων στο σύστημα-στόχο σε σύγκριση με εκείνη στο σχήμα-πηγή, καθώς και την συμβατότητα των κανόνων περιεχομένου για την συμπλήρωση των στοιχείων κάθε συστήματος.

Οι διαβάσεις είναι σημαντικές για εικονικές συλλογές, όπου οι πόροι αντλούνται από διάφορες πηγές και αναμένονται να λειτουργήσουν ως σύνολο, ίσως με την εφαρμογή μίας μόνο μηχανής αναζήτησης. Αν και οι διαβάσεις είναι το κλειδί, χρειάζονται πολύ εργατικό δυναμικό για να αναπτυχθούν και να διατηρηθούν. Η χαρτογράφηση των συστημάτων με λιγότερα στοιχεία (λιγότερη λεπτομέρεια) σε σχέση με εκείνα με περισσότερα στοιχεία (πιο λεπτομερή) είναι προβληματική.

2.4 Search engines (Google, yahoo, bing)

Ο τεράστιος όγκος πληροφοριών στο διαδίκτυο σε συνδυασμό με την άναρχη δομή και οργάνωσή του οδήγησαν πολύ γρήγορα στην ανάγκη δημιουργίας ενός μηχανισμού αναζήτησης ιστοσελίδων με θέματα που ενδιαφέρουν τους χρήστες κάθε φορά. Καθώς το **World Wide Web** εξελίσσεται στην πιο γρήγορα αναπτυσσόμενη υπηρεσία του διαδικτύου, όπου η κατασκευή και δημοσίευση ιστοσελίδων είναι μια εύκολη διαδικασία, οι μηχανές αναζήτησης πληροφοριών στις εκατοντάδες ιστοσελίδες είναι απαραίτητες.

Οι μηχανές αναζήτησης είναι προγράμματα που επιτρέπουν την αναζήτηση χρησιμοποιώντας λέξεις-κλειδιά σε τεράστιες βάσεις δεδομένων. Αυτές περιέχουν αντίγραφα εκατομμυρίων ιστοσελίδων του World Wide Web που συλλέγονται αυτόματα από ειδικά προγράμματα, τα οποία μπορεί να έχουν διάφορες ονομασίες (spider, crawler, robot κλπ.), αλλά εκτελούν ουσιαστικά την ίδια εργασία. Μία μηχανή αναζήτησης περιλαμβάνει τα εξής βασικά μέρη:

Τον Spider (ή Crawler ή Robot κλπ.): ένα ειδικό αυτόματο πρόγραμμα που επισκέπτεται ιστοσελίδες, τις «διαβάζει» και έπειτα ακολουθεί τις **υπερσυνδέσεις (hyperlinks)** των ιστοσελίδων αυτών προς άλλες ιστοσελίδες. Κατά καιρούς (π.χ. μία φορά το μήνα) ο spider επιστρέφει στις ιστοσελίδες που έχει ήδη επισκεφτεί ψάχνοντας για αλλαγές.

Το Ευρετήριο (Index): μία τεράστια βάση δεδομένων που περιέχει αντίγραφα όλων των ιστοσελίδων που επισκέφτηκε ο spider. Όταν ο spider ανακαλύψει αλλαγές σε κάποιες ιστοσελίδες, τότε ενημερώνονται και τα αντίγραφα στο ευρετήριο. Βέβαια, το τι ακριβώς αντιγράφει στο ευρετήριο ο spider εξαρτάται από κάθε μηχανή αναζήτησης. Οι περισσότερες αξιόλογες μηχανές διαθέτουν το πλήρες κείμενο των ιστοσελίδων στο ευρετήριό τους, υπάρχουν όμως και κάποιες που ευρετηριάζουν μόνο τον τίτλο μιας ιστοσελίδας και τις πρώτες γραμμές κειμένου.

Τον μηχανισμό αναζήτησης: το πρόγραμμα που ερευνάει το ευρετήριο για να βρει ιστοσελίδες που ταιριάζουν στις λέξεις-κλειδιά της αναζήτησης που έθεσε ο χρήστης. Συνήθως ιεραρχεί τα αποτελέσματα της αναζήτησης με βάση κάποια κριτήρια.

Οι μηχανές αναζήτησης έχουν δικές τους ιστοσελίδες στο διαδίκτυο. Χρηματοδοτούνται με διαφημίσεις και έτσι προσφέρουν τις υπηρεσίες τους στους χρήστες δωρεάν. Ο χρήστης μπορεί να πληκτρολογεί τις λέξεις-κλειδιά προς αναζήτηση σε ειδικά πλαίσια και η μηχανή αναζήτησης επιστρέφει τα αποτελέσματα: τους τίτλους ιστοσελίδων, συνήθως με ένα μικρό απόσπασμα του κειμένου της ιστοσελίδας (ή σπανιότερα μία μικρή περιγραφή), καθώς και με μία υπερσύνδεση που οδηγεί σε αυτήν.

Οι περισσότερες μηχανές αναζήτησης προσφέρουν επίσης καταλόγους ιστοσελίδων οργανωμένους σε θεματικές ενότητες, στους οποίους ο χρήστης μπορεί να πλοηγηθεί αναζητώντας κάτι που τον ενδιαφέρει. Επίσης, πολλές μηχανές αναζήτησης προσφέρουν μία σειρά από άλλες υπηρεσίες, όπως δωρεάν e-mail, chat, ειδήσεις, χρηματιστήριο, καιρός κλπ., οι οποίες ουσιαστικά δεν έχουν καμία σχέση με την λειτουργία των μηχανών αναζήτησης, αλλά σκοπεύουν στο να προσελκύουν επισκέπτες στις σελίδες τους.

Οι μηχανές αναζήτησης διαχωρίζονται σε απλές μηχανές αναζήτησης και μεταμηχανές αναζήτησης. Σε αντίθεση με τις απλές μηχανές αναζήτησης, οι οποίες χρησιμοποιούν έναν spider για να συγκεντρώσουν μια δική τους βάση δεδομένων, οι μεταμηχανές δεν διαθέτουν δικό τους ευρετήριο, αλλά αντλούν τα αποτελέσματά τους από τα ευρετήρια άλλων μηχανών αναζήτησης. **Οι μεταμηχανές** στέλνουν τις λέξεις-κλειδιά ταυτόχρονα σε μία σειρά από προκαθορισμένες μηχανές αναζήτησης (ή και θεματικούς καταλόγους) και παρουσιάζουν ένα μέρος από τα αποτελέσματα της κάθε μίας. Καθώς ο μηχανισμός των μεταμηχανών παραμένει λίγο χρόνο σε κάθε βάση δεδομένων ανακτά συχνά μόνο το 10% των αποτελεσμάτων από κάθε βάση.

Οι μηχανές αναζήτησης έχουν τη δυνατότητα να ιεραρχούν τα αποτελέσματα με βάση τη **συνάφεια**, το ποσοστό δηλαδή που δείχνει πόσο σχετικό είναι το περιεχόμενο μιας ιστοσελίδας με τις λέξεις-κλειδιά της αναζήτησης, ακολουθώντας μία σειρά από κανόνες, γνωστούς ως **αλγόριθμους**. Οι δύο κυριότεροι κανόνες, που ακολουθούνται από όλες τις μηχανές αναζήτησης για την ιεράρχηση των αποτελεσμάτων, αφορούν στην **τοποθεσία** και την **συχνότητα** των λέξεων-κλειδιών μέσα σε μία ιστοσελίδα. Προκειμένου να αντιμετωπίσουν αυτή την πρακτική οι μηχανές αναζήτησης βελτιώνουν συνεχώς τις τεχνικές προσδιορισμού συνάφειας χρησιμοποιώντας μία σειρά από επιπλέον κριτήρια για την ιεράρχηση των αποτελεσμάτων. Ένα από αυτά είναι η **ανάλυση υπερσυνδέσεων**. Αναλύοντας πως οι ιστοσελίδες συνδέονται μεταξύ τους, η μηχανή αναζήτησης μπορεί να προσδιορίσει το θέμα μιας σελίδας και πόσο σημαντική θεωρείται. Ένα δεύτερο κριτήριο είναι η **δημοτικότητα** μιας ιστοσελίδας, δηλαδή πόσες επισκέψεις δέχεται μία ιστοσελίδα για μία συγκεκριμένη αναζήτηση.

Πρέπει να ληφθεί υπόψη το γεγονός ότι οι μηχανές αναζήτησης δεν καλύπτουν το σύνολο των ιστοσελίδων του World Wide Web. Σύμφωνα με μελέτες ακόμη και η μεγαλύτερη μηχανή αναζήτησης καλύπτει μόλις το ένα τρίτο του World Wide Web. Βέβαια, το σύνολο των μηχανών αναζήτησης καλύπτει ένα μεγαλύτερο ποσοστό, καθώς κάποιες μηχανές έχουν στο ευρετήριό τους ιστοσελίδες που δεν υπάρχουν σε άλλες. Για το λόγο αυτό και επειδή κάθε μηχανή έχει το δικό της τρόπο ευρετηρίασης των ιστοσελίδων και ιεράρχησης των αποτελεσμάτων συνίσταται στον χρήστη η πραγματοποίηση της ίδιας αναζήτησης σε περισσότερες μηχανές. Επίσης, είναι σημαντικό ο χρήστης να γνωρίζει ποια υπηρεσία αναζήτησης να χρησιμοποιήσει για τον σκοπό του. Υπάρχουν πολλές ειδικές μηχανές αναζήτησης και θεματικοί κατάλογοι που εξειδικεύονται σε συγκεκριμένους τομείς (e-mail, newsgroups, mailing lists, επιχειρήσεις, υγεία κλπ.) και μπορεί να είναι πιο χρήσιμοι από οποιαδήποτε άλλη μηχανή ανεξάρτητα από το πόσο γνωστή ή μεγάλη είναι αυτή. Ορισμένες μηχανές αναζήτησης έχουν αναπτύξει μία τεχνική που ονομάζεται ομαδοποίηση αποτελεσμάτων (*results grouping*), όπου όλα τα αποτελέσματα από έναν τόπο συνενώνονται. Άλλες μηχανές επιτρέπουν την αναζήτηση μέσα στα αποτελέσματα της πρώτης έρευνας, δίνοντας έτσι τη δυνατότητα στον χρήστη να τα περιορίσει. Ένας άλλος τρόπος περιορισμού των αποτελεσμάτων είναι η χρήση της σύνθετης αναζήτησης. Καθοριστικής σημασίας, όχι μόνο για να περιορίζεται ο αριθμός των αποτελεσμάτων, αλλά και για να είναι τα αποτελέσματα αυτά πιο σωστά, είναι η επιλογή των σωστών λέξεων-κλειδιών που χρησιμοποιούνται για την αναζήτηση. Το σημαντικότερο είναι ο χρήστης να αναλογιστεί τι ακριβώς ψάχνει και να επιλέξει την πιο κατάλληλη λέξη που να το εκφράζει.

Ο αριθμός των μηχανών αναζήτησης αυξάνεται συνεχώς, με τις μηχανές αναζήτησης να ποικίλουν σε μέγεθος, ποιότητα και δυνατότητες. Έτσι, οι κατασκευαστές τους προσπαθούν να δημιουργήσουν στοιχεία που θα τις κάνουν ανταγωνιστικές και ελκυστικότερες για τον ερευνητή. Μπορούν, εναλλακτικά, να προσθέσουν μια βάση δεδομένων με κατηγοριοποίηση θεμάτων μαζί με ιστοσελίδες, πληροφορίες για εταιρείες, διαφημίσεις, προτροπές προς τους αναγνώστες να ψηφίσουν για τις σελίδες που βλέπουν, να μάθουν πληροφορίες για τον αριθμό των ερευνητών που έχουν επισκεφθεί την συγκεκριμένη σελίδα.

Υπάρχουν πολλές μηχανές αναζήτησης, μεταξύ των οποίων οι κυριότερες είναι αυτές της εταιρίας **Google (Google)** και της **Yahoo (Yahoo! Search)** που είναι πρώτες σε χρήση. Επίσης, σημαντική μηχανή αναζήτησης είναι και η **Bing της Microsoft**, η **MSN Search** και η **Microsoft Live Search**. Ελληνικές μηχανές αναζήτησης μεταξύ άλλων είναι αυτές της **Forthnet**, της **Otenet**, του **Cleverlook** και του **Pathfinder**.

Η **Google** (<http://www.google.gr/>) είναι μία από τις μεγαλύτερες εταιρείες διαδικτυακών υπηρεσιών που ξεκίνησε να λειτουργεί στις 27 Σεπτεμβρίου 1998. Στόχος της ήταν να οργανώσει όλες τις πληροφορίες του κόσμου και να τις κάνει παγκόσμια διαθέσιμες. Το Google ξεκίνησε σαν μια κολεγιακή εργασία από τον Larry Page και τον Sergey Brin το

1996 για μια μηχανή αναζήτησης. Σήμερα η μηχανή αναζήτησης google είναι μια από τις δημοφιλέστερες, και οι φράσεις «κάνω google», «γκουγκλάρω» και «γκουγκλίζω» είναι συνώνυμες με το «ψάχνω για πληροφορίες στο Διαδίκτυο». Αντίστοιχα, στην αγγλική γλώσσα το ρήμα "to google" έχει αποκτήσει πλέον ταυτόσημη έννοια με το ρήμα «αναζητώ». Η μηχανή Google ξεκίνησε τον Ιανουάριο του 1996, με μια καινοτομία: τη στιγμή που όλες οι συμβατικές μηχανές αναζήτησης έκαναν κατάταξη αποτελεσμάτων μετρώντας πόσες φορές οι αναζητούμενες λέξεις εμφανίζονται σε μια σελίδα, οι δύο τους, ανέπτυξαν θεωρητικά ένα σύστημα που θα αναλύει τις σχέσεις μεταξύ ιστοσελίδων. Ονόμασαν αυτή τη νέα τεχνολογία PageRank (κατάταξη σελίδας), σύμφωνα με την οποία η σημασία, η σχετικότητα και η ποιότητα μιας ιστοσελίδας εξαρτάται από τον αριθμό και την ποιότητα των άλλων ιστοσελίδων που οδηγούν πίσω σε αυτή (Backlinks). Αποθηκεύουν το σύνολο ή μέρος της ιστοσελίδας (αυτή η λειτουργία αναφέρεται ως cache), καθώς και πληροφορίες σχετικά με τις ιστοσελίδες. Το PageRank μιας ιστοσελίδας καθορίζεται κατ' επανάληψη και εξαρτάται από τον αριθμό και την τιμή του PageRank όλων των σελίδων που δείχνουν σε αυτήν. Μια σελίδα που συνδέεται με πολλές σελίδες με υψηλό PageRank, λαμβάνει η ίδια ένα υψηλό PageRank. Εάν δεν υπάρχουν σύνδεσμοι προς μια ιστοσελίδα δεν υπάρχει τιμή PageRank για αυτήν την σελίδα.

· Η **Yahoo** (<http://search.yahoo.com/>) είναι εταιρεία διαδικτυακών υπηρεσιών. Αν και ξεκίνησε ως θεματικός κατάλογος, αργότερα εξελίχτηκε και σε μια πανίσχυρη μηχανή αναζήτησης. Προσφέρει στους χρήστες του έναν μεγάλο αριθμό υπηρεσιών, που περιλαμβάνουν ηλεκτρονικό ταχυδρομείο (Yahoo! Mail), μηχανή αναζήτησης, ομάδες χρηστών, νέα, παιχνίδια, διαφημίσεις και ακόμα και ένα πρόγραμμα για άμεσα ηλεκτρονικά μηνύματα, το *Yahoo! Messenger*. Ιδρύθηκε τον Ιανουάριο του 1994 από τους τότε τελειόφοιτους του πανεπιστημίου Στάνφορντ, *David Filo* και *Jerry Yang* όταν θέλανε να ομαδοποιήσουν και να καταγράψουν διαδικτυακές τοποθεσίες μείζονος ενδιαφέροντος και να τις ταξινομήσουν σε θεματικές ενότητες, αρχικά ο θεματικός κατάλογος ήταν δημοσιευμένος στον δικτυακό τόπο του Στάνφορντ (akebono.stanford.edu/yahoo) και στις 2 Μαρτίου 1995 εισήχθη στο χρηματιστήριο.

· Το **Bing** (<http://www.bing.com/>) (στο παρελθόν Live Search, Windows Live Search, MSN Search) είναι η τρέχουσα πολυ-γλωσσική μηχανή αναζήτησης Ιστού (που διαφημίζεται ως "decision engine") της Microsoft. Παρουσιάστηκε από τον Steve Ballmer στις 28 Μαΐου 2009 στο All Things Digital συνέδριο στο Σαν Ντιέγκο. Οι σημαντικότερες αλλαγές περιλαμβάνουν τη λίστα των προτάσεων αναζήτησης σε πραγματικό χρόνο καθώς οι ερωτήσεις εισάγονται και μία λίστα σχετικών αναζητήσεων (αποκαλούμενη "Explorer pane" στη αριστερή πλευρά των αποτελεσμάτων αναζήτησης) βασισμένων στη σημασιολογική τεχνολογία από την Powerset, που η Microsoft αγόρασε το 2008. Η **Bing** που έχει ως στόχο να συμβάλει στη διαδικασία λήψης ταχύτερων και ορθότερων αποφάσεων, αξιοποιεί τα πλεονεκτήματα των σημερινών μηχανών αναζήτησης, εξασφαλίζει και επιπλέον οφέλη προσφέροντας μια νέα εμπειρία χρήσης, αλλά και ενισχυμένα εργαλεία που εστιάζουν σε τέσσερις κεντρικούς τομείς: την απόφαση αγοράς, τον προγραμματισμό ενός ταξιδιού, την έρευνα για μία ιατρική πάθηση και την αναζήτηση μίας τοπικής επιχείρησης. Η Bing περιλαμβάνει επίσης τη δυνατότητα Save & Share ιστορικών αναζήτησης μέσω των Windows Live SkyDrive, Facebook και ηλεκτρονικού ταχυδρομείου (email). Στις 29 Ιουλίου 2009, η Microsoft και το Yahoo! ανήγγειλαν μια συμφωνία στην οποία το Bing θα τροφοδοτούσε το Yahoo! Search.

«Στις μέρες μας, οι μηχανές αναζήτησης διευκολύνουν την περιήγηση στο διαδίκτυο και την εύρεση πληροφοριών, ωστόσο δεν είναι το ίδιο αποτελεσματικές όσον αφορά στη δυνατότητα αξιοποίησής τους», δήλωσε ο **Steve Ballmer, CEO της Microsoft**. «Όταν ξεκινήσαμε να «στήνουμε» το Bing, βασιστήκαμε καταρχάς στον τρόπο που ο κόσμος θέλει να χρησιμοποιεί το διαδίκτυο. Το Bing αποτελεί ένα σημαντικό πρώτο βήμα στην

προσπάθεια μας να προσφέρουμε πρωτοποριακές εμπειρίες αναζήτησης, ώστε ο κόσμος να μπορεί να βρίσκει πληροφορίες γρήγορα και να τις αξιοποιεί προκειμένου να προβαίνει στις ενέργειες που επιθυμεί, λαμβάνοντας τις σωστές αποφάσεις».

Η Microsoft είναι αφοσιωμένη στο στόχο της: τη δημιουργία καλύτερων εργαλείων που θα διευκολύνουν τη συνολική διαδικασία από το αρχικό ερώτημα αναζήτησης έως το σημείο της λήψης μίας απόφασης. Το Bing αποτελεί ένα σημαντικό πρώτο βήμα προς αυτό το όραμα και ισχυρή ένδειξη της δέσμευσης της Microsoft πως θα συνεχίσει να βελτιώνει την τεχνολογία αναζήτησης προς όφελος των πελατών της.

- Το Excite (<http://www.excite.com/>) είναι μια πύλη Διαδικτύου (Internet portal) και μηχανή αναζήτησης Παγκόσμιου Ιστού. Ξεκίνησε το 1994 ως Architext. Ήταν μία από τις σημαντικότερες "dotcom" πύλες (portals) της δεκαετίας του '90 (μαζί με τα Yahoo!, Lycos και Netscape) κι ένα από τα πιο αναγνωρισμένα εμπορικά σήματα στο διαδίκτυο. Σήμερα προσφέρει ποικίλες υπηρεσίες, συμπεριλαμβανομένης της αναζήτησης, του web ηλεκτρονικού ταχυδρομείου (webmail), του instant messaging, των αποσπασμάτων αποθεμάτων, και μιας διαμορφώσιμης από τον χρήστη αρχικής σελίδας (customizable user homepage).
- Το Ask.com (ή Ask Jeeves στο Ηνωμένο Βασίλειο) (<http://ask.com>) είναι μια μηχανή αναζήτησης που ξεκίνησε το 1996 από τον Garrett Gruener και τον David Warthen στο Μπέρκλεϋ της Καλιφόρνια. Το αρχικό λογισμικό της μηχανής αναζήτησης αναπτύχθηκε από τον Gary Chevskysky. Η Ask.com ανήκει στην IAC/InteActiveCorp.
- Η Wolfram Alpha (επίσης γράφεται ως Wolfram|Alpha) (www.wolframalpha.com) είναι μια μηχανή απάντησης (answer engine) που αναπτύσσεται από την Wolfram Research. Είναι μία υπηρεσία Διαδικτύου που απαντά στις πραγματικές ερωτήσεις άμεσα (online) με τον υπολογισμό της απάντησης από τα δομημένα δεδομένα, κι όχι με την παροχή μιας λίστας κειμένων ή ιστοσελίδων που να περιέχει την απάντηση, όπως μια μηχανή αναζήτησης. Ανακοινώθηκε τον Μάρτιο του 2009 από τον Stephen Wolfram, και κυκλοφόρησε στο κοινό στις 15 Μαΐου 2009.

Οι μηχανές αναζήτησης (search engines) δημιουργήθηκαν για να καλύψουν την ανάγκη κατηγοριοποίησης και ευρετηρίασης του τεράστιου αριθμού ιστοσελίδων. **Κατά τους Xie-Wang-Goh**, οι μηχανές αναζήτησης είναι εργαλεία αυτοματοποιημένης ανάκτησης που διευκολύνουν την ανεύρεση πληροφοριών στο Διαδίκτυο.

Όταν ο ερευνητής αποφασίσει να επιλέξει το Διαδίκτυο για την αναζήτηση των πληροφοριών που ψάχνει, πρέπει να σκεφτεί την κατάλληλη στρατηγική έρευνας (χρήση λέξεων, φράσεων, συνωνύμων που θα περιγράφουν το θέμα). Επομένως, απαιτείται να γνωρίζει τον τρόπο αναζήτησης και λειτουργίας των μεγαλύτερων μηχανών αναζήτησης, επισκεπτόμενος τις αντίστοιχες σελίδες τους. Η γνώση του περιεχομένου των μηχανών αναζήτησης θα τον βοηθήσει στην απόφαση για τη χρήση της καταλληλότερης για την έρευνά του.

Είναι πιθανό τα πρώτα αποτελέσματα της αναζήτησης να οδηγήσουν τον ερευνητή σε επαναπροσδιορισμό της στρατηγικής έρευνας με χρήση πιο συγκεκριμένων όρων. Για την ακρίβεια, στις μεγάλες μηχανές αναζήτησης με εκτενή ευρετήρια, αυξάνονται οι πιθανότητες να βρεθούν απαντήσεις που δεν σχετίζονται με το ερώτημα που έχει τεθεί. Καθώς καμιά μηχανή αναζήτησης δεν είναι πανομοιότυπη με κάποια άλλη, ενώ σε πολλές από αυτές υπάρχει ένα ποσοστό αλληλοκάλυψης, τα αποτελέσματα ποικίλουν. Η ίδια έρευνα μπορεί να επαναληφθεί σε άλλη μηχανή αναζήτησης για σύγκριση των αποτελεσμάτων ή να χρησιμοποιηθούν μεταμηχανές αναζήτησης, οι οποίες εκτελούν ταυτόχρονα μια έρευνα σε πολλές μηχανές αναζήτησης.

Τα δεδομένα σχετικά με τις ιστοσελίδες είναι αποθηκευμένα σε μια βάση δεδομένων με δείκτες για χρήση σε επόμενες αναζητήσεις. Ένα ερώτημα μπορεί να είναι μια απλή λέξη. Ο

σκοπός ενός δείκτη είναι να καταστήσει δυνατή την εύρεση της πληροφορίας το συντομότερο δυνατό. Ορισμένες μηχανές αναζήτησης, όπως το Google, αποθηκεύουν το σύνολο ή μέρος της ιστοσελίδας (αυτή η λειτουργία αναφέρεται ως cache), καθώς και πληροφορίες σχετικά με τις ιστοσελίδες, ενώ άλλες, όπως η AltaVista, αποθηκεύουν κάθε λέξη σε κάθε σελίδα που βρίσκουν. Αυτή η προσωρινά αποθηκευμένη σελίδα κρατά πάντα το πραγματικό κείμενο αναζήτησης, δεδομένου ότι είναι αυτό που πραγματικά οδήγησε στην εύρεσή του, έτσι, μπορεί να είναι πολύ χρήσιμη όταν το περιεχόμενο της τρέχουσας σελίδας ενημερωθεί και οι όροι αναζήτησης δεν είναι πλέον σε αυτό. Αυτό το πρόβλημα θα μπορούσε να θεωρηθεί ως μια ήπια μορφή linkrot (σύνδεσμοι που οδηγούν σε ιστοσελίδες, διακομιστές ή άλλους πόρους που δεν είναι πια διαθέσιμοι), και ο χειρισμός της Google αυξάνει τη χρηστικότητα ικανοποιώντας τις προσδοκίες των χρηστών, ότι οι όροι αναζήτησης θα υπάρχουν στην ιστοσελίδα που επιστρέφεται ως αποτέλεσμα. Η αυξημένη σχετικότητα αναζήτησης καθιστά αυτές τις προσωρινά αποθηκευμένες σελίδες πολύ χρήσιμες, ακόμα και πέρα από το γεγονός ότι μπορεί να περιέχουν δεδομένα που μπορεί πλέον να μην είναι διαθέσιμα αλλού.

Η χρησιμότητα μιας μηχανής αναζήτησης εξαρτάται από την σχετικότητα του συνόλου των αποτελεσμάτων που επιστρέφει. Ενώ μπορεί να υπάρχουν εκατομμύρια ιστοσελίδες που περιλαμβάνουν μια συγκεκριμένη λέξη ή φράση, μερικές σελίδες μπορεί να είναι πιο σχετικές, δημοφιλείς, ή έγκυρες από άλλες. Οι περισσότερες μηχανές αναζήτησης χρησιμοποιούν μεθόδους για την ταξινόμηση των αποτελεσμάτων για την εμφάνιση των "καλύτερων" αποτελεσμάτων πρώτων. Ο τρόπος με τον οποίο μια μηχανή αναζήτησης αποφασίζει ποιες σελίδες αντιστοιχούν καλύτερα στην αναζήτηση, και με ποια σειρά πρέπει να εμφανίζονται, διαφέρουν σημαντικά από τη μία μηχανή στην άλλη. Οι μέθοδοι επίσης αλλάζουν με την πάροδο του χρόνου καθώς αλλάζει η χρήση του Διαδικτύου και νέες τεχνικές εξελίσσονται.

Πολλές μηχανές αναζήτησης όπως το Google και το Bing παρέχουν εξατομικευμένα αποτελέσματα με βάση το ιστορικό δραστηριοτήτων του χρήστη. Αυτή η τεχνική ονομάζεται filter bubble. Ο όρος περιγράφει το φαινόμενο κατά το οποίο οι ιστοσελίδες χρησιμοποιούν αλγόριθμους για να μαντέψουν επιλεκτικά πληροφορίες που ο χρήστης θα ήθελε να δει, με βάση πληροφορίες για τον χρήστη (όπως η τοποθεσία του, το ιστορικό επισκέψεων σε σελίδες και το ιστορικό αναζήτησης). Ως αποτέλεσμα, οι ιστοσελίδες έχουν την τάση να εμφανίζουν μόνο τις πληροφορίες που συνάδουν με τις πληροφορίες για το χρήστη, απομονώνοντας ουσιαστικά το χρήστη σε μια φούσκα που τείνει να αποκλείει κάθε άλλη πληροφορία. Σύμφωνα με τον Eli Pariser, ο οποίος επινόησε τον όρο, οι χρήστες έχουν μικρότερη έκθεση σε αντικρουόμενες απόψεις και είναι απομονωμένοι πνευματικά στη δική τους πληροφοριακή φούσκα.

Σήμερα, οι μηχανές αναζήτησης είναι πολύπλοκα προγράμματα που μας επιτρέπουν να αναζητήσουμε οποιαδήποτε πληροφορία από αρχεία, κείμενα, εικόνες, βίντεο μέχρι λέξεις ή φράσεις που χρησιμοποιούμε στην καθημερινότητά μας.

2.5 Σύνδεση μεταδεδομένων με μηχανές αναζήτησης (metadata-seo)

Ένας αριθμός γλωσσών ερωτημάτων (**query languages**) έχει αναπτυχθεί για το πλαίσιο εργασίας περιγραφής πόρων και πιο γενικά για ημι-δομημένα δεδομένα. Ωστόσο, οι γλώσσες αυτές παρέχουν πολύ εκφραστικούς μηχανισμούς οι οποίοι έχουν σκοπό να καταστήσουν εύκολη τη διατύπωση πολύπλοκων ερωτημάτων. Η ύπαρξη των συγκεκριμένων μηχανισμών έκφρασης κάνει, ωστόσο, εύκολη την κατασκευή ερωτημάτων τα οποία απαιτούν την επεξεργασία ενός μεγάλου πλήθους υπολογιστικών πόρων. Κατά συνέπεια, αφού καμία σημαντική ιστοσελίδα δεν παρέχει μια διασύνδεση SQL στη σχεσιακή βάση δεδομένων, δεν περιμένει κανείς από τους δικτυακούς τόπους, ειδικά τους μεγάλους, να χρησιμοποιούν γλώσσες ερωτημάτων ως εξωτερική διασύνδεση στα δεδομένα τους. Αυτό το οποίο χρειάζεται είναι μια περισσότερο ελαφριά διασύνδεση που θα είναι πιο εύκολο να τυγχάνει υποστήριξης και το πιο σημαντικό, να παρουσιάζει προβλέψιμη συμπεριφορά. Αυτή η προβλέψιμη συμπεριφορά είναι σημαντική όχι μόνο για τον παροχέα υπηρεσιών, αλλά και για τον πελάτη των υπηρεσιών. Ένα τέτοιο ελαφρύ σύστημα ερωτημάτων είναι δυνατό να είναι συμπληρωματικό σε πιο πλήρεις γλώσσες ερωτημάτων που αναφέρθηκαν πιο πάνω και δεν αποκλείει συγκεκριμένους δικτυακούς τόπους από το να συναθροίζουν δεδομένα από πολλαπλές ιστοσελίδες και να παρέχουν πλουσιότερες διασυνδέσεις ερωτημάτων σε αυτές τις συναθροίσεις.

Η **GetData** προορίζεται να είναι μια απλή διασύνδεση ερωτημάτων σε δεδομένα τα οποία είναι προσβάσιμα στο δίκτυο. Η GetData δεν προορίζεται να είναι μια πλήρης ή εκφραστική γλώσσα ερωτημάτων όπως είναι η SQL, η Xquery, η RQL και η DQL. Σκοπός της είναι η εύκολη χρήση, η υποστήριξη και η κατασκευή της, τόσο από την προοπτική των προμηθευτών, όσο και των καταναλωτών των δεδομένων. Αυτό που επιθυμείται είναι η παροχή της δυνατότητας στις μηχανές να πραγματοποιούν ερωτήματα για δεδομένα σε απομακρυσμένους εξυπηρετητές. Η GetData, χτισμένη πάνω στο SOAP, επιτρέπει την πρόσβαση στις τιμές μιας ή περισσότερων ιδιοτήτων ενός πόρου από ένα γράφημα. Κάθε γράφημα που έχει τη δυνατότητα πραγματοποίησης ερωτημάτων έχει μια ενιαία θέση πόρου (URL – uniform resource locator), συσχετισμένη με αυτό. Κάθε ερώτημα τύπου GetData είναι ένα μήνυμα του SOAP το οποίο διευκρινίζει δύο ορίσματα, τον πόρο στις ιδιότητες του οποίου υπάρχει πρόσβαση, και τις ίδιες τις ιδιότητες στις οποίες υπάρχει πρόσβαση. Η απάντηση που επιστρέφεται σε ένα ερώτημα τύπου GetData είναι και αυτή ένα γράφημα το οποίο περιέχει τους πόρους (οι ιδιότητες του οποίου ερωτώνται), μαζί με τα τόξα τα οποία διευκρινίζονται από το ερώτημα και τους αντίστοιχους στόχους /πηγές τους. Οι διασυνδέσεις προγραμματισμού εφαρμογών κρύβουν τα μηνύματα του SOAP και την κωδικοποίηση της XML από τον προγραμματιστή.

Έτσι, όσον αφορά μια εφαρμογή χρησιμοποιώντας το σημασιολογικό ιστό, η διασύνδεση προγραμματισμού εφαρμογής (API – Application Programming Interface) έχει την ακόλουθη μορφή:

```
GetData(<resource>, <property>), => <value>
```

Καθένα από τα ερωτήματα τύπου GetData είναι ένα μήνυμα του SOAP που αντιστοιχεί στο γράφημα με τα δεδομένα. Εκτός από την διασύνδεση πυρήνα της GetData, υπάρχουν άλλες δύο διασυνδέσεις που παρέχονται από το TAP και βοηθούν στην εξερεύνηση των γραφημάτων. Η πρώτη από αυτές είναι η διασύνδεση της έρευνας η οποία παίρνει μια ακολουθία χαρακτήρων και επιστρέφει όλους τους πόρους, και η δεύτερη είναι η διασύνδεση ανάκλασης, η οποία είναι όμοια με την αντίστοιχη που χρησιμοποιείται στις γλώσσες αντικειμενοστραφούς προγραμματισμού και επιστρέφει μια λίστα από τόξα από και προς τους κόμβους. Αυτό είναι πού χρησιμεύει στην εξερεύνηση ενός γραφήματος μέσα στην εγγύτητα ενός κόμβου χωρίς γνώση του τι μπορεί να βρίσκεται γύρω από αυτόν.

2.6 Ορισμός-Σημασιολογικός Ιστός

Το **Semantic Web**, ή **Σημασιολογικός Ιστός** αναφέρεται σε ένα σύνολο τεχνολογιών και μεθόδων μέσω των οποίων οι υπολογιστές είναι σε θέση να αντιλαμβάνονται τη σημασία της πληροφορίας που διαχειρίζονται. Σύμφωνα με το όραμα των εμπνευστών του Σημασιολογικού Ιστού, η προσθήκη σημασίας στην πληροφορία του Διαδικτύου θα απελευθερώσει πλήθος δυνατοτήτων για την πιο ευφυή εκμετάλλευση της πληροφορίας αυτής. Ένας χρήστης του Διαδικτύου θα μπορεί, για παράδειγμα, μεταξύ άλλων δυνατοτήτων, να πραγματοποιεί ευφυείς αναζητήσεις, να λαμβάνει δηλαδή από μια μηχανή αναζήτησης αποτελέσματα τα οποία να είναι πιο σχετικά με αυτό που πραγματικά αναζητά.

Η ανάπτυξη του Παγκόσμιου Ιστού ή World Wide Web έγινε με στόχο να δημιουργηθεί ένα πλέγμα πληροφοριών για τους ανθρώπους. Πράγματι, οι σελίδες του Web απευθύνονται αποκλειστικά και μόνο στους ανθρώπους, με την έννοια ότι η πληροφορία που περιέχουν είναι αναγνώσιμη και κατανοητή μόνον από αυτούς.

Ο Ιστός «ανακαλύφθηκε» από τον **Tim Berners-Lee**, έναν φυσικό που εργαζόταν στο Κέντρο Πυρηνικών Μελετών και Ερευνών, CERN. Το αρχικό όραμα για τον Ιστό ήταν πιο φιλόδοξο από την πραγματικότητα του υπάρχοντος Ιστού. Οι μηχανές αναζήτησης προσπαθούν να αυτοματοποιήσουν, σε ένα βαθμό, τη διαδικασία αναζήτησης πληροφοριών στο Web. Όμως, και αυτές λειτουργούν με ορισμένες παραδοχές, όπως:

- Οι αναζητήσεις γίνονται από ανθρώπους.
- Τα αποτελέσματα των αναζητήσεων απευθύνονται σε ανθρώπους.
- Δεν υπάρχει τυποποίηση όσον αφορά τις έννοιες και τον τρόπο που αυτές εκφράζονται.

Για παράδειγμα, η λέξη «όχημα» μπορεί να έχει κυριολεκτική και μεταφορική σημασία (π.χ. όχημα για την εισαγωγή στο Χρηματιστήριο).

Αυτό που γίνεται αντιληπτό από τα παραπάνω είναι ότι 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. Αυτό έχει ως επακόλουθο κόστος σε χρόνο και κόπο, δυσχρηστία και περιορισμό του εξεταζόμενου μεγέθους εισερχομένων πληροφοριών στα ανθρώπινα μέτρα.

Για να αρθούν αυτοί οι περιορισμοί, το **W3C (World Wide Web Consortium)**, η κοινοπραξία που εξελίσσει στα πρότυπα του Web, στο πλαίσιο μιας συγκεκριμένης δράσης, αναπτύσσει τη λογική του **Σημασιολογικού Ιστού** ή **Semantic Web**. Ο Σημασιολογικός Ιστός θα επιτρέψει στην πληροφορία να είναι κατανοητή και από υπολογιστές, καθιστώντας τους ικανούς να αυτοματοποιήσουν σε μεγάλο βαθμό την επεξεργασία της πληροφορίας που υπάρχει στο Web. Έτσι, από μία σελίδα Web, οι άνθρωποι μπορούν να διαβάσουν την πληροφορία που περιέχει, αλλά και οι υπολογιστές να κατανοήσουν σημαντικό τμήμα του περιεχομένου της και να προβούν σε συγκεκριμένες ενέργειες που έχει ορίσει ο χρήστης.

Ο Σημασιολογικός Ιστός είναι μία προσπάθεια να γίνουν οι πόροι του Παγκόσμιου Ιστού προσιτοί σε αυτοματοποιημένες διεργασίες προσθέτοντας πληροφορίες που επεξηγούν ή ορίζουν το περιεχόμενό τους. Με αυτόν τον τρόπο, ο Σημασιολογικός Ιστός θα αποτελέσει ένα παγκόσμιο μέσο για ανταλλαγή πληροφοριών. Η λέξη «Σημασιολογία» έχει ρίζα τις ελληνικές λέξεις «σημάδι», «σημαίνω» και «σημαντικός» και σήμερα αναφέρεται στο νόημα συχνά σε επίπεδο γλώσσας. Πολλοί υποστηρίζουν ότι ο Σημασιολογικός Ιστός αποτελεί το μεγαλύτερο σε παγκόσμιο επίπεδο έργο έξυπνης ενσωμάτωσης συστημάτων ώστε να συνεργάζονται δια-λειτουργικά.

Ο εμπνευστής του Παγκόσμιου Ιστού υπήρξε δημιουργός και καθοδηγητής και του Σημαντικού Ιστού, ορίζοντας το Σημαντικό Ιστό ως εξής:

«Ο **Σημασιολογικός** Ιστός δεν είναι ξεχωριστός Ιστός, αλλά μία επέκταση του Παγκόσμιου Ιστού όπου η πληροφορία έχει καλά καθορισμένο νόημα, καθιστώντας τη συνεργασία μεταξύ ανθρώπων και υπολογιστών πιο αποτελεσματική.»

Όπως, αναφέρθηκε και παραπάνω στον ορισμό του Tim-Berners Lee, ο Σημασιολογικός Ιστός δεν αποτελεί κάτι εντελώς νέο, αλλά μια επέκταση του υπάρχοντος διαδικτύου, στην οποία η πληροφορία που παρουσιάζεται στο χρήστη ορίζεται με τέτοιο τρόπο ώστε να είναι κατανοητή όχι μόνο από τους ανθρώπους αλλά και από τα προγράμματα-πράκτορες, ενισχύοντας έτσι τη δια-λειτουργικότητα της επεξεργασίας των πληροφοριών μεταξύ των πρακτόρων αλλά και διευκολύνοντας τη λειτουργικότητα της χρήσης του διαδικτύου από τους ανθρώπους με τη βοήθεια των πρακτόρων. Ο Σημασιολογικός Ιστός βασίζεται στην ιδέα της οργάνωσης και διασύνδεσης της πληροφορίας που υπάρχει στο διαδίκτυο, ώστε να μπορεί να χρησιμοποιηθεί πιο αποτελεσματικά για την ανακάλυψη, αυτοματοποίηση, ομαδοποίηση και επαναχρησιμοποίηση της από διαφορετικές μεταξύ τους διαδικτυακές εφαρμογές.

Ο Σημασιολογικός Ιστός φιλοδοξεί να παρέχει στο μέλλον την υποδομή που απαιτείται για τη δημιουργία και την αξιοποίηση του συνόλου των πληροφοριών από ένα πλήθος διαδικτυακών εφαρμογών που θα περιλαμβάνουν, εκτός από τις απλές ιστοσελίδες του διαδικτύου, τις εταιρικές βάσεις δεδομένων, τις διαδικτυακές ηλεκτρονικές υπηρεσίες, τους πράκτορες, ακόμη και τις οικιακές συσκευές. Με το νέο αυτό τρόπο οργάνωσης των δεδομένων, οι ευφυείς λογισμικοί πράκτορες που μετά από αίτηση κάποιου χρήστη αναζητούν πληροφορίες ή παρεχόμενες υπηρεσίες στο διαδίκτυο, θα έχουν τη δυνατότητα να φιλτράρουν καλύτερα τα δεδομένα που συλλέγουν. Έτσι αυτά θα ανταποκρίνονται πραγματικά στις ανάγκες του χρήστη και θα παρουσιάζονται με κατανοητή μορφή. Το όραμα του Σημασιολογικού Ιστού συνεχώς ενδυναμώνεται με τη γέννηση μιας νέας γενιάς τεχνολογιών και εργαλείων.

Συγκεκριμένα, δημιουργούνται νέες γλώσσες, όπως οι XML, RDF, OWL, που αναπαριστούν την πληροφορία σε μορφή εύκολα κατανοητή και επεξεργάσιμη από τους πράκτορες.

Στο σημείο αυτό θα πρέπει να διευκρινιστεί ότι ο Σημαντικός Ιστός δεν είναι ακόμα υπαρκτός, αλλά ότι υπάρχει μια εξαιρετικά σημαντική κινητικότητα ερευνητών και εταιριών οι οποίες αναπτύσσουν τις συγκεκριμένες τεχνολογίες καθώς και κάποιες (όχι ακόμα αρκετές) εφαρμογές οι οποίες βασίζονται πάνω στις τεχνολογίες αυτές. Παρόλα αυτά το όραμα του Σημαντικού Ιστού ίσως να μην είναι και τόσο μακρινό αν γίνει ένας παραλληλισμός με τη διάδοση του ίδιου του διαδικτύου. Το 1989, όταν δηλαδή το διαδίκτυο έκανε τα πρώτα βήματα για την παγκόσμια καθιέρωσή του, η ανακάλυψη και ανάκτηση πληροφοριών από απομακρυσμένα συστήματα ήταν κάτι που μόνο ένας ειδικός μπορούσε να κάνει. Το διαδίκτυο ως τεχνολογικό υπόβαθρο υπήρχε και θεωρητικά έδινε δυνατότητα πρόσβασης σε αυτά τα συστήματα. Υπήρχαν όμως σημαντικά προβλήματα ευχρηστίας εξαιτίας των διαφορετικών πρωτοκόλλων πρόσβασης, ακόμη και μέσα στο ίδιο σύστημα.

Με την ανάπτυξη του Παγκόσμιου Ιστού (*World Wide Web*) η διασύνδεση των διαφόρων πηγών πληροφοριών έγινε πολύ εύκολη υπόθεση, χωρίς να λαμβάνονται υπόψη ποια πρωτόκολλα χρησιμοποιούνται. Η ανάπτυξη των διασυνδέσεων ιστού είναι προσανατολισμένη προς τον άνθρωπο-χρήστη. Ακόμη και αν οι πληροφορίες αυτές αντλούνται από μια καλά οργανωμένη βάση δεδομένων, η δομή αυτή δεν είναι ευδιάκριτη για ένα λογισμικό πράκτορα που αναζητά στοιχεία, με αποτέλεσμα να μην είναι εύκολη η αυτοματοποίηση κάποιων χρονοβόρων και επίπονων εργασιών στο διαδίκτυο από προγράμματα-πράκτορες.

2.6.1 Ανάγκη για επέκταση του σημερινού Ιστού

Ο Παγκόσμιος Ιστός έχει αλλάξει κατά πολύ τον τρόπο με τον οποίο διαρθρώνεται πλέον η επικοινωνία μεταξύ των ανθρώπων και ειδικά τον τρόπο με τον οποίο η πληροφορία που υπάρχει διαθέσιμη ανά τον κόσμο, διαδίδεται και ανακτάται. Από την άλλη πλευρά, η αποθήκευση όγκου πληροφοριών σε βάσεις δεδομένων οδήγησε στην εμφάνιση του εξής προβλήματος: την καθιέρωση και διατήρηση της σημασιολογίας των δεδομένων που είναι αποθηκευμένα στις βάσεις δεδομένων. Αυτό το πρόβλημα σημασιολογίας δεδομένων παρέμενε ελέγξιμο εφόσον ελέγξιμες ήταν και οι αλλαγές που θα μπορούσαν να προκληθούν στη κάθε βάση, αφού οι πληροφορίες ήταν διαθέσιμες και μπορούσαν να αλλοιωθούν μόνο από συγκεκριμένο αριθμό ατόμων και συγκεκριμένο αριθμό εφαρμογών.

Με την εμφάνιση λοιπόν του Παγκόσμιου Ιστού το τοπίο αυτό άλλαξε. Αμέτρητοι πλέον χρήστες και εφαρμογές μπορούν και έχουν πρόσβαση στις βάσεις δεδομένων που είναι διαθέσιμες στον Ιστό. Υπό αυτές τις συνθήκες, η σημασιολογία κάθε πληροφορίας πρέπει να είναι διαθέσιμη στον κάθε χρήστη μαζί με την ίδια την πληροφορία. Όταν ως χρήστης εννοείται κάποιο φυσικό πρόσωπο, αυτό μπορεί να επιτευχθεί με την κατάλληλη επιλογή κάποιας σχηματικής παρουσίασης για σημασιολογικά δεδομένα. Όταν, όμως, πρόκειται για κάποια εφαρμογή που θα αποκτήσει πρόσβαση στη βάση, η σημασιολογία πρέπει να είναι δομημένη σε μορφή που θα είναι προσπελάσιμη και κατανοητή από τη μηχανή που θα αναλάβει την επεξεργασία της. Ως εκ τούτου, κρίθηκε απαραίτητη η επέκταση του σημερινού Ιστού, η οποία κατέληξε στη δημιουργία του Σημασιολογικού Ιστού, γνωστό και ως Semantic Web.

Ο Σημασιολογικός Ιστός, λοιπόν, αποτελεί μια επέκταση του σημερινού Ιστού, η οποία έχει ως σκοπό την αυτοματοποίηση των λειτουργιών και των εφαρμογών του διαδικτύου. Η αυτοματοποίηση αυτή μπορεί να επιτευχθεί εφόσον η γνώση και η πληροφορία που υπάρχει αποθηκευμένη και δημοσιευμένη αυτή την στιγμή στον σημερινό Παγκόσμιο Ιστό αποκτήσει τυπικό νόημα και σημασιολογία και δομηθεί με έναν τέτοιο τρόπο ώστε να γίνεται κατανοητή από τις μηχανές που την επεξεργάζονται. Υπάρχει, πλέον, παγκοσμίως η διάθεση συνεργασίας μεταξύ των, ανά τον κόσμο, χρηστών του ιστού με σκοπό να διασπείρουν πληροφορίες εμπλουτισμένες από τη γνώση όχι μόνο ενός ατόμου, αλλά πολλών. Αυτό οδήγησε στη δημιουργία των wikis, ιστοσελίδων οι οποίες μπορούν να εμπλουτιστούν και να επανεκδοθούν από τον καθένα. Συνδυάζοντας τις τεχνολογίες του Σημασιολογικού Ιστού και των wikis, έχουμε οδηγηθεί στη δημιουργία των Semantic wikis.

Σήμερα, επιπλέον, υπάρχει μια τάση από τον μεμονωμένο χρήστη να μετατοπίσει τη γνώση από συστήματα που βασίζονται στην επιφάνεια εργασίας σε συστήματα βασισμένα στον Σημασιολογικό Ιστό. Καθένας, λοιπόν, μπορεί να αποθηκεύσει κάποια πληροφορία ως έναν πόρο Σημασιολογικού Ιστού που χαρακτηρίζεται από μια μοναδική σειρά χαρακτήρων, το URI του. Η συσκευή στην οποία κάποιος αποθηκεύει τις πληροφορίες αυτές ονομάζεται Semantic Desktop.

Αν και ο Παγκόσμιος Ιστός είναι η μεγαλύτερη αποθήκη πληροφοριών που δημιουργήθηκε ποτέ, με τα περιεχόμενά του να επεκτείνονται σε διάφορες γλώσσες και πεδία γνώσεων, μακροπρόθεσμα, είναι εξαιρετικά δύσκολο να βγει νόημα από το περιεχόμενό του. Από το 2000 έως το 2006 (χονδρικά) ένας αριθμός τεχνικών καινοτομιών, το RDF το οποίο είναι για τα δεδομένα ό, τι η HTML για τα έγγραφα και η Web Ontology Language (OWL) η οποία μας επιτρέπει να εκφράσουμε το πώς συνδέονται μεταξύ τους οι πηγές δεδομένων, σε συνδυασμό με τις πιο ανοιχτές πρακτικές διαμοιρασμού πληροφοριών, μετακινούν τον Παγκόσμιο Ιστό προς αυτό που ονομάζουμε Σημασιολογικό Ιστό.

Παρόλο που η αναζήτηση μιας λέξης θα μας αποφέρει χιλιάδες άμεσα συσχετιζόμενες πληροφορίες, η ύπαρξη επίσης χιλιάδων πληροφοριών που δεν σχετίζονται άμεσα με το πεδίο

της αναζήτησης μας αλλά τυγχάνει να περιέχουν τη λέξη κλειδί που έχουμε χρησιμοποιήσει, δημιουργούν ένα φόρτο γνώσης, που ο μέσος χρήστης είναι συχνά ανήμπορος να διαχειριστεί αποτελεσματικά. Άλλωστε, πολλές φορές η ύπαρξη πολλών και άχρηστων πληροφοριών είναι πιο αναποτελεσματική από την ελλιπή πληροφόρηση. Και τελικά, ακόμη και η άμεσα συσχετιζόμενη με την αναζήτηση του χρήστη γνώση είναι διασκορπισμένη σε σελίδες του Ιστού οι οποίες δεν έχουν καμία αλληλεξάρτηση, κάτι το οποίο δυσχεραίνει την πλοήγηση του μέσου χρήστη στον ιστό.

Γίνεται, λοιπόν, αντιληπτό πως η αναπαράσταση της γνώσης στον ιστό από μια γλώσσα αντιληπτή από τον χρήστη αλλά όχι από την μηχανή που αναζητά και αποδίδει την πληροφορία σε αυτόν, οδηγεί σε δυσλειτουργίες. Και εδώ γεννιέται η ιδέα δημιουργίας του Σημασιολογικού Ιστού, στον οποίο όλες οι πληροφορίες θα είναι μεταφρασμένες σε μια γλώσσα εύκολα αντιληπτή από τις μηχανές. Ο Σημασιολογικός Ιστός δε θα αποτελεί έναν παγκόσμιο ιστό που θα λειτουργεί παράλληλα με τον ήδη υπάρχοντα ιστό, αλλά θα αποτελεί μια μετεξέλιξή του. Πρώτο στάδιο για την δημιουργία του έργου αυτού, η αναπαράσταση της γνώσης σε γλώσσα αντιληπτή από τη μηχανή.

Προκειμένου η γνώση και η πληροφορία να περιγραφεί με έναν τυπικό τρόπο, ο οποίος θα δηλώνει την σημασία της, πρέπει να χρησιμοποιήσουμε γλώσσες αναπαράστασης γνώσης. Προκειμένου όμως να χρησιμοποιήσουμε τέτοιες τεχνολογίες στο διαδίκτυο θα πρέπει να αναθεωρήσουμε και να τροποποιήσουμε κάποια από τα συστατικά τους. Πιο συγκεκριμένα, όπως είναι γνωστό, στον σημερινό Ιστό κατά ένα πολύ μεγάλο ποσοστό η πληροφορία δομείται με τη χρήση της γλώσσας XML. Έτσι, λοιπόν, αφενός στον Σημασιολογικό Ιστό πρέπει να περιγράψουμε τη γνώση με τη χρήση κάποιας γλώσσας αναπαράστασης γνώσης, αφετέρου η σύνταξη της γλώσσας που θα χρησιμοποιήσουμε θα πρέπει να βασίζεται στη γλώσσα XML. Για το λόγο αυτό η W3C, η οποία είναι ο οργανισμός που ασχολείται με την ανάπτυξη και προτυποποίηση τεχνολογιών για τον Παγκόσμιο Ιστό, έχει αναπτύξει δύο γλώσσες αναπαράστασης γνώσης. Οι γλώσσες αυτές είναι η RDF(S) και η OWL. Χρειαζόμαστε δύο γλώσσες για την αναπαράσταση γνώσης στον σημασιολογικό ιστό εξαιτίας της αρχιτεκτονικής του. Στην αρχιτεκτονική αυτή παρατηρούμε ότι ο Σημασιολογικός Ιστός αποτελείται από στρώματα (layers). Κάθε στρώμα υλοποιεί μια λειτουργικότητα (functionality), χρησιμοποιώντας και επεκτείνοντας τη λειτουργικότητα και τις τεχνολογίες που παρέχονται από τα χαμηλότερα στρώματα. Έτσι, λοιπόν, στα χαμηλά επίπεδα υλοποιούνται λειτουργίες οι οποίες είναι πολύ κοντά στον Παγκόσμιο Ιστό και στις μηχανές, όπως είναι οι τεχνολογίες που ασχολούνται με τον καθορισμό και την αναγνώριση των πόρων (URIs), ενώ καθώς ανεβαίνουμε στην ιεραρχία των επιπέδων συναντάμε επίπεδα τα οποία υλοποιούν λειτουργίες αναπαράστασης γνώσης, πολύπλοκου συλλογισμού που πλησιάζουν στην ανθρώπινη γνώση και σκέψη.

Η ύπαρξη του Σημασιολογικού Ιστού δίνει τη δυνατότητα τοποθέτησης ετικετών σε όλο το περιεχόμενο του Ιστού, περιγραφής κάθε πληροφορίας και απόδοση σημασιολογικού νοήματος στο περιεχόμενο του αντικειμένου. Έτσι, οι μηχανές αναζήτησης θα γίνουν πιο αποτελεσματικές απ' ό,τι είναι τώρα και οι χρήστες θα μπορούν να βρουν τις ακριβείς πληροφορίες που ψάχνουν. Οι οργανισμοί που παρέχουν διάφορες υπηρεσίες μπορούν να βάλουν χαρακτηριστικές ετικέτες σε αυτές τις υπηρεσίες. Χρησιμοποιώντας πράκτορες λογισμικού βασισμένους στον Ιστό, μπορούμε να βρούμε δυναμικά αυτές τις υπηρεσίες και να τις χρησιμοποιήσουμε προς όφελός μας ή σε συνεργασία με άλλες υπηρεσίες. Η πρόοδος προς την καλύτερη ολοκλήρωση των δεδομένων θα συντελεστεί μέσω της χρήσης της τεχνολογίας-κλειδί που έκανε τον Παγκόσμιο Ιστό τόσο επιτυχημένο: τον σύνδεσμο.

Η δύναμη του Ιστού σήμερα, συμπεριλαμβανομένης της δυνατότητας εύρεσης της σελίδας που ψάχνουμε, έγκειται στο γεγονός ότι τα έγγραφα τοποθετούνται στον Ιστό σε πρότυπη μορφή και έπειτα συνδέονται μεταξύ τους. Ο Σημασιολογικός Ιστός θα δώσει τη δυνατότητα καλύτερης ολοκλήρωσης των δεδομένων, επιτρέποντας στον καθένα ο οποίος

τοποθετεί μεμονωμένα στοιχεία δεδομένων στον Ιστό, να τα συνδέσει με άλλα δεδομένα, χρησιμοποιώντας πρότυπες μορφοποιήσεις. Το αποτέλεσμα της χρήσης των τεχνολογιών του Σημασιολογικού Ιστού για να δημοσιεύσουμε δομημένα δεδομένα στον Ιστό και να θέσουμε συνδέσμους μεταξύ δεδομένων από μια πηγή δεδομένων προς δεδομένα μέσα σε άλλες πηγές δεδομένων είναι τα Συνδεδεμένα Δεδομένα (Linked Data).

Ως αποτέλεσμα θα προκύψει ο Ιστός των Δεδομένων (Web of Data). Με τη διασύνδεση συνόλων δεδομένων, οι μηχανές θα μπορούν να διασχίζουν έναν ανεξάρτητο ιστό δομημένων πληροφοριών έτσι ώστε να συλλέξουν σημασιολογική γνώση από αυθαίρετες οντολογίες και τομείς. Το αποτέλεσμα θα είναι μια μαζική βάση γνώσης με ελεύθερη πρόσβαση η οποία θα αποτελέσει τα θεμέλια μιας νέας γενιάς εφαρμογών και υπηρεσιών.

2.6.2 Οι Στόχοι του Σημασιολογικού Ιστού

Ο σημασιολογικός ιστός (semantic web) αποτελεί μια επέκταση του ήδη υπάρχοντος ιστού, στον οποίο δίνεται στην πληροφορία μια έννοια καθορισμένη με σαφήνεια, επιτρέποντας τη βέλτιστη συνεργασία μεταξύ ανθρώπου και υπολογιστή.

Η βασική ιδέα είναι η ύπαρξη δεδομένων στον ιστό, ορισμένων και συνδεδεμένων με έναν τρόπο που να επιτρέπει την αποδοτική ανακάλυψη, αυτοματοποίηση, ενσωμάτωση και επαναχρησιμοποίησή τους μέσα σε διάφορες εφαρμογές. Πιο συγκεκριμένα, ο σημασιολογικός ιστός θα περιέχει πόρους, οι οποίοι θα αντιστοιχούν όχι μόνο σε αντικείμενα πολυμέσων (ιστοσελίδες, εικόνες, ακουστικά αποσπάσματα κτλ.), όπως ο παραδοσιακός ιστός, αλλά και σε αντικείμενα όπως είναι φυσικά πρόσωπα, τοποθεσίες, οργανισμοί και γεγονότα.

Επιπρόσθετα, ο σημασιολογικός ιστός δε θα περιέχει μόνο ένα είδος σχέσης (hyperlink) ανάμεσα στους πόρους, αλλά πολλά και διαφορετικά είδη σχέσεων ανάμεσα στα διάφορα είδη πόρων. Η γενική ιδέα θεωρεί ότι τα δεδομένα στον σημασιολογικό ιστό μοντελοποιούνται ως ένα κατευθυνόμενο γράφημα με σήμανση, όπου κάθε κόμβος αντιστοιχεί σε έναν πόρο και κάθε τόξο επισημαίνεται με έναν τύπο ιδιότητας (property type). Αν και υπάρχουν διάφορες προτάσεις για την αναπαράσταση των πόρων και των αμοιβαίων σχέσεών τους στον σημασιολογικό ιστό βασισμένες στην XML, ένα ιδιαίτερο σύστημα πρέπει να αναλάβει τη δέσμευση για ένα ή περισσότερα σχήματα και πρωτόκολλα ανταλλαγής των πληροφοριών αυτών. Το περιγραφόμενο σύστημα χρησιμοποιεί το πλαίσιο περιγραφής πόρων του W3C με το λεξιλόγιο σχημάτων που παρέχεται από το RDFS ως ένα μέσο για την περιγραφή των πόρων και των μεταξύ τους σχέσεων. Το SOAP, χρησιμοποιείται ως πρωτόκολλο για την ερώτηση και την ανταλλαγή αυτών των στιγμιαίων δεδομένων RDF μεταξύ μηχανών.

Εκτός από τον εμπλουτισμό της παραδοσιακής αναζήτησης με δεδομένα που προέρχονται από τον σημασιολογικό ιστό, θα ήταν επιθυμητή η δυνατότητα χρησιμοποίησης του σημασιολογικού ιστού για τη βελτίωση των αποτελεσμάτων της αναζήτησης κειμένου. Η αναζήτηση κειμένου θα πρέπει να μπορεί να εκμεταλλευτεί την κατανόηση της επιθυμητής, από την πλευρά του χρήστη, πληροφορίας. Είναι πιθανό να υπάρχουν αρκετοί διαφορετικοί τρόποι με τους οποίους τα δεδομένα από τον σημασιολογικό ιστό είναι δυνατό να χρησιμοποιηθούν προκειμένου να φιλτραριστούν τα αποτελέσματα της αναζήτησης κειμένου. Πιο κάτω περιγράφεται η προσπάθεια επίλυσης ενός συγκεκριμένου προβλήματος. Η μηχανή αναζήτησης του Google, παρουσιάζει περίπου 136,000 αποτελέσματα στην αναζήτηση για τον μουσικό “Yo-Yo Ma”. Μια χειροκίνητη ανάλυση των πρώτων 500 αποτελεσμάτων δείχνει ότι στο σύνολό τους αναφέρονται στο συγκεκριμένο μουσικό. Από την άλλη πλευρά, η αναζήτηση στην ίδια μηχανή αναζήτησης για το πρόσωπο “Eric Miller” παράγει περίπου

1,400,000 αποτελέσματα, από τα οποία τα 20 πρώτα έχουν να κάνουν με 16 διαφορετικά πρόσωπα που διαθέτουν το συγκεκριμένο όνομα. Το πιο πιθανό είναι ότι ο χρήστης επιθυμεί την εύρεση πληροφοριών για κάποιο συγκεκριμένο πρόσωπο με το όνομα αυτό. Δυστυχώς, δεν υπάρχει κάποιος εύκολος τρόπος για να διαβιβαστεί κάτι τέτοιο στο σύστημα.

Ο σκοπός είναι να επιτραπεί στη μηχανή αναζήτησης να κατανοήσει ότι διαφορετικές εκδοχές της ίδιας ακολουθίας χαρακτήρων προσδιορίζουν διαφορετικές έννοιες και να φιλτράρει, να αξιολογήσει και να προβάλει τα αποτελέσματα των εγγράφων που ανταποκρίνονται στον επιθυμητό προσδιορισμό. Η αρχική εστίαση γίνεται πάνω στα ερωτήματα αναζήτησης φυσικών προσώπων. Θεωρείται χρήσιμη η παροχή στο χρήστη ενός απλού μηχανισμού για την αναγνώριση του κατάλληλου προσδιορισμού. Σε κάποιες περιπτώσεις (π.χ. στην λέξη “Jaguar”), υπάρχει ένας μικρός αριθμός αρχικών προσδιορισμών (το ζώο, το αυτοκίνητο, ο αθλητικός σύλλογος), έτσι ώστε να μπορούμε να επιλέξουμε έναν και να απαριθμήσουμε τους υπόλοιπους προκειμένου να επιτραπεί στο χρήστη να κάνει μια επιλογή του επιθυμητού μεταξύ αυτών. Στην περίπτωση που υπάρχουν χιλιάδες εν δυνάμει προσδιορισμοί (όπως στην περίπτωση της αναζήτησης στο Google του ανθρώπου “Eric Miller”), η προσέγγιση της απαρίθμησης όλων των πιθανών προσδιορισμών ξεχωριστά από τα αποτελέσματα αναζήτησης δεν είναι δυνατή. Έτσι, πρέπει να τροποποιήσουμε την παρουσίαση των αποτελεσμάτων της αναζήτησης έτσι ώστε κάθε αποτέλεσμα να έχει μια επιπρόσθετη σύνδεση δίπλα του, με τη χρήση της οποίας ο χρήστης μπορεί να καθορίσει στη μηχανή αναζήτησης ότι αυτός είναι ο επιθυμητός προσδιορισμός.

Ο Σημασιολογικός Ιστός αποτελεί ένα σχετικά καινούριο όραμα για την κοινωνία της πληροφορίας του μέλλοντος, όπως το εμπνεύστηκε ο πατέρας του σύγχρονου παγκόσμιου Ιστού, Berners-Lee (1998). Ο ίδιος υποστήριξε πως η κατανόηση του περιεχομένου που είναι διαθέσιμο στον Παγκόσμιο Ιστό από τους ηλεκτρονικούς υπολογιστές μπορεί να προσφέρει μία επανάσταση από νέες δυνατότητες και πληροφορικές εφαρμογές (Berners-Lee 2001). Στη διαμόρφωση της κοινωνίας της πληροφορίας του μέλλοντος συμμετέχουν ενεργά οι σύγχρονοι διαχειριστές της, οι ειδικοί στο χώρο των επιστημών της πληροφορίας.

Ο Σημασιολογικός Ιστός αποτελεί πρωτοβουλία της Κοινοπραξίας του Παγκοσμίου Ιστού ή World Wide Web Consortium ή W3C (<http://www.w3.org>). Η ιδέα στην οποία βασίζεται αφορά την οργάνωση και τη διασύνδεση της διαθέσιμης πληροφορίας στο Διαδίκτυο ώστε να χρησιμοποιηθεί πιο αποτελεσματικά στην ανακάλυψη, στην αυτοματοποίηση, στην ομαδοποίηση και στην επαναχρησιμοποίησή της από διαφορετικές μεταξύ τους διαδικτυακές εφαρμογές. Δηλαδή στοχεύει στη μετεξέλιξη του σημερινού ιστού ώστε να μετατραπεί σε ένα περιβάλλον όπου οι πληροφορίες που υπάρχουν και διακινούνται να είναι επεξεργάσιμες από τους ηλεκτρονικούς υπολογιστές. Κινείται προς την κατεύθυνση της επέκτασης και της βελτίωσης της δόμησης της πληροφορίας για να είναι προσπελάσιμη από προγράμματα υπολογιστών με τελικό στόχο την αυτοματοποίηση πολλών λειτουργιών στον Παγκόσμιο Ιστό. Σε αυτό το πλαίσιο, τα σημερινά κείμενα στις σελίδες του Ιστού θα εμπλουτιστούν με δομημένα κείμενα και δεδομένα σε μορφή XML και RDF. Από αυτή την άποψη ο Σημασιολογικός Ιστός είναι ένα πρόβλημα αναπαράστασης γνώσης από και για τους υπολογιστές και απαιτεί ένα μηχανισμό που θα επιτρέπει την επεξεργασία αυτής της γνώσης. Αυτός ο μηχανισμός θα πρέπει να υποστηρίζει τη δυνατότητα λογικής επεξεργασίας των πληροφοριών με σκοπό την εξαγωγή συμπερασμάτων, τη δημιουργία νέας γνώσης, την υποστήριξη στη λήψη αποφάσεων και την αυτόματη εκτέλεση ενεργειών.

Ο Σημασιολογικός Ιστός αποτελεί σήμερα μια από τις πιο σημαντικές προσπάθειες για την αυτόματη ενοποίηση των συστημάτων και των εφαρμογών ώστε να συνεργάζονται διαλειτουργικά σε παγκόσμιο επίπεδο. Ο Tim Berners-Lee οραματιζόταν έναν ιστό δεδομένων αυτόματα επεξεργάσιμων από τις εφαρμογές βάσει του νοήματος και όχι της μορφής της πληροφορίας. Αυτό θα έχει σαν αποτέλεσμα από το απλό κείμενο που το καταλαβαίνει πλήρως μόνο ο άνθρωπος να παρατηρείται κίνηση προς τις ημιδομημένες

(semi-structured) και δομημένες (structured) πληροφορίες που μπορούν να γίνουν κατανοητές από τις διαδικτυακές εφαρμογές, όπως, για παράδειγμα, συμβαίνει στα συστήματα ηλεκτρονικού εμπορίου (e-commerce), ηλεκτρονικού επιχειρείν (e-business) και ηλεκτρονικής μάθησης (e-learning).

Η πραγματική δύναμη του Σημασιολογικού Ιστού θα αποκαλυφθεί όταν οι άνθρωποι δημιουργήσουν συστήματα, τα οποία θα συλλέγουν τα περιεχόμενα του Ιστού από διαφορετικές πηγές, θα επεξεργάζονται τις πληροφορίες και θα ανταλλάσσουν τα αποτελέσματα με άλλους ανθρώπους ή μηχανές. Με αυτόν τον τρόπο, η αποτελεσματικότητα του Σημασιολογικού Ιστού θα βελτιωθεί δραστικά αφού μεγαλύτερο ποσοστό περιεχομένου του Ιστού θα είναι «αναγνώσιμο» από τους υπολογιστές και περισσότερες αυτοματοποιημένες υπηρεσίες θα είναι διαθέσιμες. Δύο σημαντικές τεχνολογίες που στοχεύουν στην ανάπτυξη του Σημασιολογικού Ιστού είναι ήδη σε ισχύ: η XML (eXtensible Markup Language) και το RDF (Resource Description Framework).

Η χρήση των οντολογιών μπορεί να αποδειχθεί ιδιαίτερα κρίσιμη παράμετρος προς αυτή την κατεύθυνση. Μια οντολογία, όπως αναφέρθηκε, είναι μια αυστηρή περιγραφή των αντικειμένων-πόρων και των σχέσεων μεταξύ τους. Οι οντολογίες δίνουν τη δυνατότητα σε μια κοινότητα χρηστών να έχουν κοινή ονοματολογία και κοινή αντίληψη για τα αντικείμενα-πόρους τα οποία δηλώνουν ή χρησιμοποιούν. Επίσης, τους επιτρέπουν να περιγράψουν τον «κόσμο» που χρησιμοποιούν (που τους αφορά) με έναν τέτοιο τρόπο, ώστε οι υπολογιστές να μπορούν να επεξεργαστούν και να «συνδέσουν» τα δεδομένα από διαφορετικούς κόσμους. Η δυσκολία στις οντολογίες έγκειται στο γεγονός ότι οι κοινότητες των χρηστών θα πρέπει να συμφωνήσουν στην οντολογική περιγραφή του πεδίου ενδιαφέροντός τους.

Ο τρόπος με τον οποίο θα αναπτυχθεί ο Σημασιολογικός Ιστός προβλέπεται να είναι ανάλογος με αυτόν του σύγχρονου Παγκόσμιου Ιστού. Το πρώτο βήμα στη διαμόρφωση του Παγκόσμιου Ιστού του μέλλοντος είναι η δημιουργία νησίδων πληροφορίας οργανωμένης σημασιολογικά. Οι νησίδες αυτές σιγά-σιγά θα διασυνδεθούν μεταξύ τους προσφέροντας περισσότερες δυνατότητες για την ανάπτυξη προηγμένων εφαρμογών. Το επόμενο βήμα είναι η αξιοποίηση του πλέγματος πληροφοριών, που δημιουργείται, σύμφωνα με τις προοπτικές που παρέχει η σημασιολογική τους οργάνωση.

Το καθένα από τα στάδια εξέλιξης του Σημασιολογικού Ιστού χαρακτηρίζεται από συγκεκριμένες τεχνολογικές δυνατότητες. Η διάθεση μέσω του Παγκόσμιου Ιστού δεδομένων που έχουν οριστεί και διασυνδεθεί μεταξύ τους αποτελεί τη βασική υποδομή για ό, τι υπόσχεται ο μελλοντικός Παγκόσμιος Ιστός: αποτελεσματικότερη αναζήτηση δεδομένων και πληροφοριών, αυτοματοποίηση και ολοκλήρωση υπηρεσιών, αξιοποίηση της υπάρχουσας γνώσης σε ποικίλες εφαρμογές.

Σταδιακά, η χρήση του Σημασιολογικού Ιστού επεκτείνεται σε συσκευές (π.χ. τηλεοράσεις, κινητά τηλέφωνα κ.α.) που συνδέονται στο Διαδίκτυο. Καθώς οι συσκευές αυτές θα χρησιμοποιούν τεχνολογίες Σημασιολογικού Ιστού, θα περιγράφουν τις λειτουργίες τους και θα τις ανακοινώνουν στο Διαδίκτυο. Έτσι, θα προκύψουν νέες δυνατότητες και αποδοτικότερη συνεργασία μεταξύ των συσκευών αυτών και του Ιστού. Για παράδειγμα, στο μέλλον η τηλεόραση θα μας προτείνει προγράμματα σύμφωνα με κριτήρια που θέσαμε στον Σημασιολογικό Ιστό. Ένα σημαντικό χαρακτηριστικό των μελλοντικών συστημάτων διαχείρισης πελατειακών σχέσεων (CRM) θα είναι οι «συσκευές διεπαφής χρήστη», οι οποίες θα μπορούν να «αναγνωρίζουν» τον χρήστη και να «κατανοούν» τη διάθεση και τα συναισθήματά του. Στη συνέχεια, θα του προτείνουν εξατομικευμένες υπηρεσίες και προϊόντα.

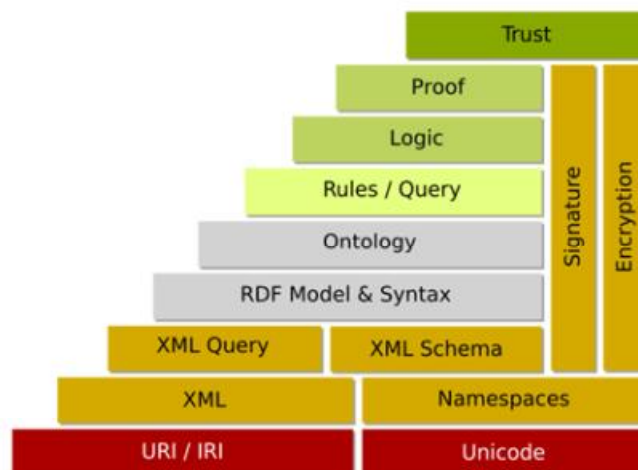
2.6.3 Αρχιτεκτονική Σημασιολογικού Ιστού

Ο Σημασιολογικός Ιστός έχει εξελιχθεί σε έναν αριθμό συγκεκριμένων τεχνολογιών που παρέχουν υποστήριξη για τις σημασιολογικές υπηρεσίες. Η Εικόνα 1 δείχνει την «Στοιβά του Σημασιολογικού Ιστού», η οποία δείχνει σαφώς τον κεντρικό ρόλο των τεχνολογιών του πυρήνα (π.χ. RDF, OWL) στην συνολική αρχιτεκτονική του ιστού.

Η στοιβά του Σημασιολογικού Ιστού είναι μια απεικόνιση της ιεραρχίας των γλωσσών, όπου κάθε στρώμα αξιοποιεί και χρησιμοποιεί τις δυνατότητες των πιο κάτω στρωμάτων. Δείχνει πώς οι τεχνολογίες που είναι τυποποιημένες για τον Σημασιολογικό Ιστό είναι οργανωμένες για να καταστήσουν την ύπαρξη του Σημασιολογικού Ιστού δυνατή. Δείχνει, επίσης, πώς Σημασιολογικός Ιστός είναι μια επέκταση (και όχι αντικατάσταση) του κλασικού Ιστού υπερκειμένων.

Το πρώτο στρώμα, URI και Unicode, ακολουθεί τα σημαντικά χαρακτηριστικά του υπάρχοντος Παγκόσμιου Ιστού. Το Unicode είναι ένα πρότυπο κωδικοποίησης διεθνών συνόλων χαρακτήρων και επιτρέπει σε όλες τις ανθρώπινες γλώσσες να μπορούν να χρησιμοποιηθούν (για γραφή και ανάγνωση) στο Διαδίκτυο, χρησιμοποιώντας μια τυποποιημένη μορφή. Το URI είναι μια ακολουθία σε τυποποιημένη μορφή που επιτρέπει τη μοναδική αναγνώριση των πόρων (π.χ. εγγράφων). Ένα υποσύνολο του URI είναι το Uniform Resource Locator (URL), το οποίο περιέχει μηχανισμό πρόσβασης και την (δικτυακή) τοποθεσία του εγγράφου. Η χρήση του URI είναι σημαντική για ένα κατακεμημένο σύστημα στο διαδίκτυο, καθώς παρέχει κατανοητή αναγνώριση όλων των πόρων.

Το στρώμα της Extensible Markup Language (XML) με τους ορισμούς των XML namespaces και XML Schema εξασφαλίζει ότι υπάρχει μια κοινή σύνταξη που χρησιμοποιείται στον σημασιολογικό ιστό. Η XML είναι μια γλώσσα σήμανσης γενικού σκοπού για έγγραφα που περιέχουν δομημένες πληροφορίες. Ένα έγγραφο XML περιέχει στοιχεία που μπορούν να εμφωλιαστούν και μπορούν να έχουν χαρακτηριστικά και περιεχόμενο. Τα XML namespaces επιτρέπουν τον καθορισμό διαφορετικών λεξιλογίων σήμανσης σε ένα έγγραφο XML. Το XML Schema εξυπηρετεί στην έκφραση σχήματος για ένα συγκεκριμένο σύνολο εγγράφων XML.



Εικόνα 1: Αρχιτεκτονική Semantic Web

<http://nemertes.lis.upatras.gr>

Μια μορφή αναπαράστασης δεδομένων για τον Σημασιολογικό Ιστό είναι το Resource Description Framework (RDF). Το RDF είναι ένα πλαίσιο αναπαράστασης πληροφοριών σχετικών με τους πόρους σε μορφή γραφήματος. Προοριζόταν πρωτίστως για την αναπαράσταση μεταδεδομένων για τους πόρους του Παγκόσμιου Ιστού, όπως ο τίτλος, ο

συγγραφέας και η ημερομηνία τροποποίησης μιας ιστοσελίδας, αλλά μπορεί να χρησιμοποιηθεί για την αποθήκευση οποιωνδήποτε άλλων δεδομένων. Βασίζεται σε τριάδες υποκείμενο-κατηγορημα-αντικείμενο που σχηματίζουν γραφήματα δεδομένων. Όλα τα δεδομένα στον σημασιολογικό ιστό χρησιμοποιούν το RDF ως την κύρια γλώσσα αναπαράστασης. Το κανονιστικό πρότυπο σύνταξης για σειριοποίησης RDF είναι η XML σε RDF / XML μορφή. Ορίζεται, επίσης, η τυπική σημασιολογία της RDF.

Το RDF εξυπηρετεί το ίδιο ως περιγραφή ενός γραφήματος που σχηματίζεται από τριάδες. Οποιοσδήποτε μπορεί να ορίσει το λεξιλόγιο των όρων που χρησιμοποιούνται για πιο λεπτομερή περιγραφή. Για να καταστεί δυνατή η τυποποιημένη περιγραφή των ταξινομήσεων και άλλων οντολογικών κατασκευασμάτων, δημιουργήθηκε ένα RDF Schema (RDFS) μαζί με την επίσημη σημασιολογία του μέσα στο RDF. Το RDFS μπορεί να χρησιμοποιηθεί για να περιγράψει ταξινομήσεις κλάσεων και ιδιοτήτων και να τις χρησιμοποιήσει για να δημιουργήσει ελαφριές οντολογίες.

Πιο λεπτομερείς οντολογίες μπορούν να δημιουργηθούν με την Web Ontology Language (OWL). Η OWL είναι μια γλώσσα που προέρχεται από λογικές περιγραφής, και προσφέρει περισσότερες κατασκευές πάνω στο RDFS. Είναι συντακτικά ενσωματωμένο στο RDF, οπότε όπως και το RDFS, παρέχει επιπλέον τυποποιημένο λεξιλόγιο. Η OWL έχει τρία είδη – την OWL Lite για ταξινομήσεις και απλούς περιορισμούς, την OWL DL για πλήρη υποστήριξη περιγραφικής λογικής, και την OWL Full για μέγιστη εκφραστικότητα και συντακτική ελευθερία του RDF. Δεδομένου ότι η OWL βασίζεται στη περιγραφική λογική, δεν είναι περίεργο που μια επίσημη σημασιολογία ορίζεται για αυτήν τη γλώσσα.

Για το RDFS και την OWL έχει οριστεί σημασιολογία και αυτή η σημασιολογία μπορεί να χρησιμοποιηθεί για τη λογική στις οντολογίες και τις βάσεις γνώσης που περιγράφονται χρησιμοποιώντας αυτές τις γλώσσες. Για να θεσπίζονται κανόνες πέρα από τις κατασκευές που διατίθενται από αυτές τις γλώσσες, τυποποιούνται και γλώσσες κανόνων για τον Σημασιολογικό Ιστό. Δύο πρότυπα αναδύονται, τα RIF και SWRL.

Για τις υποβολές ερωτημάτων (querying) σε δεδομένα RDF καθώς και οντολογίες RDFS και OWL με βάσεις γνώσης, είναι διαθέσιμη η SPARQL (Simple Protocol and RDF Query Language). Η SPARQL είναι μια γλώσσα σαν την SQL, αλλά χρησιμοποιεί τις τριάδες και τους πόρους του RDF, τόσο για το μέρος αντιστοίχισης του ερωτήματος, όσο και για την επιστροφή των αποτελεσμάτων του ερωτήματος. Δεδομένου ότι τόσο το RDFS όσο και η OWL είναι χτισμένα σε RDF, η SPARQL μπορεί, επίσης, να χρησιμοποιηθεί άμεσα για την υποβολή ερωτημάτων σε οντολογίες και βάσεις γνώσεων. Η SPARQL δεν είναι μόνο γλώσσα ερωτημάτων, είναι επίσης πρωτόκολλο για την πρόσβαση σε δεδομένα RDF.

Είναι αναμενόμενο ότι όλοι οι κανόνες και η σημασιολογία θα πρέπει να εκτελούνται στα κάτω στρώματα. Η επίσημη απόδειξη μαζί με τις έμπιστες εισόδους για την απόδειξη, θα σημαίνουν ότι τα αποτελέσματα είναι αξιόπιστα, κάτι το οποίο εμφανίζεται στο ανώτερο στρώμα της παραπάνω εικόνας. Για αξιόπιστες εισόδους, πρόκειται να χρησιμοποιηθούν κρυπτογραφικά μέσα, όπως ψηφιακές υπογραφές για την επαλήθευση της προέλευσης των πηγών. Στην κορυφή αυτών των στρωμάτων, μπορεί να κατασκευαστεί μια εφαρμογή με διεπαφή χρήστη.

2.7 Το όραμα του Σημασιολογικού Ιστού

Η λέξη “σημασιολογία” έχει ως ρίζα τις Ελληνικές λέξεις “σημάδι”, “σημαίνω”, “σημαντικός” και σήμερα αναφέρεται στην απόδοση νοήματος στα σύμβολα μιας γλώσσας. Ο Σημασιολογικός Ιστός (Semantic Web) αποτελεί πρωτοβουλία της Κοινοπραξίας του Παγκοσμίου Ιστού και είναι η μεγαλύτερη προσπάθεια αυτόματης ενοποίησης συστημάτων, ώστε αυτά να συνεργάζονται δια-λειτουργικά σε παγκόσμιο επίπεδο.

Ο Tim Berners-Lee, πρόεδρος του W3C, είχε το όραμα ενός Ιστού δεδομένων που η επεξεργασία τους θα γίνεται αυτόματα από τις εφαρμογές, βάσει του νοήματος και όχι της μορφής της πληροφορίας. Ο Ιστός αυτός βασίζεται στην ιδέα της οργάνωσης και διασύνδεσης της πληροφορίας που υπάρχει στο διαδίκτυο, ώστε να μπορεί να χρησιμοποιηθεί πιο αποτελεσματικά για την ανακάλυψη, αυτοματοποίηση, ομαδοποίηση και επαναχρησιμοποίησή της από διαφορετικές μεταξύ τους διαδικτυακές εφαρμογές.

Σημαντικό είναι να αναφερθεί ότι ο Σημασιολογικός Ιστός αποτελεί επέκταση και βελτίωση του υπάρχοντος Παγκόσμιου Ιστού, όπου η πληροφορία αποκτά καλά ορισμένο νόημα, κάνοντας έτσι αποτελεσματικότερη τη συνεργασία ανθρώπου-υπολογιστή, αφού πλέον θα υπάρχει μια κοινή γλώσσα επικοινωνίας ανάμεσά τους (Berners-Lee et al., 2001). Πιο συγκεκριμένα, δίνει τη δυνατότητα καλύτερης πρόσβασης σε μεγάλο όγκο πηγών πληροφορίας, καθώς και πιο αποτελεσματικής διακίνησης των πληροφοριών, χρησιμοποιώντας δεδομένα που τις περιγράφουν και ονομάζονται μεταδεδομένα.

Το όραμα του Σημασιολογικού Ιστού συνεχώς ενδυναμώνεται με τη γέννηση μιας νέας γενιάς τεχνολογιών και εργαλείων. Συγκεκριμένα, δημιουργούνται νέες γλώσσες, όπως οι XML, RDF, OWL, που αναπαριστούν την πληροφορία σε μορφή εύκολα κατανοητή και επεξεργάσιμη.

2.7.1 Παραδείγματα χρήσης semantic

Υπάρχουν σημαντικοί ιατρικοί οργανισμοί που ασχολούνται ερευνητικά με την ανάπτυξη ιατρικών συστημάτων βασισμένων στον σημασιολογικό ιστό.

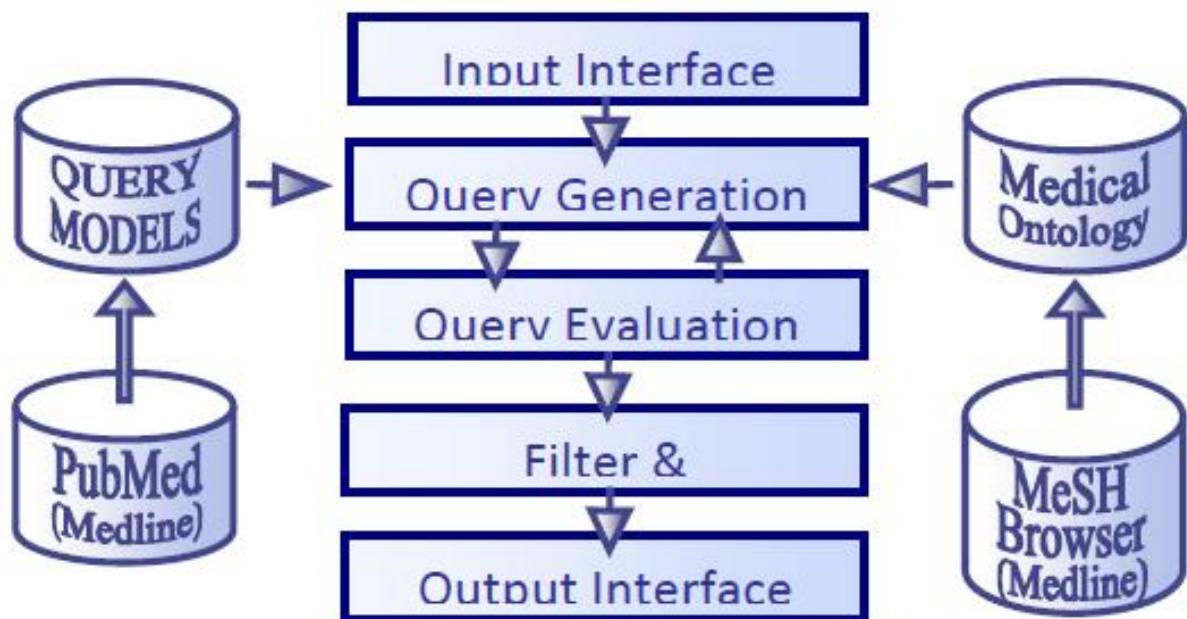
MELISA (Medical Literature Search Agent)

Ο πράκτορας MELISA υλοποιήθηκε στο Ινστιτούτο Έρευνας στην Τεχνητή Νοημοσύνη στην Ισπανία και αποτελεί ένα πρότυπο για την ανάκτηση πληροφοριών με βάση τις οντολογίες. Ο στόχος αυτού του έργου είναι η ανάπτυξη ενός συστήματος που θα δουλεύει πάνω σε ετερογενείς πηγές πληροφόρησης.

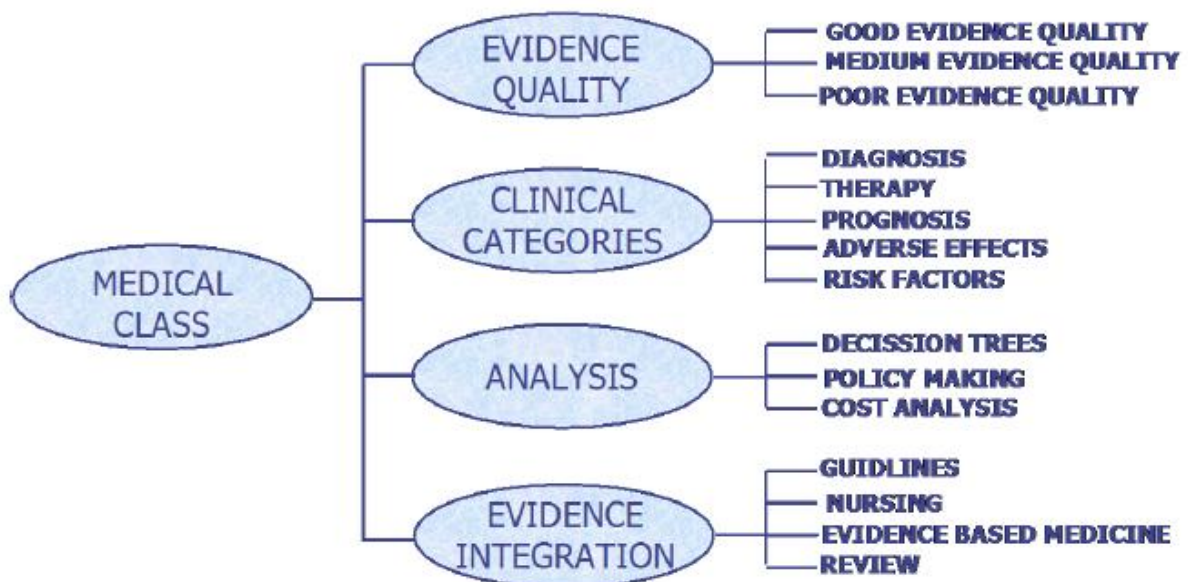
Ο πράκτορας επί του παρόντος είναι εξειδικευμένος στην ανάκτηση ιατρικού περιεχομένου πληροφορίας, όμως μπορεί να επεκταθεί εύκολα και σε άλλα πεδία. Ο MELISA χρησιμοποιεί την πληροφορία από τη βάση δεδομένων MEDLINE και την αντίστοιχη οντολογία MeSH, όμως είναι δυνατό μέσω της ιεραρχικής δομής του, με τα διάφορα επίπεδα αφαιρετικότητας, να χρησιμοποιήσει στο μέλλον διαφορετικές βάσεις δεδομένων και μηχανές αναζήτησης. Η οντολογία της MELISA είναι σχεδιασμένη από τη σκοπιά του Evidence Based Medicine και απευθύνεται κυρίως σε ειδικούς. Τα δύο πρώτα επίπεδα της ιεραρχικής οντολογίας παρουσιάζονται στην Εικόνα 2. Στην πράξη, οι όροι MeSH που διατυπώνονται ως keywords το τρίτο επίπεδο της ιεραρχίας και συνδυάζονται με τα άλλα δύο ώστε να επιστρέφονται τα σχετικά άρθρα με τις κατάλληλες ιατρικές κατηγορίες. Τέλος ο χρήστης είναι δυνατό να επιλέξει κατάλληλα φίλτρα για να φράξει την πληροφορία που αναζητά.

Το κύριο χαρακτηριστικό του MELISA είναι πως μέσω μίας αξιολόγησης των αποτελεσμάτων από τις διάφορες επερωτήσεις, είναι ικανός να επαναπροσδιορίσει την

ερώτηση να προτείνει εναλλακτικές διατυπώσεις μέσα από μια συλλογή από πρότυπων ερωτήσεων. Στην Εικόνα 3 παρουσιάζεται το σχεδιάγραμμα της δομής του πράκτορα MELISA και στη συνέχεια αναλύονται τα επιμέρους επίπεδα λειτουργικότητας του.



Εικόνα 2: Τα δύο πρώτα επίπεδα της οντολογίας που χρησιμοποιείται στον πράκτορα MELISA
<https://dspace.lib.uom.gr>



Εικόνα 3: Το σχεδιάγραμμα της δομής του πράκτορα MELISA
<https://dspace.lib.uom.gr>

Διεπαφή Χρήστη (Input Interface): Το επίπεδο αυτό επιτρέπει στο χρήστη να επιλέξει διάφορα κεντρικά θέματα και ιατρικές κατηγορίες προς αναζήτηση και να καθορίσει κάποιες πληροφορίες (κυρίως μεταδεδομένα) όπως ημερομηνία έκδοσης, ποιότητα των ιατρικών ενδείξεων κλπ. Ουσιαστικά το επίπεδο αυτό καθορίζει μία πολύ υψηλού επιπέδου επερώτηση (consultation level query) η οποία είναι ανεξάρτητη της βάσης δεδομένων.

Παραγωγή Επερώτησης και Επαναδιατύπωση (Query Generation & Reformulation): Το επίπεδο αυτό μετατρέπει την υψηλού επιπέδου επερώτηση σε μία συλλογή από χαμηλού επιπέδου εξαρτώμενες από τη βάση δεδομένων συγκεκριμένες επερωτήσεις. Επιπλέον, επαναδιατυπώνει τις επερωτήσεις που δεν απέδωσαν ικανοποιητικά αποτελέσματα. Η έξοδος από αυτό το επίπεδο αποτελεί ως επί το πλείστο μία συλλογή από βιβλιογραφικές παραπομπές στις εξειδικευμένες πλέον επερωτήσεις.

Αξιολόγηση Επερωτήσεων (Query Evaluation) Αυτό το επίπεδο βαθμολογεί τα αποτελέσματα από τις επερωτήσεις. Τα κριτήρια της αξιολόγησης είναι η συνέπεια ως προς τις απαιτήσεις του χρήστη, η ποιότητα των ενδείξεων που αναφέρονται σε αυτές τις παραπομπές κ.α. Τα αποτελέσματα των εξειδικευμένων ερωτήσεων ομαδοποιούνται σε θεματικές ενότητες.

Φιλτράρισμα και συνδυασμός (Filter and Combination): Η φύση των επερωτήσεων καθιστά αναπόφευκτη την επανάληψη των πόρων. Συνεπώς, τα αποτελέσματα συγκρίνονται και φιλτράρονται ώστε να απομακρυνθούν αφ' ενός αναφορές στους ίδιους πόρους και αφ' εταίρου αναφορές που δεν ικανοποιούν το κριτήριο ελάχιστης ικανοποίησης.

Διεπαφή εξόδου (Output Interface): Τέλος, τα αποτελέσματα συνδυάζονται διαδραστικά ώστε να παρουσιαστούν σύμφωνα με τις απαιτήσεις του χρήστη.

Μοντέλα Επερωτήσεων (Query models): αποτελούν σχήματα πληροφορίας που επιτρέπουν τη σύνταξη επερωτήσεων σε διάφορα επίπεδα αφαιρετικότητας, ξεκινώντας από επίπεδο consultation, έως το χαμηλότερο επίπεδο των εξειδικευμένων επερωτήσεων. Τα μοντέλα αυτά καθιστούν πράκτορα ανεξάρτητο από τις λέξεις των επερωτήσεων σε διακριτές μηχανές αναζήτησης όπως η PubMed. Η χρήση των προσδιοριστών αναζήτησης της PubMed επιτρέπει τη διενέργεια πολλαπλών επερωτήσεων πάνω στον ίδιο όρο της MeSH εξετάζοντας την αντίστοιχη οντολογία ώστε να ξεχωρίσει τους επιτρεπούς προσδιοριστές.

Στην Εικόνα 4 παρουσιάζεται η διεπαφή χρήστη του πράκτορα MELISA και ένα συγκεκριμένο παράδειγμα όπου ο χρήστης έχει επιλέξει την ακόλουθη πληροφορία.

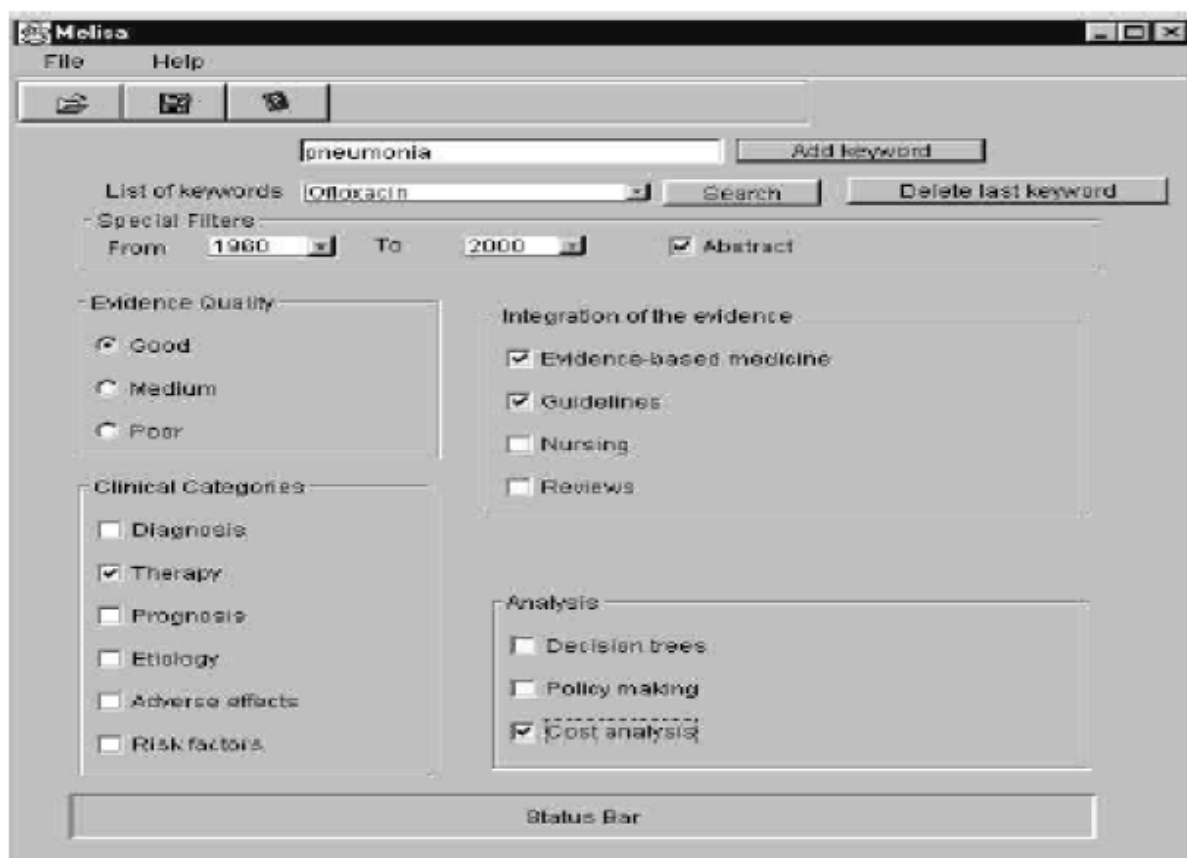
keywords
'Levofloxacin'
'Pneumonia'

Medical Categories:

'Good Evidence Quality'
'Therapy'
'Recommendations based on evidence'
'Guidelines'
'Cost analysis'

Special Filters

-Documents Accepted from 1960 to 2000
-Only retrieve articles with abstract



Εικόνα 4: Η διεπαφή χρήστη του πράκτορα MELISA
www.dmst.aueb.gr

2.7.2 Χαρακτηριστικά και επεξήγηση λειτουργίας

Η αξία και η σημασία του Παγκόσμιου Ιστού (World Wide Web ή WWW ή απλά Web) στην σχέση του ανθρώπου με τον υπολογιστή είναι πάρα πολύ μεγάλη. Παρόλα αυτά, το WWW έχει ένα βασικό μειονέκτημα: βασίζεται στην συντακτική περιγραφή του περιεχομένου (content), κύριος στόχος της οποίας είναι η καταληπτότητά του από τον άνθρωπο. Το χαρακτηριστικό αυτό του WWW δεν επιτρέπει την αυτόματη επεξεργασία του περιεχομένου. Τέτοια επεξεργασία προϋποθέτει μια καταλληλότερη περιγραφή του δημοσιευμένου περιεχομένου και αλγόριθμους που να προσδίδουν την επιθυμητή ευφυΐα στο υπολογιστικό περιβάλλον.

Τον παραπάνω περιορισμό του WWW, καθώς και τη λύση για την άρση του, συνέλαβε πρώτος ο ιδρυτής του, Tim Berners-Lee, ο οποίος όρισε ένα εξελικτικό στάδιο του WWW: το «Σημασιολογικό Ιστό» (Semantic Web). Ο Σημασιολογικός Ιστός έχει σαν στόχο να προσδώσει δομή στο νόημα του περιεχομένου των ιστοσελίδων, δημιουργώντας ένα περιβάλλον όπου οι πράκτορες λογισμικού (software agents) περιπλανώμενοι από σελίδα σε σελίδα θα μπορούν να εκτελούν προηγμένες εργασίες για τους χρήστες.

Ο Σημασιολογικός Ιστός (ΣΙ) αποτελεί τη βάση για μια πλήρως κατανεμημένη μορφή τεχνητής νοημοσύνης. Όπως είναι γνωστό, η τεχνητή νοημοσύνη ασχολείται κυρίως με δύο θέματα: την αναπαράσταση γνώσης (knowledge representation) και τις μεθόδους αναζήτησης και συμπερασμού (reasoning). Έτσι, για να λειτουργήσει ο Σημασιολογικός Ιστός θα πρέπει οι υπολογιστές να έχουν πρόσβαση σε δομημένες βάσεις γνώσης και σε κανόνες συμπερασμού τους οποίους να μπορούν να χρησιμοποιήσουν για να διενεργήσουν αυτόματη συλλογιστική (reasoning/inference).

Όσον αφορά στην αναπαράσταση γνώσης, το κύριο «εργαλείο» για την αναπαράσταση γνώσης στον ΣΙ είναι οι οντολογίες. Μάλιστα, η μεγαλύτερη συμβολή της πρωτοβουλίας του ΣΙ μέχρι σήμερα είναι η προτυποποίηση που παρείχε σε γλώσσες και τεχνολογίες ανάπτυξης και χρήσης οντολογιών. Πιο συγκεκριμένα, σήμερα η πιο διαδεδομένη γλώσσα για δημιουργία οντολογιών είναι η Web Ontology Language (OWL). Το συντακτικό της γλώσσας αυτής βασίζεται στην XML και στην RDF/RDF Schema (Resource Description Framework). Η εκφραστικότητα και η σημασιολογία της OWL καθορίζεται κυρίως από τις Περιγραφικές Λογικές (Description Logics). Τέλος, ο συμπερασμός γίνεται με τεχνολογίες κανόνων (σε συνδυασμό με οντολογίες) και με κυριότερη την Semantic Web Rule Language (SWRL). Πιο πολλές λεπτομέρειες για τα παραπάνω δίνονται στις επόμενες παραγράφους.

Οι οντολογίες αποτελούν μία σχετικά σύγχρονη μέθοδος αναπαράστασης γνώσης, αλλά και ερευνητικό πεδίο της τεχνητής νοημοσύνης γενικότερα. Ένας εύστοχος και περιεκτικός ορισμός της οντολογίας είναι ο εξής:

«Ontology is a formal, explicit specification of a shared conceptualization»

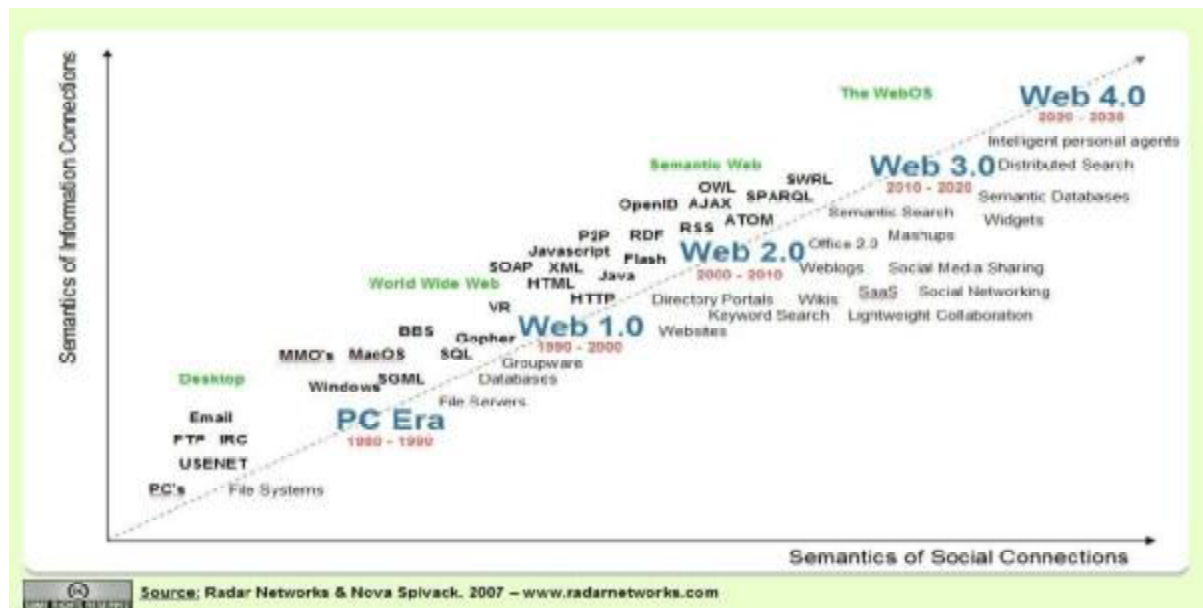
(R. Studer 1998, πρωτότυπος ορισμός από τον T. Gruber το 1993)

Ο ορισμός αυτός θα μπορούσε να αποδοθεί στα ελληνικά ακολούθως:

«Οντολογία είναι μια τυπική, ρητή προδιαγραφή μιας κοινής εννοιολογικής θεώρησης ενός φαινομένου».

Μια οντολογία (ontology) είναι η αυστηρά μαθηματική περιγραφή ενός πεδίου γνώσης και περιλαμβάνει ένα σύνολο από όρους και συσχετίσεις μεταξύ τους. Οι όροι περιγράφουν κλάσεις αντικείμενων, δηλαδή έννοιες σχετικές με αντικείμενα. Οι συσχετίσεις συνήθως αφορούν ιεραρχικές εξαρτήσεις μεταξύ των όρων. Άλλες πληροφορίες που μπορεί να υπάρχουν σε μία οντολογία είναι οι ιδιότητες των εννοιών, περιορισμοί γύρω από αυτές, σχέσεις ισοδυναμίας, καθώς και σημασιολογικοί συσχετισμοί μεταξύ των εννοιών με τη χρήση της λογικής. Η οντολογία που σχεδιάζουμε για ένα πεδίο ενδιαφέροντος (domain), χρησιμοποιείται τόσο για να υπάρξει ένα κοινά αποδεκτό λεξιλόγιο του πεδίου όσο και για να μπορούμε να εξάγουμε συμπεράσματα εκμεταλλευόμενοι τα στοιχεία μοντελοποίησης και τη σημασιολογία τους.

2.8 Το μέλλον του semantic



Εικόνα 5: Semantic Web

blog.law.cornell.edu

Ο σημασιολογικός ιστός (semantic web) αποτελεί μια επέκταση του ήδη υπάρχοντος ιστού, στον οποίο δίνεται στην πληροφορία μια έννοια καθορισμένη με σαφήνεια, επιτρέποντας τη βέλτιστη συνεργασία μεταξύ ανθρώπου και υπολογιστή. Η βασική ιδέα είναι η ύπαρξη δεδομένων στον ιστό, ορισμένων και συνδεδεμένων με έναν τρόπο που να επιτρέπει την αποδοτική ανακάλυψη, αυτοματοποίηση, ενσωμάτωση και επαναχρησιμοποίησή τους μέσα σε διάφορες εφαρμογές. Πιο συγκεκριμένα, ο σημασιολογικός ιστός θα περιέχει πόρους οι οποίοι θα αντιστοιχούν όχι μόνο σε αντικείμενα πολυμέσων (ιστοσελίδες, εικόνες, ακουστικά αποσπάσματα κτλ.), όπως ο παραδοσιακός ιστός, αλλά και σε αντικείμενα όπως είναι φυσικά πρόσωπα, τοποθεσίες, οργανισμοί και γεγονότα. Επιπρόσθετα, ο σημασιολογικός ιστός δε θα περιέχει μόνο ένα είδος σχέσης (hyperlink) ανάμεσα στους πόρους, αλλά πολλά και διαφορετικά είδη σχέσεων ανάμεσα στα διάφορα είδη πόρων. Η γενική ιδέα θεωρεί ότι τα δεδομένα στο σημασιολογικό ιστό μοντελοποιούνται ως ένα κατευθυνόμενο γράφημα με σήμανση, όπου κάθε κόμβος αντιστοιχεί σε έναν πόρο και κάθε τόξο επισημαίνεται με έναν τύπο ιδιότητας (property type). Αν και υπάρχουν διάφορες προτάσεις για την αναπαράσταση των πόρων και των αμοιβαίων σχέσεών τους στο σημασιολογικό ιστό βασισμένες στην XML, ένα ιδιαίτερο σύστημα πρέπει να αναλάβει τη δέσμευση για ένα ή περισσότερα σχήματα και πρωτόκολλα ανταλλαγής των πληροφοριών αυτών. Το περιγραφόμενο σύστημα χρησιμοποιεί το πλαίσιο περιγραφής πόρων του W3C με το λεξιλόγιο σχημάτων που παρέχεται από το RDFS ως ένα μέσο για την περιγραφή των πόρων και των μεταξύ τους σχέσεων. Το SOAP, χρησιμοποιείται ως πρωτόκολλο για την ερώτηση και την ανταλλαγή αυτών των στιγμιαίων δεδομένων RDF μεταξύ μηχανών.

Εκτός από τον εμπλουτισμό της παραδοσιακής αναζήτησης με δεδομένα που προέρχονται από το σημασιολογικό ιστό, θα ήταν επιθυμητή η δυνατότητα χρησιμοποίησης του σημασιολογικού ιστού για τη βελτίωση των αποτελεσμάτων της αναζήτησης κειμένου. Η αναζήτηση κειμένου θα πρέπει να μπορεί να εκμεταλλευτεί την κατανόηση της επιθυμητής, από την πλευρά του χρήστη, πληροφορίας. Είναι πιθανό να υπάρχουν αρκετοί διαφορετικοί τρόποι με τους οποίους τα δεδομένα από το σημασιολογικό ιστό είναι δυνατό να χρησιμοποιηθούν προκειμένου να φιλτραριστούν τα αποτελέσματα της αναζήτησης κειμένου.

Πιο κάτω περιγράφεται η προσπάθεια επίλυσης ενός συγκεκριμένου προβλήματος. Η μηχανή αναζήτησης του Google, παρουσιάζει περίπου 136,000 αποτελέσματα στην αναζήτηση για τον μουσικό “Yo-Yo Ma”. Μια χειροκίνητη ανάλυση των πρώτων 500 αποτελεσμάτων δείχνει ότι στο σύνολό τους αναφέρονται στο συγκεκριμένο μουσικό. Από την άλλη πλευρά, η αναζήτηση στην ίδια μηχανή αναζήτησης για το πρόσωπο “Eric Miller” παράγει περίπου 1,400,000 αποτελέσματα, από τα οποία τα 20 πρώτα έχουν να κάνουν με 16 διαφορετικά πρόσωπα που διαθέτουν το συγκεκριμένο όνομα. Το πιο πιθανό είναι ότι ο χρήστης επιθυμεί την εύρεση πληροφοριών για κάποιο συγκεκριμένο πρόσωπο με το όνομα αυτό. Δυστυχώς, δεν υπάρχει κάποιος εύκολος τρόπος για να διαβιβαστεί κάτι τέτοιο στο σύστημα.

Ο σκοπός είναι να επιτραπεί στην μηχανή αναζήτησης να κατανοήσει ότι διαφορετικές εκδοχές της ίδιας ακολουθίας χαρακτήρων προσδιορίζουν διαφορετικές έννοιες και περαιτέρω να φιλτράρει, να αξιολογήσει και να προβάλει τα αποτελέσματα των εγγράφων που ανταποκρίνονται στον επιθυμητό προσδιορισμό. Η αρχική εστίαση γίνεται πάνω στα ερωτήματα αναζήτησης φυσικών προσώπων. Θεωρείται χρήσιμη η παροχή στο χρήστη ενός απλού μηχανισμού για την αναγνώριση του κατάλληλου προσδιορισμού. Σε κάποιες περιπτώσεις (π.χ. στην λέξη “Jaguar”), υπάρχει ένας μικρός αριθμός αρχικών προσδιορισμών (το ζώο, το αυτοκίνητο, ο αθλητικός σύλλογος), έτσι ώστε να μπορούμε να επιλέξουμε έναν και να απαριθμήσουμε τους υπόλοιπους προκειμένου να επιτραπεί στο χρήστη να κάνει μια επιλογή του επιθυμητού μεταξύ αυτών. Στην περίπτωση που υπάρχουν χιλιάδες εν δυνάμει προσδιορισμοί (όπως στην περίπτωση της αναζήτησης στο Google του ανθρώπου “Eric Miller”), η προσέγγιση της απαρίθμησης όλων των πιθανών προσδιορισμών ξεχωριστά από τα αποτελέσματα αναζήτησης δεν είναι δυνατή. Έτσι, πρέπει να τροποποιήσουμε την παρουσίαση των αποτελεσμάτων της αναζήτησης έτσι ώστε κάθε αποτέλεσμα να έχει μια επιπρόσθετη σύνδεση δίπλα του, χρησιμοποιώντας την οποία ο χρήστης μπορεί να καθορίσει στην μηχανή αναζήτησης ότι αυτός είναι ο επιθυμητός προσδιορισμός.

Ο Σημασιολογικός Ιστός αποτελεί ένα σχετικά καινούριο όραμα για την κοινωνία της πληροφορίας του μέλλοντος, όπως το εμπνεύστηκε ο πατέρας του σύγχρονου παγκόσμιου Ιστού, Berners-Lee (1998). Ο ίδιος υποστήριξε πως η κατανόηση του περιεχομένου που είναι διαθέσιμο στον Παγκόσμιο Ιστό από τους ηλεκτρονικούς υπολογιστές μπορεί να προσφέρει μία επανάσταση από νέες δυνατότητες και πληροφορικές εφαρμογές (Berners-Lee 2001). Στη διαμόρφωση της κοινωνίας της πληροφορίας του μέλλοντος συμμετέχουν ενεργά οι σύγχρονοι διαχειριστές της, οι ειδικοί από το χώρο των επιστημών της πληροφορίας

Ο Σημασιολογικός Ιστός αποτελεί πρωτοβουλία της Κοινοπραξίας του Παγκοσμίου Ιστού ή World Wide Web Consortium ή W3C (<http://www.w3.org>). Η ιδέα στην οποία βασίζεται αφορά στην οργάνωση και στη διασύνδεση της διαθέσιμης πληροφορίας στο Διαδίκτυο ώστε να χρησιμοποιηθεί πιο αποτελεσματικά στην ανακάλυψη, στην αυτοματοποίηση, στην ομαδοποίηση και στην επαναχρησιμοποίησή της από διαφορετικές μεταξύ τους διαδικτυακές εφαρμογές. Δηλαδή στοχεύει στη μετεξέλιξη του σημερινού ιστού ώστε να μετατραπεί σε ένα περιβάλλον όπου οι πληροφορίες που υπάρχουν και διακινούνται να είναι επεξεργάσιμες από τους ηλεκτρονικούς υπολογιστές. Κινείται προς την κατεύθυνση της επέκτασης και της βελτίωσης της δόμησης της πληροφορίας για να είναι προσπελάσιμη από προγράμματα υπολογιστών με τελικό στόχο την αυτοματοποίηση πολλών λειτουργιών στον Παγκόσμιο Ιστό. Σε αυτό το πλαίσιο, τα σημερινά κείμενα στις σελίδες του Web θα εμπλουτιστούν με δομημένα κείμενα και δεδομένα σε μορφή XML και RDF. Από αυτή την άποψη ο Σημασιολογικός Ιστός είναι ένα πρόβλημα αναπαράστασης γνώσης από και για τους υπολογιστές και απαιτεί ένα μηχανισμό που θα επιτρέπει την επεξεργασία αυτής της γνώσης. Αυτός ο μηχανισμός θα πρέπει να υποστηρίζει τη δυνατότητα λογικής επεξεργασίας των πληροφοριών με σκοπό την εξαγωγή συμπερασμάτων, τη δημιουργία νέας γνώσης, την υποστήριξη στη λήψη αποφάσεων και την αυτόματη εκτέλεση ενεργειών.

Ο Σημασιολογικός Ιστός αποτελεί σήμερα μια από τις πιο σημαντικές προσπάθειες για την αυτόματη ενοποίηση των συστημάτων και των εφαρμογών ώστε να συνεργάζονται διαλειτουργικά σε παγκόσμιο επίπεδο. Ο Tim Berners-Lee οραματιζόταν ένα ιστό δεδομένων αυτόματα επεξεργάσιμων από τις εφαρμογές βάσει του νοήματος και όχι της μορφής της πληροφορίας. Αυτό θα έχει σαν αποτέλεσμα από το απλό κείμενο που το καταλαβαίνει πλήρως μόνο ο άνθρωπος να παρατηρείται κίνηση προς τις ημιδομημένες (semi-structured) και δομημένες (structured) πληροφορίες που μπορούν να γίνουν κατανοητές από τις διαδικτυακές εφαρμογές όπως για παράδειγμα συμβαίνει στα συστήματα ηλεκτρονικού εμπορίου (e-commerce), ηλεκτρονικού επιχειρείν (e-business) και ηλεκτρονικής μάθησης (e-learning).

Η πραγματική δύναμη του Σημασιολογικού Ιστού θα αποκαλυφθεί όταν οι άνθρωποι δημιουργήσουν συστήματα τα οποία θα συλλέγουν τα περιεχόμενα του Ιστού από διαφορετικές πηγές, θα επεξεργάζονται τις πληροφορίες και θα ανταλλάσσουν τα αποτελέσματα με άλλους ανθρώπους ή μηχανές. Με αυτό τον τρόπο, η αποτελεσματικότητα του Σημασιολογικού Ιστού θα βελτιωθεί δραστικά αφού μεγαλύτερο ποσοστό περιεχομένου του Ιστού θα είναι «αναγνώσιμο» από τους υπολογιστές και περισσότερες αυτοματοποιημένες υπηρεσίες θα είναι διαθέσιμες. Δύο σημαντικές τεχνολογίες που στοχεύουν στην ανάπτυξη του Σημασιολογικού Ιστού είναι ήδη σε ισχύ: η XML (eXtensible Markup Language) και το RDF (Resource Description Framework).

Η χρήση των οντολογιών μπορεί να αποδειχθεί ιδιαίτερα κρίσιμη παράμετρος προς αυτή την κατεύθυνση. Μια οντολογία, όπως θα αναφερθεί και παρακάτω, είναι μια αυστηρή περιγραφή των αντικειμένων-πόρων και των σχέσεων μεταξύ τους. Οιοντολογίες δίνουν τη δυνατότητα σε μια κοινότητα χρηστών να έχουν κοινή ονοματολογία και κοινή αντίληψη για τα αντικείμενα-πόρους τα οποία δηλώνουν ή χρησιμοποιούν. Επίσης, τους επιτρέπουν να περιγράψουν τον «κόσμο» που χρησιμοποιούν (που τους αφορά) με ένα τέτοιο τρόπο, ώστε οι υπολογιστές να μπορούν να επεξεργαστούν και να «συνδέσουν» τα δεδομένα από διαφορετικούς κόσμους. Η δυσκολία στις οντολογίες έγκειται στο γεγονός ότι οι κοινότητες των χρηστών θα πρέπει να συμφωνήσουν στην οντολογική περιγραφή του πεδίου ενδιαφέροντός τους.

Ο τρόπος με τον οποίο θα αναπτυχθεί ο Σημασιολογικός Ιστός προβλέπεται (Semaview 2002: 6) να είναι ανάλογος με αυτόν του σύγχρονου Παγκόσμιου Ιστού. Το πρώτο βήμα στη διαμόρφωση του Παγκόσμιου Ιστού του μέλλοντος είναι η δημιουργία νησίδων πληροφορίας οργανωμένης σημασιολογικά. Οι νησίδες αυτές σιγά-σιγά θα διασυνδεθούν μεταξύ τους προσφέροντας περισσότερες δυνατότητες για την ανάπτυξη προηγμένων εφαρμογών. Το επόμενο βήμα είναι η αξιοποίηση του πλέγματος πληροφοριών, που δημιουργείται, σύμφωνα με τις προοπτικές που παρέχει η σημασιολογική τους οργάνωση.

Το καθένα από τα στάδια εξέλιξης του Σημασιολογικού Ιστού χαρακτηρίζεται από συγκεκριμένες τεχνολογικές δυνατότητες. Η διάθεση μέσω του Παγκόσμιου Ιστού δεδομένων που έχουν οριστεί και διασυνδεθεί μεταξύ τους αποτελεί τη βασική υποδομή για ό,τι υπόσχεται ο μελλοντικός Παγκόσμιος Ιστός: αποτελεσματικότερη αναζήτηση δεδομένων και πληροφοριών, αυτοματοποίηση και ολοκλήρωση υπηρεσιών, αξιοποίηση της υπάρχουσας γνώσης σε ποικίλες εφαρμογές (Hendler 2002).

Σταδιακά, η χρήση του Σημασιολογικού Ιστού επεκτείνεται σε συσκευές (π.χ. τηλεοράσεις, κινητά τηλέφωνα κ.α.) που συνδέονται στο Διαδίκτυο. Καθώς οι συσκευές αυτές θα χρησιμοποιούν τεχνολογίες Σημασιολογικού Ιστού, θα περιγράφουν τις λειτουργίες τους και θα τις ανακοινώνουν στο Διαδίκτυο. Έτσι, θα προκύψουν νέες δυνατότητες και αποδοτικότερη συνεργασία μεταξύ των συσκευών αυτών και του Web. Για παράδειγμα, στο μέλλον η τηλεόραση θα μας προτείνει προγράμματα σύμφωνα με κριτήρια που θέσαμε στον Σημασιολογικό Ιστό. Ένα σημαντικό χαρακτηριστικό των μελλοντικών συστημάτων διαχείρισης πελατειακών σχέσεων (CRM) θα είναι οι «συσκευές διεπαφής χρήστη», οι οποίες

θα μπορούν να «αναγνωρίζουν» τον χρήστη και να «κατανοούν» τη διάθεση και τα συναισθήματά του. Στη συνέχεια, θα του προτείνουν εξατομικευμένες υπηρεσίες και προϊόντα.

2.8.1 Είναι υλοποιήσιμο-εφικτό ένα τέτοιο εγχείρημα

Όπως γνωρίζουμε οι κυριότερες δραστηριότητες των ανθρώπων στο διαδίκτυο είναι η αναζήτηση πληροφοριών, η επικοινωνία και το ηλεκτρονικό εμπόριο. Η αναζήτηση γίνεται από τις μηχανές αναζήτησης με λέξεις- κλειδιά που επιστρέφουν τεράστιες ποσότητες πολλές φορές ακατάλληλης πληροφορίας. Αυτό συμβαίνει γιατί ο υπολογιστής δεν έχει τη δυνατότητα να αντιλαμβάνεται τη σημασία των λέξεων που ανακτά, δεν μπορεί δηλαδή να ερμηνεύσει το περιεχόμενο των δεδομένων αυτών και θα πρέπει ένας άνθρωπος μέσα από κάποιες ιστοσελίδες να τον καθοδηγήσει. Στον Σημασιολογικό Ιστό ένας υπολογιστής για να πραγματοποιήσει λειτουργίες όπως αναζητήσεις, συγκρίσεις, διαμοιρασμό κλπ. δεν θα χρειάζεται ανθρώπινη διαμεσολάβηση. Με την τεχνολογία που θα διαθέτει (αυτό σημαίνει ότι κατανοεί τη σημασία των λέξεων) θα έχει τη δυνατότητα να καταλάβει τη σημασία της πληροφορίας και να την επεξεργαστεί. Σε πολλούς η διαδικασία θα θυμίσει ταινίες επιστημονικής φαντασίας στις οποίες ένας υπολογιστής με νοημοσύνη (και όνομα συνήθως) γνωρίζει και εκτελεί όλες τις σημαντικές ενέργειες. Η ουσία δεν απέχει πολύ.

Ας πούμε για παράδειγμα, ότι θέλουμε να κλείσουμε ένα ιατρικό ραντεβού. Αφού εισάγουμε τα δεδομένα μας, στις απαντήσεις που θα πάρουμε από τους ευφυείς πράκτορες (intelligent agents), θα έχει ληφθεί υπόψη ότι, ο γιατρός που αναζητάμε θα έχει την ειδικότητα που μας ενδιαφέρει, θα είναι συμβεβλημένος με το ταμείο μας, θα βρίσκεται στην περιοχή που μένουμε ή εργαζόμαστε και θα έχει ελεύθερες ώρες σε σχέση με το δικό μας πρόγραμμα.. Πριν το ραντεβού ο ιατρικός μας φάκελος θα μπορούσε να σταλεί στο γιατρό, έτσι ώστε να είναι ήδη ενήμερος για την κατάστασή μας και να προχωρήσει στην κλινική εξέταση.

Άλλο ένα παράδειγμα θα μπορούσε να είναι η οργάνωση ενός ταξιδιού. Σήμερα αν θέλουμε να προγραμματίσουμε ένα ταξίδι, θα πρέπει να δοκιμάσουμε διάφορες ιστοσελίδες και στο τέλος να συγκεντρώσουμε όλες αυτές τις πληροφορίες για να καταλήξουμε στο συνδυασμό που θέλουμε (εισιτήρια, ξενοδοχείο, εισιτήρια για μουσεία, συναυλίες κλπ). Γνωρίζουμε ότι οι απαντήσεις που θα πάρουμε είναι ένα μέρος μόνο του υπάρχοντος ιστού αφού πολλές χρήσιμες πληροφορίες βρίσκονται σε βάσεις δεδομένων, αρχεία XML, EXCEL κλπ (deep web), ενώ εμείς έχουμε πρόσβαση σε ότι οι σχεδιαστές των ιστοσελίδων μας επιτρέπουν να δούμε. Τα εξειδικευμένα sites συγκεντρώνουν και συνδέουν κατά κάποιον τρόπο δεδομένα από διαφορετικές πηγές αλλά πάλι ελέγχουν το πώς και τι θα δούμε. Στο ΣΙ όλη αυτή η διαδικασία θα γίνεται από τον υπολογιστή εισάγοντας μόνο τα δικά μας δεδομένα, ενώ θα χρειάζεται μόνο να επικυρώσουμε το τελικό πλάνο. Απαραίτητες έννοιες τα Μεταδεδομένα (Metadata) είναι οι περιγραφικές πληροφορίες μιας ψηφιακής πηγής. Παράδειγμα: δεδομένα είναι τα προσωπικά στοιχεία ενός ανθρώπου και μεταδεδομένα η κωδικοποίησή τους. Με τα μεταδεδομένα γίνεται δυνατή η αναγνώριση και η ανάκτηση της πληροφορίας από τις αρχικές της πηγές (Δεδομένα - Μεταδεδομένα)

Όνομα Ευθυμίου Γεώργιος
Διεύθυνση Αμερικής 60
Πόλη Αθήνα
Ταχυδρομικός κώδικας 45500

Οντολογίες (ontologies) είναι οι αναπαραστάσεις των εννοιών μιας θεματικής περιοχής π.χ. Τουρισμός, Εκπαίδευση κλπ. έτσι ώστε να μπορεί να είναι επεξεργάσιμη από

υπολογιστή. Με τη βοήθεια της οντολογίας προσδίδουμε στις εκφράσεις σημασία, περιεχόμενο δηλαδή που έχει νόημα. Οι οντολογίες ερμηνεύουν λοιπόν την πληροφορία και επικοινωνούν με άλλα εργαλεία

Λογική (Logic) Με τη βοήθεια εργαλείων εξάγουμε λογικά συμπεράσματα για τις έννοιες που αναπτύσσονται μέσα στην οντολογία και τις σχέσεις μεταξύ τους δηλαδή επεξεργάζεται η πληροφορία και εξάγονται τα συμπεράσματα.

Πράκτορες (intelligent agents) είναι μικρό αυτόνομο λογισμικό που συλλέγει, οργανώνει και παρουσιάζουν την βέλτιστη πληροφορία στον χρήστη, κάτι σαν προσωπική μηχανή αναζήτησης

Διαλειτουργικότητα (interoperability) η δυνατότητα συνεργασίας μεταξύ διαφορετικών προγραμμάτων. Οτιδήποτε λοιπόν θέλουμε να περιγράψουμε, χαρακτηρίζεται με τη βοήθεια των μεταδεδομένων. Τα μεταδεδομένα συνδέονται ώστε να αποκτήσουν νόημα και να περιγράψουν μια θεματική περιοχή (οντολογίες).

Βασική προϋπόθεση για τη χρήση οντολογιών στα πλαίσια του ΣΙ είναι η δυνατότητα αποδοτικής εξαγωγής συμπερασμάτων από αυτές μέσω μιας λογικής διαδικασίας. Όπως όλα τα φιλόδοξα σχέδια ο Σημασιολογικός Ιστός έχει πολλούς που δεν συμμερίζονται τον ενθουσιασμό των ερευνητών του. Επισημαίνουν τεχνικές ατέλειες στα πρότυπα, αδυναμία ανάλυσης των οντολογιών λόγω όγκου και πολυπλοκότητας κ.ά. Πολλοί ερευνητές κυρίως σε ερευνητικά και εκπαιδευτικά ιδρύματα έχουν δημιουργήσει εφαρμογές ΣΙ, αλλά η μαζική χρήση του μάλλον δεν θα είναι εφικτή σε σύντομο χρονικό διάστημα. Οι επιστήμονες εργάζονται και εμείς οι υπόλοιποι απλοί χρήστες περιμένουμε τη μέρα, που οι μηχανές θα μας δώσουν πίσω τον πολυπόθητο ελεύθερο χρόνο που σήμερα τους αφιερώνουμε αδιαμαρτύρητα.

2.8.2 Ποιοι πρέπει να το υιοθετήσουν για να εφαρμοστεί (Google,e-bay,κτλ)

Η Google, Bing και Yahoo συμφώνησαν σε ένα κοινό λεξικό, ένα πρότυπο για την απόδοση εννοιών στο περιεχόμενο του Παγκόσμιου Ιστού που θα διευκολύνει την αναζήτηση με πολλαπλά κριτήρια, εντούτοις προϋποθέτει από webmaster και δημιουργούς περιεχομένου να υιοθετήσουν το πρότυπο και να χαρακτηρίσουν το περιεχόμενο για να επιτύχει η προσπάθεια. Το αποτέλεσμα της συναίνεσής τους είναι η σύσταση του Schema.org. Οι δημιουργοί περιεχομένου αποκτούν ένα λεξικό το οποίο μπορούν να αξιοποιήσουν για να αποδίδουν όλες οι μηχανές την ίδια «σημασία» στα δεδομένα, εάν θέλουν να εμφανίζονται στα αποτελέσματά τους. Το πρόβλημα ως τώρα ήταν ότι πολλές φορές η πληροφορία υπήρχε δομημένη σε βάσεις δεδομένων - με την ένταξη των δεδομένων σε πεδία (π.χ. Όνομα, Ημερομηνία γέννησης κ.λπ.). Αυτή η σημασία όμως, χανόταν όταν γινόταν η μεταφορά σε HTML. Το πρότυπο που υιοθετούν οι τρεις μεγαλύτερες μηχανές αναζήτησης ονομάζεται microdata και αφορά στις προδιαγραφές που προστίθενται βαθιά στην HTML5.0.

Έτσι, για παράδειγμα όταν αναφέρονται στοιχεία για ένα πρόσωπο, η microdata προδιαγράφει ότι για να περιγραφούν τα χαρακτηριστικά του μπορούν να χρησιμοποιηθούν, για παράδειγμα, στην HTML οι ετικέτες itemprop="name", itemprop="birthdate", itemprop="jobtitle" για να αποδοθεί η έννοια του ονόματος, της ημερομηνίας γέννησης και του επαγγέλματος αντίστοιχα κάθε φορά που αναφέρονται αυτά τα δεδομένα στο περιεχόμενο. Το αποτέλεσμα είναι ότι οι μηχανές αναζήτησης στο εξής θα στέλνουν τα ρομποτάκια τους να συλλέξουν στοιχεία από τον λεγόμενο Σημασιολογικό Ιστό (το λεγόμενο Semantic Web κατά τον Τιμ Μπέρνερς Λι, τον «πατέρα του World Wide Web»). Έτσι, οι

μηχανές θα μπορούν να παρέχουν υπηρεσίες αναζήτησης με σαφή κριτήρια και θεωρητικά τουλάχιστον εγγυημένα αποτελέσματα -έχει ήδη γίνει λόγος για εκμετάλλευση.

Η Google έχει κάνει το πείραμα με τις συνταγές: ο επισκέπτης μπορεί να αναζητήσει συνταγές με βάση το βασικό συστατικό τους και να παραμετροποιήσει περαιτέρω την αναζήτησή του με πρόσθετα χαρακτηριστικά, όπως άλλα συστατικά, χρόνο μαγειρέματος και άλλα. Για να φέρει αποτελέσματα, η μηχανή αναζήτησης συνταγών του Google αντλεί *πληροφορία* από το σημασιολογικό ιστό, δηλαδή site μαγειρικής που έχουν συμπεριλάβει ήδη στην HTML των σελίδων τους τις ετικέτες όπως αυτές περιγράφονται στο πρότυπο της microdata.

2.9 HAKIA



Η HAKIA είναι μια μηχανή αναζήτησης στο Διαδίκτυο. Η εταιρεία έχει εφεύρει την QDEXing τεχνολογία, μια εναλλακτική λύση νέων υποδομών για τη δημιουργία ευρετηρίου που χρησιμοποιεί SemanticRank αλγόριθμο, ένα μίγμα λύση από τον κλάδο της οντολογικής σημασιολογίας, ασαφής λογική, υπολογιστική γλωσσολογία, και τα μαθηματικά. Ιδρύθηκε το 2004, η εταιρεία είναι ιδιωτική εταιρεία και εδρεύει στη Νέα Υόρκη.

Η hakia ιδρύθηκε από τον **Riza Berkan**, ένας τυρηνικός επιστήμονας με εκπαίδευση, με ειδίκευση στην τεχνητή νοημοσύνη και ασαφής λογική, και την **Pentti Kourή**, οικονομολόγο. Ο **καθηγητής Βίκτωρ Raskin, ο πατέρας του οντολογικής σημασιολογίας** που σημείωσε διεθνούς αρχής στον τομέα της υπολογιστικής γλωσσολογίας, χρησιμεύει ως επιστημονικός σύμβουλος της Hakia. Τα μέλη του διοικητικού της συμβουλίου περιλαμβάνουν πρώην γεροϋσιαστής Bill Bradley, Pentti Kouri, Riza C. Berkan, Ryszard Krauze, Anuj Mathur, Murat Vargi and John Grzymala. Η Hakia έχει προκαλέσει 21 εκατομμύρια δολάρια από ιδιώτες επενδυτές.

Οι Hakia οργανώνει τα αποτελέσματα αναζήτησης σε καρτέλες: Αποτελέσματα Web, αξιόπιστες ιστοσελίδες, εικόνες και ειδήσεις. Αξιόπιστες ιστοσελίδες προσφέρουν αποτελέσματα από site όπου έχουν εξετασθεί προσεκτικά από βιβλιοθηκονόμους και άλλους επαγγελματίες της πληροφόρησης και προκαλεί την Hakia με αυτό τον τρόπο να προσδιορίσει αξιόπιστες ιστοσελίδες.

2.9.1 Πράγματα που κάνουν να αναζητούμε μία σημασιολογική αναζήτηση

1. Χειρισμός – μορφολογικές παραλλαγές

Μια σημασιολογική μηχανή αναζήτησης αναμένεται να χειριστεί όλες τις μορφολογικές μεταβολές (όπως χρόνους, πληθυντικούς, κ.λπ.) σε σταθερή βάση.

2. Χειρισμός εννοιών που ταιριάζουν

Ίσως η μεγαλύτερη πρόκληση της λειτουργικότητας μεταξύ όλων, μιας σημασιολογικής μηχανής αναζήτησης, είναι να αναγνωρίζει έννοιες και να φέρνει τα σχετικά αποτελέσματα.

3. Χειρισμός της φυσικής γλώσσας με απορίες και ερωτήσεις

Μια σημασιολογική μηχανή αναζήτησης αναμένεται να ανταποκριθεί λογικά όταν το ερώτημα είναι σε μορφή ερώτησης (τι, πού, πώς, γιατί, κλπ.) Σημειώστε ότι, μια "μηχανή αναζήτησης" είναι διαφορετική από ένα "ερωταποκρίσεων" σύστημα. Κύριο καθήκον των μηχανών αναζήτησης είναι η κατάταξη των αποτελεσμάτων αναζήτησης με τον πιο λογικό και συναφή τρόπο, ενώ το σύστημα ερωταποκρίσεων μπορεί να παράγει ένα ενιαίο αποτέλεσμα ονομαζόμενο οντότητα.

4. Δυνατότητα λειτουργίας χωρίς να στηρίζει σε στατιστικά στοιχεία και σε άλλα τεχνικά μέσα, τη συμπεριφορά των χρηστών

Μια σημασιολογική μηχανή αναζήτησης αναμένεται να αποφέρει σημαντικά αποτελέσματα από την ανάλυση του περιεχομένου της σελίδας (ή εγγράφου), την προέλευση και τους συγγραφείς. Στηριζόμενη στις παραπομπές, στους συνδέσμους, στη συμπεριφορά των χρηστών / tagging, και σε άλλα τεχνητά μέσα, μπορεί να έχει καλά αποτελέσματα όταν τα δεδομένα αυτά είναι διαθέσιμα, αλλά είναι έξω από τη σφαίρα της σημασιολογικής αναζήτησης. Με το να μην βασίζονται σε τεχνητή είσοδο, η σημασιολογική τεχνολογία αναζήτησης είναι πιο καθολική, ισχύει για οποιαδήποτε κατάσταση, ιδίως για τις επιχειρήσεις παρέχοντας περιεχόμενο σε πραγματικό χρόνο.

5. Ικανότητα να εντοπίζει το μέγεθος της απόδοσης

Όταν δεν υπάρχει σημασιολογική ανάλυση του περιεχομένου σε έναν αλγόριθμο αναζήτησης, τα αποτελέσματα σχετικότητας αναφέρονται σε τεχνητές μετρήσεις, όπως το πόσο δημοφιλής είναι η σελίδα.

2.10 Συστήματα διαχείρισης Γνώσης βασισμένα σε οντολογίες

Σε περιπτώσεις όπου η γνώση χρειάζεται να μοντελοποιηθεί, να δομηθεί και να διασυνδεθεί, οι οντολογίες παρέχουν έναν ικανοποιητικό μηχανισμό αναπαράστασης και διαμοίρασης της σε ομάδες ανθρώπων.

Στη παράγραφο αυτή γίνεται ανάλυση του τρόπου με τον οποίο θα μπορούσαν τα συστήματα διαχείρισης γνώσης να στηρίζονται σε οντολογίες.

Προτείνεται ένα μοντέλο με δύο επιμέρους διαδικασίες και γίνεται ανάλυση των επιμέρους φάσεων της κάθε διαδικασίας. Σύμφωνα με το μοντέλο αυτό γίνεται διάκριση μεταξύ διαδικασίας γνώσης και μετά-διαδικασίας γνώσης. Οι δύο αυτές διαδικασίες είναι αναδραστικές, δηλαδή οι διάφορες φάσεις από τις οποίες αποτελούνται επαναλαμβάνονται κατά τη διάρκεια του χρόνου, και είναι ορθογώνιες μεταξύ τους. Η μετά-διαδικασία γνώσης αποτελεί τη διαδικασία η οποία θα πρέπει να ακολουθηθεί ώστε να δημιουργηθεί η κατάλληλη οντολογία και εμφανίζεται στο μοντέλο ως κάθετος άξονας, ενώ η διαδικασία γνώσης αποτελεί τη διαδικασία η οποία ακολουθείται ώστε να έχουμε ένα σύστημα διαχείρισης γνώσης βασισμένο στην οντολογία και να εμφανίζεται στο μοντέλο ως οριζόντιος άξονας

2.10.1 Μετά - Διαδικασία Γνώσης

Η μετά- διαδικασία γνώσης αφορά τη διαδικασία ορισμού της οντολογίας, η οποία θα πρέπει να γίνεται με έναν επίσημο τρόπο και αποτελείται από τις παρακάτω πέντε φάσεις:

1. Φάση Επιτευξιμότητας (Feasibility Study)
2. Φάση Έναρξης (Kickoff Phase)
3. Φάση Τελειοποίησης (Refinement Phase)
4. Φάση Αξιολόγησης (Evaluation Phase)
5. Φάση Συντήρησης (Maintenance Phase)

Μελέτη Επιτευξιμότητας

Για να λειτουργήσει σωστά ένα σύστημα διαχείρισης γνώσης θα πρέπει να έχει ολοκληρωθεί με το σωστό τρόπο. Υπάρχουν πολλοί παράγοντες, εκτός της ίδιας της τεχνολογίας οι οποίοι μπορούν να καθορίσουν την επιτυχία ή την αποτυχία. Η μελέτη της επιτευξιμότητας περιλαμβάνει την ανάλυση όλων αυτών των παραγόντων, ώστε να διερευνηθούν όλες οι πτυχές του προβλήματος, οι περιοχές ευκαιριών και οι πιθανές λύσεις. Η ανάλυση αυτή θα πρέπει να λάβει υπ' όψιν μία ευρύτερη προοπτική για τον οργανισμό. Στη φάση αυτή καθορίζεται η οικονομική και τεχνική επιτευξιμότητα του έργου και γίνεται η επιλογή της κατάλληλης λύσης για τα πιθανά προβλήματα. Η φάση αυτή λειτουργεί ως βάση για την επόμενη φάση και για το λόγο αυτό θα πρέπει να ολοκληρωθεί πριν ξεκινήσει η φάση έναρξης (Kick off Ph ase).

Στη φάση αυτή γίνεται περιγραφή των εννοιών που θα πρέπει να υποστηρίζει η οντολογία και σχεδιογραφείται η περιοχή εφαρμογής της οντολογίας. Το αποτέλεσμα της φάσης αυτής είναι ένα έντυπο με τη προδιαγραφή των απαιτήσεων της οντολογίας.

Η προδιαγραφή των απαιτήσεων είναι το εργαλείο του μηχανικού ο οποίος θα πρέπει να αποφασίσει ποιες έννοιες θα εισαχθούν και ποιες θα εξαχθούν στην ιεραρχική δομή των εννοιών της οντολογίας. Στο επίπεδο αυτό επίσης μπορεί να από φασιστεί η ανάγκη χρησιμοποίησης άλλων διαθέσιμων οντολογιών.

Αναφορικά η **φάση Έναρξης** θα πρέπει να είναι σε θέση να ορίζει καθαρά πληροφορία σχετικά με:

- Το σκοπό της οντολογίας
- Το πεδίο ορισμού και την εμβέλεια του συστήματος
- Τις εφαρμογές που μπορούν να υποστηριχτούν
- Τις πηγές πληροφορίας
- Τους πιθανούς χρήστες και τα πιθανά σενάρια χρήσης

Επίσης στη φάση αυτή αναπτύσσεται και ένα ερωτηματολόγιο πληρότητας της οντολογίας το οποίο θα πρέπει να ελεγχθεί κατά τη φάση αξιολόγησης.

Τέλος γίνεται σχολιασμός για πιθανόν ύπαρξη αυτόνομων οντολογιών, οι οποίες μπορούν να επαναχρησιμοποιηθούν.

Φάση Τελειοποίησης

Ο σκοπός της φάσης αυτής είναι να παράγει μία ώριμη οντολογία στην οποία να μπορούν να στηριχτούν εφαρμογές, σύμφωνα με τις προδιαγραφές των απαιτήσεων οι οποίες προέκυψαν στην προηγούμενη φάση. Η φάση τελειοποίησης μπορεί να διαιρεθεί στις παρακάτω επιμέρους φάσεις:

- Συλλογή των εννοιών της φάσης Έναρξης και υλοποίηση βασικής ταξινόμιας
- Εξαγωγή γνώσης από ειδικούς βασισμένη στη βασική ταξινόμια με σκοπό την υλοποίηση μίας καλά ορισμένης ταξινόμιας με σχετικές έννοιες και σχέση μεταξύ τους.
- Περιγραφή της τελικής οντολογίας με χρήση τυπικών γλωσσών αναπαράστασης (Frame Logic, Description Logic, Conceptual Graphs)

Φάση Αξιολόγησης

Η φάση αυτή χρησιμοποιείται για να αποδειχθεί εάν πράγματι η οντολογία που έχει αναπτυχθεί ικανοποιεί τις προδιαγραφές των απαιτήσεων και εάν απαντά στις ερωτήσεις πληρότητας όπως αυτές είχαν περιγραφεί κατά τη φάση Έναρξης. Σε αυτή τη φάση επίσης, γίνεται έλεγχος του περιβάλλοντος εφαρμογής, για να διαπιστωθεί η λειτουργικότητα του συστήματος από τους χρήστες. Ο τρόπος με τον οποίο οι χρήστες περιπλανιούνται μέσα στην οντολογία αναζητώντας έννοιες γίνεται πηγή μελέτης για την αναδιάταξη των εννοιών. Πολλές φορές τα αποτελέσματα της φάσης αυτής οδηγούν στην προηγούμενη φάση της Τελειοποίησης, όπου η οντολογία εξελίσσεται, και επαναξιολογείται.

Φάση Συντήρησης

Καθώς ο πραγματικό κόσμος εξελίσσεται υπάρχει η ανάγκη αναθεώρησης της οντολογίας ώστε αυτή να συνεχίσει να αναπαριστά σωστά τις έννοιες που μοντελοποιεί. Η ανάγκη για συντήρηση της οντολογίας είναι συχνή και για αυτό θα πρέπει να υπάρχουν αυστηροί κανόνες οι οποίοι να ορίζουν τον τρόπο με τον οποίο γίνεται ενημέρωση, διαγραφή ή εισαγωγή εννοιών στην οντολογία. Θα πρέπει έπειτα να εξετάζεται με ποιον τρόπο η αλλαγή στον ορισμό της οντολογίας επηρεάζει τις εφαρμογές που στηρίζονται σε αυτή και να γίνονται οι κατάλληλες αλλαγές.

2.10.2 Διαδικασία Γνώσης

Η διαδικασία γνώσης περιγράφει τη διαδικασία ανάπτυξης ενός συστήματος διαχείρισης γνώσης δεδομένης μίας οντολογίας και περιλαμβάνει τα παρακάτω βήματα:

1. Δημιουργία Γνώσης

2. Εισαγωγή Γνώσης
3. Δέσμευση Γνώσης
4. Ανάκτηση και Προσπέλαση Γνώσης
5. Χρησιμοποίηση Γνώσης

Δημιουργία Γνώσης

Η δημιουργία γνώσης στην οποία έχει πρόσβαση ο υπολογιστής (computer-accessible knowledge), διακρίνεται σε τυπικά και μη τυπικά έγγραφα. Με τον όρο τυπικός και μη τυπικός εκφράζεται ο βαθμός ελευθερίας της δομής ενός εγγράφου. Είναι θεμιτό ο υπολογιστής να μπορεί να έχει πρόσβαση σε έγγραφα των οποίων η δομή εμπίπτει μεταξύ αυτών των δύο άκρων. Για το σκοπό αυτό εισάγεται η ιδέα των “document templates”, δηλαδή περιγράμματα εγγράφων. Τα document templates είναι ημι-δομημένα δεδομένα τα οποία αποτελούνται από “knowledge items”. Ως Knowledge item ορίζεται ένα αυτόνομο αντικείμενο γνώσης για την περιγραφή κάποιας συγκεκριμένης διαδικασίας ή κάποιου συγκεκριμένου αντικειμένου.

Εισαγωγή Γνώσης

Σε κάθε σύστημα διαχείρισης Γνώσης, η δυνατότητα εισαγωγής γνώσης είναι εξίσου ή και περισσότερο σημαντική από τη δημιουργία γνώσης. Κατά το βήμα αυτό θα πρέπει να προσεχθεί ώστε η καινούρια γνώση να μπορέσει επιτυχώς να διασυνδεθεί και να ενοποιηθεί με την ήδη υπάρχουσα γνώση. Καθώς οι δομές είναι πλέον πολύ πιο πλούσιες και σύνθετες αυτό χρειάζεται ιδιαίτερη προσοχή, κυρίως όταν η γνώση είναι ξένη, οπότε υπάρχει ελλιπής κατανόηση των δεδομένων που εισάγονται.

Δέσμευση Γνώσης

Από τη στιγμή που έχουν δημιουργηθεί τα “knowledge items”, το επόμενο βήμα είναι η δέσμευση των σημ. αντικών τους περιεχομένων. Για το σκοπό αυτό μπορούν να χρησιμοποιηθούν διάφορες τεχνικές αφαίρεσης ή δεικτοδότησης, γνωστές από την επιστήμη των βιβλιοθηκών ή άλλα εργαλεία όπως το Onto Annotate, για δέσμευση αποσπασμάτων από έγγραφα και της μεταξύ τους διασύνδεση.

Ανάκτηση και Προσπέλαση Γνώσης

Συνήθως η ανάκτηση και προσπέλαση της γνώσης πραγματοποιείται μέσω των παραδοσιακών GUI. Πέρα από αυτά μπορεί να χρησιμοποιηθεί η οντολογία για την εξαγωγή επιπλέον όψεων της γνώσης. Πιο συγκεκριμένα η οντολογία μπορεί να χρησιμοποιηθεί για περιήγηση, προσφέροντας διάφορες δομές. Οι χρήστες μπορούν να ανακαλύψουν την υπάρχουσα γνώση χωρίς να χρειάζεται να εκτελούν συγκεκριμένες επερωτήσεις, κάτι το οποίο είναι δύσκολο, κυρίως για τους μη ειδικούς. Επιπλέον, η χρήση της οντολογίας επιτρέπει την εξαγωγή χρήσιμων συμπερασμάτων σχετικά με τη διασύνδεση και τη περιγραφή των εννοιών, πολλές φορές χωρίς να χρειάζεται να υπάρχουν στιγμιότυπα της πληροφορίας, απλά και μόνο την περιγραφή τους.

Χρησιμοποίηση Γνώσης

Η Χρησιμοποίηση της γνώσης είναι συνήθως παραμελημένο μέρος, ενώ είναι ουσιαστικά ο λόγος για τον οποίο αναπτύσσονται τα συστήματα διαχείρισης γνώσης. Πολλά συστήματα διαχείρισης γνώσης θεωρούν πως όλα τελειώνουν από τη στιγμή που ένα έγγραφο θα βρεθεί. Παρόλα αυτά θέματα όπως προσπέλαση υπό συνθήκη, προσωποποίηση και κυρίως ολοκλήρωση με άλλες εφαρμογές παίζουν σημαντικό ρόλο για την αποτελεσματική επαναχρησιμοποίηση της γνώσης. πολύ συχνά δεν είναι η γνώση η οποία έχει μεγαλύτερο ενδιαφέρον αλλά οι εξαγωγείς που μπορούν να γίνουν από αυτήν. Επίσης τις περισσότερες φορές χρήσιμο να γίνεται συνδυασμός των “knowledge items”, έτσι ώστε να υπάρχει πλήρης εικόνα γύρω από κάποιο θέμα και όχι η γνώση που προκύπτει από ένα απομονωμένο “knowledge item”.

ΚΕΦΑΛΑΙΟ 3^ο

3.1 Τι είναι μια οντολογία και γιατί τη χρειαζόμαστε

Στη διεθνή βιβλιογραφία είναι δυνατό να εντοπιστούν πάρα πολλοί ορισμοί για το τι είναι η οντολογία. Ένας ορισμός της, που προσεγγίζει τη φιλοσοφική πλευρά του όρου, ορίζει την οντολογία ως την επιστήμη που μελετά την ύπαρξη, τη φύση και τις σχέσεις των όντων.

Ιστορικά, ο όρος οντολογία (από το ελληνικό ὄν, γενική ὄντος: από την ύπαρξη (ουδέτερη μετοχή του εἶναι: για να είναι) και -λογία, -logia: η επιστήμη, η μελέτη, η θεωρία) είναι η φιλοσοφική μελέτη της φύσης της υπόστασης, της ύπαρξης ή της πραγματικότητας γενικά, καθώς επίσης και των βασικών κατηγοριών ύπαρξης και σχέσεών τους.

Παραδοσιακά, η οντολογία, που αποτελεί μέρος του σημαντικότερου κλάδου της φιλοσοφίας, γνωστής ως μεταφυσική, εξετάζει ζητήματα σχετικά με το ποιες οντότητες υπάρχουν ή μπορεί να ειπωθεί ότι υπάρχουν, και το πώς τέτοιες οντότητες μπορούν να ομαδοποιηθούν, να ενταχθούν σε μια ιεραρχία, και να υποδιαιρεθούν σύμφωνα με ομοιότητες και διαφορές.

Εντούτοις, τα τελευταία χρόνια, η οντολογία είναι μία από τις πολλές λέξεις, την οποία έχει δανειστεί η επιστήμη της πληροφορικής από το καθημερινό μας λεξιλόγιο ή από άλλους επιστημονικούς κλάδους και της έχει αποδώσει ένα συγκεκριμένο τεχνικό νόημα που είναι μάλλον διαφορετικό από το αρχικό. Στη διεθνή βιβλιογραφία είναι δυνατό να εντοπιστούν πάρα πολλοί ορισμοί για το τι είναι η οντολογία, μερικοί από αυτούς παρατίθενται παρακάτω.

Ίσως ο πιο διαδεδομένος και αξιόπιστος ορισμός της οντολογίας, όσον αφορά τη χρήση της στην επιστήμη των υπολογιστών εν γένει, είναι εκείνος που διατυπώθηκε από τον Gruber:

«Η Οντολογία είναι ένας τυπικός, σαφής προσδιορισμός μιας διαμοιρασμένης εννοιολογικής αναπαράστασης.»

Ο Fensel επιχειρώντας να αναλύσει αυτόν το σύνθετο εννοιολογικά ορισμό, αναγνωρίζει τέσσερις βασικές έννοιες που εμπλέκονται σε αυτόν:

1. «**εννοιολογική αναπαράσταση**»: ένα αφηρημένο μοντέλο ενός φαινομένου που ορίζεται ως, μια ακριβής,
2. «**διαμοιρασμένη**»: η ύπαρξη μιας συμφωνίας μεταξύ εκείνων που χρησιμοποιούν τις οντολογίες,
3. «**τυπικός**»: μία ακριβής, μαθηματικά, περιγραφή
4. «**σαφής**»: η ακρίβεια των εννοιών και των εμφανώς ορισμένων σχέσεών τους.

Αν και ο ορισμός από τον Gruber είναι ο πιο διαδεδομένος, όπως αναφέρθηκε και παραπάνω, η ανάγκη για έναν πιο επίσημο ορισμό ήταν επιτακτική. Στόχος ήταν ένας πιο τυπικός ορισμός ο οποίος δεν θα άφηνε περιθώρια πολλών διαφορετικών ερμηνειών του όρου και τελικώς θα έκανε πλήρως κατανοητό τι αποτελεί μία οντολογία. Ο Guarino επιχειρεί να εκλεπτύνει τον παραπάνω ορισμό, απομακρύνοντας την οντολογία από την εννοιολογική αναπαράσταση:

«Μια οντολογία είναι μια λογική θεωρία η οποία ερμηνεύει την επιδιωκόμενη έννοια ενός τυπικού λεξιλογίου, δηλ. την οντολογική του δέσμευση για μια συγκεκριμένη σύλληψη του κόσμου. Τα επιδιωκόμενα πρότυπα μιας λογικής γλώσσας που χρησιμοποιεί ένα τέτοιο λεξιλόγιο περιορίζονται από αυτή την οντολογική δέσμευση. Μια οντολογία απεικονίζει έμμεσα αυτήν την υποχρέωση (και την υποκείμενη εννοιολογική σύλληψη), με το να προσεγγίζει αυτά τα επιδιωκόμενα πρότυπα.»

Οι ορισμοί, που δόθηκαν παραπάνω, δύσκολα μπορούν να έχουν πρακτική αξία και κυρίως στα πλαίσια υλοποίησης. Για το λόγο αυτό κρίνεται σκόπιμο να αποδοθεί στον όρο οντολογία ένας πιο συγκεκριμένος ορισμός. Ο ορισμός αυτός χωρίς να είναι αντικρουόμενος με τους παραπάνω, θα αποτυπώνει σαφώς τόσο τον τρόπο όσο και την σημασία με την οποία οι οντολογίες χρησιμοποιούνται στην πράξη και κυρίως σε εφαρμογές του Σημαντικού Ιστού.

Η λέξη οντολογία έχει χρησιμοποιηθεί για να περιγράψει μία ποικιλία εννοιών με διαφορετικό βαθμό δομής, από απλές ταξινομίες (όπως η ιεραρχία του Yahoo), έως σχήματα μεταδομένων (όπως το Dublin Core), και λογικές θεωρίες. Εντούτοις, ο Σημαντικός Ιστός χρειάζεται τις οντολογίες με αυξημένο βαθμό δομής. Πιο συγκεκριμένα, μία οντολογία ορίζεται ως:

«Μια οντολογία καθορίζει εκείνους τους όρους που χρησιμοποιούνται για την περιγραφή και την αναπαράσταση ενός πεδίου γνώσης. Περιλαμβάνει ορισμούς ικανούς να χρησιμοποιηθούν από τη μηχανή για τις βασικές έννοιες του συγκεκριμένου πεδίου γνώσης και για τις σχέσεις μεταξύ τους. Η οντολογία κωδικοποιεί γνώση για ένα πεδίο όπως επίσης και γνώση που συνδέει πεδία. Κατά αυτό τον τρόπο, παράγει γνώση επαναχρησιμοποιήσιμη. Πιο συγκεκριμένα, μια οντολογία παρέχει περιγραφές για:

1. Ιδιότητες (ή χαρακτηριστικά) που περιγράφουν τις κλάσεις και συνήθως τις συσχετίζουν μεταξύ τους.
2. Στιγμιότυπα των κλάσεων αυτών.
3. Κλάσεις (γενικές έννοιες) στο πεδίο αναφοράς.»²

Μια οντολογία καθορίζει ένα κοινό λεξιλόγιο για τους ερευνητές που πρέπει να μοιραστούν πληροφορίες σε μια περιοχή. Περιλαμβάνει ερμηνεύσιμους από τους υπολογιστές ορισμούς των βασικών εννοιών στην περιοχή και τις σχέσεις μεταξύ τους.

Γιατί κάποιος να επιθυμεί να αναπτύξει μια οντολογία; Μερικοί από τους λόγους είναι:

1. Διαμοιρασμός της κοινής αντίληψης της δομής της πληροφορίας μεταξύ των ανθρώπων ή των πρακτόρων λογισμικού. Είναι ίσως ο πιο σημαντικός στόχος στην ανάπτυξη οποιασδήποτε οντολογίας. Για παράδειγμα, αν υποθέσουμε ότι διάφοροι ιστότοποι περιέχουν ιατρική πληροφορία, διαμοιράζονται και δημοσιεύουν υλικό που βασίζεται πάνω στην ίδια οντολογία, τότε οι πράκτορες μπορούν να ανακτήσουν και να συνθέσουν πληροφορία από όλες αυτές τις διαφορετικές πηγές.

2. Διευκόλυνση της επαναχρησιμοποίησης της γνώσης ενός πεδίου: Αποτελεί μία από τις κινητήριες δυνάμεις που βρίσκονται πίσω από τις τρέχουσες εξελίξεις στην έρευνα για τις οντολογίες. Εάν μια ερευνητική ομάδα αναπτύσσει μια οντολογία λεπτομερώς, κάποιες άλλες ομάδες μπορούν απλά να την επαναχρησιμοποιήσουν στο δικό τους αντικείμενο. Επιπρόσθετα, εάν χρειαστεί να αναπτυχθεί μια ευρύτερη οντολογία, αυτό μπορεί να γίνει με τη συνένωση ήδη υπάρχοντων επιμέρους οντολογιών. Επίσης, μια γενική οντολογία μπορεί να επαναχρησιμοποιηθεί με σκοπό να επεκταθεί ώστε να περιγραφούν νέα πεδία ενδιαφέροντος.

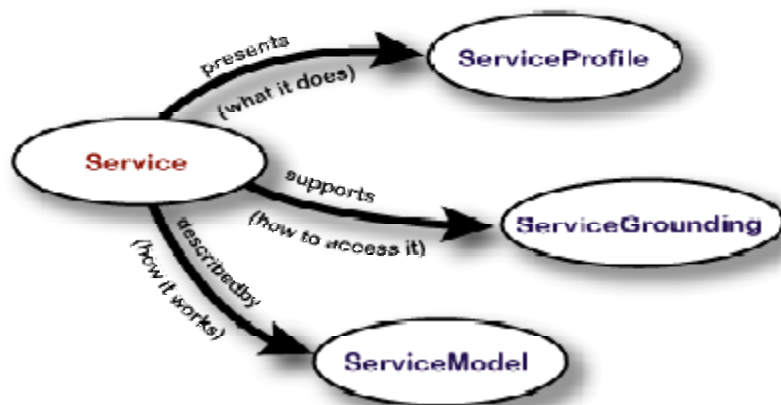
3. Δημιουργία ρητών υποθέσεων για κάθε γνωστικό πεδίο, έτσι ώστε αν χρειαστεί να αλλάξει κάποια από τις υποθέσεις, εφόσον αλλάξει η γνώση που έχουμε για το πεδίο αυτό, να μπορεί να υλοποιηθεί πολύ εύκολα η αλλαγή αυτή. Επιπλέον, βοηθούν το νέο χρήστη με περιορισμένη γνώση, ώστε να μπορέσει να αντιληφθεί ευκολότερα και με σαφήνεια την έννοια και τη σημασία των όρων του πεδίου.

² Liyang Yu, 2007, Semantic Web and Semantic Web Services

4. Διαχωρισμός της γνώσης του κάθε πεδίου από τη λειτουργική γνώση: Μπορεί, για παράδειγμα, να γίνει η περιγραφή μιας εργασίας για τη συναρμολόγηση ενός προϊόντος από τα συστατικά του σύμφωνα με κάποιες προδιαγραφές, και να υλοποιηθεί και ένα πρόγραμμα το οποίο κάνει τη διαδικασία συναρμολόγησης και ρύθμισης ανεξάρτητη από τα εκάστοτε συστατικά.

5. Ανάλυση της γνώσης του κάθε πεδίου: Αυτό μπορεί να γίνει πολύ εύκολα σε περίπτωση που είναι διαθέσιμες και σαφείς οι προδιαγραφές για την ορολογία που χρησιμοποιείται.

Συχνά στόχος δεν είναι η ίδια η οντολογία ενός συγκεκριμένου πεδίου. Η ανάπτυξη μιας οντολογίας είναι όμοια με το καθορισμό ενός συνόλου δεδομένων και της δομής τους για την χρησιμοποίησή τους από άλλα προγράμματα. Οι μέθοδοι επίλυσης προβλήματος, οι εφαρμογές ανεξαρτήτου πεδίου, και οι πράκτορες λογισμικού χρησιμοποιούν τις οντολογίες και τις βάσεις γνώσεων που χτίζονται από τις οντολογίες ως δεδομένα.



Εικόνα 6: Επίπεδα ανάπτυξης μιας οντολογίας
<http://nefeli.lib.teicrete.gr/>

3.1.1 Η μορφή μιας οντολογίας και τα βασικά συστατικά της

Η γνώση στις οντολογίες τυποποιείται χρησιμοποιώντας 5 κατηγορίες συστατικών³:

1. **τις κλάσεις (classes):** Οι κλάσεις αναπαριστώνται με τις έννοιες. Οι έννοιες χρησιμοποιούνται με την ευρεία σημασία τους. Μια έννοια μπορεί να είναι οτιδήποτε για κάτι που λέγεται και γι' αυτό το λόγο θα μπορούσε επίσης να είναι η περιγραφή μιας εργασίας, μιας λειτουργίας, μιας ενέργειας, μιας ιδέας, μιας κρίσης κλπ.

Οι έννοιες είναι δυνατόν να διαιρεθούν σε δύο κατηγορίες:

α) τις πρωταρχικές έννοιες (primitive concepts), οι οποίες έχουν μόνο απαραίτητες συνθήκες (ως προς τις ιδιότητές τους), για να είναι μέλος μιας κλάσης,

β) τις έννοιες εξ' ορισμού (defined concepts), των οποίων η περιγραφή είναι ικανή και αναγκαία συνθήκη, για να είναι ένα αντικείμενο μέλος της κλάσης.

2. **τις σχέσεις (relations):** Οι σχέσεις εκφράζουν ένα είδος αλληλεπίδρασης μεταξύ των εννοιών ενός πεδίου (π.χ. subclass-of, is-a).

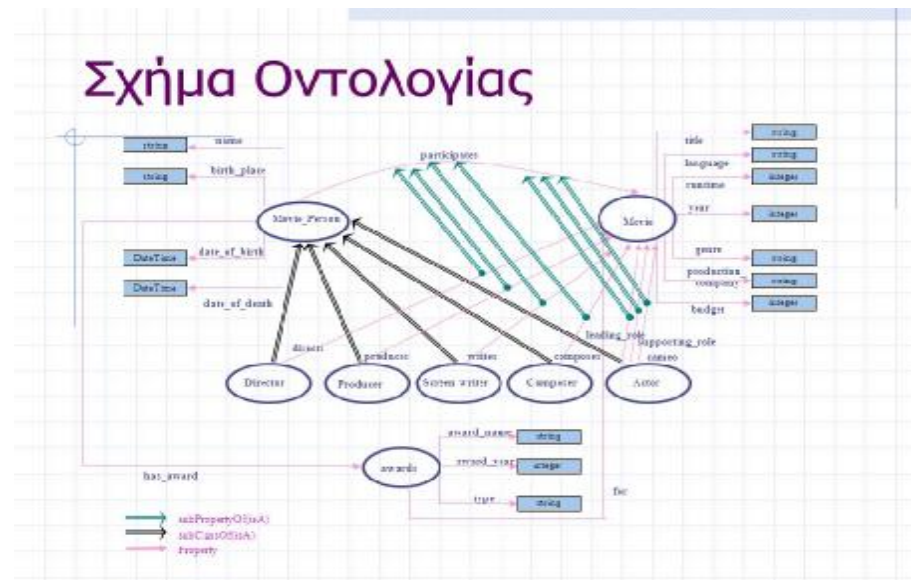
3. **τις συναρτήσεις (functions):** Οι συναρτήσεις εκπροσωπούν μια ειδική περίπτωση σχέσης, στην οποία το n -οστό στοιχείο της σχέσης προσδιορίζεται μοναδικά από τα $n-1$ προηγούμενα στοιχεία. Για παράδειγμα, η τιμή-μεταχειρισμένου- αυτοκινήτου μπορεί να

³ T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5): 34-43, 2001

προσδιορίζεται ως συνάντηση της αρχικής τιμής του καινούριου αυτοκινήτου, του μοντέλου και των χαρακτηριστικών του αυτοκινήτου, καθώς και των χιλιομέτρων που έχει διανύσει.

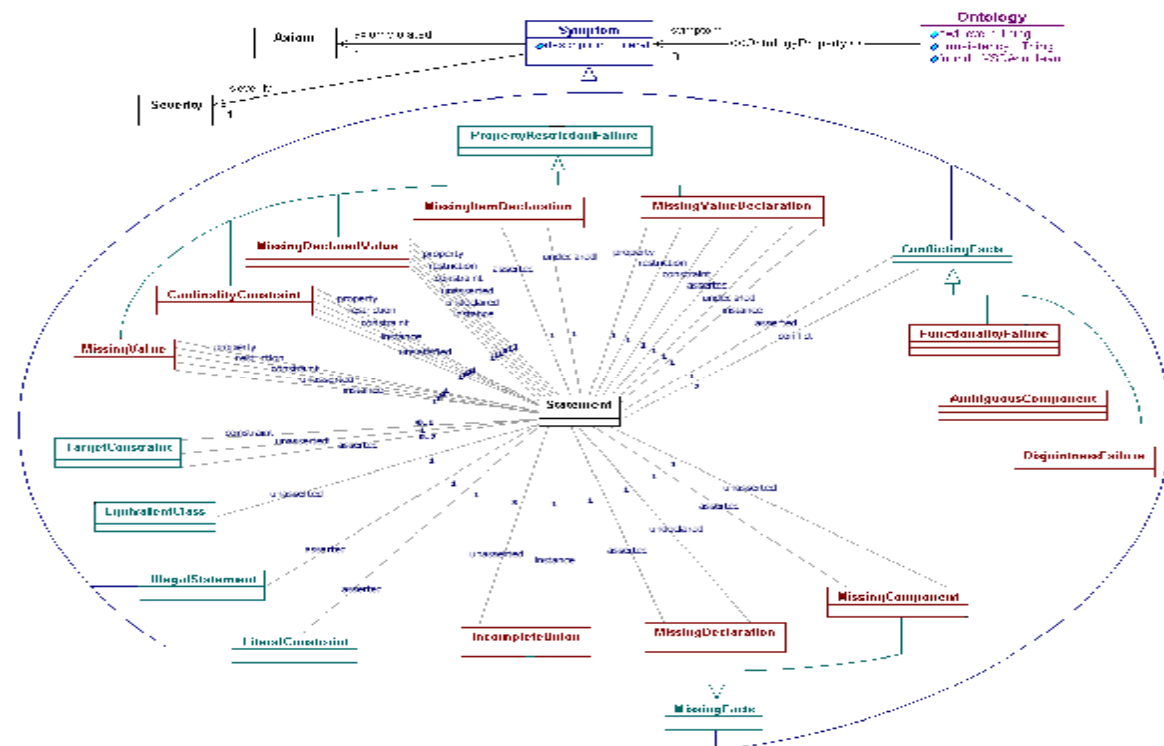
4. **τα αξιώματα (axioms):** Τα αξιώματα χρησιμοποιούνται προκειμένου να αναπαριστούν προτάσεις που είναι πάντοτε αληθείς. Για παράδειγμα, αν ο X είναι δευτεροετής φοιτητής, τότε μπορεί να παρακολουθήσει το επιλεγόμενο μάθημα M.

5. **τα στιγμιότυπα (instances):** Τέλος, τα στιγμιότυπα εκφράζουν συγκεκριμένα στοιχεία, π.χ. ο φοιτητής με το όνομα “Πέτρος” είναι στιγμιότυπο της κλάσης “φοιτητής”.



Εικόνα 7: Σχήμα οντολογίας για κινηματογραφικές ταινίες

<http://p-comp.di.uoa.gr>



Εικόνα 8: Σχήμα 2

<http://p-comp.di.uoa.gr>

3.2 Μεθοδολογία ανάπτυξης οντολογιών

Ακολούθως αναφέρουμε συνοπτικά βασικές μεθοδολογίες ανάπτυξης οντολογιών.

Η μεθοδολογία Uschold και King's

Αυτή η μεθοδολογία είναι βασισμένη στην εμπειρία που αποκτήθηκε κατά τη διάρκεια ανάπτυξης της οντολογίας με το όνομα *Enterprise Ontology*. Η οντολογία αυτή αφορά την μοντελοποίηση διαδικασιών μέσα σε μια επιχείρηση (Uschold and King, 1995). Αυτή η μεθοδολογία παρέχει οδηγίες για την ανάπτυξη οντολογιών, που είναι οι εξής:

1. Αναγνώριση του σκοπού (*purpose identification*): Είναι σημαντικό να είναι ξεκάθαρος ο λόγος για τον οποίο κατασκευάζεται μια οντολογία και ποιοι είναι οι προτιθέμενοι χρήστες της.

2. Χτίσιμο της οντολογίας (*ontology build*):

2.1 Σύλληψη (*capture*) της οντολογίας. Το στάδιο αυτό περιέχει:

- i. Αναγνώριση των σημαντικών εννοιών και των σχέσεών τους μέσα στο πεδίο ενδιαφέροντος, δηλ. ορισμός πεδίου δράσης.
- ii. Παραγωγή με ακρίβεια και σαφήνεια ορισμών κειμένου για τις έννοιες και τις σχέσεις.
- iii. Αναγνώριση όρων που αναφέρονται σε τέτοιες έννοιες και σχέσεις.
- iv. Συμφωνία όλων των παραπάνω.

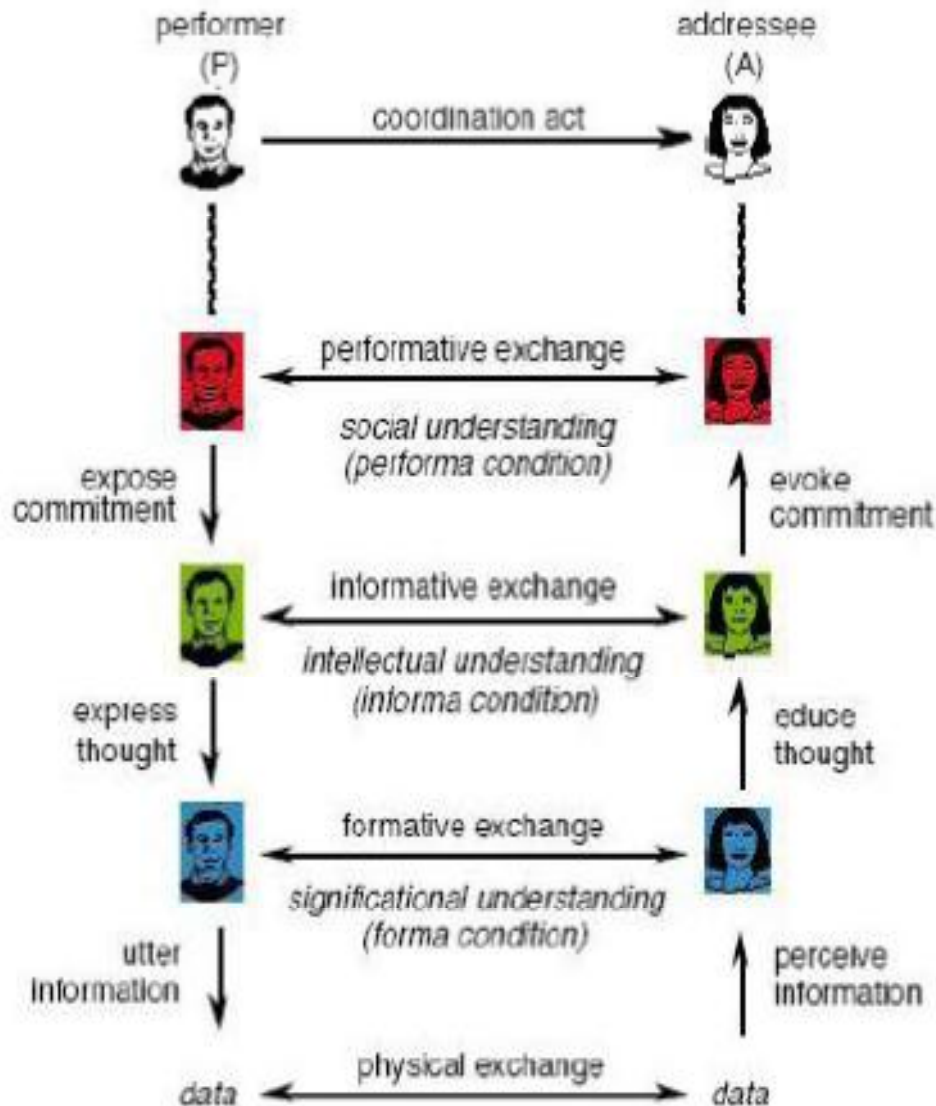
2.2 Κωδικοποίηση (*coding*): Αποσκοπεί στην σαφή αναπαράσταση της γνώσης που αποκτιέται στο βήμα 2.1 με χρήση τυπικής γλώσσας.

2.3 Ενοποίηση (*integration*) υπάρχοντων οντολογιών: Κατά τη διάρκεια των διαδικασιών της σύλληψης και/ή της κωδικοποίησης, υπάρχει το ερώτημα για το αν και πως χρησιμοποιεί κανείς οντολογίες που ήδη υπάρχουν. Δεν περιγράφεται όμως η διαδικασία ενοποίησης με σαφή βήματα.

3. Αξιολόγηση (*evaluation*): Κριτική αξιολόγηση των οντολογιών, του περιβάλλοντος ανάπτυξης οντολογιών, και των κειμένων τεκμηρίωσης, σε σχέση με ένα πλαίσιο αναφοράς. Το πλαίσιο αναφοράς μπορεί να είναι «καθορισμός απαιτήσεων» (*requirements specification*), «ερωτήσεις επάρκειας» (*competency questions*), ή/και «ο πραγματικός κόσμος» (*the world*) (Gómez-Perez et al., 1995).

4. Τεκμηρίωση (*documentation*): Τελική τεκμηρίωση των οντολογιών.

Όπως ειπώθηκε, το σημαντικότερο έργο που αναπτύχθηκε με τη χρήση αυτής της μεθοδολογίας είναι η «Enterprise Ontology», η οποία είναι μια συλλογή από όρους και ορισμούς σχετικούς με οικονομικές εταιρείες. Το σημαντικότερο εργαλείο που αναπτύχθηκε με τη χρήση αυτής της μεθοδολογίας είναι το Enterprise Toolset το οποίο αποτελείται από ένα κατασκευαστή διαδικασιών (*Procedure Builder*) για την σύλληψη μοντέλων διαδικασιών, ένα σύνολο από εργαλεία για Πράκτορες (*Agent Toolkit*) για την υποστήριξη της ανάπτυξης πρακτόρων, ένας διαχειριστής έργων (*Task Manager*) για ενοποίηση, απεικόνιση, και υποστήριξη της θέσπισης διαδικασιών, και τέλος μία (επιχειρηματική) οντολογία για επικοινωνία.



Εικόνα 9: Enterprise Ontology

http: [//www.theenterprisearchitect.eu/blog/2009/10/10/modeling-an-organization-using-enterprise-ontology/](http://www.theenterprisearchitect.eu/blog/2009/10/10/modeling-an-organization-using-enterprise-ontology/)

Η μεθοδολογία Gröninger and Fox's

Αυτή η μέθοδος είναι βασισμένη στην εμπειρία που αποκτήθηκε κατά τη διάρκεια ανάπτυξης της οντολογίας για το έργο TOVE [Gröninger et al., 1995] σχετικά με επιχειρηματικές διαδικασίες και δραστηριότητες μοντελοποίησης. Πρώτα, χτίζεται μία άτυπη περιγραφή των εννοιών της οντολογίας, και μετά αυτή η περιγραφή γίνεται τυπική. Τα βήματα που προτείνονται είναι τα εξής:

1. Σύλληψη των σεναρίων κινήτρων (motivation scenarios capture): Με βάση τους Gröninger and Fox, η ανάπτυξη των οντολογιών παρακινείται από σεναρία που εμφανίζονται στην εφαρμογή. Τα σεναρία κινήτρων είναι προβλήματα-ιστορίες ή παραδείγματα που δεν είναι ικανοποιητικά διευθετημένα από υπάρχουσες οντολογίες. Ένα υποκινούμενο σενάριο επίσης παρέχει ένα σύνολο από διαισθητικά πιθανές λύσεις στα προβλήματα του σεναρίου. Αυτές οι λύσεις παρέχουν μία μη τυπική προτιθέμενη ερμηνεία για τα αντικείμενα και τις σχέσεις τους που αργότερα θα συμπεριληφθούν στην οντολογία. Κάθε πρόταση για

καινούργια οντολογία ή για προέκταση υπάρχουσας θα πρέπει να περιγράφει ένα ή περισσότερα σενάρια κινήτρων, και το σύνολο των προτιθέμενων λύσεων των προβλημάτων που παρουσιάζονται στα σενάρια.

2. Τυποποίηση των άτυπων ερωτημάτων επάρκειας (formulation of competency questions): Αυτά είναι βασισμένα στα σενάρια που έχουν δημιουργηθεί στο προηγούμενο βήμα και μπορούν να θεωρηθούν ως απαιτήσεις εκφραστικότητας που έχουν μορφή. Μία οντολογία πρέπει να μπορεί να απαντήσει σε αυτές τις ερωτήσεις χρησιμοποιώντας την ορολογία της, και να μπορεί να περιγράφει τις απαντήσεις χρησιμοποιώντας τα αξιώματα και τους ορισμούς. Αυτές είναι οι μη τυπικές ερωτήσεις επάρκειας, αφού ακόμα δεν έχουν εκφραστεί στην τυπική γλώσσα της οντολογίας. Οι ερωτήσεις επάρκειας τοποθετούνται σε ιεραρχία έτσι ώστε η απάντηση σε μία ερώτηση να μπορεί να χρησιμοποιηθεί για να απαντηθούν πιο γενικές ερωτήσεις από την ίδια ή από άλλη οντολογία, με τη χρήση τελεστών σύνθεσης και αποσύνθεσης. Αυτός είναι ένας τρόπος για την ταυτοποίηση της γνώσης που έχει ήδη αναπαρασταθεί με σκοπό την επαναχρησιμοποίηση και ενοποίηση των οντολογιών. Οι ερωτήσεις παίζουν το ρόλο των περιορισμών για το τι μπορεί να περιγράφει μία οντολογία, παρά του καθορισμού ενός συγκεκριμένου σχεδίου για τις αντίστοιχες οντολογικές δεσμεύσεις. Δεν υπάρχει μία μοναδική οντολογία που είναι συνδεδεμένη με ένα σύνολο από ερωτήσεις επάρκειας. Αντίθετα, οι ερωτήσεις επάρκειας χρησιμοποιούνται για την αξιολόγηση των οντολογικών δεσμεύσεων που έχουν γίνει ώστε να γίνει γνωστό αν η οντολογία τηρεί τις απαιτήσεις.

3. Ορισμός της ορολογίας της οντολογίας με τυπική γλώσσα (Ontology specification with formal language). Τα βήματα που ακολουθούνται είναι τα εξής:

I. Αναζήτηση της άτυπης ορολογίας. Όταν οι άτυπες ερωτήσεις επάρκειας είναι διαθέσιμες, το σύνολο των όρων που χρησιμοποιείται μπορεί να εξαχθεί από αυτές. Αυτοί οι όροι αποτελούν την βάση για τον ορισμό της ορολογίας σε μία τυπική γλώσσα.

II. Ορισμός της τυπικής ορολογίας. Αφού έχουν διατυπωθεί οι άτυπες ερωτήσεις επάρκειας, η ορολογία της οντολογίας ορίζεται με τη χρήση μίας τυπικής γλώσσας όπως είναι η KIF (Genesereth and Fikes, 1992). Αυτή η ορολογία θα επιτρέψει να εκφραστούν οι ορισμοί και οι περιορισμοί αργότερα με τη χρήση αξιωμάτων.

4. Τυποποίηση των τυπικών ερωτήσεων επάρκειας με τη χρήση της ορολογίας της οντολογίας. Αφού οι ερωτήσεις επάρκειας έχουν διατυπωθεί με άτυπο τρόπο και η ορολογία της οντολογίας έχει οριστεί, οι ερωτήσεις επάρκειας ορίζονται τώρα με τυπικό τρόπο.

5. Καθορισμός αξιωμάτων και ορισμών για την ορολογία της οντολογίας με τυπική γλώσσα. Τα αξιώματα καθορίζουν τους ορισμούς των όρων της οντολογίας και περιορίζουν την ερμηνεία τους. Εάν τα προτεινόμενα αξιώματα δεν επαρκούν για να αναπαραστήσουν τις τυπικές ερωτήσεις επάρκειας και να χαρακτηρίσουν τις απαντήσεις των ερωτήσεων, τότε πρέπει να προστεθούν συμπληρωματικά αξιώματα ή αντικείμενα στην οντολογία. Η ανάπτυξη των αξιωμάτων για την οντολογία με βάση τις ερωτήσεις επάρκειας είναι μία επαναληπτική διαδικασία.

6. Καθορισμός συνθηκών για τον καθορισμό της πληρότητας της οντολογίας (ontology completeness). Όταν οι ερωτήσεις επάρκειας έχουν και τυπικά οριστεί, πρέπει να οριστούν οι συνθήκες κάτω από τις οποίες οι απαντήσεις στις ερωτήσεις είναι πλήρεις. Η μεθοδολογία αυτή χρησιμοποιήθηκε για την κατασκευή των οντολογιών στο έργο TOVE (Toronto Virtual Enterprise) στο πανεπιστήμιο του Τορόντο.

Η μεθοδολογία ON-TO-KNOWLEDGE

Η μεθοδολογία On – To – Knowledge αναπτύχθηκε και εφαρμόστηκε στο ευρωπαϊκό έργο Ev IST-199910132 On – To – Knowledge (<http://www.ontoknowledge.org>) με σκοπό να εισάγει και να συντηρεί επιχειρηματικές εφαρμογές διαχείρισης γνώσης βασισμένες σε οντολογίες. Σε αντίθεση με άλλες μεθοδολογίες μηχανικής οντολογιών οι οποίες περιορίζουν την προσοχή τους μέσα στην ίδια την οντολογία, η μεθοδολογία On – To – Knowledge εστιάζει στην ανάπτυξη των οντολογιών με γνώμονα την εφαρμογή (Application driven) διαχείρισης γνώσης βασισμένη σε οντολογίες (ontology - based knowledge management application). Η μεθοδολογία αυτή καλύπτει θέματα από τις αρχικές φάσεις ανάπτυξης ενός έργου διαχείρισης γνώσης μέχρι τη τελική φάση δημιουργίας ενός συστήματος διαχείρισης γνώσης βασισμένο σε οντολογίες. Η μεθοδολογία εστιάζει σε διαδικασίες γνώσης (Knowledge Processes) και μετά-διαδικασίες γνώσης (Knowledge Meta-processes).

Οι πρώτες διαδικασίες έχουν να κάνουν με τη χρήση των οντολογιών, και οι δεύτερες με την καθοδήγηση για την κατασκευή τους. Υπάρχουν διάφορες μεθοδολογίες που υποστηρίζουν τη συστηματική εύρεση λύσεων διαχείρισης γνώσης μέσα σε επιχειρήσεις. Μία από τις πιο αξιοπρόσεκτες είναι η CommonKADS (Schreiber et al., 1999) η οποία δίνει έμφαση σε μία πρώιμη μελέτη σκοπιμότητας (feasibility study) όπως επίσης και στη κατασκευή πολλών μοντέλων που καταγράφουν διαφορετικού τύπου γνώση που χρειάζεται για την δημιουργία μίας λύσης διαχείρισης γνώσης.

Η μεθοδολογία On – To – Knowledge αποτελείται από πέντε βασικά βήματα. Τα πέντε βασικά βήματα-φάσεις που τελικά οδηγούν σε μία εφαρμογή διαχείρισης γνώσης βασισμένη σε οντολογία είναι:

1. Μελέτη Σκοπιμότητας (Feasibility Study): Πολλοί είναι οι παράγοντες που μπορούν να προσδιορίσουν την επιτυχία ή την αποτυχία ενός συστήματος διαχείρισης γνώσης. Για να αναλυθούν αυτοί οι παράγοντες, πρέπει κανείς να εκπονήσει αρχικά μία μελέτη σκοπιμότητας. Να αναγνωρίσει δηλαδή τα προβλήματα, τις ευκαιρίες, τις πιθανές λύσεις και τις απειλές, και στη συνέχεια να τα τοποθετήσει στην προοπτική εξέλιξης του οργανισμού. Η μελέτη σκοπιμότητας εξυπηρετεί την υποστήριξη στη λήψη αποφάσεων για την οικονομική και τεχνολογική σκοπιμότητα του έργου, με στόχο την επιλογή του πεδίου ενδιαφέροντος για το οποίο θα αναπτυχθεί το σύστημα βασισμένο σε οντολογία. Η μελέτη σκοπιμότητας στηρίζεται στην μελέτη σκοπιμότητας που περιγράφεται από την μεθοδολογία CommonKADS. Παράλληλα με την αναγνώριση του πεδίου ενδιαφέροντος, βοηθάει και στην αναγνώριση των ανθρώπων που εμπλέκονται στην ανάπτυξη και χρήση του συστήματος, δηλαδή στους ειδικούς του πεδίου ενδιαφέροντος, στους χρήστες και στους υποστηρικτές του συστήματος.

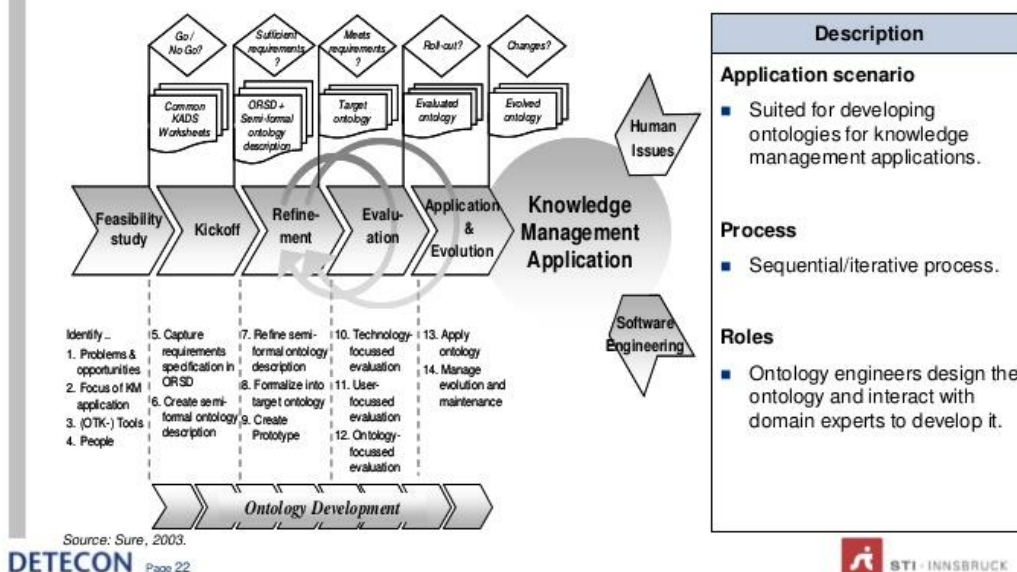
2. Εκκίνηση (Kickoff): Το επόμενο βήμα είναι η σύλληψη και καταγραφή των απαιτήσεων που περιγράφει το τι πρέπει να αναπαριστά η οντολογία. Η καταγραφή αυτή θα πρέπει να καθοδηγεί τον μηχανικό οντολογιών να αποφασίσει για το τι θα αποτελεί την ορολογία και την ιεραρχική δομή των εννοιών μέσα στην οντολογία.

3. Επεξεργασία (Refinement): Ο στόχος της φάσης αυτής είναι η παραγωγή μίας ώριμης και βασισμένης στην εφαρμογή (application oriented) οντολογίας-στόχου (target ontology) με βάση τον ορισμό που δόθηκε στην φάση Εκκίνησης. Η φάση αυτή αποτελείται από δύο διαδικασίες, (α) τη διαδικασία εξόρυξης γνώσης από τους ειδικούς του πεδίου ενδιαφέροντος, (β) τη διαδικασία τυποποίησης για την μετατροπή της αρχικής οντολογίας σε οντολογία-στόχο, εκφρασμένη σε τυπική γλώσσα αναπαράστασης γνώσης.

Ontology Engineering Methodologies

Methodologies Related to Knowledge Management Systems

The On-To-Knowledge methodology takes a pragmatic approach to ontology engineering and contains many useful tips to support non-experts to build an ontology.



Εικόνα 10: ON-TO-KNOWLEDGE

<http://www.mkbergman.com/906/a-brief-survey-of-ontology-development-methodologies/>

4. Αξιολόγηση (Evaluation): Η φάση αξιολόγησης εξυπηρετεί την απόδειξη της χρησιμότητας των αναπτυγμένων οντολογιών. Στα πρώτα βήματα της φάσης, ο μηχανικός οντολογιών ελέγχει αν η οντολογία-στόχος ικανοποιεί τις οντολογικές απαιτήσεις και αν η οντολογία υποστηρίζει ή απαντά στις ερωτήσεις επάρκειας που αναλύθηκαν στην φάση Εκκίνησης. Σε ένα δεύτερο βήμα, η οντολογία δοκιμάζεται μέσα στο περιβάλλον της εφαρμογής. Η ανάδραση πληροφορίας από χρήστες-ελεγκτές αποτελεί χρήσιμη γνώση για περαιτέρω επεξεργασία και εξέλιξη της οντολογίας.

5. Συντήρηση (Maintenance): Στο πραγματικό κόσμο τα πράγματα εξελίσσονται, και το ίδιο συμβαίνει με τη γνώση για τα πράγματα αυτά και με τις οντολογίες που αναπαριστούν τη γνώση αυτή. Για να ενσωματώνονται αυτές οι αλλαγές, οι οντολογίες πρέπει να συντηρούνται συχνά, όπως γίνεται και με τη συντήρηση ενός προγράμματος (software maintenance). Η συντήρηση και διαχείριση εκδόσεων των οντολογιών είναι πρωταρχική διαδικασία ενός οργανισμού.

Η μεθοδολογία αυτή αναπτύχθηκε και εφαρμόστηκε στο έργο On – To – Knowledge. Μία από τις βασικές συνεισφορές της μεθοδολογίας είναι η σύνδεση της διαθέσιμης υποστηρίξης εργαλείων του On – To – Knowledge με τις μελέτες περίπτωσης δείχνοντας το πώς και πότε πρέπει να χρησιμοποιούνται τα εργαλεία κατά τη διάρκεια της ανάπτυξης και της εκτέλεσης των εφαρμογών διαχείρισης γνώσης που βασίζονται σε οντολογίες. Μία άλλη βασική συνεισφορά είναι η υποστήριξη της από το OntoEdit , ένα προηγμένο περιβάλλον οντολογικής μηχανικής που όπως θα δούμε παρακάτω υποστηρίζει κάθε μία από τις φάσεις της οντολογίας.

3.3 Βασικές αρχές ανάπτυξης οντολογιών

Με τα όσα έχουν παρουσιαστεί έως τώρα για τις οντολογίες γίνεται αντιληπτό στον αναγνώστη ότι δεν υπάρχει κανένας απόλυτα «σωστός» τρόπος ή μεθοδολογία για την ανάπτυξη των οντολογιών. Στην παρούσα ενότητα, θα γίνει αναφορά στους τρόπους, στις μεθοδολογίες και στα κριτήρια που έχουν προταθεί από τους ερευνητές για τη διαδικασία ανάπτυξης μιας οντολογίας.

Στην παρούσα φάση κρίνεται σκόπιμο να επισημανθούν μερικοί θεμελιώδεις κανόνες που αφορούν στο σχεδιασμό της οντολογίας. Αυτοί οι κανόνες μπορούν να φανούν μάλλον δογματικοί αλλά θα βοηθήσουν αρκετές φορές στο να λάβουμε κάποιες αποφάσεις στο σχεδιασμό.

1. Δεν υπάρχει κανένας σωστός τρόπος στο να διαμορφωθεί ένα πεδίο γνώσης - υπάρχουν πάντα βιώσιμες εναλλακτικές λύσεις. Η καλύτερη λύση σχεδόν πάντα εξαρτάται από την εφαρμογή που έχει κατά νου ο δημιουργός και τις επεκτάσεις που πιθανόν συμβούν.

2. Η ανάπτυξη οντολογίας είναι απαραίτητως μια επαναληπτική διαδικασία.

3. Οι έννοιες στην οντολογία πρέπει να είναι κοντά στα αντικείμενα (φυσικά ή λογικά) και οι σχέσεις στο πεδίο ενδιαφέροντος του συγκεκριμένου πεδίο γνώσης. Στις προτάσεις που περιγράφουν ένα πεδίο γνώσης, συνήθως, για την περιγραφή των αντικειμένων (objects) χρησιμοποιούμε ουσιαστικά (αντικείμενα), ενώ για την περιγραφή των σχέσεων (relationships) χρησιμοποιούμε ρήματα.

Η απόφαση για το που πρόκειται να χρησιμοποιήσουμε την οντολογία και πόσο λεπτομερής ή γενική θα είναι η οντολογία πρόκειται να καθοδηγήσει τον δημιουργό σε πολλές από τις αποφάσεις διαμόρφωσης που θα προκύψουν κατά τη διαδικασία ανάπτυξης. Μεταξύ διάφορων βιώσιμων εναλλακτικών λύσεων στην ανάπτυξη της οντολογίας, θα πρέπει να καθορίσει ο δημιουργός ποια λύση θα λειτουργούσε καλύτερα για την εργασία που προορίζεται, ποια θα ήταν πιο ενορατική λύση, πιο εξελίξιμη και πιο συντηρήσιμη.

Χρειάζεται, επίσης, ο δημιουργός μιας οντολογίας να θυμάται ότι μια οντολογία είναι ένα μοντέλο της πραγματικότητας του κόσμου και οι έννοιες στην οντολογία πρέπει να απεικονίσουν αυτήν την πραγματικότητα. Αφότου καθοριστεί μια αρχική έκδοση της οντολογίας, έπεται η αξιολόγησή της και η διόρθωσή της με τη χρησιμοποίηση της στις εφαρμογές ή με μεθόδους επίλυσης προβλημάτων (problem – solving methods) ή με κάποια πιθανή συζήτηση με τους εμπειρογνώμονες του συγκεκριμένου τομέα.

Σχεδόν, πάντα, μετά από τις παραπάνω διαδικασίες θα πρέπει να γίνεται αναθεώρηση της αρχικής οντολογίας. Αυτή η διαδικασία του επαναληπτικού σχεδιασμού θα συνεχιστεί πιθανώς σε ολόκληρο τον κύκλο της ζωής της οντολογίας.

Οι επίσημες οντολογίες σχεδιάζονται, όπως αναφέρει ο Thomas Gruber στο Concepts and Abstract Syntax⁴. Όταν επιλέγουμε πώς να αντιπροσωπεύσουμε κάτι σε μια οντολογία, λαμβάνουμε αποφάσεις σχεδιασμού. Για να καθοδηγήσουμε και να αξιολογήσουμε τις ενέργειες αυτές, χρειαζόμαστε αντικειμενικά κριτήρια που θεμελιώνονται με σκοπό το δημιουργήμα, παρά στις a priori έννοιες της φυσικότητας ή της αλήθειας. Εδώ προτείνουμε ένα αρχικό σύνολο κριτηρίων σχεδιασμού για τις οντολογίες που στοχεύουν στη διανομή γνώσης και στη διαλειτουργικότητα μεταξύ των προγραμμάτων που βασίζονται σε μια κοινή εννοιολογική αναπαράσταση.

1. Σαφήνεια (Clarity): Μια οντολογία πρέπει να εκφράζει αποτελεσματικά την έννοια των καθορισμένων όρων. Οι ορισμοί πρέπει να είναι αντικειμενικοί. Ενώ το κίνητρο για τον καθορισμό μιας έννοιας πρέπει να προκύπτει από τις κοινωνικές καταστάσεις ή τις

⁴ "Resource Description Framework (RDF): Concepts and Abstract Syntax", W3C Recommendation, 10 February 2004, <http://travesia.mcu.es/portallnb/jspu/bitstream/10421/2427/1/rdf-concepts%20and%20abstract%20syntax.pdf>

υπολογιστικές απαιτήσεις, ο καθορισμός πρέπει να είναι ανεξάρτητος από το κοινωνικό ή υπολογιστικό πλαίσιο. Ο φορμαλισμός είναι ένας τρόπος για τον σκοπό αυτό. Όπου είναι δυνατόν, ένας πλήρης καθορισμός (ένα κατηγορημα που καθορίζεται από τους απαραίτητους και ικανοποιητικούς όρους) προτιμάται πέρα από έναν μερικό καθορισμό (που καθορίζεται από μόνο τους απαραίτητους ή ικανοποιητικούς όρους). Όλοι οι ορισμοί θα πρέπει να τεκμηριώνονται με τη φυσική γλώσσα.

2. Συνοχή (Coherence): Μια οντολογία πρέπει να είναι συνεπής: δηλαδή πρέπει να εγκρίνει τα συμπεράσματα που είναι σύμφωνα με τους ορισμούς. Σε κάθε περίπτωση, τα αξιώματα καθορισμού πρέπει να είναι λογικά συνεπή. Η συνοχή πρέπει επίσης να ισχύει για τις έννοιες που καθορίζονται ανεπίσημα, όπως εκείνοι που περιγράφονται στην τεκμηρίωση και τα παραδείγματα φυσικής γλώσσας. Εάν μια πρόταση που μπορεί να προκύψει από τα αξιώματα έρχεται σε αντίθεση με έναν καθορισμό ή ένα παράδειγμα που δίνεται ανεπίσημα, τότε η οντολογία είναι ασυνάρτητη.

3. Επεκτασιμότητα (Extendibility): Μια οντολογία θα πρέπει να σχεδιάζεται με τέτοιο τρόπο ώστε να προβλέπει τις χρήσεις του κοινού λεξιλογίου. Πρέπει να προσφέρει μία εννοιολογική θεμελίωση για μια σειρά προσδοκώμενων στόχων, και η αναπαράσταση πρέπει να επεξεργαστεί έτσι ώστε να μπορεί κάποιος να επεκτείνει και να ειδικεύσει την οντολογία μονοτονικά. Με άλλα λόγια, κάποιος πρέπει να είναι σε θέση να καθορίσει νέους όρους για πρόσθετες χρήσεις που βασίζονται στο υπάρχον λεξιλόγιο, με τέτοιο τρόπο ώστε αυτός να μην απαιτεί την αναθεώρηση των υπάρχοντων ορισμών.

4. Ελάχιστο εύρος κωδικοποίησης (Minimal Encoding bias): Η κωδικοποίηση της οντολογίας, θα πρέπει να γίνεται με τον απλούστερο δυνατό τρόπο, απαιτώντας όσο το δυνατόν μικρότερους και απλούστερους κώδικες για την αναπαράστασή της.

5. Ελάχιστη οντολογική δέσμευση (Minimal Ontological Commitment): Μία οντολογία θα πρέπει να απαιτεί την ελάχιστη οντολογική δέσμευση, η οποία θα είναι ικανή να υποστηρίξει τις επιθυμητές λειτουργίες διαμοιρασμού της γνώσης. Θα πρέπει να έχει κάποιες αξιώσεις σχετικά με το χώρο ο οποίος μοντελοποιείται, επιτρέποντας στα μέρη που δεσμεύονται στην ελευθερία της οντολογίας να εξειδικεύουν και να εγκαθιστούν την οντολογία όπως πραγματικά χρειάζεται.

Η μηχανική ανάπτυξης οντολογιών είναι ακόμα ένας «ανώριμος επιστημονικός κλάδος» και διάφορες ερευνητικές ομάδες προτείνουν διάφορες μεθόδους πιο γνωστές ως μεθοδολογίες για την ανάπτυξη οντολογιών. Εντούτοις, οι μεταβλητές είναι τόσες πολλές που μπορεί να είναι αδύνατον να βρεθεί μια ενιαία μέθοδος που να είναι επαρκής για όλες τις καταστάσεις. Πιθανώς, η καλύτερη λύση θα είναι μια επιλογή μεταξύ διαφορετικών δυνατοτήτων ή μια σύνθεση των διαφορετικών μεθόδων ανάπτυξης οντολογίας.

3.4 Περιγραφή της Οντολογικής Γλώσσας OWL

Η γλώσσα οντολογίας Ιστού OWL είναι μια επίσημη γλώσσα αναπαράστασης οντολογιών στον Σημασιολογικό Ιστό. Η OWL είναι μια γλώσσα οντολογίας για τον Σημασιολογικό Ιστό, που αναπτύχθηκε από την ομάδα εργασίας οντολογιών Ιστού (Web Ontology Working Group) της Κοινότητας Παγκόσμιου Ιστού (W3C). Η OWL σχεδιάστηκε, κυρίως, για την αναπαράσταση πληροφοριών σχετικών με κατηγορίες αντικειμένων και του τρόπου διασύνδεσης των αντικειμένων, δηλαδή το είδος των πληροφοριών που ονομάζουμε οντολογίες. Η OWL έχει χαρακτηριστικά από διάφορες οικογένειες γλωσσών αναπαράστασης, περιλαμβάνοντας, κυρίως, Περιγραφικές Λογικές και πλαίσια. Η OWL μοιράζεται, επίσης, πολλά χαρακτηριστικά με το with RDF. Σε αυτήν την ενότητα περιγράφονται η φιλοσοφία και τα χαρακτηριστικά της OWL και πώς μπορούν να εντοπιστούν σε παλιότερους φορμαλισμούς, με μετατροπές λόγω ορισμένων περιορισμών στην OWL. Έχουν προκύψει διάφορα προβλήματα που προκαλούν ενδιαφέρον, λόγω των οποίων αυτές οι επιρροές έρχονταν σε διαφωνία με τη γλώσσα. Η OWL μπορεί, επίσης, να αναπαριστά πληροφορίες σχετικά με τα ίδια τα αντικείμενα, το είδος, δηλαδή, των πληροφοριών που ονομάζουμε δεδομένα.

Η OWL δεν σχεδιάστηκε εν κενώ. Υπήρχαν πολλές επιρροές στον σχεδιασμό της. Καθώς η OWL είναι μία προσπάθεια της δραστηριότητας του W3C για τον Σημασιολογικό Ιστό, έπρεπε να ταιριάζει στο όραμα του Σημασιολογικού Ιστού μιας στοίβας γλωσσών που περιλαμβάνει την XML και το RDF. Καθώς η OWL προοριζόταν ως μια γλώσσα οντολογίας, έπρεπε να μπορεί να αναπαριστά ένα χρήσιμο σύνολο χαρακτηριστικών οντολογίας. Καθώς υπήρχαν ήδη διάφορες γλώσσες οντολογίας σχεδιασμένες για χρήση στον Ιστό, η OWL έπρεπε να διατηρεί τόση συμβατότητα, όση ήταν δυνατή από τις ισχύουσες γλώσσες οντολογίας, συμπεριλαμβάνοντας τις SHOE, OIL και DAML+OIL.

Οι πολλαπλές επιδράσεις στην OWL οδήγησαν σε μερικούς δύσκολους συμβιβασμούς. Επίσης, και κάπως αναμενόμενο, έπρεπε να εκτελεστεί σημαντικό τεχνικό έργο για να σχεδιάσουν την OWL με τέτοιο τρόπο, ώστε να φανεί ότι έχει διάφορα επιθυμητά χαρακτηριστικά, εξασφαλίζοντας, παράλληλα, επαρκή συμβατότητα με τις ρίζες του. Στο επόμενο κεφάλαιο περιγράφονται μερικοί από τους συμβιβασμούς και οι αποφάσεις σχεδιασμού που έπρεπε να παρθούν από την ομάδα εργασίας οντολογιών Ιστού κατά τον σχεδιασμό της OWL. Αν και πολλές από αυτές τις αποφάσεις βασίστηκαν στις απαιτήσεις που είχαν καθοριστεί για την OWL και σε στέρεες επιστημονικές γνώσεις, μερικές από αυτές βασίστηκαν, αναγκαστικά, σε πιο ήπιες κρίσεις, και μερικές ήταν απλά ζήτημα προτιμήσεων. Στο επόμενο κεφάλαιο γίνεται εκτενέστερη περιγραφή της γλώσσας και των επιρροών της.

3.4.1 Επισκόπηση της OWL

Στο πλαίσιο του Σημασιολογικού Ιστού, οι οντολογίες διαδραματίζουν σημαντικό ρόλο, βοηθώντας τις αυτοματοποιημένες διαδικασίες (τους λεγόμενους «ευφυείς πράκτορες») να αποκτήσουν πρόσβαση σε πληροφορίες. Ειδικότερα, οι οντολογίες αναμένεται να χρησιμοποιηθούν για την παροχή διαρθρωμένων λεξιλογίων που αναπτύσσουν τις σχέσεις μεταξύ των διαφορετικών όρων, επιτρέποντας στους ευφυείς πράκτορες (και ανθρώπους) να ερμηνεύσουν την σημασία τους ευέλικτα αλλά με σαφήνεια. Για παράδειγμα, μια κατάλληλη οντολογία «πίτσα» θα μπορούσε να περιλαμβάνει τις πληροφορίες ότι η Mozzarella και η Gorgonzola είναι είδη τυριών, ότι το τυρί δεν είναι είδος κρέατος ή ψαριού, και ότι μια πίτσα για χορτοφάγους είναι εκείνη της οποίας τα υλικά δεν περιλαμβάνουν κρέας ή ψάρι. Αυτές οι πληροφορίες επιτρέπουν στον όρο "πίτσα με υλικά (μόνο) Mozzarella και Gorgonzola" να ερμηνεύεται με σαφήνεια (από, π.χ., έναν πράκτορα παραγγελίας πίτσα) ως εξειδίκευση του όρου «πίτσα για χορτοφάγους».

Όροι των οποίων η έννοια καθορίζεται με οντολογίες μπορούν να χρησιμοποιηθούν σε σημασιολογικές σημάνσεις που περιγράφουν το περιεχόμενο και τη λειτουργικότητα των προσπελάσιμων από τον Ιστό πόρων. Οι οντολογίες και η σημασιολογική σήμανση με βάση τις οντολογίες μπορούν να χρησιμοποιηθούν:

-στο ηλεκτρονικό εμπόριο, όπου μπορούν να διευκολύνουν την επικοινωνία μεταξύ των πρακτόρων αγοράς και πώλησης παρέχοντας ένα κοινό λεξιλόγιο για την περιγραφή αγαθών και υπηρεσιών,

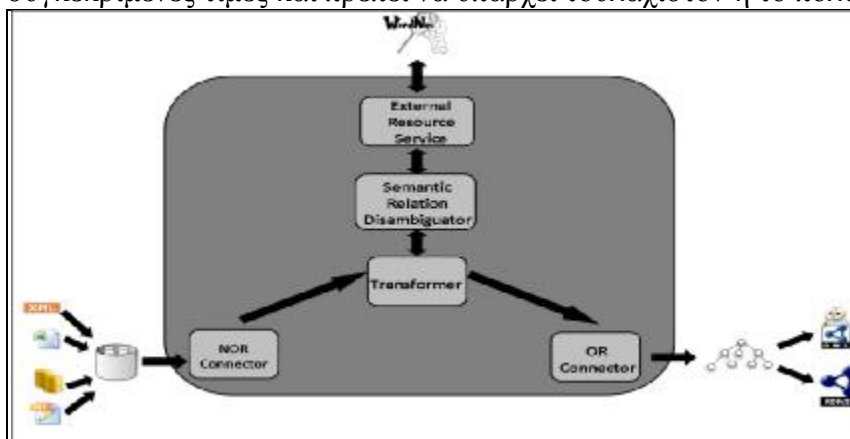
-στις μηχανές αναζήτησης, όπου μπορούν να βοηθήσουν στην εύρεση σελίδων που περιέχουν σημασιολογικά παρόμοιες αλλά συντακτικά διαφορετικές λέξεις και φράσεις και

-στις υπηρεσίες ιστού (*web services*),² όπου μπορούν να προσφέρουν πλούσιες περιγραφές υπηρεσιών που μπορούν να βοηθήσουν στον εντοπισμό των κατάλληλων υπηρεσιών.

Για να υποστηρίξει αυτά και άλλα σενάρια χρήσης, η OWL παίρνει τη βασική δυνατότητα δήλωσης γεγονότων του RDF και τις δυνατότητες δομής κλάσεων και ιδιοτήτων του RDF Schema και τις επεκτείνει με σημαντικούς τρόπους. Η OWL μπορεί να δηλώνει κλάσεις και να τις οργανώνει ιεραρχικά ("σε υποκατηγορίες"), όπως το RDF Schema. Οι κλάσεις της OWL μπορούν να οριστούν ως λογικοί συνδυασμοί (τομές, ενώσεις ή συμπληρώματα) άλλων κλάσεων, ή ως απαριθμήσεις συγκεκριμένων αντικειμένων, επεκτείνοντας τις δυνατότητες του RDFS. Η OWL μπορεί, επίσης, να δηλώνει ιδιότητες, να οργανώνει αυτές τις ιδιότητες σε μια ιεραρχία «υπό-ιδιοτήτων», και να παρέχει τομείς και εύρος για αυτές τις ιδιότητες, και πάλι όπως το RDFS. Οι τομείς των ιδιοτήτων της OWL είναι οι κλάσεις της OWL, και το εύρος μπορεί να είναι είτε κλάσεις της OWL ή εξωτερικά καθορισμένοι τύποι δεδομένων, όπως συμβολοσειρές ή ακέραιοι αριθμοί. Η OWL μπορεί να δηλώσει ότι μια ιδιότητα είναι μεταβατική, συμμετρική, λειτουργική, ή είναι η αντίστροφη μιας άλλης ιδιότητας, επεκτείνοντας, εδώ, και πάλι το RDFS.

Η OWL μπορεί να εκφράσει ποια αντικείμενα ανήκουν σε ποιες κλάσεις, και ποιες είναι οι τιμές των ιδιοτήτων για συγκεκριμένα αντικείμενα. Οι δηλώσεις ισοδυναμίας ισχύουν για τις κλάσεις και τις ιδιότητες, οι δηλώσεις ασυνδετότητας (*disjointness*) μπορούν να εφαρμοστούν μεταξύ των κλάσεων, και η ισότητα και η ανισότητα μπορεί να υποστηριχθεί από τα αντικείμενα.

Ωστόσο, η σημαντική επέκταση του RDFS είναι η ικανότητα της OWL να παρέχει περιορισμούς σχετικά με την συμπεριφορά των ιδιοτήτων τοπικά σε μια κλάση. Η OWL μπορεί να καθορίσει κλάσεις όπου μια συγκεκριμένη ιδιότητα είναι περιορισμένη έτσι, ώστε όλες οι τιμές για την ιδιότητα στα στιγμιότυπα της κλάσης να πρέπει να ανήκουν σε μια συγκεκριμένη κλάση (ή τύπο δεδομένων). Τουλάχιστον μία τιμή πρέπει να προέρχεται από μια συγκεκριμένη κλάση (ή τύπο δεδομένων). Πρέπει να υπάρχουν τουλάχιστον ορισμένες συγκεκριμένες τιμές και πρέπει να υπάρχει τουλάχιστον ή το πολύ ορισμένες διακριτές τιμές.



Εικόνα 11: Παράδειγμα αναπαράστασης οντολογιών

<http://www.csd.uoc.gr>

Για παράδειγμα, χρησιμοποιώντας RDFS μπορούμε:

- Να δηλώσουμε κατηγορίες όπως Country, Person, Student και Canadian,
- Να δηλώσουμε ότι το Student είναι μια υπό-κλάση του Person,
- Να δηλώσουμε ότι τα Canada και England είναι δύο στιγμιότυπα της κλάσης Country,
- Να δηλώσουμε το Nationality ως ιδιότητα που σχετίζεται με τις κλάσεις Person (τον τομέα του) και Country (το εύρος του),
 - Να δηλώσουμε ότι το age είναι μια ιδιότητα, με το Person ως τομέα της και ακέραιος ως το εύρος της, και
 - Να δηλώσουμε ότι ο Πέτρος είναι ένα στιγμιότυπο της κλάσης Canadian, και ότι το age του έχει τιμή 48.

Με την OWL μπορούμε επιπρόσθετα:

- Να δηλώσουμε ότι τα Country και Person είναι ασύνδετες κλάσεις,
- Να δηλώσουμε ότι τα Canada και England είναι διακριτά αντικείμενα,
- Να δηλώσουμε το HasCitizen ως την αντίστροφη ιδιότητα του Nationality,
- Να δηλώσουμε ότι η κλάση Stateless ορίζεται ακριβώς ως εκείνα τα μέλη της κλάσης Person που δεν έχουν τιμές για την ιδιότητα Nationality,
 - Να δηλώσουμε ότι η κλάση MultipleNationals ορίζεται ακριβώς ως εκείνα τα μέλη της κλάσης Person που έχουν τουλάχιστον δύο τιμές για την ιδιότητα Nationality,
 - Να δηλώσουμε ότι η κλάση Canadian ορίζεται ακριβώς ως εκείνα τα μέλη της κλάσης Person που έχουν το Canada ως τιμή για την ιδιότητα Nationality και
 - Να δηλώσουμε ότι το age είναι μια λειτουργική ιδιότητα.

Τα παραπάνω δείχνουν ότι η OWL είναι μια αρκετά δύσκολη γλώσσα. Η OWL έχει τόσο RDF / XML σύνταξη, όσο και μια αφηρημένη σύνταξη παρόμοια πλαισίου και έχει τρεις ονομασμένες υπό-γλώσσες. Αυτή η πολλαπλότητα είναι το άμεσο αποτέλεσμα της προσπάθειας να ικανοποιηθεί μεγάλο αριθμό ενίοτε αλληλοσυγκρουόμενων επιρροών και απαιτήσεων⁵.

3.4.2 Επιρροές της OWL

Όπως προαναφέρθηκε, ο σχεδιασμός της OWL αποτέλεσε αντικείμενο μιας ποικιλίας επιρροών. Αυτές περιλαμβάνουν επιρροές από καθιερωμένους φορμαλισμούς και παραδείγματα αναπαράστασης γνώσης, επιρροές από τις υπάρχουσες γλώσσες οντολογίας, και επιρροές από τις υπάρχουσες γλώσσες του Σημασιολογικού Ιστού.

Μερικές από τις πιο σημαντικές επιρροές στο σχεδιασμό της OWL προέρχονται από τον προκάτοχό της DAML + OIL , από τις Περιγραφικές Λογικές, από το παράδειγμα πλαισίων και από το RDF. Πιο συγκεκριμένα, η επίσημη προδιαγραφή της γλώσσας επηρεάστηκε από τις Περιγραφικές Λογικές, η δομή της επιφάνειας της γλώσσας (όπως φαίνεται στην αφηρημένη σύνταξη) επηρεάστηκε από το παράδειγμα πλαισίων, και η RDF / XML σύνταξη ανταλλαγής επηρεάστηκε από την απαίτηση για την προς τα πάνω συμβατότητα με το RDF.

3.4.3 Βασικά στοιχεία μοντελοποίησης της OWL

Η OWL (Ontology Web Language – Γλώσσα Οντολογίας Ιστού) είναι η πιο πρόσφατη εξέλιξη στις γλώσσες οντολογιών για το Σημαντικό Ιστό. Αποτελεί μία σημασιακή γλώσσα

⁵ Athanasios G.Malamos, Paraskevi V. Sympa, Georgios S.Mamakis, “Xml Annotation Of Conceptual Characteristics In Interior Decoration”, 6th International Conference, New Horizons in Industry, Business and Education (NHIBE 2009), 27 - 28 August 2009, Santorini

σήμανσης για τη δημιουργία και τη διανομή οντολογιών στο διαδίκτυο. Είναι σημαντικό να αναφερθεί ότι η OWL (που αναπτύσσεται από το W3C) έχει οριστικοποιηθεί πολύ πρόσφατα, μόλις τον Φεβρουάριο του 2004 πέρασε σε κατάσταση Σύστασης (W3C Recommendation).

Σύνταξη της OWL

Η OWL χρησιμοποιεί τη σύνταξη της RDF που στηρίζεται στην XML (RDF/XML). Βέβαια, η RDF/XML σύνταξη δεν παρέχει μια πολύ αναγνώσιμη σύνταξη ως εκ τούτου, άλλες συντακτικές μορφές για την OWL έχουν καθοριστεί επίσης:

- Σύνταξη βασισμένη στην XML, ανεξάρτητη της RDF/XML.
- Μία αφηρημένη σύνταξη πιο συμπαγής και ευανάγνωστη από τις αντίστοιχες XML και RDF/XML.
- Μία γραφική σύνταξη που στηρίζεται στην UML. Στην συνέχεια, θα γίνει μία σύντομη περιγραφή στη σύνταξη:

Πληροφορίες Επικεφαλίδας (Header Information)

Τα OWL έγγραφα καλούνται συνήθως OWL οντολογίες και είναι RDF έγγραφα. Το στοιχείο ρίζας μιας OWL οντολογίας είναι ένα rdf: RDF στοιχείο, το οποίο διευκρινίζει διάφορα namespaces:

```
<rdf: RDF
xmlns: owl = "http://www.w3.org/2002/07/owl#"
xmlns: rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns: rdfs = "http://www.w3.org/2000/01/rdf-schema#"
xmlns: xsd = "http://www.w3.org/2001/XMLSchema#">
```

Όσον αφορά τη δομή γενικά ενός OWL εγγράφου, να σημειωθεί ότι όλες οι δηλώσεις, στις οποίες θα αναφερθούμε παρακάτω, θα βρίσκονται εμφωλευμένες στο παραπάνω στοιχείο-ρίζα rdf: RDF του εγγράφου.

Μια OWL οντολογία μπορεί να αρχίσει με μια συλλογή ισχυρισμών οι οποίοι εμπεριέχουν την έννοια της πληροφορίας για την ίδια την οντολογία που πρόκειται να αναπτυχθεί στη συνέχεια. Οι ισχυρισμοί αυτοί ομαδοποιούνται κάτω από στοιχείο owl: Ontology, και περιέχουν σχόλια (δήλωση rdfs: comment), τον έλεγχο έκδοσης (δήλωση owl: priorVersion, στο συγκεκριμένο θέμα θα γίνει αναφορά και παρακάτω), και το συνυπολογισμό άλλων οντολογιών (δήλωση owl: imports) και άλλα.

Για παράδειγμα,

```
<owl: Ontology rdf: about="">
<rdfs: comment>An example OWL ontology</rdfs: comment>
<owl: priorVersion
rdf: resource="http://www.mydomain.org/uni-ns-old"/>
<owl: imports
rdf: resource="http://www.mydomain.org/persons"/>
<rdfs: label>University Ontology</rdfs: label>
</owl: Ontology>
```

Μόνο ένας από αυτούς τους ισχυρισμούς έχει οποιαδήποτε συνέπεια για τη λογική έννοια της οντολογίας, η δήλωση owl: imports, η οποία απαριθμεί άλλες οντολογίες των οποίων το περιεχόμενο υποτίθεται ότι είναι μέρος της τρέχουσας οντολογίας. Η δήλωση αυτή

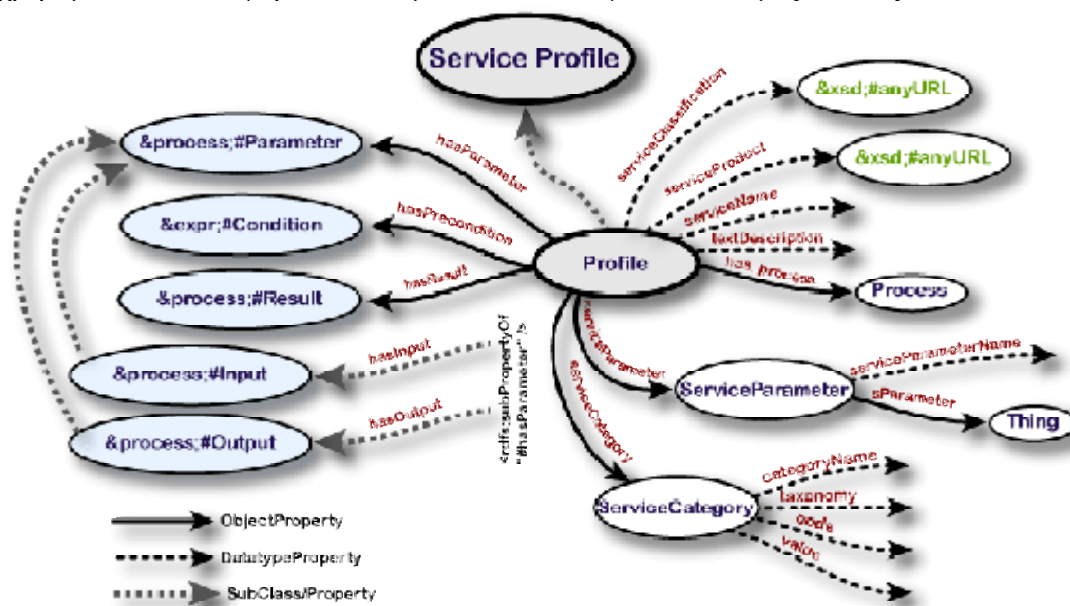
εισάγεται σε μία οντολογία και αναφέρεται σε μία άλλη για να δηλώσει ότι κάποιος περιορισμός που περιέχεται στην δεύτερη χρησιμοποιείται στην πρώτη. Είναι σημαντικό να αναφερθεί ότι ο συνυπολογισμός είναι μεταβατική ιδιότητα. Αυτό σημαίνει ότι εάν η οντολογία A συνυπολογίζει την B και η B συνυπολογίζει την Γ, τότε η οντολογία A συνυπολογίζει την B αλλά και την Γ.

OWL Κλάσεις

Οι OWL κλάσεις ορίζονται με την δήλωση owl: Class. Οι κλάσεις επίσης μπορούν να οργανωθούν σε μία ιεραρχία χρησιμοποιώντας τη δήλωση rdfs: subClass. Οι ιεραρχίες των κλάσεων μπορούν να οριστούν χρησιμοποιώντας μία ή περισσότερες δηλώσεις που ορίζουν ότι μία κλάση είναι υπο-κλάση μιας άλλης κλάσης.

Με τη δήλωση owl: Thing, ορίζουμε την υπερ-κλάση όλων, την κλάση Thing. Κάθε κλάση, λοιπόν, που ορίζεται από το δημιουργό μιας οντολογίας είναι υποκλάση αυτής, στην ιεραρχία των κλάσεων βρίσκεται κάτω από αυτήν. Αντίθετα, με την δήλωση owl: Nothing, ορίζουμε την κλάση Nothing η οποία είναι η κενή κλάση και είναι η υποκλάση κάθε κλάσης.

Η ισοδυναμία δύο κλάσεων ορίζεται με την δήλωση owl: equivalentClass. Οι ισοδύναμες κλάσεις έχουν τις ίδιες οντότητες. Οι δηλώσεις ισοδυναμίας κλάσεων είναι χρήσιμες όταν, για παράδειγμα, επιθυμεί κανείς να συνδυάσει οντολογίες που ενδεχομένως χρησιμοποιούν διαφορετικά ονόματα κλάσεων για ισοδύναμες έννοιες.



Εικόνα 12: Selected classes and properties of the Profile

<http://www.w3.org/Submission/OWL-S/>

OWL Ιδιότητες

Στην OWL, υπάρχουν δύο είδη ιδιοτήτων, οι ιδιότητες αντικειμένων (object properties) και οι ιδιότητες τύπων δεδομένων (datatype properties). Να σημειωθεί ότι η OWL δεν έχει προκαθορισμένους τύπους δεδομένων και για αυτό επιτρέπει τη χρήση τύπων δεδομένων της XML. Οι ιδιότητες αντικειμένων χρησιμοποιούνται για να συνδέσουν δύο άτομα μεταξύ τους, όπως για παράδειγμα η ιδιότητα έχωΠατερα συνδέει ένα άτομο με ένα άλλο που αντιπροσωπεύει τον πατέρα του, ενώ οι ιδιότητες τύπων δεδομένων συνδέουν ένα άτομο με μια τιμή τύπου δεδομένων, όπως για παράδειγμα η ιδιότητα έχωΗλικια η οποία συνδέει ένα

άτομο με μια τιμή που αντιπροσωπεύει την ηλικία του. Στη συνέχεια, παρατίθενται ένα σύνολο δηλώσεων σχετικών με τις ιδιότητες:

- δήλωση owl: ObjectProperty: Χρησιμοποιείται για να δηλώσει μία ιδιότητα αντικειμένων.

- δήλωση owl: DatatypeProperty: Χρησιμοποιείται για να δηλώσει μία ιδιότητα τύπων δεδομένων.

- δήλωση rdfs: subPropertyOf: Χρησιμοποιείται για να δηλώσει ότι μία ιδιότητα είναι υποπερίπτωση μιας άλλης ιδιότητας.

- δήλωση rdfs: domain: δηλώνει το πεδίο ορισμού μιας ιδιότητας.

- δήλωση rdfs: range: δηλώνει το πεδίο τιμών μιας ιδιότητας.

- δήλωση owl: equivalentProperty: δύο ιδιότητες μπορούν να δηλωθούν ως ισοδύναμες.

- δήλωση owl: inverseOf: Μία ιδιότητα μπορεί να δηλωθεί ως αντίστροφη μιας άλλης. Αν η ιδιότητα P1 δηλωθεί ως αντίστροφη της ιδιότητας P2 και η οντότητα X σχετίζεται με μία άλλη οντότητα Y μέσω της ιδιότητας P1, τότε και η οντότητα Y σχετίζεται με την οντότητα X μέσω της ιδιότητας P2.

- δήλωση owl: TransitiveProperty: Μία ιδιότητα μπορεί να δηλωθεί μεταβατική. Δηλαδή, αν η ιδιότητα P δηλωθεί μεταβατική, και αν $X P Y$ και αν $Y P Z$, τότε μία μηχανή συμπερασμού μπορεί να συμπεράνει ότι $X P Z$.

- δήλωση owl: SymmetricProperty: Μία ιδιότητα μπορεί να δηλωθεί ως συμμετρική. Για παράδειγμα, αν η ιδιότητα P δηλωθεί συμμετρική, και αν $X P Y$, τότε μία μηχανή συμπερασμού μπορεί να συμπεράνει ότι $Y P X$.

- δήλωση owl: FunctionalProperty: Αν μία ιδιότητα δηλωθεί ως συναρτησιακή, τότε έχει το πολύ μία τιμή για κάθε οντότητα του πεδίου ορισμού. Η δήλωση αυτή αποτελεί ουσιαστικά μία συντόμευση που παρέχει η OWL, καθώς είναι ισοδύναμη με το να δηλώσουμε ταυτόχρονα τους περιορισμούς πληθάριμου `minCardinality 0` και `maxCardinality 1` για αυτήν την ιδιότητα.

- δήλωση owl: inverseFunctionalProperty: Ορίζει μία ιδιότητα για την οποία δύο διαφορετικά αντικείμενα δεν μπορούν να έχουν την ίδια τιμή.

OWL Στιγμιότυπα

Τα στιγμιότυπα μιας κλάσης δηλώνονται ακριβώς όπως και στην RDF ή ισοδύναμα:

```
<rdf: Description rdf: ID="949352">
<rdf: type rdf: resource="#academicStaffMember"/>
</rdf: Description>
ή ισοδύναμα:
<academicStaffMember rdf: ID="949352"/>
ή με περισσότερες λεπτομέρειες:
<academicStaffMember rdf: ID="949352"/>
<uni: age rdf: datatype="xsd:integer">39</uni: age>
</academicStaffMember>
```

Με τη δήλωση owl: sameAs, δύο στιγμιότυπα μπορούν να δηλωθούν για να είναι τα ίδια. Έτσι μπορούν να χρησιμοποιηθούν για να δημιουργήσουν διάφορα διαφορετικά ονόματα που αναφέρονται στο ίδιο στιγμιότυπο.

Από την άλλη, μπορεί να απαιτείται να δηλωθεί ότι ένα στιγμιότυπο να είναι διαφορετικό από άλλα στιγμιότυπα και αυτό επιτυγχάνεται με την δήλωση owl: differentFrom. Η ρητή δήλωση ότι δύο στιγμιότυπα είναι διαφορετικά μπορεί να είναι σημαντική για μία γλώσσα όπως η OWL, καθώς αυτή δεν απαιτεί κάθε στιγμιότυπο να έχει ένα μόνο όνομα και επομένως μία μηχανή συμπερασμού δεν θα καταλήγει σε αξιόπιστα συμπεράσματα.

Ένας αριθμός στιγμιότυπων μπορούν να δηλωθούν ως αμοιβαία διαφορετικά σε μία owl: AllDifferent. Η δήλωση αυτή είναι ιδιαίτερα χρήσιμη όταν υπάρχουν σύνολα ευδιάκριτων αντικειμένων και όταν θέλουμε να εισαγάγουμε την υπόθεση της μοναδικότητας των ονομάτων μεταξύ των αντικειμένων του συνόλου. Πάντοτε χρησιμοποιείται από κοινού με την δήλωση owl: distinctMembers για να δηλώσει ότι όλα τα μέλη μιας λίστας είναι ξεχωριστά και ευδιάκριτα μεταξύ τους.

ΑΛΛΑ ΣΤΟΙΧΕΙΑ

Πολλές ερευνητικές ομάδες σε Ευρώπη και Αμερική είχαν διαγνώσει την ανάγκη ύπαρξης μιας νέας ισχυρής γλώσσας οντολογιών, καθώς η RDF και το RDF Schema δεν παρείχαν την απαραίτητη εκφραστική ισχύ. Με τη χρήση του λεξιλογίου της RDF και του RDF Schema είναι δυνατόν να οριστούν στιγμιότυπα τάξεων, ιεραρχίες τάξεων και ιδιοτήτων καθώς και καθολικοί περιορισμοί που αφορούν το πεδίο ορισμού και το πεδίο τιμών ιδιοτήτων. Εντούτοις, διάφορα άλλα χαρακτηριστικά γνωρίσματα λείπουν. Μερικά από αυτά είναι τα εξής:

1. Τοπική εμβέλεια ιδιοτήτων
2. Μη επικάλυψη κλάσεων: δεν είναι δυνατή η δήλωση κλάσεων ξένων μεταξύ τους παρά μόνο η δήλωση σχέσεων υποκλάσεων.
3. Λογικοί συνδυασμοί κλάσεων: Η δημιουργία νέων κλάσεων από το συνδυασμό υπαρχουσών δεν είναι δυνατή, δηλαδή κλάσεις που να προκύπτουν από τομές, ενώσεις και συμπληρώματα άλλων κλάσεων δεν είναι έγκυρες.
4. Περιορισμοί πληθυκότητας: δεν είναι δυνατή η επιβολή περιορισμών στο πλήθος των διακριτών τιμών μιας ιδιότητας. Για παράδειγμα, δεν είναι δυνατό να οριστεί ότι ένας άνθρωπος έχει ακριβώς μία μητέρα.
5. Ειδικά χαρακτηριστικά ιδιοτήτων: δεν υπάρχει η δυνατότητα δήλωσης ειδικών χαρακτηριστικών για τις ιδιότητες, π.χ. μία ιδιότητα είναι μεταβατική, μοναδική ή αντίστροφη μιας άλλης.

Οι παραπάνω περιορισμοί της RDF και του RDF Schema οδήγησαν στον ορισμό της γλώσσας DAML+OIL, η οποία αποτέλεσε σημείο εκκίνησης για τον ορισμό της OWL.

Η OWL αναπτύχθηκε ως επέκταση του RDF Schema και ενσωμάτωσε λύσεις για προβλήματα που εμφανίστηκαν κατά τον ορισμό και τη χρησιμοποίηση της γλώσσας DAML+OIL. Η OWL, καλύπτοντας όλες τις παραπάνω αδυναμίες, διαθέτει μεγαλύτερη εκφραστικότητα η οποία σαφώς είναι αναγκαία για το Σημαντικό Ιστό καθώς και επιτρέπει τη χρήση συστημάτων συλλογισμού για την διεξαγωγή αποτελεσματικών συμπερασμών πάνω στις οντολογίες που αναπαρίσταται στον Ιστό με αυτή τη γλώσσα.

Γενικά ισχύει ότι μεγαλύτερη εκφραστική δύναμη μιας γλώσσας συνεπάγεται τη διεξαγωγή συμπερασμών αυξανόμενης πολυπλοκότητας έως και μη αποφασισιμότητα. Για το λόγο αυτό, η OWL ορίζεται από το Web Ontology Working Group του W3C ως τρεις υπογλώσσες (OWL Lite, OWL DL, OWL Full) με διαφορετική εκφραστική ισχύ, οι οποίες παρουσιάζονται παρακάτω ξεκινώντας από αυτή που έχει την μικρότερη (OWL Lite) και καταλήγοντας σε αυτή που έχει τη μεγαλύτερη εκφραστική ισχύ (OWL Full). Κάθε υπογλώσσα αποτελεί επέκταση της υπογλώσσας με την αμέσως μικρότερη εκφραστική ισχύ.

3.4.4 Οι υπο - γλώσσες της OWL

OWL Lite

Η OWL DL σχετίζεται με τη SHOIN(D), μια πολύ εκφραστική Περιγραφική Λογική. Αυτή η Περιγραφική Λογική είναι κάπως δύσκολη στην παρουσίαση σε άπειρους χρήστες, δεδομένου ότι είναι δυνατή η δόμηση πολύπλοκων boolean περιγραφών χρησιμοποιώντας, για παράδειγμα, ένωση και συμπλήρωμα. Η SHOIN(D) είναι επίσης δύσκολη στην συλλογιστική, καθώς βασικά προβλήματα σχετικά με την συμπερασματολογία έχουν πολύπλοκότητα NExpTime.

Για τους λόγους αυτούς, έχει προσδιοριστεί ένα υποσύνολο της OWL DL που πρέπει να είναι ευκολότερο για όλες τις παραπάνω μετρήσεις. Αυτό το υποσύνολο ονομάζεται OWL Lite. Η OWL Lite απαγορεύει τις ενώσεις και τα συμπληρώματα, περιορίζει τις τομές των εξυπακουόμενων τομών στα αξιώματα των κλάσεων που μοιάζουν με πλαίσια, περιορίζει όλες τις ενσωματωμένες περιγραφές στα ονόματα εννοιών, δεν επιτρέπει στα αντικείμενα να εμφανίζονται σε περιγραφές ή αξιώματα κλάσεων, και περιορίζει τα cardinalities σε 0 ή 1.

Οι περιορισμοί αυτοί καθιστούν την OWL Lite παρόμοια με την Περιγραφική Λογική SHIF(D). Όπως η SHIF(D), η βασική συμπερασματολογία στην OWL Lite μπορεί να υπολογιστεί στη χειρότερη περίπτωση εκθετικού χρόνου (ExpTime), και υπάρχουν ήδη αρκετοί βελτιστοποιημένοι reasoners για λογικές ισοδύναμες με την OWL Lite. Η βελτίωση αυτή στην βολικότητα έρχεται με σχετικά μικρή απώλεια σε εκφραστική ισχύ αν και η σύνταξη της OWL Lite είναι πιο περιορισμένη από αυτή της OWL DL, είναι ακόμα δυνατό να εκφραστούν πολύπλοκες περιγραφές με την εισαγωγή νέων ονομάτων κλάσεων και την αξιοποίηση των αρνήσεων που εισάγονται από τα αξιώματα ασυνδετότητας. Χρησιμοποιώντας αυτές τις τεχνικές, όλες οι περιγραφές της OWL DL μπορούν να συλληφθούν στην OWL Lite, εκτός από εκείνες που περιέχουν είτε ονόματα αντικειμένων είτε cardinalities μεγαλύτερα από 1.

OWL DL

Η OWL DL σχεδιάστηκε για τους χρήστες που επιθυμούν τη μέγιστη δυνατή εκφραστικότητα διατηρώντας:

I. Την υπολογιστική πληρότητα (όλα τα συμπεράσματα να είναι εγγυημένα υπολογίσιμα) και

II. Την αποφασισιμότητα (η διεξαγωγή συμπερασμών να πραγματοποιείται σε πεπερασμένο χρόνο).

Η OWL DL ονομάζεται έτσι λόγω της αντιστοιχίας της με τις Λογικές Περιγραφές. Οι Λογικές Περιγραφές είναι μία οικογένεια λογικών γλωσσών που επιτρέπουν την εύκολη περιγραφή τάξεων αντικειμένων, καθώς και την αποδοτική διενέργεια ελέγχων υπαγωγής και ταξινόμησης, δηλαδή ελέγχων του κατά πόσον μία τάξη αποτελεί υπό-τάξη μιας άλλης και του κατά πόσον μία οντότητα ανήκει σε μία τάξη.

Το πλεονέκτημα της υπογλώσσας αυτής είναι η αποδοτική υποστήριξη της συλλογιστικής. Από την άλλη όμως χάνεται η πλήρης συμβατότητα με την RDF. Δηλαδή, ένα RDF κείμενο θα πρέπει να επεκταθεί ή να περιοριστεί σε κάποια σημεία του ώστε να γίνει νόμιμο OWL DL κείμενο⁶.

⁶ Alan Rector, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang, Chris Wroe. OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns. *14th International Conference on Knowledge Engineering and Knowledge*

OWL Full*

Οι OWL DL και OWL Lite είναι επεκτάσεις μιας περιορισμένης χρήσης των RDF και RDFS, διότι, σε αντίθεση με τα RDF και RDFS, δεν επιτρέπουν να χρησιμοποιούνται οι κλάσεις ως αντικείμενα, και οι κατασκευαστές γλώσσας δε μπορούν να εφαρμοστούν στην ίδια τη γλώσσα. Για τους χρήστες που χρειάζονται αυτές τις δυνατότητες, έχει σχεδιαστεί μια έκδοση της OWL που είναι συμβατή προς τα πάνω με τα RDF και RDFS. Αυτή η έκδοση ονομάζεται OWL Full. Η OWL Full, επιτρέπει όλους τους συνδυασμούς RDF και RDFS.

Η OWL Full περιέχει την OWL DL, αλλά ξεφεύγει πολύ από το τυποποιημένο πλαίσιο της Περιγραφικής Λογικής. Το τίμημα που πρέπει να πληρωθεί εδώ είναι διπλό. Πρώτον, η συλλογιστική στην OWL Full είναι ανίκανη για λήψεις αποφάσεων (λόγω περιορισμών που απαιτούνται για να διατηρηθεί η ικανότητα λήψης αποφάσεων της OWL DL, δεν ισχύουν για την OWL Full). Δεύτερον, η αφηρημένη σύνταξη της OWL DL είναι ανεπαρκής για την OWL Full, καθώς και πρέπει να χρησιμοποιείται η επίσημη σύνταξη ανταλλαγής της OWL η RDF / XML.

Η πλήρης γλώσσα καλείται OWL Full. Χρησιμοποιεί όλες τις θεμελιώδεις αρχές των υπο-γλωσσών OWL. Επιτρέπει το συνδυασμό των αρχών αυτών με την RDF και το RDF Schema. Έχει το πλεονέκτημα ότι είναι προς τα πάνω συμβατή με την RDF, τόσο συντακτικά όσο και σημασιολογικά:

- κάθε νόμιμο RDF κείμενο είναι και νόμιμο OWL Full κείμενο και,
- κάθε έγκυρο RDF/RDF Schema συμπέρασμα είναι και έγκυρο OWL συμπέρασμα.

Παράλληλα, η OWL Full προορίζεται για χρήστες που επιθυμούν μέγιστη εκφραστικότητα και την πλήρη εκφραστική ελευθερία του RDF χωρίς όμως εγγυήσεις επιλυσιμότητας. Για παράδειγμα η OWL FULL επιτρέπει σε μια βάση να ορίζεται ως συλλογή ατόμων και συγχρόνως να θεωρείται άτομο η ίδια (πράγμα που οδηγεί σε κυκλικούς ορισμούς εννοιών και επομένως σε μη αποφασισιμότητα). Θεωρείται απίθανο ότι κάποιο σύστημα συλλογισμού θα μπορέσει ποτέ να υποστηρίξει όλα τα χαρακτηριστικά της OWL Full.

Συμβατότητα μεταξύ των τριών υπο-γλωσσών

Κάθε μια από αυτές τις υπογλώσσες είναι η επέκταση του απλούστερου προκατόχου της, σχηματικά αυτό φαίνεται στη παρακάτω εικόνα. Ισχύει το ακόλουθο σύνολο σχέσεων μεταξύ των τριών υπογλωσσών, αλλά τα αντίστροφα τους όχι.

1. Κάθε έγκυρη οντολογία της OWL Lite είναι μια έγκυρη οντολογία της OWL DL.
2. Κάθε έγκυρη οντολογία της OWL DL είναι μια έγκυρη οντολογία της OWL Full.
3. Κάθε έγκυρο συμπέρασμα της OWL Lite είναι ένα έγκυρο συμπέρασμα της OWL DL.
4. Κάθε έγκυρο συμπέρασμα OWL DL είναι ένα έγκυρο συμπέρασμα της OWL Full.

Η OWL εξακολουθεί να χρησιμοποιεί την RDF και το RDF Schema σε μεγάλο βαθμό:

1. Όλα τα είδη της OWL χρησιμοποιούν την RDF για τη σύνταξή τους.
2. Τα στιγμιότυπα ορίζονται όπως στην RDF, χρησιμοποιώντας τις RDF περιγραφές και εισάγοντας την πληροφορία.
3. Constructors της OWL όπως owl: Class owl: DatatypeProperty και owl: ObjectProperty αποτελούν εξειδικεύσεις των αντίστοιχων της RDF.

* Management (EKAW), Whittlebury Hall, UK, 2004

Οι υπεύθυνοι για την ανάπτυξη οντολογιών που υιοθετούν την OWL πρέπει να εξετάσουν το ποια υπογλώσσα ανταποκρίνεται καλύτερα στις ανάγκες τους. Η επιλογή μεταξύ της OWL Lite και της OWL DL εξαρτάται από το βαθμό στον οποίο οι χρήστες απαιτούν περισσότερο εκφραστικά κατασκευάσματα που παρέχεται 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 διευκολύνσεις της μετα- διαμόρφωσης της RDF Schema. Κατά τη χρησιμοποίηση της OWL Full σε σύγκριση με την OWL DL, η υποστήριξη συλλογισμών είναι λιγότερο προβλέψιμη, μιας και δεν υπάρχουν ακόμη πλήρεις εφαρμογές.

Η OWL Full μπορεί να θεωρηθεί ως επέκταση της RDF, ενώ η OWL Lite και η OWL DL ως επεκτάσεις μιας περιορισμένης άποψης της RDF. Κάθε έγγραφο OWL (Lite, DL, Full) είναι ένα έγγραφο RDF, και κάθε έγγραφο RDF είναι ένα έγγραφο OWL Full, αλλά μόνο μερικά έγγραφα RDF θα είναι έγκυρα έγγραφα OWL Lite ή OWL DL.

Για το λόγο αυτό, χρειάζεται προσοχή όταν θέλει ένας χρήστης να θεωρήσει ένα έγγραφο RDF ως έγγραφο OWL. Όταν η εκφραστικότητα της OWL DL ή της OWL Lite κρίνεται κατάλληλη, μερικές προφυλάξεις πρέπει να ληφθούν για να εξασφαλίσουν ότι το αρχικό έγγραφο RDF συμμορφώνεται με τους πρόσθετους περιορισμούς από την OWL DL και την OWL Lite. Μεταξύ των άλλων, κάθε URI που χρησιμοποιείται σαν όνομα κατηγορίας πρέπει να βεβαιωθεί ρητά ότι είναι του τύπου owl: Class και κάθε αντικείμενο πρέπει να βεβαιωθεί ότι ανήκει τουλάχιστον σε μια κατηγορία.

Σημασιολογία για την OWL Full

Στην OWL Full έχει δοθεί ένα μοντέλο σημασιολογίας που είναι η επέκταση ενός λεξιλογίου του θεωρητικού μοντέλου του RDF. Η αντιστοιχία μεταξύ αυτής της σημασιολογίας και της σημασιολογίας της OWL DL έχει επίσης καθιερωθεί. Έχει αποδειχθεί ότι το μοντέλο θεωρίας για την OWL DL έχει τις ίδιες συνέπειες με αυτό το μοντέλο θεωρίας τύπου RDF για εκείνες τις οντολογίες OWL που μπορούν να γραφτούν στην αφηρημένη σύνταξη της OWL DL.

Η αντιστοιχία σημαίνει ότι, δεδομένων δύο οντολογιών OWL DL O1 και O2, γραμμένες σε αφηρημένη σύνταξη, η O1 συνεπάγεται την O2, σύμφωνα με το μοντέλο θεωρίας της OWL DL αν και μόνο αν η αντιστοίχιση της O1 στις τριάδες RDF συνεπάγεται την αντιστοίχιση της O2 στις τριάδες RDF σύμφωνα με το μοντέλο θεωρίας της OWL Full. Η απόδειξη αυτής της αντιστοιχίας είναι αρκετά πολύπλοκη και θα μπορούσε να διασπαστεί, για παράδειγμα, ως αποτέλεσμα (προφανώς) μικρών αλλαγών στις προδιαγραφές του RDF ή της OWL. Λόγω της σχετικής αστάθειας αυτής της αντιστοιχίας, και προκειμένου να αποφευχθεί ενδεχόμενη σύγχυση ως προς την έννοια της OWL DL, το μοντέλο θεωρίας της OWL Full έχει "μη κανονιστικό" χαρακτήρα (δηλαδή, δεν είναι μόνο πληροφοριακό) για τις οντολογίες OWL που μπορούν να γραφτούν σε αφηρημένη σύνταξη. Αυτό σημαίνει ότι το μοντέλο θεωρίας της OWL DL θα λαμβάνεται ως οριστικό αν η αντιστοιχία διαλυθεί ή αποδειχθεί ότι είναι ελλιπής.

Η σημασιολογία της OWL Full αποδίδεται απευθείας στα RDF/XML στοιχεία της OWL, όπως είναι το allValuesFrom και άλλα. Αυτό συνεπάγεται ότι προκειμένου να ερμηνευτούν κάποιες περίπλοκες και περιττές δομές που χρησιμοποιούνται στην RDF/XML σύνταξη της OWL η σημασιολογία αυτή περιπλέκεται χωρίς να υπάρχει κάποιος ιδιαίτερος λόγος. Για το λόγο αυτό συνήθως για να αποδώσουμε ερμηνεία στα δομικά στοιχεία, τα αξιώματα και τους κατασκευαστές της OWL, ανάγουμε τους κατασκευαστές και τα δομικά στοιχεία της OWL σε εκφραστικές Περιγραφικές Λογικές. Εκτός από την παροχή σημασιολογίας στους κατασκευαστές και τα αξιώματα της OWL η αναγωγή αυτή μας προσφέρει ακόμα ένα πλεονέκτημα. Μπορούμε να ανάγουμε τα προβλήματα εξαγωγής

συμπερασμάτων των OWL οντολογιών στα προβλήματα εξαγωγής συμπερασμάτων των Περιγραφικών Λογικών για τα οποία υπάρχουν υλοποιημένοι και βελτιστοποιημένοι αλγόριθμοι.

3.5 Περιγραφικές Λογικές

Οι Περιγραφικές Λογικές είναι μια οικογένεια φορμαλισμών αναπαράστασης γνώσης βασισμένων σε κλάσεις. Χαρακτηρίζονται από τη χρήση διαφόρων κατασκευαστών για την κατασκευή πολύπλοκων κλάσεων από απλούστερες, την έμφαση στην ικανότητα λήψης αποφάσεων σε βασικά προβλήματα αιτιολόγησης, και την παροχή ορθών, πλήρων και (εμπειρικά) ήπιων υπηρεσιών αιτιολόγησης. Οι Περιγραφικές Λογικές, καθώς και στοιχεία που προέκυψαν από την έρευνα της Περιγραφικής Λογικής, είχαν ισχυρή επιρροή στον σχεδιασμό της OWL, ιδίως όσον αφορά την τυποποίηση της σημασιολογίας, την επιλογή των γλωσσών κατασκευαστών και την ολοκλήρωση των τύπων δεδομένων και των τιμών δεδομένων. Στην πραγματικότητα, οι OWL DL και OWL Lite (δύο από τα τρία είδη της OWL) μπορούν να θεωρηθούν ως εκφραστικές Περιγραφικές Λογικές, με μια οντολογία να είναι ισοδύναμη με μια βάση γνώσεων Περιγραφικής Λογικής.

Ένα βασικό χαρακτηριστικό των Περιγραφικών Λογικών είναι ότι είναι λογικές, δηλαδή, η επίσημη γλώσσα με σαφώς καθορισμένη σημασιολογία. Η τυποποιημένη τεχνική για τον προσδιορισμό της έννοιας μιας Περιγραφικής Λογικής είναι μέσω ενός θεωρητικού μοντέλου σημασιολογίας, σκοπός του οποίου είναι να εξηγήσει την σχέση ανάμεσα στην σύνταξη της γλώσσας και το επιδιωκόμενο μοντέλο του τομέα. Ένα μοντέλο αποτελείται από έναν τομέα και μία λειτουργία διερμηνείας, όπου ο τομέας είναι ένα σύνολο αντικειμένων και η λειτουργία διερμηνείας είναι μια χαρτογράφηση από τα ονόματα των αντικειμένων, των κλάσεων και των ιδιοτήτων προς τα στοιχεία του τομέα, τα υποσύνολα του τομέα και τις δυαδικές σχέσεις στον τομέα, αντίστοιχα.

Τα αντικείμενα στον τομέα δεν έχουν καθαυτά κάποιο νόημα, ούτε η επιλογή ενός συγκεκριμένου συνόλου αντικειμένων που απαρτίζουν τον τομέα. Αυτό που είναι σημαντικό είναι οι σχέσεις μεταξύ των αντικειμένων και των συνόλων αντικειμένων.

Μια βάση γνώσεων Περιγραφικής Λογικής αποτελείται από ένα σύνολο αξιωμάτων, για παράδειγμα, ότι μία κλάση είναι υποκλάση μιας άλλης, ή ότι ένα αντικείμενο είναι ένα στιγμιότυπο της συγκεκριμένης κλάσης. Το νόημα αυτών των αξιωμάτων δίνεται από αντίστοιχους περιορισμούς στα μοντέλα. Αν, για παράδειγμα, η βάση γνώσης περιέχει ένα αξίωμα που δηλώνει ότι το Person είναι υπό-κλάση του Animal, τότε, σε ένα μοντέλο της βάσης γνώσεων, η ερμηνεία του Person πρέπει πάντα να είναι υποσύνολο της ερμηνείας του Animal. Η έννοια μιας βάσης γνώσεων προέρχεται από χαρακτηριστικά και σχέσεις που είναι κοινές σε όλα τα πιθανά μοντέλα. Αν, για παράδειγμα, η ερμηνεία μιας κλάσης πρέπει να είναι πάντα το κενό σύνολο, τότε η κλάση αυτή λέγεται ότι είναι ασυνεπής, ενώ αν δεν υπάρχουν πιθανές ερμηνείες, η ίδια η βάση γνώσης λέγεται ότι είναι ασυνεπής. Αν η σχέση που ορίζεται από ένα συγκεκριμένο αξίωμα πρέπει να ισχύει σε όλες τις ερμηνείες μιας βάσης γνώσεων, τότε, αυτό το αξίωμα λέγεται ότι προκύπτει από τη βάση γνώσεων, και αν μία βάση γνώσεων συνεπάγεται κάθε αξίωμα σε άλλη βάση γνώσεων, τότε η πρώτη βάση γνώσεων λέμε ότι συνεπάγεται τη δεύτερη βάση γνώσεων. Μια βάση γνώσεων που περιέχει το αξίωμα «το Person είναι υπό-κλάση του Animal», για παράδειγμα, συνεπάγεται ότι η τομή του Male και του Person είναι επίσης μια υποκατηγορία του Animal. Αυτή η συνεπαγωγή είναι αρκετά τετριμμένη, αλλά με μια γλώσσα τόσο σύνθετη όσο η OWL, ο έλεγχος των συνεπαγωγών μπορεί, σε γενικές γραμμές, να είναι ένα πολύ δύσκολο έργο.

Όπως η OIL και η DAML + OIL, η OWL χρησιμοποιεί ένα μοντέλο θεωρίας σε στυλ Περιγραφικής Λογικής για να τυποποιήσει την ερμηνεία της γλώσσας. Αυτό αναγνωρίστηκε ως ένα βασικό χαρακτηριστικό και στις τρεις γλώσσες, καθώς επιτρέπει οι οντολογίες, και το λεξιλόγιο που χρησιμοποιεί πληροφορίες που ορίζονται από οντολογίες, να μοιράζονται και να ανταλλάσσονται χωρίς διαφωνίες για την ακριβή έννοια. Η ανάγκη για τέτοιου είδους διατύπωση ενισχύθηκε από την εμπειρία με παλιότερες εκδόσεις προδιαγραφών του RDF και

του RDFS, όπου η έλλειψη διατύπωσης σύντομα οδήγησε σε διαφωνίες ως προς την ερμηνεία των γλωσσικών δομών, όπως οι περιορισμοί τομέα και εύρους. Προκειμένου να αποφευχθούν τέτοια προβλήματα, η ερμηνεία του RDF τώρα, επίσης, ορίζεται από την άποψη της θεωρίας μοντέλου.

Ένα άλλο πλεονέκτημα της τυποποίησης της ερμηνείας της γλώσσας με αυτό τον τρόπο είναι ότι οι αυτοματοποιημένες τεχνικές συλλογιστικής μπορούν να χρησιμοποιηθούν για τον έλεγχο της συνέπειας των κλάσεων και των οντολογιών, και τον έλεγχο των σχέσεων συνεπαγωγής. Αυτό είναι σημαντικό αν πρόκειται να αξιοποιηθεί η πλήρης ισχύς των οντολογιών από ευφυείς πράκτορες, και η δυνατότητα παροχής αυτής της υποστήριξης συλλογιστικής ήταν βασικός στόχος του σχεδιασμού της OWL.

Η εκφραστική ισχύς μιας γλώσσας όπως είναι η OWL καθορίζεται από τους κατασκευαστές κλάσεων (και ιδιοτήτων) που υποστηρίζονται, καθώς και από τα είδη των αξιωμάτων που μπορούν να συμβούν σε μια οντολογία. Φυσικά η αυξημένη εκφραστική ισχύς αναπόφευκτα σημαίνει αύξηση της υπολογιστικής πολυπλοκότητας για βασικά προβλήματα, όπως η συνεπαγωγή.

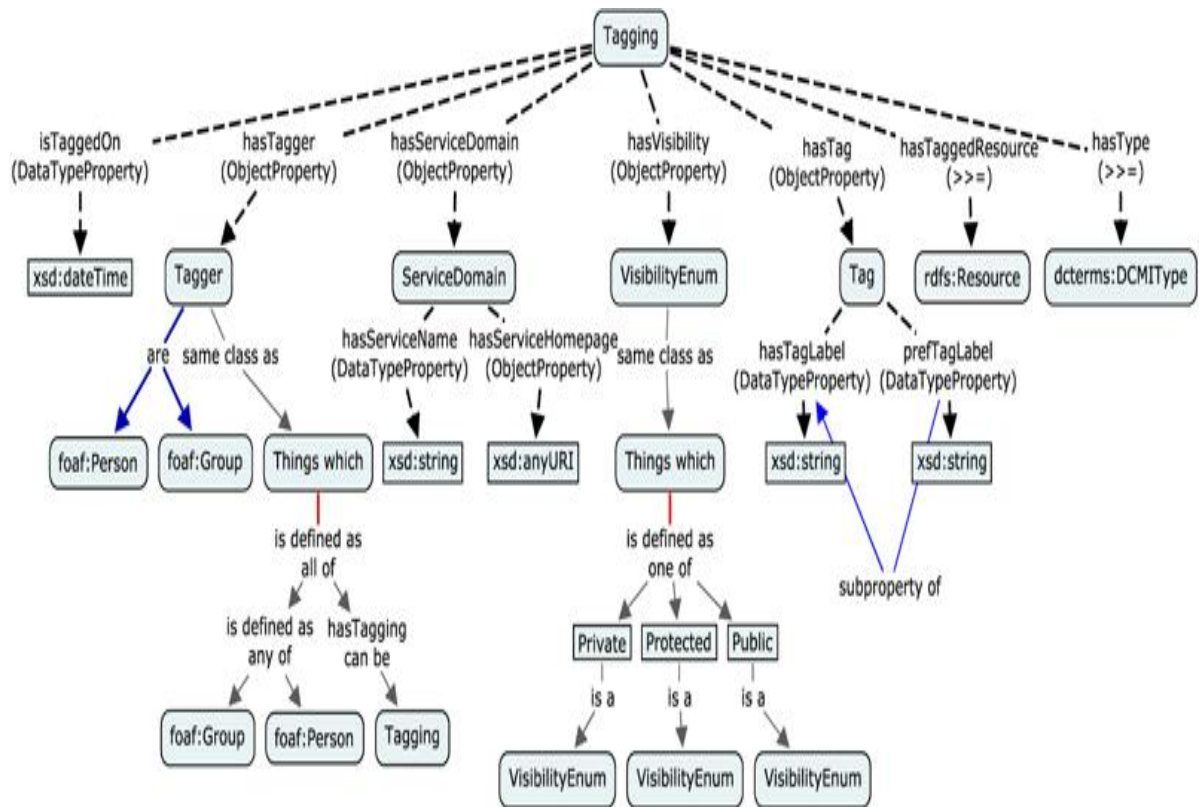
Ο σχεδιασμός της OWL εμπνεύστηκε από περισσότερο από 10 χρόνια έρευνας της Περιγραφικής Λογικής, η οποία έχει καθορίσει με αρκετή λεπτομέρεια το τοπίο της πολυπλοκότητας για ένα ευρύ φάσμα κατασκευαστών και αξιωμάτων, και των διάφορων συνδυασμών τους. Η γνώση αυτή επέτρεψε να επιλεγούν προσεκτικά το σύνολο των κατασκευαστών και των αξιωμάτων που υποστηρίζονται από την OWL, έτσι ώστε να εξισορροπηθούν οι εκφραστικές απαιτήσεις των τυπικών εφαρμογών με την απαίτηση για την αξιόπιστη και αποτελεσματική υποστήριξη συλλογιστικής.

Ένας ιδιαίτερος στόχος αυτής της διαδικασίας σχεδιασμού ήταν να διασφαλιστεί ότι η συνεπαγωγή της OWL θα ήταν τουλάχιστον ικανή για λήψη αποφάσεων, δηλαδή, ότι θα ήταν δυνατό να σχεδιάσουμε έναν αλγόριθμο που θα μπορούσε να καθορίσει εάν ή όχι μία οντολογία OWL συνεπάγεται μια άλλη (ένας τέτοιος αλγόριθμος συχνά καλείται μία διαδικασία απόφασης). Η διαθεσιμότητα των πρακτικών διαδικασιών αποφάσεων (συνεπαγωγής), και ακόμη και τα συστήματα που εφαρμόζονται, ήταν επίσης μια σημαντική εξέταση.

Εκτός από την αντιμετώπιση «αφηρημένων» κλάσεων όπως Person και Animal, πολλές πρακτικές εφαρμογές χρειάζεται να αναπαριστούν και να αιτιολογούν τύπους δεδομένων και τιμές, όπως ακέραιοι και συμβολοσειρές. Η ολοκλήρωση των τύπων δεδομένων στη γλώσσα OWL είναι και αυτή πολύ επηρεασμένη από την έρευνα της Περιγραφικής Λογικής, η οποία έχει επιδείξει ότι απαιτείται προσοχή ώστε να αποφευχθεί η έκρηξη πολυπλοκότητας ή ακόμα και η αδυναμία λήψης αποφάσεων λόγω των τύπων δεδομένων. Στην Περιγραφική Λογική SHOQ(D) φάνηκε πως αυτό θα μπορούσε να επιτευχθεί με σαφή διαχωρισμό της ερμηνείας των τύπων δεδομένων και των τιμών από αυτή των κλάσεων και των αντικειμένων.

Αυτό το σχέδιο έχει το πλεονέκτημα ότι η συλλογιστική με τύπους δεδομένων και τιμές μπορούν να διαχωριστούν σχεδόν εξ ολοκλήρου από την συλλογιστική με τις κλάσεις και τα αντικείμενα. Επιπλέον, η γλώσσα εξακολουθεί να έχει την ικανότητα λήψης αποφάσεων εάν η συλλογιστική τύπων δεδομένων και τιμών έχει την ικανότητα λήψης αποφάσεων. Αυτό μπορεί εύκολα να επιτευχθεί, για μια σειρά κοινών τύπων δεδομένων, όπως ακέραιοι αριθμοί, δεκαδικά ψηφία, και συμβολοσειρές.

Εκτός αυτών των πρακτικών ζητημάτων, μπορεί επίσης να υποστηριχθεί ότι ο διαχωρισμός των κλάσεων και των τύπων δεδομένων είναι λογικός από φιλοσοφική άποψη, καθώς οι τύποι δεδομένων είναι ήδη δομημένοι από ενσωματωμένα κατηγορήματα, όπως το μεγαλύτερο-από και λιγότερο-από. Από αυτή την άποψη, δεν έχει νόημα η χρήση των αξιωμάτων οντολογιών για την πρόσθεση επιπλέον δόμησης στους τύπους δεδομένων ή για την κατασκευή «υβριδικών» κλάσεων.



Εικόνα 13: Παράδειγμα σχεδιασμού της OWL

http://ai-group.ds.unipi.gr/kotis/sites/default/files/Lecture_in_OE_2013_lowRes.pdf

3.6 Πρότυπο Πλαισίων

Στα πλαίσια του Σημασιολογικού Ιστού, όπου οι χρήστες με ευρύ φάσμα γνώσεων αναμένεται να δημιουργήσουν ή να τροποποιήσουν οντολογίες, η αναγνωσιμότητα και η γενική ευκολία στη χρήση αποτελούν σημαντικούς παράγοντες για μια γλώσσα οντολογίας⁷. Κατά τον σχεδιασμό της OIL, μίας από τις γλώσσες στις οποίες βασίζεται η OWL, οι απαιτήσεις αυτές καλύφθηκαν από την παροχή σύνταξης επιφάνειας βασισμένης στο πρότυπο πλαισίων. Τα πλαίσια συγκεντρώνουν πληροφορίες για κάθε κλάση, κάνοντας τις οντολογίες πιο εύκολα αναγνώσιμες και κατανοητές, ιδίως για τους χρήστες που δεν είναι εξοικειωμένοι με τις (Περιγραφικές) Λογικές. Το πρότυπο πλαισίων έχει χρησιμοποιηθεί σε μια σειρά από γνωστά συστήματα αναπαράστασης γνώσης, συμπεριλαμβανομένου του εργαλείου σχεδιασμού οντολογιών Protégé και το μοντέλο γνώσεων OKBC. Ο σχεδιασμός της OIL επηρεάστηκε από την XOL.

Στις γλώσσες που είναι βασισμένες σε πλαίσια, κάθε κλάση περιγράφεται από ένα πλαίσιο. Το πλαίσιο περιλαμβάνει το όνομα της κλάσης, αναγνωρίζει τη γενικότερη κλάση (ή κλάσεις) που εξειδικεύει, και απαριθμεί μια σειρά από " slots ". Ένα slot μπορεί να αποτελείται από ένα ζεύγος τιμών ιδιοτήτων, ή έναν περιορισμό στις τιμές που μπορούν να λειτουργήσουν ως «fillers» (στο πλαίσιο αυτό, η τιμή σημαίνει είτε ένα αντικείμενο ή μια τιμή δεδομένων). Η δομή αυτή χρησιμοποιήθηκε στη γλώσσα OIL, με κάποιον εμπλουτισμό της σύνταξης για τον καθορισμό κλάσεων και τους περιορισμούς στα slots έτσι ώστε να καταστεί δυνατή η πλήρης ισχύς μιας γλώσσας σε στυλ Περιγραφικής Λογικής. Επιπλέον, πλαίσια ιδιοτήτων χρησιμοποιούνται για την περιγραφή ιδιοτήτων, π.χ., προσδιορισμός πιο γενικών ιδιοτήτων, περιορισμοί εύρους και τομέα, μεταβατικότητα και σχέσεις αντίστροφων ιδιοτήτων.

Ένα πλαίσιο κλάσης είναι σημασιολογικά ισοδύναμο με ένα αξίωμα Περιγραφικής Λογικής που υποστηρίζει ότι η κλάση που περιγράφεται από το πλαίσιο είναι μια υποκλάση καθεμιάς από τις κλάσεις που ειδικεύει και καθενός από τους περιορισμούς ιδιοτήτων που αντιστοιχούν στα slots. Εκτός από μια πιο πλούσια σύνταξη για slot, η OIL προσέφερε επίσης τη δυνατότητα ισχυρισμού ότι η κλάση που περιγράφεται από το πλαίσιο ήταν ακριβώς ισοδύναμο με την σχετική κλάση τομής. Ένα πλαίσιο ιδιοτήτων είναι ισοδύναμο με ένα σύνολο αξιωμάτων που ισχυρίζονται τους περιορισμούς σχετικά με τις σχέσεις υποδιοτήτων, το εύρος και του τομέα κ.λπ. Η OIL έχει σχεδιαστεί έτσι ώστε τα πλαίσιά της να μπορούν εύκολα να αντιστοιχηθούν σε ισοδύναμα αξιώματα στην Περιγραφική Λογική SHOQ (D).

Η επίσημη προδιαγραφή και η σημασιολογία της OWL δίνονται από μια αφηρημένη σύνταξη που έχει επηρεαστεί σε μεγάλο βαθμό από τα πλαίσια γενικότερα και από τον σχεδιασμό της OIL ειδικότερα. Στην αφηρημένη σύνταξη, τα αξιώματα είναι σύνθετες κατασκευές που μοιάζουν πολύ σε ένα πλαίσιο όπως η OIL. Για τις κλάσεις, αποτελούνται από το όνομα της κλάσης που περιγράφεται, μία τροπικότητα «μερική» ή «πλήρης» (που δηλώνει ότι το αξίωμα ισχυρίζεται μια υποκλάση ή μια σχέση ισοδυναμίας, αντίστοιχα), και μια ακολουθία περιορισμών ιδιοτήτων και ονομάτων των γενικότερων κλάσεων. Ομοίως, ένα αξίωμα ιδιότητας καθορίζει το όνομα της ιδιότητας και διάφορα χαρακτηριστικά του.

Το ύφος πλαισίου της αφηρημένης σύνταξης το καθιστά πολύ ευκολότερο στην ανάγνωση (σε σύγκριση με την RDF / XML σύνταξη), και επίσης ευκολότερο στην κατανόηση και τη χρήση. Επιπλέον, τα αξιώματα της αφηρημένης σύνταξης έχουν άμεση αντιστοιχία με τα αξιώματα της Περιγραφικής Λογικής και μπορούν, επίσης, να αντιστοιχιστούν σε ένα σύνολο RDF τριάδων.

⁷ From SHIQ and RDF to OWL: The Making of a Web Ontology Language", Ian Horrocks, Peter F. Patel-Schneider, and Frank van Harmelen

ΚΕΦΑΛΑΙΟ 4^ο

Οι Γλώσσες του Σημασιολογικού Ιστού

4.1 XML (Extensible Markup Language)

Η XML, η οποία έχει ονομαστεί από τα αρχικά των λέξεων Extensible Markup Language (Επεκτάσιμη Γλώσσα Σήμανσης), ορίστηκε από την ομάδα XML Working Group της Κοινοπραξίας Παγκόσμιου Ιστού (World Wide Web Consortium, ή αλλιώς W3C). Η ομάδα αυτή όρισε τη γλώσσα XML ως εξής:

«Η XML αποτελεί ένα υποσύνολο της SGML (Standard Generalized Markup Language – Πρότυπη Γενικευμένη Γλώσσα Σήμανσης). Σκοπός της είναι να διευκολύνει την αποστολή, τη λήψη και την επεξεργασία της γενικευμένης γλώσσας SGML στον Ιστό όπως ακριβώς γίνεται τώρα με την HTML. Η XML έχει σχεδιαστεί με στόχο την ευκολία της υλοποίησης και τη δυνατότητα παράλληλης χρήσης της με τις άλλες γλώσσες SGML και HTML.» Αυτό είναι ένα απόσπασμα από την έκδοση 1.0 της επίσημης προδιαγραφής της XML, η οποία ολοκληρώθηκε το Φεβρουάριο του 1998.

Όπως είναι φανερό, η XML είναι μια γλώσσα σήμανσης που σχεδιάστηκε ειδικά για τη μετάδοση πληροφοριών στον Παγκόσμιο Ιστό, όπως ακριβώς και η HTML (HyperText Markup Language – Γλώσσα Σήμανσης Υπερκειμένου), η οποία αποτέλεσε την κύρια γλώσσα δημιουργίας ιστοσελίδων από την έναρξη λειτουργίας του Ιστού.

4.1.1 Η ανάγκη για την XML

Η HTML παρέχει ένα σταθερό σύνολο προκαθορισμένων στοιχείων με τα οποία συμβολίζονται τα περιεχόμενα μιας τυπικής ιστοσελίδας γενικού περιεχομένου.

Παραδείγματα τέτοιων στοιχείων είναι οι επικεφαλίδες, οι ενότητες, οι λίστες, οι πίνακες, οι εικόνες και οι σύνδεσμοι. Για παράδειγμα, με την HTML είναι πολύ εύκολη η δημιουργία μιας προσωπικής αρχικής σελίδας, όπως δείχνει το παρακάτω παράδειγμα:

```
<HTML>
<HEAD>
<TITLE>Home Page</TITLE>
</HEAD>
<BODY>
<H1><IMG SRC="MainLogo.gif">Michael Young's Home Page</H1>
<P><EM>Welcome to my Web site!</EM></P>
<H2>Web Site Contents</H2>
<P>Please choose one of the following topics: </P>
<UL>
<LI><A HREF="writing.htm"><B>Writing</B></LI>
<LI><A HREF="family.htm"><B>Family</B></LI>
<LI><A HREF="photos.htm"><B>Photo Gallery</B></LI>
</UL>
<H2>Other Interesting Web Sites</H2>
<P>Click one of the following to explore another Web Site: </P>
<UL>
<LI><A HREF="http://www.yahoo.com/"><B>Yahoo Search Engine</B></LI>
```

```

<LI><A HREF="http://www.amazon.com/"><B>Amazon Bookstore</B></LI>
<LI><A HREF="http://mspress.microsoft.com/"><B>Microsoft Press</B></LI>
</UL>
</BODY>
</HTML>

```

Κάθε στοιχείο ξεκινά με μια ετικέτα αρχής: ένα τμήμα κειμένου που περικλείεται σε μια αριστερή γωνιακή αγκύλη (<) και σε μια δεξιά γωνιακή αγκύλη (>) και περιέχει το όνομα του στοιχείου και, πιθανώς, μερικές ακόμα πληροφορίες. Τα περισσότερα στοιχεία τελειώνουν με μια ετικέτα τέλους, ανάλογη με την ετικέτα αρχής μόνο που σε αυτή περιλαμβάνεται μια κάθετος (/) ακολουθούμενη από το όνομα του στοιχείου. Το περιεχόμενο του στοιχείου είναι το κείμενο (αν υπάρχει) που βρίσκεται ανάμεσα στην ετικέτα αρχής και στην ετικέτα τέλους. Αξίζει να παρατηρηθεί ότι πολλά από τα στοιχεία περιέχουν ένθετα στοιχεία.

Η σελίδα HTML του παραδείγματος περιέχει τα παρακάτω στοιχεία:

Στοιχείο HTML	Συστατικό σελίδας που σημαίνεται
HTML	Ολόκληρη η σελίδα
HEAD	Πληροφορίες επικεφαλίδας
TITLE	Τίτλος σελίδας που εμφανίζεται στη γραμμή τίτλου του περιηγητή
BODY	Το κύριο σώμα του κειμένου που εμφανίζει ο περιηγητής
H1	Επικεφαλίδα ανωτάτου επιπέδου
H2	Επικεφαλίδα δεύτερου επιπέδου
P	Ενότητα κειμένου
UL	Λίστα με κουκίδες
LI	Ένα συγκεκριμένο στοιχείο της λίστας
IMG	Εικόνα
A	Σύνδεσμος προς κάποιο άλλο σημείο ή σελίδα
EM	Τμήμα κειμένου με πλάγιους χαρακτήρες
B	Τμήμα κειμένου με έντονους χαρακτήρες

Ο περιηγητής που εμφανίζει την σελίδα HTML αναγνωρίζει αυτά τα βασικά στοιχεία και γνωρίζει πώς να τα μορφοποιήσει και να τα εμφανίσει. Αν και το σύνολο των προκαθορισμένων στοιχείων HTML έχει εμπλουτιστεί σημαντικά από την πρώτη έκδοση της HTML, η γλώσσα αυτή εξακολουθεί να είναι ακατάλληλη για τον προσδιορισμό πολλών ειδών εγγράφων. Ακολουθούν ορισμένα παραδείγματα εγγράφων που δε μπορούν να προσδιοριστούν επαρκώς με την HTML:

- Έγγραφα που δεν αποτελούνται από τυπικά στοιχεία (επικεφαλίδες, ενότητες, λίστες, πίνακες, κλπ). Για παράδειγμα, η HTML δεν έχει τα απαραίτητα στοιχεία που χρειάζονται για την σήμανση μιας μουσικής παρτιτούρας ή ενός συνόλου μαθηματικών εξισώσεων.
- Βάσεις Δεδομένων, όπως για παράδειγμα ένας κατάλογος βιβλίων. Μια σελίδα HTML μπορεί να χρησιμοποιηθεί για να αποθηκεύσει και να εμφανίσει στατικές πληροφορίες (όπως μια λίστα με περιγραφές βιβλίων). Όμως, σε περίπτωση ταξινόμησης, επιλογής, εντοπισμού και επεξεργασίας των πληροφοριών με κάποιο διαφορετικό τρόπο, για κάθε πληροφορία θα πρέπει να οριστεί και από μία ετικέτα. Η HTML δε διαθέτει τα απαραίτητα στοιχεία για κάτι τέτοιο.
- Έγγραφα που θέλουν ιεραρχική οργάνωση σε μορφή δένδρου. Ας υποθέσουμε, για παράδειγμα, ότι γράφετε ένα βιβλίο και θέλετε να το χωρίσετε σε μέρη, κεφάλαια,

ενότητες Α, ενότητες Β, ενότητες Γ, και ούτω καθεξής. Σε αυτήν την περίπτωση, ένα πρόγραμμα θα μπορούσε να χρησιμοποιήσει αυτό το δομημένο έγγραφο για να δημιουργήσει έναν πίνακα περιεχομένων, να εμφανίσει τη διάρθρωσή του με διάφορα επίπεδα λεπτομερειών, να εξαγάγει κάποιες συγκεκριμένες ενότητες και να διαχειριστεί με άλλους τρόπους τις πληροφορίες. Το στοιχείο επικεφαλίδας της HTML όμως επισημαίνει μόνο το κείμενο της επικεφαλίδας. Το πραγματικό κείμενο και τα στοιχεία που ανήκουν σε μια ενότητα του εγγράφου δεν περιέχονται μέσα στο στοιχείο της αντίστοιχης επικεφαλίδας, γι' αυτό και τα στοιχεία αυτά δε μπορούν να χρησιμοποιηθούν για να αποδώσουν με σαφήνεια την ιεραρχική δομή του εγγράφου.

Λύση σε τέτοιου είδους περιορισμούς δίνει η XML.

4.1.2 Η Λύση της XML

Ο ορισμός της XML αποτελείται από ένα στοιχειώδες μόνο συντακτικό. Όταν δημιουργείτε ένα έγγραφο XML, αντί να χρησιμοποιήσετε ένα περιορισμένο σύνολο προκαθορισμένων στοιχείων, δημιουργείτε τα δικά σας στοιχεία και τους δίνετε όποιες ονομασίες θέλετε. Αυτή είναι και η έννοια του όρου «επεκτάσιμη» στην ονομασία της γλώσσας. Επομένως, μπορείτε να χρησιμοποιήσετε την XML για την περιγραφή οποιουδήποτε είδους εγγράφου, από μουσικές παρτιτούρες μέχρι βάσεις δεδομένων. Για παράδειγμα, υπάρχει η δυνατότητα περιγραφής μίας λίστας βιβλίων, όπως στο έγγραφο XML που ακολουθεί.

```
<?xml version="1.0"?>
<INVENTORY>
<BOOK>
<TITLE>The Adventures of Huckleberry Finn</TITLE>
<AUTHOR>Mark Twain</AUTHOR>
<BINDING>mass market paperback</BINDING>
<PAGES>298</PAGES>
<PRICE>5.49</PRICE>
</BOOK>
<BOOK>
<TITLE>Moby-Dick</TITLE>
<AUTHOR>Herman Melville</AUTHOR>
<BINDING>trade paperback</BINDING>
<PAGES>605</PAGES>
<PRICE>4.95</PRICE>
</BOOK>
<BOOK>
<TITLE>The Scarlet Letter</TITLE>
<AUTHOR>Nathaniel Hawthorne</AUTHOR>
<BINDING>trade paperback</BINDING>
<PAGES>253</PAGES>
<PRICE>4.25</PRICE>
</BOOK>
</INVENTORY>
```

Είναι σημαντικό να κατανοήσουμε ότι τα ονόματα στοιχείων σε ένα έγγραφο XML δεν αποτελούν μέρος του ορισμού της XML. Στην πραγματικότητα, τα ονόματα δημιουργούνται παράλληλα με τη δημιουργία του εγγράφου. Μπορεί να γίνει επιλογή οποιωνδήποτε έγκυρων ονομάτων για τα στοιχεία.

Όπως φαίνεται και στο προηγούμενο παράδειγμα, ένα έγγραφο XML είναι δομημένο ιεραρχικά σε μορφή δένδρου, με στοιχεία που περικλείονται μέσα σε άλλα και με ένα στοιχείο ανωτάτου επιπέδου, γνωστό ως στοιχείο εγγράφου ή βασικό στοιχείο (root element), το οποίο περιέχει όλα τα υπόλοιπα.

4.1.3 Δημιουργία εγγράφων XML

Επειδή η XML δεν περιέχει προκαθορισμένα στοιχεία, ίσως να δίνει την εντύπωση ενός σχετικά ανεπίσημου προτύπου. Διαθέτει, όμως, αυστηρά καθορισμένο συντακτικό. Για παράδειγμα, σε αντίθεση με την HTML, κάθε στοιχείο XML πρέπει να έχει μια ετικέτα αρχής και μια ετικέτα τέλους (ή μια ειδική ετικέτα κενού στοιχείου). Επίσης, οποιοδήποτε ένθετο στοιχείο πρέπει να περιέχεται ολόκληρο στο στοιχείο που το περικλείει.

Για την ακρίβεια, η ίδια η ευκολία δημιουργίας στοιχείων προϋποθέτει την τήρηση ενός αυστηρού συντακτικού. Γι' αυτό και η προσαρμόσιμη φύση των εγγράφων XML απαιτεί προσαρμόσιμα προγράμματα χειρισμού και εμφάνισης των πληροφοριών που περιέχουν αυτά τα έγγραφα. Το αυστηρό συντακτικό της XML δίνει στα έγγραφα μια προβλέψιμη μορφή, κάνοντας έτσι πιο εύκολη τη γραφή αυτών των προγραμμάτων. Άλλωστε, η εύκολη εφαρμογή είναι ένας από τους πρωταρχικούς στόχους αυτής της γλώσσας.

4.1.4 Μέθοδος XML – XML schema

Μία μέθοδος XML (XML schema) είναι μια περιγραφή του τύπου του εγγράφου XML, που συνήθως εκφράζεται όσον αφορά τους περιορισμούς σχετικά με τη δομή και το περιεχόμενο των εγγράφων αυτού του τύπου, πάνω και πέρα από τους βασικούς συντακτικούς περιορισμούς που επιβάλλονται από την ίδια την XML. Αυτοί οι περιορισμοί εκφράζονται γενικά με τη χρήση κάποιου συνδυασμού γραμματικών κανόνων που διέπουν την σειρά των στοιχείων, τα κατηγορήματα Boolean που πρέπει να πληρεί το περιεχόμενο, τους τύπους δεδομένων που διέπουν το περιεχόμενο των στοιχείων και των χαρακτηριστικών, και πιο εξειδικευμένων κανόνων, όπως οι περιορισμοί μοναδικότητας και αναφορικής ακεραιότητας.

Υπάρχουν γλώσσες που έχουν αναπτυχθεί ειδικά για να εκφράσουν τις μεθόδους XML. Ο Ορισμός Τύπου Εγγράφου (Document Type Definition - DTD) είναι μια γλώσσα μεθόδου εγγενής στην προδιαγραφή XML που είναι σχετικά περιορισμένων δυνατοτήτων, αλλά έχει και άλλες χρήσεις στην XML εκτός από την έκφραση των μεθόδων. Δύο πιο εκφραστικές γλώσσες μεθόδου XML σε ευρεία χρήση είναι το XML Schema (με κεφαλαίο S) και το Relax NG.

Ο μηχανισμός για τον συσχετισμό ενός εγγράφου XML με μία μέθοδο XML ποικίλει ανάλογα με τη γλώσσα μεθόδου. Ο συσχετισμός μπορεί να επιτευχθεί μέσω σήμανσης μέσα στο ίδιο το XML έγγραφο, ή με κάποια εξωτερικά μέσα.

4.1.5 Αναπαράσταση Γνώσης με χρήση XML

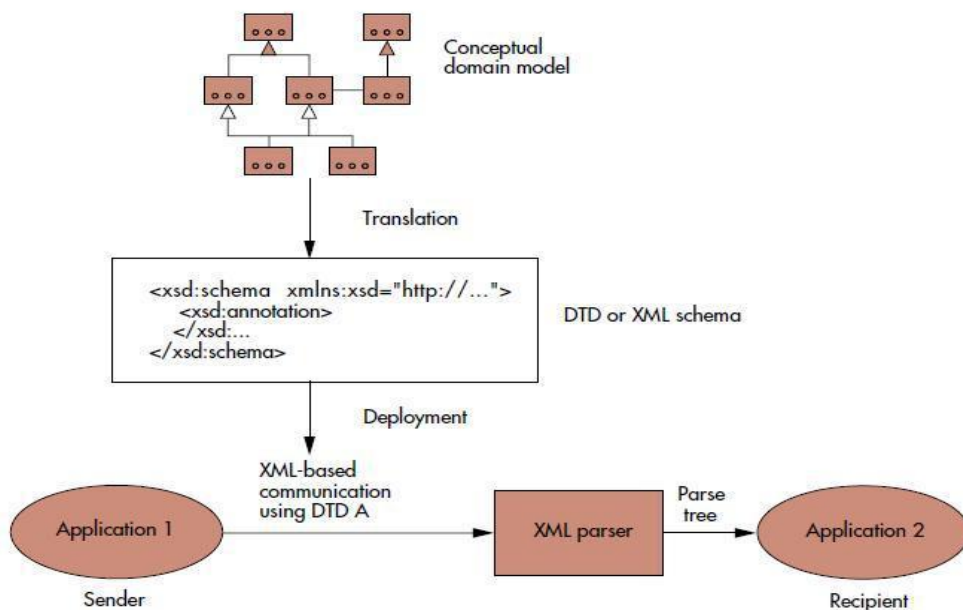
Ο Ιστός είναι η πρώτη μορφή ανταλλαγής δεδομένων μεταξύ πολλών που έχει αξιοποιηθεί ευρέως, και δημιουργεί νέες απαιτήσεις για οποιαδήποτε μορφή ανταλλαγής.

- Παγκόσμια Εκφραστική Δύναμη. Επειδή δεν είναι δυνατή η πρόβλεψη κάθε πιθανής χρήσης, μια μορφή ανταλλαγής βασισμένη στον Ιστό πρέπει να είναι σε θέση να εκφράζει οποιαδήποτε μορφή δεδομένων.
- Συντακτική Διαλειτουργικότητα. Οι εφαρμογές πρέπει να έχουν τη δυνατότητα να διαβάζουν τα δεδομένα και να φτιάχνουν μια αναπαράσταση που μπορεί να αξιοποιηθεί. Τα συστατικά λογισμικού, όπως προγράμματα ανάλυσης και APIs ερωτημάτων, για παράδειγμα, θα πρέπει να είναι όσο το δυνατόν επαναχρησιμοποιήσιμοι ανάμεσα σε διάφορες εφαρμογές. Η συντακτική διαλειτουργικότητα είναι υψηλή όταν τα προγράμματα ανάλυσης και τα APIs που είναι απαραίτητα για τον χειρισμό των δεδομένων είναι άμεσα διαθέσιμα.
- Σημασιολογική Διαλειτουργικότητα. Μία από τις πιο σημαντικές απαιτήσεις για μια μορφή ανταλλαγής δεδομένων είναι τα δεδομένα να είναι κατανοητά. Ενώ η διαλειτουργικότητα σχετίζεται με την ανάλυση των δεδομένων, η σημασιολογική διαλειτουργικότητα έχει να κάνει με τον προσδιορισμό των αντιστοιχίσεων μεταξύ των όρων μέσα στα δεδομένα, κάτι το οποίο απαιτεί ανάλυση περιεχομένου.

Η XML πληρεί την απαίτηση παγκόσμιας εκφραστικής δύναμης, καθώς για οτιδήποτε μπορεί να ορίσει η γραμματική, υπάρχει η δυνατότητα κωδικοποίησής του σε XML. Πληρεί, επίσης, την απαίτηση συντακτικής διαλειτουργικότητας, καθώς ο αναλυτής XML μπορεί να αναλύσει όλα τα δεδομένα XML, και συνήθως είναι επαναχρησιμοποιήσιμο στοιχείο. Όσον αφορά την σημασιολογική διαλειτουργικότητα, όμως, η XML έχει μειονεκτήματα.

Ο σημαντικός περιορισμός της XML είναι ότι απλώς περιγράφει γραμματικές. Δεν υπάρχει κανένας τρόπος να αναγνωριστεί μια σημασιολογική μονάδα από έναν συγκεκριμένο τομέα καθώς η XML αποσκοπεί στη δομή του εγγράφου και δεν επιβάλλει καμία κοινή ερμηνεία των στοιχείων που περιέχονται στο έγγραφο.

Το επόμενο σχήμα (Εικόνα 14) απεικονίζει δύο εφαρμογές που προσπαθούν να επικοινωνήσουν μεταξύ τους. Και οι δύο συμφωνούν σχετικά με τη χρήση και την έννοια της δομής του εγγράφου που δίνεται από το DTD A, αλλά πρέπει να κατασκευαστεί ένα μοντέλο για να αποσαφηνιστεί το είδος των δεδομένων που αποστέλλονται πριν τη δυνατότητα ανταλλαγής των δεδομένων. Τότε κατασκευάζεται ένα DTD ή μια μέθοδος XML από το μοντέλο του domain.



Εικόνα 14: Μοντέλο DTD

<http://www.w3.org/TR/xmlsig-core/>

Η XML είναι χρήσιμη για την ανταλλαγή δεδομένων μεταξύ εφαρμογών που αμφότερες γνωρίζουν τι είναι τα δεδομένα, αλλά όχι για τις περιπτώσεις όπου νέοι εταίροι επικοινωνίας προστίθεται συχνά. Στο διαδίκτυο, νέες πηγές πληροφόρησης είναι συνεχώς διαθέσιμες και νέοι επιχειρηματικοί εταίροι εντάσσονται σε υπάρχουσες σχέσεις. Επομένως, είναι σημαντικό να μειωθεί το κόστος της προσθήκης των εταίρων επικοινωνίας όσο το δυνατόν περισσότερο.

Για την ανταλλαγή εγγράφων XML, οι αντιστοιχίσεις των τομέων πρέπει να μεταφράζονται με διαδικασίες αντιστοίχισης, όπως οι XSL Transformations (XSLT). Αυτό απαιτεί μεγάλη προσπάθεια και εξαρτάται από την κωδικοποίηση που έχει χρησιμοποιηθεί για την κατασκευή των αρχικών DTDs. Πρόσθετες προσπάθειες απαιτούνται για την μετάφραση του αναδιοργανωμένου μοντέλου τομέα σε XML DTD και για τη δημιουργία διαδικασιών χαρτογράφησης των XML εγγράφων που βασίζονται στις καθιερωμένες χαρτογραφήσεις τομέων. Η χρήση ενός πιο κατάλληλου φορμαλισμού από την καθαρή XML για τη μεταφορά δεδομένων μπορεί να εξοικονομήσει μεγάλο μέρος αυτής της επιπλέον προσπάθειας.

4.2 Resource Description Framework (RDF)

Το RDF, που αναπτύχθηκε από το W3C για την περιγραφή πόρων στον Ιστό, επιτρέπει τις προδιαγραφές της σημασιολογίας των δεδομένων που βασίζονται σε XML με τυποποιημένο, διαλειτουργικό τρόπο. Παρέχει, επίσης, μηχανισμούς για να εκπροσωπεί ρητά τις υπηρεσίες, τις διαδικασίες, και τα επιχειρηματικά μοντέλα, ενώ επιτρέπει την αναγνώριση μη σαφών πληροφοριών.

Το RDF μοντέλο δεδομένων είναι ισοδύναμο με το φορμαλισμό των σημασιολογικών δικτύων. Αποτελείται από τρία είδη αντικειμένων: οι πόροι περιγράφονται από RDF εκφράσεις και ονομάζονται πάντα από τα URIs και προαιρετικά από αναγνωριστικά. Οι ιδιότητες καθορίζουν συγκεκριμένες πτυχές, χαρακτηριστικά, ιδιότητες, ή σχέσεις που χρησιμοποιούνται για να περιγράψουν έναν πόρο. Τέλος, οι δηλώσεις εκχωρούν μια τιμή σε μια ιδιότητα ενός συγκεκριμένου πόρου (η τιμή αυτή μπορεί να είναι μια άλλη δήλωση RDF).

Η ανάπτυξη του RDF έχει κίνητρο από τις ακόλουθες χρήσεις, μεταξύ των άλλων:

- Μεταδεδομένα Ιστού: παροχή πληροφοριών σχετικά με τους πόρους του Ιστού και τα συστήματα που τους χρησιμοποιούν (π.χ. διαβάθμιση του περιεχομένου, περιγραφές δυνατοτήτων, προτιμήσεις απορρήτου, κλπ.)
- Εφαρμογές που απαιτούν ανοιχτά μοντέλα πληροφοριών (π.χ. προγραμματισμός δραστηριοτήτων, περιγραφή οργανωτικών διαδικασιών, σχολιασμός πόρων του Παγκόσμιου Ιστού, κ.λπ.)
- Για να καταστήσει δυνατό για τις πληροφορίες που υποβάλλονται σε επεξεργασία από μηχανές ό, τι κατέστησε δυνατό ο Παγκόσμιος Ιστός για τα υπερκείμενα: να επιτρέψει την επεξεργασία των δεδομένων έξω από το συγκεκριμένο περιβάλλον μέσα στο οποίο δημιουργήθηκαν, με τρόπο που να μπορεί να λειτουργήσει σε διαδικτυακή κλίμακα.
- Συνεργασία μεταξύ των εφαρμογών: συνδυασμός στοιχείων από διάφορες εφαρμογές για δημιουργία νέων πληροφοριών.
- Αυτοματοποιημένη επεξεργασία πληροφοριών στον Ιστό από πράκτορες λογισμικού: ο Παγκόσμιος Ιστός μεταβαίνει προς ένα παγκόσμιο δίκτυο συνεργαζόμενων διαδικασιών. Το RDF παρέχει μια παγκόσμια γλώσσα για αυτές τις διαδικασίες.

Το RDF έχει σχεδιαστεί για να αναπαριστά πληροφορίες με έναν ελάχιστα περιοριστικό, ευέλικτο τρόπο. Μπορεί να χρησιμοποιηθεί σε μεμονωμένες εφαρμογές αλλά η γενικότητα του RDF παρέχει μεγαλύτερη αξία με την ανταλλαγή. Επομένως, η αξία των πληροφοριών αυξάνεται όσο αποκτά προσβασιμότητα σε περισσότερες εφαρμογές σε ολόκληρο το Διαδίκτυο.

Ο σχεδιασμός του RDF στοχεύει να καλύψει τους ακόλουθους στόχους:

1. Να έχει ένα απλό μοντέλο δεδομένων.

Το RDF έχει ένα απλό μοντέλο δεδομένων που καθιστά εύκολη για τις εφαρμογές την επεξεργασία και τη διαχείριση. Το μοντέλο δεδομένων είναι ανεξάρτητο οποιασδήποτε συγκεκριμένης σύνταξης σειριοποίησης.

2. Να διαθέτει επίσημη σημασιολογία και αποδείξιμα συμπεράσματα.

Το RDF έχει μια επίσημη σημασιολογία που παρέχει μια αξιόπιστη βάση για την αιτιολόγηση σχετικά με την έννοια μιας έκφρασης RDF. Ειδικότερα, υποστηρίζει αυστηρά τις ορισμένες έννοιες συνεπαγωγής που παρέχουν μια βάση για ορισμό αξιόπιστων κανόνων συμπερασμάτων σε δεδομένα RDF.

3. Να χρησιμοποιεί ένα επεκτάσιμο λεξιλόγιο βασισμένο στο URI.

Το λεξιλόγιο είναι πλήρως επεκτάσιμο, βασιζόμενο σε URIs με προαιρετικά τμήματα αναγνωριστικών(αναφορές URI, ή URIsrefs). Οι αναφορές URI χρησιμοποιούνται για την ονομασία όλων των ειδών πραγμάτων στο RDF.

4. Να χρησιμοποιεί σύνταξη βασισμένη στην XML.

Το RDF έχει μία συνιστώμενη μορφή σειριοποίησης XML, η οποία μπορεί να χρησιμοποιηθεί για να κωδικοποιησει το μοντέλο δεδομένων για ανταλλαγή πληροφοριών μεταξύ των εφαρμογών.

5. Να υποστηρίζει τη χρήση των τύπων δεδομένων του XML σχήματος.

Το RDF μπορεί να χρησιμοποιήσει τιμές που παρουσιάζονται σύμφωνα με τους τύπους δεδομένων του XML σχήματος, βοηθώντας, έτσι, την ανταλλαγή πληροφοριών μεταξύ RDF και άλλων XML εφαρμογών.

6. Να επιτρέπει σε οποιονδήποτε να κάνει δηλώσεις σχετικά με οποιαδήποτε πηγή.

Για να διευκολύνει τη λειτουργία σε κλίμακα του Διαδικτύου, το RDF είναι ένα ανοικτό παγκόσμιο πλαίσιο που επιτρέπει σε οποιονδήποτε να προβεί σε δηλώσεις σχετικά με οποιαδήποτε πηγή.

Σε γενικές γραμμές, δεν είναι δεδομένο ότι είναι διαθέσιμες ολοκληρωμένες πληροφορίες σχετικά με οποιοδήποτε πόρο. Το RDF δεν αποτρέπει κανέναν από το να κάνει ισχυρισμούς που είναι παράλογοι ή δεν συνάδουν με άλλες δηλώσεις, ή τον κόσμο όπως τον βλέπουν οι άνθρωποι. Οι σχεδιαστές εφαρμογών που χρησιμοποιούν το RDF θα πρέπει να το γνωρίζουν αυτό και μπορούν να σχεδιάζουν τις εφαρμογές τους έτσι, ώστε να αντιμετωπίζουν τις μη πλήρεις ή ασυνεπείς πηγές πληροφοριών.

Το RDF χρησιμοποιεί τις ακόλουθες βασικές έννοιες:

Μοντέλο δεδομένων γραφημάτων

Λεξιλόγιο με βάση το URI

Τύποι δεδομένων

Λεκτικές Σταθερές

Σύνταξη XML σειριοποίησης

Έκφραση απλών γεγονότων

Συνεπαγωγή

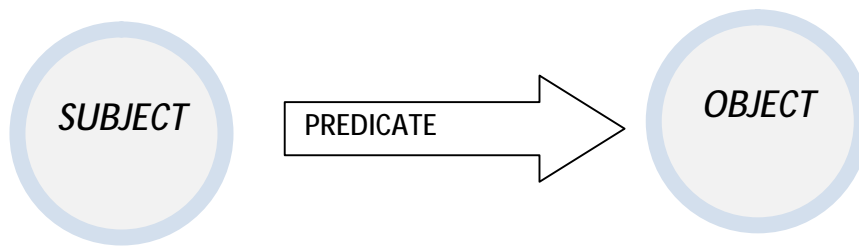
Μοντέλο δεδομένων γραφημάτων

Η βασική δομή της κάθε έκφρασης σε RDF είναι μια συλλογή από τριάδες, η καθεμία αποτελούμενη από ένα υποκείμενο, ένα κατηγορημα και ένα αντικείμενο. Ένα σύνολο τέτοιων τριάδων αποτελεί ένα γράφημα RDF. Αυτό μπορεί να απεικονιστεί από ένα διάγραμμα με κόμβους και κατευθυνόμενα τόξα, στο οποίο κάθε τριάδα αναπαρίσταται ως ένας σύνδεσμος κόμβου-τόξου-κόμβου.

Κάθε τριάδα αντιπροσωπεύει μια δήλωση μιας σχέσης μεταξύ των πραγμάτων που συμβολίζονται με τους κόμβους που ενώνει. Κάθε τριάδα αποτελείται από τρία μέρη:

1. ένα υποκείμενο,
2. ένα αντικείμενο, και

3. ένα κατηγορημα, που ορίζει μία σχέση.



Η κατεύθυνση του τόξου είναι σημαντική και είναι πάντοτε στραμμένη προς το αντικείμενο. Οι κόμβοι ενός γραφήματος RDF είναι τα υποκείμενα και τα αντικείμενά του.

Ο ισχυρισμός μιας RDF τριάδας λέει ότι κάποια σχέση, που υποδεικνύεται από το κατηγορημα, συνδέει τα πράγματα που έχουν οριστεί ως το υποκείμενο και το αντικείμενο της τριάδας.

Λεξιλόγιο με βάση το URI και Αναγνώριση Κόμβου

Ένας κόμβος μπορεί να είναι ένα URI με ένα προαιρετικό αναγνωριστικό τμήμα (URIref), μία κυριολεκτική έννοια, ή κενός.

Μία αναφορά URI ή μία λεκτική σταθερά χρησιμοποιημένες ως κόμβοι, προσδιορίζουν τι αντιπροσωπεύει ο κόμβος. Μία αναφορά URI, χρησιμοποιημένη ως κατηγορημα, προσδιορίζει μια σχέση μεταξύ των πραγμάτων που εκπροσωπούνται από τους κόμβους που συνδέει. Μία αναφορά URI, χρησιμοποιημένη ως κατηγορημα μπορεί επίσης να είναι κόμβος στο διάγραμμα.

Ένας κενός κόμβος είναι ένας κόμβος που δεν είναι μια αναφορά URI ή λεκτική σταθερά. Στην αφηρημένη RDF σύνταξη, ένας κενός κόμβος είναι απλώς ένας μοναδικός κόμβος που μπορεί να χρησιμοποιηθεί σε μια ή περισσότερες δηλώσεις RDF, αλλά δεν έχει ουσιαστική ονομασία.

Μια σύμβαση που χρησιμοποιείται από γραμμικές παραστάσεις ενός γραφήματος RDF για να επιτρέψει σε πολλές δηλώσεις να αναφέρουν τον ίδιο απροσδιόριστο πόρο είναι η χρήση ενός κενού κόμβου αναγνωριστικού, ο οποίος είναι ένα τοπικό αναγνωριστικό που μπορεί να διακριθεί από όλα τα URIs και τις λεκτικές σταθερές.

Όταν συγχωνεύονται γραφήματα, οι κενοί κόμβοι πρέπει να διαχωρίζονται, εφόσον η έννοια πρέπει να διατηρηθεί. Αυτό μπορεί να απαιτήσει εκ νέου κατανομή των κενών αναγνωριστικών κόμβων.

Τύποι Δεδομένων

Οι τύποι δεδομένων χρησιμοποιούνται από το RDF στην αναπαράσταση τιμών όπως ακέραιοι αριθμοί, αριθμοί κινητής υποδιαστολής και ημερομηνίες.

Το RDF προκαθορίζει έναν μόνο τύπο δεδομένων rdf: XMLLiteral8, που χρησιμοποιείται για την ενσωμάτωση της XML σε RDF.

Το RDF δεν προβλέπει κανένα μηχανισμό για ορισμό νέων τύπων δεδομένων. Το XML Schema Datatypes παρέχει ένα πλαίσιο επέκτασης κατάλληλο για ορισμό νέων τύπων δεδομένων για χρήση σε RDF.

Λεκτικές Σταθερές

Οι Λεκτικές σταθερές χρησιμοποιούνται για τον προσδιορισμό τιμών, όπως αριθμοί και οι ημερομηνίες μέσω μιας λεκτικής αναπαράστασης. Οτιδήποτε προσδιορίζεται με μια λεκτική

σταθερά θα μπορούσε επίσης να προσδιορίζεται από ένα URI, αλλά είναι συχνά πιο κατάλληλη και εύκολη η χρήση λεκτικών σταθερών. Μια λεκτική σταθερά μπορεί να είναι το αντικείμενο μιας δήλωσης RDF, αλλά όχι το υποκείμενο ή το κατηγορημα.

Έκφραση Απλών Γεγονότων

Μερικά απλά γεγονότα υποδεικνύουν μια σχέση ανάμεσα σε δύο πράγματα. Ένα τέτοιο γεγονός μπορεί να αντιπροσωπεύεται από μια RDF τριάδα στην οποία το κατηγορημα ονομάζει την σχέση, και το υποκείμενο και το αντικείμενο δηλώνουν τα δύο πράγματα. Μια γνωστή αναπαράσταση ενός τέτοιου γεγονότος μπορεί να είναι η γραμμή ενός πίνακα σε μια σχεσιακή βάση δεδομένων. Ο πίνακας έχει δύο στήλες, που αντιστοιχούν στο αντικείμενο και το υποκείμενο της RDF τριάδας. Το όνομα του πίνακα αντιστοιχεί στο κατηγορημα.

Οι σχεσιακές βάσεις δεδομένων επιτρέπουν σε έναν πίνακα να έχουν αυθαίρετο αριθμό στηλών, μια σειρά του οποίου εκφράζει πληροφορίες που αντιστοιχούν σε ένα κατηγορημα στη λογική πρώτης τάξης με έναν αυθαίρετο αριθμό θέσεων. Μια τέτοια σειρά, ή κατηγορημα, πρέπει να αποσυντεθεί για την αναπαράστασή της σε RDF τριάδες. Μια απλή μορφή αποσύνθεσης εισάγει ένα νέο κενό κόμβο, που αντιστοιχεί στη γραμμή, και μια νέα τριάδα εισάγεται για κάθε κελί της γραμμής. Το υποκείμενο της κάθε τριάδας είναι ο νέος κενός κόμβος, το κατηγορημα αντιστοιχεί στο όνομα της στήλης, και το αντικείμενο αντιστοιχεί στην τιμή του κελιού. Ο νέος κενός κόμβος μπορεί να έχει επίσης μια ιδιότητα `rdf:type` η τιμή του οποίου αντιστοιχεί στο όνομα του πίνακα.

Αυτές οι πληροφορίες θα μπορούσαν να αντιστοιχούν στη γραμμή ενός πίνακα με το όνομα "STAFFADDRESSES", με primary key STAFFID και στήλες STREET, STATE, CITY και POSTALCODE.

Ένα πιο σύνθετο γεγονός εκφράζεται σε RDF χρησιμοποιώντας σύζευξη (λογικό- ΚΑΙ) απλών δυαδικών σχέσεων. Το RDF δεν παρέχει τα μέσα για την έκφραση άρνησης (NOT) ή διάζευξης (OR).

Μέσω της χρήσης του επεκτάσιμου λεξιλογίου με βάση το URI, το RDF επιτρέπει την έκφραση γεγονότων σχετικά με αυθαίρετα θέματα. Ένα URI μπορεί να κατασκευαστεί για κάθε πράγμα που μπορεί να ονομαστεί, έτσι τα RDF γεγονότα μπορούν να αφορούν οποιαδήποτε τέτοιου είδους πράγματα.

Συνεπαγωγή

Οι ιδέες για την σημασία και το συμπέρασμα σε RDF υποστηρίζεται από την επίσημη έννοια της συνεπαγωγής. Εν συντομία, μια RDF έκφραση A λέγεται ότι συνεπάγεται μια άλλη RDF έκφραση B εάν κάθε πιθανή διάταξη των πραγμάτων στον κόσμο που κάνει την A αληθινή κάνει και τη B αληθινή. Βάσει αυτών, αν η αλήθεια του A ή θεωρείται δεδομένη ή αποδεικνύεται, τότε συνεπάγεται και η αλήθεια του B.

Το RDF μοντέλο δεδομένων δεν παρέχει μηχανισμούς για τον προσδιορισμό των σχέσεων μεταξύ των ιδιοτήτων και των πόρων. Αυτός είναι ο ρόλος του RDFS. Το RDFS προσφέρει αρχέτυπα για τον καθορισμό μοντέλων γνώσης που είναι πιο κοντά σε προσεγγίσεις βασισμένες σε πλαίσια. Το RDF (S) χρησιμοποιείται ευρέως ως μια μορφή αναπαράστασης σε πολλά εργαλεία και προγράμματα, όπως τα (Amaya, Protégé, Mozilla, SilR)I, και ούτω καθεξής.

4.2.1 Αναπαράσταση Δεδομένων σχετικά με Δεδομένα

Η XML παρέχει μια σύνταξη για την κωδικοποίηση των δεδομένων. Το RDF είναι ένας μηχανισμός που λέει πράγματα σχετικά με τα δεδομένα. Όπως υποδηλώνει το όνομα, δεν είναι μια γλώσσα, αλλά ένα μοντέλο για την αναπαράσταση των δεδομένων σχετικά με «τα πράγματα στον Ιστό». Αυτό το είδος των δεδομένων σχετικά με τα δεδομένα ονομάζεται μεταδεδομένα. Τα "πράγματα" είναι οι πόροι στο λεξιλόγιο RDF. Το βασικό μοντέλο δεδομένων του RDF είναι απλό.

Εκτός από τους πόρους, περιέχει τις ιδιότητες και τις δηλώσεις. Μια ιδιότητα είναι ένα συγκεκριμένο θέμα, χαρακτηριστικό, ιδιότητα ή σχέση που περιγράφει έναν πόρο. Μια δήλωση αποτελείται από έναν συγκεκριμένο πόρο με την ονομασμένη ιδιότητα συν την τιμή της εν λόγω ιδιότητας για αυτόν τον πόρο. Αυτή η τιμή μπορεί να είναι άλλος πόρος ή μια λεκτική σταθερά. Συνολικά, μία περιγραφή RDF είναι μια λίστα τριάδων: ένα υποκείμενο (ένας πόρος), ένα κατηγορημα (μια ιδιότητα), και ένα αντικείμενο (μια τιμή που μπορεί να είναι ένας πόρος ή ελεύθερο κείμενο). Για παράδειγμα, ο πίνακας 1 παρουσιάζει τις τρεις τριάδες που είναι απαραίτητες για να διευκρινίσουν ότι μια συγκεκριμένη ιστοσελίδα δημιουργήθηκε από κάποιον με το όνομα "John" και με αριθμό τηλεφώνου «47782».

Subject	Predicate	Object
http://www.w3.org/	created_by	#anonymous_resource1
#anonymous_resource1	name	"John"
#anonymous_resource1	phone	"47782"

Πίνακας 1: Μια περιγραφή RDF αποτελούμενη από τρεις τριάδες

Η απεικόνιση ενός μοντέλου RDF είναι εύκολη με ένα κατευθυνόμενο γράφημα. Κάθε πόρος αναπαρίσταται με ένα οβάλ και κάθε κατηγορημα με ένα τόξο, και οι λεκτικές τιμές αναπαρίστανται από κουτιά με τιμές. Το σχήμα δείχνει ένα τέτοιο γράφημα για τα τρίκλινα στον πίνακα 1.

Το παράδειγμα αποκαλύπτει ότι το RDF αγνοεί την σύνταξη. Παρέχει μόνο ένα μοντέλο για την αναπαράσταση μεταδεδομένων. Η λίστα με τις τριάδες είναι μία πιθανή αναπαράσταση, όπως είναι, επίσης, και το επισημασμένο γράφημα, και άλλες συντακτικές αναπαραστάσεις. Φυσικά, η XML θα ήταν μια προφανής υποψηφιότητα για μια εναλλακτική αναπαράσταση. Οι προδιαγραφές του μοντέλου δεδομένων περιλαμβάνουν μια τέτοια βασισμένη σε XML κωδικοποίηση του RDF.

Όπως με την XML, ένα μοντέλο RDF δεν ορίζει την σημασιολογία της κάθε εφαρμογής τομέα ή κάνει υποθέσεις σχετικά με έναν συγκεκριμένο τομέα. Απλά προσφέρει έναν μηχανισμό ουδέτερου τομέα για την περιγραφή μεταδεδομένων. Ο ορισμός των ιδιοτήτων του τομέα και της σημασιολογίας του απαιτεί πρόσθετες εγκαταστάσεις.

4.2.2 RDF Schema

Βασικά, το RDF Schema είναι ένα απλό σύστημα τύπου για το RDF. Παρέχει ένα μηχανισμό ορισμού ιδιοτήτων σχετικά με τους τομείς και τάξεις των πόρων στους οποίους μπορούν να εφαρμοστούν αυτές οι ιδιότητες.

Οι βασικές αρχές μοντέλου στο RDF Schema είναι ορισμοί κλάσεων και υποκατηγορίες των δηλώσεων (που από κοινού επιτρέπουν τον ορισμό των ιεραρχιών κλάσεων), των

ορισμών ιδιοτήτων και των υπό-ιδιοτήτων των δηλώσεων (για την κατασκευή ιεραρχιών ιδιοτήτων), των τομέων και του εύρους δηλώσεων (για τον περιορισμό των πιθανών συνδυασμών ιδιοτήτων και κλάσεων), καθώς και των δηλώσεων τύπου (για τη δήλωση ενός πόρου ως στιγμιότυπο μιας συγκεκριμένης κλάσης). Με αυτές τις αρχές είναι δυνατή η δημιουργία ενός Schema για έναν συγκεκριμένο τομέα. Στο παράδειγμα μπορεί να οριστεί ένα Schema που δηλώνει δύο κλάσεις πόρων, Person και WebPage, και δύο ιδιότητες, name και phone, αμφότερες με τον τομέα Person και λεκτικό (Literal) εύρος. Αυτό το Schema θα μπορούσε να χρησιμοποιηθεί για να ορίσει τον πόρο <http://www.w3.org/> ως ένα στιγμιότυπο της κλάσης WebPage και το anonymous resource ως ένα στιγμιότυπο της κλάσης Person. Αυτό θα δώσει κάποια ερμηνεία και δυνατότητες επικύρωσης στα δεδομένα RDF.

Το RDF Schema είναι αρκετά απλό σε σύγκριση με τις πλήρεις γλώσσες αναπαράστασης γνώσης. Επίσης, εξακολουθεί να μην παρέχει ακριβή σημασιολογία. Ωστόσο, η παράλειψη αυτή είναι εν μέρει εκ προθέσεως. Το W3C προβλέπει και τάσσεται υπέρ της περαιτέρω επέκτασης στο RDF Schema.

Επειδή οι προδιαγραφές του RDF Schema είναι επίσης ένα είδος μεταδεδομένων, μπορεί να χρησιμοποιηθεί το RDF για την κωδικοποίησή του. Αυτό ακριβώς συμβαίνει στο έγγραφο προδιαγραφής RDF Schema. Επιπλέον, η προδιαγραφή παρέχει ένα έγγραφο RDF Schema που καθορίζει τις ιδιότητες και τις κλάσεις που εισήγαγαν οι προδιαγραφές του RDF Schema. Όπως και με την προδιαγραφή του XML Schema, ένας τέτοιος αναδρομικός ορισμός του RDF Schema φαίνεται κάπως συγκεχυμένος.

4.2.3 Αναπαράσταση Γνώσης με χρήση RDF

Η ένθετη «υποκείμενο-κατηγορημα-αντικείμενο» δομή του RDF ικανοποιεί την απαίτηση παγκόσμιας εκφραστικής δύναμης για μια μορφή ανταλλαγής, αν και αυτό δεν είναι εύκολο να φανεί. Μεταγλωττιστές RDF ανεξάρτητοι εφαρμογών είναι επίσης διαθέσιμοι, έτσι το RDF πληρεί και την απαίτηση για συντακτική διαλειτουργικότητα.

Όσον αφορά την σημασιολογική διαλειτουργικότητα, το RDF έχει σημαντικά πλεονεκτήματα έναντι της XML. Η διάρθρωση αντικείμενο-κατηγορημα παρέχει φυσικές σημασιολογικές μονάδες, καθώς όλα τα αντικείμενα είναι ανεξάρτητες οντότητες. Ένα μοντέλο τομέα, για τον καθορισμό των σχέσεων και των αντικειμένων, μπορεί να αναπαρασταθεί φυσικά στο RDF, κι έτσι, δεν είναι απαραίτητα τα βήματα μετάφρασης όπως στην XML. Για την εύρεση αντιστοιχίσεων μεταξύ δυο περιγραφών RDF, τεχνικές από την έρευνα στην αναπαράσταση της γνώσης είναι άμεσα εφαρμόσιμες. Φυσικά, αυτό δε λύνει το πρόβλημα της γενικής διαλειτουργικότητας της εύρεσης σημασιολογίας που να διατηρεί τις αντιστοιχίσεις μεταξύ των αντικειμένων, αλλά χρησιμοποιώντας RDF για ανταλλαγή δεδομένων ανεβαίνει το επίπεδο της δυναμικής επαναχρησιμοποίησης των στοιχείων του λογισμικού πολύ πέρα από την επαναχρησιμοποίηση του μεταγλωττιστή, το οποίο αποτελεί ό, τι έχει να προσφέρει η XML. Επιπλέον, το μοντέλο RDF (και του λογισμικό που χρησιμοποιεί το μοντέλο RDF), μπορεί ακόμα να χρησιμοποιηθεί ακόμη και αν αλλάξει ή εξαφανιστεί η σύνταξη XML, καθώς το RDF περιγράφει ένα στρώμα ανεξάρτητο της XML.

4.3 Γλώσσες Οντολογίας - Προκάτοχοι της OWL

Οι οντολογίες έχουν αποδειχθεί ουσιαστικό στοιχείο πολλών εφαρμογών. Χρησιμοποιούνται σε συστήματα πρακτόρων (agent systems)¹⁰, συστήματα διαχείρισης γνώσης, και πλατφόρμες ηλεκτρονικού εμπορίου. Μπορούν, επίσης, να παράγουν φυσική γλώσσα, να ενσωματώσουν ευφυείς πληροφορίες, να παρέχουν σημασιολογική πρόσβαση στο Διαδίκτυο, και να εξάγουν πληροφορίες από κείμενα, πέραν της χρήσης τους σε πολλές άλλες εφαρμογές με σκοπό να δηλώσουν ρητά τη γνώση που είναι ενσωματωμένη σε αυτά.

Ωστόσο, οι οντολογίες δεν είναι μόνο χρήσιμες στις εφαρμογές στις οποίες η γνώση παίζει σημαντικό ρόλο, αλλά μπορούν, επίσης, να προκαλέσουν σημαντικές αλλαγές στο τωρινό περιεχόμενο του Ιστού. Η αλλαγή αυτή οδηγεί στην τρίτη γενιά του Ιστού- γνωστή ως ο Σημασιολογικός Ιστός, η οποία έχει οριστεί ως «η εννοιολογική διάρθρωση του Ιστού με σαφές αναγνώσιμο από μηχανές τρόπο. Ο ορισμός αυτός δε διαφέρει πολύ από εκείνον που χρησιμοποιείται για να ορίσει μια οντολογία: “Μια οντολογία είναι μια ρητή, αναγνώσιμη από μηχανές προδιαγραφή μιας κοινής εννοιολογικής σύλληψης”.

Στην πραγματικότητα, νέες, βασισμένες στην οντολογία εφαρμογές και αρχιτεκτονικές γνώσεις έχουν αναπτυχθεί για αυτόν το νέο Ιστό. Μια κοινή αξίωση όλων αυτών των προσεγγίσεων είναι η ανάγκη να αντιπροσωπεύουν οι γλώσσες τις σημασιολογικές πληροφορίες που απαιτεί αυτός ο Ιστός, επιλύοντας τις ετερογενείς ανταλλαγές δεδομένων σε αυτό το ετερογενές περιβάλλον.

Πολλές γλώσσες οντολογίας έχουν αναπτυχθεί κατά τα τελευταία χρόνια, και σίγουρα θα γίνουν γλώσσες οντολογίας στα πλαίσια του Σημασιολογικού Ιστού. Κάποιες από αυτές βασίζονται σε σύνταξη XML, όπως η XOL (XML-based Ontology Exchange Language), η SHOE (η οποία βασίζεται στην HTML), και η OML (Ontology Markup Language), ενώ η RDF (Resource Description Framework) και RDF Schema είναι γλώσσες που δημιουργήθηκαν από το World Wide Web Consortium (W3C). Τέλος, δύο επιπλέον γλώσσες αναπτύσσονται πάνω από την RDF(S) με σκοπό να βελτιώσουν τα χαρακτηριστικά της: η OIL (Ontology Inference Layer) και DAML + OIL.

4.3.1 XML-based Ontology Exchange Language (XOL)

Η αμερικανική κοινότητα βιοπληροφορικής σχεδίασε την XOL για την ανταλλαγή ορισμών οντολογιών μεταξύ ετερογενών συνόλων συστημάτων λογισμικού στον τομέα (domain) τους. Οι ερευνητές το δημιούργησαν ύστερα από μελέτη των αναγκών αναπαράστασης των εμπειρογνομόνων στη βιοπληροφορική. Επέλεξαν τις Ontolingua και OML ως βάση για τη δημιουργία της XOL, συγχωνεύοντας την υψηλή εκφραστικότητα της OKBC-Lite, ένα υποσύνολο του πρωτοκόλλου Open Knowledge Based Connectivity, καθώς και την σύνταξη της OML, που βασίζεται στην XML.

Η σύνταξη της XOL βασίζεται στην XML, η οποία είναι μια γλώσσα για τη σύνταξη εγγράφων για τον Παγκόσμιο Ιστό. Η σύνταξη της XML επιλέχθηκε επειδή είναι εύλογα εύκολο να αναλυθεί, η σύνταξή της είναι σαφώς ορισμένη, είναι αναγνώσιμη από τον άνθρωπο, φαίνεται ότι η XML θα χρησιμοποιείται ευρύτατα, και φαίνεται, επίσης, ότι πολλά εργαλεία λογισμικού για ανάλυση και χειρισμό της XML θα είναι σύντομα διαθέσιμα. Η σημασιολογία της XOL βασίζεται στην OKBC-Lite, η οποία είναι μια απλουστευμένη μορφή του μοντέλου γνώσης για την OKBC (Open Knowledge Base Connectivity). Η OKBC είναι ένα API για την πρόσβαση σε συστήματα αναπαράστασης γνώσης πλαισίου. Το μοντέλο γνώσης της υποστηρίζει τα χαρακτηριστικά που εντοπίζονται συχνότερα στα συστήματα αναπαράστασης γνώσης, στις βάσεις δεδομένων αντικειμένων και τις σχεσιακές βάσεις δεδομένων (αν και έως τώρα δεν υποστηρίζονται όλοι οι τύποι δεδομένων για τις βάσεις

δεδομένων αντικειμένων και τις σχεσιακές). Η OKBC-Lite έχει αποσπάσει τα περισσότερα από τα ουσιώδη χαρακτηριστικά της OKBC, ενώ δεν περιλαμβάνει μερικές από τις πιο σύνθετες πτυχές του.

Η XOL είναι παρόμοια με άλλες γλώσσες ανταλλαγής οντολογιών. Η ανάπτυξη της εμπνεύστηκε από την Ontolingua και την OML. Η XOL διαφέρει από την Ontolingua στην σύνταξη, καθώς βασίζεται σε XML και όχι σε Lisp. Η σημασιολογία της OKBC-Lite είναι πολύ παρόμοια με εκείνη της Ontolingua. Η XOL διαφέρει από την OML με την έννοια ότι η σημασιολογία της OML βασίζεται σε εννοιολογικά γραφήματα, τα οποία έχουν ορισμένες διαφορές από την OKBC-Lite.

Το μοντέλο γνώσης OKBC-Lite παρέχει ακριβείς ορισμούς για τις αντιλήψεις, τις σχέσεις, τα αντικείμενα και τους περιορισμούς. Το OKBC-Lite μοντέλο μπορεί να εκφράζει τα σχήματα που χρησιμοποιούνται σε ορισμένους τύπους συστημάτων διαχείρισης πληροφοριών, όπως τα σχήματα σχεσιακών βάσεων δεδομένων, τα σχήματα βάσεων δεδομένων αντικειμένων και τις βάσεις γνώσης που χρησιμοποιούνται σε συστήματα αναπαράστασης γνώσης πλαισίου. Το OKBC-Lite συναθροίζει, επίσης, λιγότερο σύνθετους τύπους οντολογιών, συμπεριλαμβανομένων των ελεγχόμενων λεξιλογίων και των ταξινομήσεων.

Πέρα από τις πληροφορίες σχήματος, η οντολογία περιλαμβάνει, επίσης, άλλες πληροφορίες, όπως μεμονωμένα αντικείμενα μέσα σε έναν διαχειριστή βάσης δεδομένων αντικειμένων, ή μεμονωμένες γραμμές μέσα σε μια σχεσιακή βάση δεδομένων. Η συμπερίληψη και των σχημάτων και των δεδομένων στα πλαίσια του ορισμού της οντολογίας, δεν είναι ασυνήθιστη δεδομένης της κεντρικής σημασίας της επαναχρησιμοποίησης στην ανάπτυξη των οντολογιών. Είναι σύνηθες για τους προγραμματιστές εφαρμογών να θέλουν να χρησιμοποιήσουν ξανά τα δεδομένα καθώς και τις πληροφορίες σχήματος στις εφαρμογές.

Η ανταλλαγή οντολογιών είναι σημαντική για πολλούς λόγους.

Πρώτον, η ανάπτυξη οντολογιών που αποτελεί προϋπόθεση για τον σχεδιασμό βάσεων δεδομένων και για την ανάπτυξη συστημάτων βασισμένων στη γνώση είναι δύσκολη και χρονοβόρα. Διαφορετικές ομάδες ή οργανισμοί που επιθυμούν την ανάπτυξη βάσεων δεδομένων για τους ίδιους τύπους πληροφοριών συχνά θα καταλήξουν σε μια λύση πιο γρήγορα υιοθετώντας μια υπάρχουσα οντολογία παρά με την ανάπτυξη μιας νέας οντολογίας.

Δεύτερον, αν διάφορες βάσεις δεδομένων που καλύπτουν τους ίδιους τύπους δεδομένων (π.χ., αρχεία εργαζομένων), χρησιμοποιούν την ίδια οντολογία, απλοποιούν το πρόβλημα της ολοκλήρωσης των βάσεων δεδομένων, δηλαδή, την επεξεργασία ερωτημάτων σε πολλαπλές βάσεις δεδομένων. Η ύπαρξη διαφορετικών οντολογιών για τους ίδιους τύπους δεδομένων δημιουργεί ένα σημασιολογικό πρόβλημα αναντιστοιχίας που περιπλέκει το πρόβλημα αυτό.

Τρίτον, οι προγραμματιστές πολλών βάσεων δεδομένων επιθυμούν να κοινοποιήσουν τα σχήματά τους στους χρήστες τους σε μια τυπικά καθορισμένη, εύκολα κατανοητή μορφή, έτσι, ώστε οι χρήστες να έχουν πλήρη κατανόηση της σημασιολογίας αυτών των βάσεων δεδομένων.

Τέταρτον, η κοινή χρήση οντολογιών είναι σημαντική, διότι οι ίδιες οι οντολογίες αποτελούν μια μορφή γνώσης που κάποιες κοινότητες (όπως οι επιστημονικοί κλάδοι) επιθυμούν να μοιραστούν.

Η γλώσσα XOL έχει σχεδιαστεί για να παρέχει ένα μηχανισμό για την κωδικοποίηση οντολογιών μέσα σε ένα απλό αρχείο που μπορεί εύκολα να δημοσιευτεί στον Παγκόσμιο Ιστό για ανταλλαγή μεταξύ ενός συνόλου προγραμματιστών εφαρμογών. Η γλώσσα έχει σχεδιαστεί έτσι, ώστε να είναι αναγνώσιμη από τον άνθρωπο και να είναι εύκολα μεταφράσιμη από τα προγράμματα μέτριας πολυπλοκότητας. Έχει σχεδιαστεί ώστε να είναι εύκολα κατανοητή από ανθρώπους που δεν έχουν πλήρη κατανόηση της XML. Έχει, επίσης,

σχεδιαστεί για να είναι εκφραστική, έτσι, ώστε να μπορεί να συλλάβει μια πλούσια ποικιλία οντολογιών.

Η XOL είναι μια γλώσσα για την ανταλλαγή οντολογιών. Λέγοντας γλώσσα ανταλλαγής, εννοούμε ότι η XOL προορίζεται να χρησιμοποιηθεί ως ενδιάμεση γλώσσα για τη μεταφορά οντολογιών μεταξύ διαφόρων συστημάτων βάσεων δεδομένων, εργαλείων ανάπτυξης οντολογιών ή προγραμμάτων εφαρμογών.

Για παράδειγμα, μια ομάδα που αναπτύσσει μια επιστημονική βάση δεδομένων θα μπορούσε να χρησιμοποιήσει το σύστημα διαχείρισης βάσεων δεδομένων (ΣΔΒΔ) της Oracle για να υλοποιήσει τη ΒΔ. Ωστόσο, η ομάδα θα μπορούσε να μεταφράσει το σχήμα ΣΔΒΔ από SQL σε XOL, και στη συνέχεια να δημοσιεύσει το αρχείο που προκύπτει στον Παγκόσμιο Ιστό για παραπομπή στους χρήστες της βάσης δεδομένων, ή σε άλλες ομάδες που αναπτύσσουν παρόμοιες βάσεις δεδομένων. Η ομάδα μπορεί να μετατρέψει το σχήμα της σε XOL, χρησιμοποιώντας έναν υπάρχοντα από SQL σε XOL μεταφραστή. Ή, αν η ομάδα σχεδιάσει το σχήμα ΣΔΒΔ με χρήση γραφικού εργαλείου, όπως ένα εργαλείο UML, θα μπορούσε να χρησιμοποιήσει έναν προϋπάρχοντα μεταφραστή από UML σε XOL για να μεταφράσει το σχήμα σε XOL.

Μια άλλη επιστημονική ομάδα ανάπτυξης βάσης δεδομένων θα μπορούσε να βρει την παραπάνω οντολογία σε μορφή XOL στον Παγκόσμιο Ιστό. Η ομάδα θα μπορούσε να χρησιμοποιήσει ένα άλλο πρόγραμμα μετάφρασης για να μετατρέψει την οντολογία στη μορφή που χρησιμοποιείται από το ΣΔΒΔ της ή το εργαλείο ανάπτυξης οντολογιών.

Η XOL μπορεί, επίσης, να χρησιμοποιηθεί για ανταλλαγή στιγμιότυπων βάσεων δεδομένων. Για παράδειγμα, όλα τα στιγμιότυπα αντικειμένων μέσα σε μια αντικειμενοστρεφή βάση δεδομένων, ή μια σχεσιακή βάση δεδομένων, θα μπορούσαν να αποθηκευτούν σε ένα αρχείο XOL για μεταφορά σε άλλο οργανισμό που χρησιμοποιεί είτε το ίδιο ΣΔΒΔ με τον αρχικό οργανισμό, είτε ένα διαφορετικό ΣΔΒΔ. Και πάλι, για να είναι εφικτή η ανταλλαγή πρέπει να κατασκευαστούν μεταφραστές μεταξύ κάθε ΣΔΒΔ και της γλώσσας XOL.

Η XOL δεν είναι ένα περιβάλλον λογισμικού για την ανάπτυξη οντολογιών, όπως η Ontolingua ή το GKB-Editor. Ωστόσο, δεδομένου ότι τα XOL αρχεία είναι αρχεία κειμένου, ένα πρόγραμμα επεξεργασίας κειμένου ή επεξεργασίας XML μπορεί να χρησιμοποιηθεί για συγγραφή αρχείων XOL. Παρόλο που ορισμένοι XOL χρήστες χρησιμοποιούν αυτήν την προσέγγιση, αυτή η προδιαγραφή αποθαρρύνει αυτήν την πρακτική, διότι η χειρόγραφη συγγραφή των XOL αρχείων είναι πιθανόν να οδηγήσει σε αρχεία που είναι συντακτικά ή σημασιολογικά ακατάλληλα, ή και τα δύο. Ωστόσο, δεδομένου ότι η γλώσσα XML είναι σε πολύ πρώιμο στάδιο ανάπτυξης, αναμένεται να είναι σύντομα διαθέσιμα εργαλεία τέτοια, ώστε η χειρωνακτική συγγραφή XOL εγγράφων να είναι αναγκαία μόνο για το άμεσο μέλλον.

Ο σχεδιασμός της XOL χρησιμοποιεί εσκεμμένα μια γενική προσέγγιση για τον ορισμό οντολογιών, που σημαίνει ότι το ενιαίο σύνολο ετικετών XML (που ορίζονται από ένα XML DTD) που ορίζεται για την XOL μπορεί να περιγράψει οποιαδήποτε και κάθε οντολογία. Η προσέγγιση αυτή έρχεται σε αντίθεση με τις προσεγγίσεις που λαμβάνονται από άλλες γλώσσες XML σχήματος, στις οποίες συνήθως χρησιμοποιείται ένα γενικό σύνολο ετικετών για τον καθορισμό του σχήματος της οντολογίας, και το ίδιο το σχήμα χρησιμοποιείται για να δημιουργήσει μια δεύτερη σειρά ετικετών για εφαρμογές (και ένα DTD για εφαρμογές) που με την σειρά τους χρησιμοποιούνται για να κωδικοποιήσουν ένα ξεχωριστό αρχείο XML που περιέχει το τμήμα δεδομένων της οντολογίας.

Ας εξετάσουμε τους ακόλουθους XOL ορισμούς:

```
<class>  
<name>person</name>
```

```

</class>
<slot>
<name>age</name>
<domain>person</domain>
<value-type>integer</value-type>
<numeric-max>150</numeric-max>
</slot>
<individual>
<name>fred</name>
<type>person</type>
<slot-values>
<name>age</name>
<value>35</value>
</slot-values>
</individual>

```

Όλα τα στοιχεία XML, όπως "class", "individual", και "name" είναι γενικά, δεν αφορούν την οντολογία Person που ορίζεται σε αυτό το παράδειγμα, αφορούν όλες τις οντολογίες. Όλες οι πληροφορίες που σχετίζονται με τις οντολογίες βρίσκονται στο τμήμα κειμένου του αρχείου XML, δηλαδή, μεταξύ των ζευγών στοιχείων.

Αντίθετα, θα μπορούσαμε να φανταστούμε το ακόλουθο είδος σήμανσης XML για να καθορίσει το άτομο Fred:

```

<person>
<name>fred</name>
<age>35</age>
</person>

```

Το προηγούμενο κείμενο δεν αποτελεί έγκυρη προδιαγραφή XOL επειδή οι ετικέτες που χρησιμοποιεί (όπως το "person" και "age"), δεν ορίζονται από την XOL. Αυτό το κείμενο είναι έγκυρο έγγραφο XML, αλλά δεν είναι έγκυρο XOL. Δεν αποτελούν όλα τα έγκυρα XML έγκυρα XOL.

Η γενική μορφή ενός εγγράφου XOL έχει ως εξής. Το έγγραφο αρχίζει με ένα στοιχείο module, το οποίο προσδιορίζει τη μοναδική οντολογία που περιέχεται σε αυτό το XOL αρχείο. Ακολουθούν μια σειρά από στοιχεία class, τα οποία καθορίζουν τις κλάσεις εντός της εν λόγω οντολογίας. Μια σειρά από στοιχεία slots απαριθμεί τα slots (κάθε slot ορίζει τα χαρακτηριστικά μιας κλάσης ή την σχέση μεταξύ κλάσεων) που ορίζονται σε αυτές τις κλάσεις. Στην συνέχεια, μια σειρά από επιμέρους στοιχεία καθορίζουν τα αντικείμενα στο εσωτερικό της οντολογίας.

```

<module>
<class>
...
</class>
<class>
...
</class>
...
<slot>
...

```

```

</slot>
<slot>
...
</slot>
...
<individual>
...
</individual>
<individual>
...
</individual>
...
</module>

```

Το πρώτο στοιχείο του εγγράφου μπορεί να είναι ένα από τα επόμενα πέντε ονόματα:

- module
- ontology
- kb
- database
- dataset

Αυτά τα ονόματα στοιχείων είναι συνώνυμα. Η XOL παρέχει σκόπιμα πολλά συνώνυμα για χρήση εντός των ορισμών του επειδή οι διαφορετικές κοινότητες είναι συνηθισμένες στη χρήση διαφορετικών όρων για αυτά που είναι ουσιαστικά οι ίδιες βασικές έννοιες της επιστήμης των υπολογιστών.

Ολόκληρη η περιγραφή της οντολογίας δίνεται μεταξύ των ετικετών <module> και </module >. Το στοιχείο <name> καθορίζει το όνομα της οντολογίας και είναι υποχρεωτικό.

4.3.2 Simple HTML Ontology Extension (SHOE)

Η SHOE συνδυάζει τα χαρακτηριστικά γλωσσών σήμανσης, αναπαράστασης γνώσης και οντολογιών σε μια προσπάθεια να αντιμετωπίσει τα ιδιαίτερα προβλήματα σημασιολογίας στον Ιστό. Υποστηρίζει την απόκτηση γνώσης στον Παγκόσμιο Ιστό αυξάνοντάς την με ετικέτες που παρέχουν σημασιολογική έννοια. Η βασική δομή αποτελείται από οντολογίες, οι οποίες ορίζουν τους κανόνες που καθορίζουν το είδος των ισχυρισμών που μπορούν να γίνουν και το είδος των συμπερασμάτων που μπορούν να εξαχθούν, και στιγμιότυπα που εξάγουν τους ισχυρισμούς με βάση αυτούς τους κανόνες.

Η SHOE μπορεί να ενσωματωθεί άμεσα σε HTML έγγραφα ή να χρησιμοποιηθεί σε έγγραφα XML. Υπάρχουν αρκετά πλεονεκτήματα για τη χρήση συντακτικού XML. Πρώτον, αν και οι συντάξεις πιο τυπικών αναπαραστάσεων γνώσης θα μπορούσαν να ενσωματωθούν μεταξύ ενός ζεύγους οριοθετημένων ετικετών, οι συγγραφείς του Ιστού είναι άνετοι με γλώσσες βασισμένες σε ετικέτες σε στυλ XML. Δεύτερον, η σύνταξη XML επιτρέπει στις πληροφορίες SHOE να αναλύονται και να υποβάλλονται σε επεξεργασία με χρήση του Document Object Model (DOM), επιτρέποντας σε λογισμικό που γνωρίζει XML, αλλά δε γνωρίζει SHOE να εξακολουθεί να χρησιμοποιεί τις πληροφορίες με πιο περιορισμένο, αλλά παρ' όλα αυτά ισχυρό τρόπο. Ο τρίτος λόγος για τη χρήση συντακτικού XML είναι ότι τα έγγραφα SHOE μπορούν στην συνέχεια να χρησιμοποιήσουν το πρότυπο XSLT stylesheet για να αναπαραστήσουν τις πληροφορίες για ανθρώ διότι μπορεί να εξαλείψει τον πλεονασμό από την ύπαρξη ξεχωριστών συνόλων ετικετών για ανάγνωση της γνώσης από τον άνθρωπο και για ανάγνωση από μηχανή.

Η SHOE χρησιμοποιεί οντολογίες για να καθορίσει τα έγκυρα στοιχεία που μπορούν να χρησιμοποιηθούν για την περιγραφή οντοτήτων. Κάθε οντολογία μπορεί να επαναχρησιμοποιήσει άλλες οντολογίες επεκτείνοντάς τες. Μια οντολογία αποθηκεύεται σε ένα αρχείο HTML ή XML και διατίθεται στους συγγραφείς των εγγράφων και πράκτορες SHOE τοποθετώντας τη σε έναν διακομιστή ιστού. Αυτό το αρχείο περιλαμβάνει ετικέτες που προσδιορίζουν την οντολογία, δηλώνει ποιες οντολογίες (αν υπάρχουν) έχουν επεκταθεί, και καθορίζει τα διάφορα στοιχεία της οντολογίας.

Στο συντακτικό της SHOE μια οντολογία εμφανίζεται μεταξύ των ετικετών <ONTOLOGY ID=id VERSION=version> και </ONTOLOGY>, και προσδιορίζεται από τον συνδυασμό των id και version. Μια οντολογία μπορεί να καθορίσει κατηγορίες, σχέσεις και άλλα στοιχεία με τη συμπερίληψη ειδικών ετικετών για τους σκοπούς αυτούς.

Η ετικέτα <DEF-CATEGORY> μπορεί να χρησιμοποιηθεί για να ορίσει τις κατηγορίες στις οποίες διάφορα στιγμιότυπα θα μπορούσαν να καταταγούν. Οι κατηγορίες μπορούν να ομαδοποιηθούν ως υποκατηγορίες σε μία ή περισσότερες υπερ-κατηγορίες, προσδιορίζοντας ουσιαστικά μια σχέση is-a που χρησιμοποιείται ευρέως σε σημασιολογικά δίκτυα και συστήματα πλαισίου. Η χρήση των κατηγοριών επιτρέπει την κατασκευή ταξινομιών από πάνω προς τα κάτω ανακατανέμοντας τις γνωστές κλάσεις σε μικρότερα σύνολα.

Η ετικέτα <DEF-RELATION> μπορεί να χρησιμοποιηθεί για να κάνει σχεσιακούς ορισμούς που καθορίζουν τη μορφή η σχεσιακών αξιώσεων που μπορεί να προκύψουν από στιγμιότυπα που αφορούν στιγμιότυπα και άλλα δεδομένα. Μια σχέση μπορεί να ανήκει σε έναν από τους τέσσερις βασικούς τύπους (συμβολοσειρά, αριθμός, ημερομηνία, ή λογική τιμή).

Η SHOE χρησιμοποιεί συμπερασματικούς κανόνες, που υποδεικνύονται από την ετικέτα <DEF-INFERENCE>, για να παρέχει πρόσθετα αξιώματα. Ένας συμπερασματικός κανόνας SHOE αποτελείται από ένα σύνολο από υποθέσεις και ένα σύνολο συνεπειών. Οι ετικέτες <INF-IF> και <INF-THEN> αναφέρονται στις υποθέσεις και τις συνέπειες του συμπεράσματος, αντίστοιχα. Υπάρχουν τρεις τύποι υποπροτάσεων συμπερασμάτων: κατηγορία, σχέση και σύγκριση. Τα επιχειρήματα κάθε υποπρότασης μπορούν να είναι σταθερές ή μεταβλητές, όπου οι μεταβλητές υποδεικνύονται από τη ποικίλη (VAR) χρήση. Αυτός είναι ίσως ο σημαντικότερος λόγος για την σύνταξη XML, διότι μπορεί να εξαλείψει τον πλεονασμό από την ύπαρξη ξεχωριστών συνόλων ετικετών για ανάγνωση της γνώσης από τον άνθρωπο και για ανάγνωση από μηχανή.

Η SHOE χρησιμοποιεί συμπερασματικούς κανόνες, που υποδεικνύονται από την ετικέτα <DEF-INFERENCE>, για να παρέχει πρόσθετα αξιώματα. Ένας συμπερασματικός κανόνας SHOE αποτελείται από ένα σύνολο από υποθέσεις και ένα σύνολο συνεπειών. Οι ετικέτες <INF-IF> και <INF-THEN> αναφέρονται στις υποθέσεις και τις συνέπειες του συμπεράσματος, αντίστοιχα. Υπάρχουν τρεις τύποι υποπροτάσεων συμπερασμάτων: κατηγορία, σχέση και σύγκριση. Τα επιχειρήματα κάθε υποπρότασης μπορούν να είναι σταθερές ή μεταβλητές, όπου οι μεταβλητές υποδεικνύονται από τη λέξη VAR. Οι σταθερές πρέπει να συνδυάζονται ακριβώς και οι μεταβλητές με το ίδιο όνομα πρέπει να συνδέονται με την ίδια τιμή.

Η SHOE κάνει διάκριση μεταξύ του τι μπορεί να ειπωθεί σε μια οντολογία και τι μπορεί να ειπωθεί σε μια αυθαίρετη ιστοσελίδα. Οι συνηθισμένες ιστοσελίδες δηλώνουν ένα ή περισσότερα στιγμιότυπα που εκπροσωπούν οντότητες SHOE, και κάθε στιγμιότυπο περιγράφει το ίδιο ή άλλα στιγμιότυπα χρησιμοποιώντας κατηγορίες και σχέσεις. Η σύνταξη για τα στιγμιότυπα περιλαμβάνει ένα στοιχείο <INSTANCE> που έχει ένα χαρακτηριστικό για ένα KEY12 που προσδιορίζει μοναδικά το στιγμιότυπο. Ένα στιγμιότυπο δεσμεύεται με μια συγκεκριμένη οντολογία μέσω της ετικέτας <USE ONTOLOGY>, η οποία έχει την ίδια λειτουργία με το στοιχείο που έχει το ίδιο όνομα και χρησιμοποιείται στις οντολογίες.

Η SHOE σχεδιάστηκε ειδικά για τις ανάγκες των κατανεμημένων πρακτόρων του διαδικτύου. Ένα βασικό πρόβλημα στα κατανεμημένα συστήματα είναι η διαλειτουργικότητα. Η SHOE επιχειρεί να μεγιστοποιήσει τη διαλειτουργικότητα μέσω της χρήσης κοινών οντολογιών, προκαθορισμένης ονομασίας, της πρόληψης αντιθέσεων, και της τοπικότητας των συμπερασματικών κανόνων. Η γλώσσα έχει σχεδιαστεί προσεκτικά για να εξαλείψει τη πιθανότητα αντιφάσεων μεταξύ των ισχυρισμών των πρακτόρων. Το κάνει αυτό με τέσσερις(4) τρόπους:

1. Επιτρέπει μόνο τους ισχυρισμούς, όχι ανακλήσεις.
2. Δεν επιτρέπει τη λογική άρνηση.
3. Δεν έχει σχεσιακά σύνολα που μπορούν να έχουν μόνο μία τιμή.
4. Δεν επιτρέπει την προδιαγραφή ασύνδετων κλάσεων.

Παρά το γεγονός ότι υπάρχουν πολλοί πιθανοί τρόποι χρήσης της γλώσσας, ο απλούστερος είναι εκείνος που παραλληλίζει τον τρόπο που λειτουργεί ο Ιστός σήμερα. Υπάρχουν πολλά εργαλεία που μπορούν να χρησιμοποιηθούν για την παραγωγή περιεχομένου, και αυτό το περιεχόμενο δημοσιεύεται στο Διαδίκτυο μέσω ενός διακομιστή. Ένα web-crawler μπορεί να συγκεντρώσει στην συνέχεια τις σχετικές σελίδες και να τις αποθηκεύσει σε ένα αποθετήριο, στο οποίο μπορεί στην συνέχεια να τεθούν ερωτήματα μέσω κάποιου τύπου διεπαφής χρήστη. Το βασικό στοιχείο ενός συστήματος SHOE είναι ότι τόσο το περιεχόμενο, όσο και οι οντολογίες που παρέχουν σημασιολογία για το περιεχόμενο δημοσιεύονται στον Παγκόσμιο Ιστό. Δεδομένου ότι αυτή η πληροφορία έχει δομηθεί και έχει σημασιολογία, το αποθετήριο θα πρέπει να είναι μια βάση γνώσεων και όχι ένα σύστημα ανάκτησης πληροφοριών.

Το πρώτο βήμα στη χρήση της SHOE είναι ο εντοπισμός ή ο σχεδιασμός μιας κατάλληλης οντολογίας. Για τη διευκόλυνση αυτής της διαδικασίας, θα πρέπει να υπάρχει ένα κεντρικό αποθετήριο οντολογιών. Ένα απλό αποθετήριο θα μπορούσε να είναι ένα σύνολο ιστοσελίδων που κατηγοριοποιούν οντολογίες, ενώ ένα πιο σύνθετο θα μπορούσε να συνδέει μια σειρά από κοινά χαρακτηριστικά με κάθε οντολογία έτσι, ώστε να υλοποιούνται συγκεκριμένες αναζητήσεις. Αν διαπιστωθεί ότι δεν υπάρχουν κατάλληλες οντολογίες διαθέσιμες, κάθε νεοσύστατη οντολογία θα πρέπει να επεκτείνει πάντα διαθέσιμες οντολογίες που περιέχουν σχετικές έννοιες.

Η διαδικασία προσθήκης σημασιολογικών ετικετών σε μια ιστοσελίδα ονομάζεται σχολιασμός (annotation). Υπάρχουν μια σειρά από εργαλεία που μπορούν να χρησιμοποιηθούν σε αυτήν τη διαδικασία, από απλούς συντάκτες κειμένου, σε συντάκτες GUI, έως ημι-ή πλήρως αυτοματοποιημένες τεχνικές. Οι συντάκτες κειμένου έχουν το πλεονέκτημα ότι είναι κοινοί, αλλά απαιτούν εξοικείωση των χρηστών με την σύνταξη της SHOE. Έχει αναπτυχθεί ένας συντάκτης GUI¹³ για τη SHOE που ονομάζεται Knowledge Annotator και επιτρέπει στο χρήστη να δημιουργήσει σήμανση επιλέγοντας από καταλόγους και συμπληρώνοντας φόρμες. Ωστόσο, υπάρχουν περιπτώσεις όπου ο χρήστης μπορεί να θέλει να δημιουργήσει σήμανση που αντιστοιχεί σε πληροφορίες από μεγάλους καταλόγους ή πίνακες σε προϋπάρχουσες ιστοσελίδες. Για τέτοιες περιπτώσεις, μπορεί να χρησιμοποιηθεί το εργαλείο Running SHOE για να δημιουργήσει γρήγορα ένα περιτύλιγμα για ένα υπάρχον έγγραφο, προσδιορίζοντας τα σημεία που υποδεικνύουν εγγραφές και πεδία, και στη συνέχεια χαρτογραφώντας αυτές τις εγγραφές και τα πεδία σε τάξεις και σχέσεις από κάποια οντολογία. Αυτό το εργαλείο μπορεί να χρησιμοποιηθεί για να δημιουργήσει ένα μεγάλο αριθμό σημασιολογικών σημάνσεων από απλά έγγραφα μέσα σε λίγα λεπτά. Άλλες προσεγγίσεις δημιουργίας σήμανσης μπορεί να περιλαμβάνουν μηχανική μάθηση, εξόρυξη πληροφοριών, και, σε περιορισμένες περιπτώσεις, τεχνικές επεξεργασίας φυσικών γλωσσών.

Μόλις το απαραίτητο περιεχόμενο και τα έγγραφα οντολογιών δημοσιεύονται στο Διαδίκτυο, μπορούν να συλλεχθούν από έναν ανιχνευτή Ιστού. Το Exposé, ο αναζητητής SHOE, ψάχνει για ιστοσελίδες με σήμανση SHOE και αποθηκεύει τις πληροφορίες σε μια βάση γνώσης. Όταν μια σελίδα δεσμεύεται με μια οντολογία άγνωστη στο σύστημα, αυτή η οντολογία επίσης ανακτάται και αποθηκεύεται στη βάση γνώσης. Το μεγαλύτερο μειονέκτημα στην προσέγγιση με αναζητητή είναι ότι οι πληροφορίες είναι τόσο πρόσφατες, όσο και η τελευταία επίσκεψη του αναζητητή στην ιστοσελίδα. Σε ορισμένες εφαρμογές αυτό το χρονικό διάστημα μπορεί να μην είναι αποδεκτό. Μια ενδιαφέρουσα κατεύθυνση για μελλοντική έρευνα είναι η ανάπτυξη εστιασμένων αναζητητών που ψάχνουν απαντήσεις σε συγκεκριμένες ερωτήσεις σε πραγματικό χρόνο.

Όπως προαναφέρθηκε, ο αναζητητής χρειάζεται μια βάση γνώσεων για να αποθηκεύει τα αποτελέσματα των προσπαθειών του. Μια βάση γνώσεων παρέχει μόνιμη αποθήκευση πληροφοριών και την ικανότητα χρήσης των γνώσεων για εξαγωγή συμπερασμάτων.

4.3.3 Ontology Markup Language (OML)

Η OML, που αναπτύχθηκε στο Πανεπιστήμιο της Ουάσιγκτον, βασίζεται εν μέρει στη SHOE. Στην πραγματικότητα, αρχικά θεωρήθηκε ως η XML σειριοποίηση της SHOE. Ως εκ τούτου, η OML και η SHOE έχουν πολλά κοινά στοιχεία.

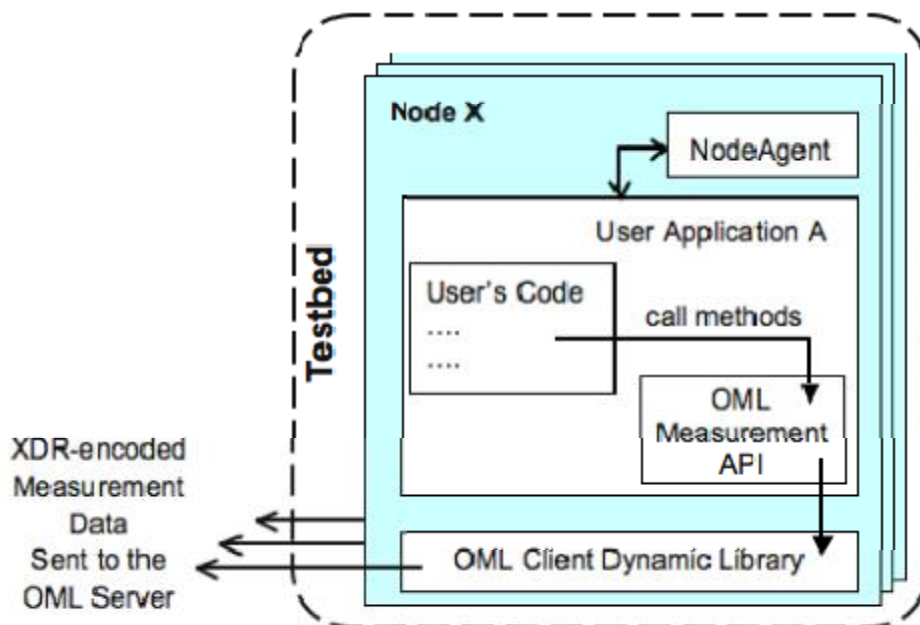
Η OML είναι υποσύνολο της Conceptual Knowledge Markup Language (CKML), είναι μία αυτόνομη γλώσσα σήμανσης και ακολουθεί τις αρχές και τις ιδέες των εννοιολογικών γραφημάτων (Conceptual Graphs). Χρησιμοποιείται για τη διάρθρωση των προδιαγραφών και των αξιωμάτων των μεταδεδομένων στις οντολογίες. Η OML ενσωματώνει τις ιδέες των εννοιολογικών γραφημάτων για την έννοια, την εννοιολογική σχέση, το εννοιολογικό πλαίσιο, τους συμμετέχοντες και την οντολογία. Ο σύνδεσμος της OML με την CKML είναι η διαδικασία της εννοιολογικής κλιμάκωσης, η οποία είναι η ερμηνευτική μεταμόρφωση της οντολογικά δομημένης γνώσης σε εννοιολογικά δομημένη γνώση.

Είναι πολύ σημαντικό θέμα το πώς και πόσο καλά μια γλώσσα αναπαράστασης γνώσης εκφράζει περιορισμούς. Η OML έχει τρία επίπεδα για την έκφραση περιορισμών. Το πάνω επίπεδο αντιπροσωπεύει τους θεωρητικούς περιορισμούς στη ροή της πληροφορίας, το μεσαίο επίπεδο προκύπτει τόσο από την πρακτική σημασία των περιορισμών των δυαδικών σχέσεων, όσο και από τον προσανατολισμό της θεωρίας των κατηγοριών στην ταξινόμηση-προβολή της σημασιολογίας του διαγράμματος και το κάτω επίπεδο αντιστοιχεί στα εννοιολογικά γραφήματα. Πράγματι, οι OML ισχυρισμοί (κλειστές εκφράσεις) αντιστοιχούν ακριβώς στα εννοιολογικά γραφήματα.

Υπάρχουν τέσσερα διαφορετικά επίπεδα OML: Η OML Core σχετίζεται με τις λογικές πτυχές της γλώσσας και περιλαμβάνεται από τα υπόλοιπα στρώματα. Η Simple OML οδηγεί απευθείας στην RDF(S). Η Abbreviated OML περιλαμβάνει τα χαρακτηριστικά των εννοιολογικών γραφημάτων και η Standard OML είναι η πιο εκφραστική εκδοχή της OML.

Ο σημασιολογικός πυρήνας της CKML συλλαμβάνεται σε ένα υποσύνολο της OML που ονομάζεται Simple OML. Επειδή η Simple OML βασίζεται στη θεμελιώδη σημασιολογία ταξινόμησης-προβολής, η κανονική έκφραση των τύπων και των στιγμιότυπων μοιάζει με τα πλαίσια. Μια βασική πτυχή της Open Knowledge Base Connectivity (OKBC) είναι στενά συνδεδεμένη με την Simple OML.

Δεν υπάρχουν συγκεκριμένα εργαλεία για την συγγραφή οντολογιών OML εκτός από τα υπάρχοντα γενικά σκοπού εργαλεία της XML.



Εικόνα 15: OML Framework

<http://www.orbit-lab.org/wiki/Software/dOML>

4.3.4 Ontology Interchange Language (OIL)

Η αποτελεσματική και αποδοτική εργασία με τις οντολογίες πρέπει να υποστηρίζεται από προηγμένα εργαλεία που θα επιτρέψουν την πλήρη ισχύ αυτής της τεχνολογίας. Συγκεκριμένα, χρειαζόμαστε μια προηγμένη γλώσσα οντολογίας έκφρασης και αναπαράστασης οντολογιών. Μια τέτοια γλώσσα οντολογίας πρέπει να πληρεί τρεις σημαντικές προϋποθέσεις:

- Πρέπει να είναι πολύ διαισθητική για τον ανθρώπινο χρήστη. Λαμβάνοντας υπόψη την τρέχουσα επιτυχία του βασισμένου σε πλαίσιο και αντικειμενοστρεφούς προτύπου μοντελοποίησης, θα πρέπει να έχουν την εμφάνιση και την αίσθηση πλαισίου.
- Θα πρέπει να έχει μια καλά καθορισμένη επίσημη σημασιολογία με καθιερωμένες ιδιότητες συλλογιστικής όσον αφορά την πληρότητα, την ορθότητα και την αποτελεσματικότητα.
- Θα πρέπει να έχει μια σωστή σύνδεση με τις ήδη υπάρχουσες γλώσσες Ιστού όπως η XML και το RDF διασφαλίζοντας, έτσι, τη διαλειτουργικότητα.

Η Ontology Inference Layer (OIL) ανταποκρίνεται στα κριτήρια που αναφέρονται ανωτέρω. Η OIL συνδυάζει τρεις σημαντικές πτυχές που παρέχονται από διαφορετικές κοινότητες: Επιστημολογικά πλούσιες αρχές μοντελοποίησης, όπως προβλέπεται από την Κοινότητα Πλαισίου, επίσημη σημασιολογία και αποτελεσματική υποστήριξη συλλογιστικής, όπως προβλέπεται από τις Λογικές Περιγραφές και μια τυπική πρόταση για την συντακτική ανταλλαγή συμβολισμών, όπως προβλέπεται από την Κοινότητα του Ιστού.

Είναι απίθανο μια μόνο γλώσσα οντολογίας να μπορεί να εκπληρώσει όλες τις ανάγκες του μεγάλου εύρους των χρηστών και των εφαρμογών του Σημασιολογικού Ιστού. Ως εκ τούτου, η OIL έχει οργανωθεί σε ολόένα αυξανόμενα στρώματα υπό-γλωσσών. Κάθε επιπλέον στρώμα προσθέτει λειτουργικότητα και πολυπλοκότητα στο προηγούμενο στρώμα. Αυτό γίνεται έτσι, ώστε οι πράκτορες (άνθρωποι ή μηχανές) που μπορούν να επεξεργαστούν μόνο ένα χαμηλότερο στρώμα να μπορούν να καταλάβουν εν μέρει τις οντολογίες που

εκφράζονται σε οποιοδήποτε από τα ανώτερα στρώματα. Μια πρώτη και πολύ σημαντική εφαρμογή της αρχής αυτής είναι η σχέση μεταξύ της OIL και του RDF Schema.

- Η Core OIL συμπίπτει σε μεγάλο βαθμό με το RDF Schema (με εξαίρεση τα χαρακτηριστικά «υλικοποίησης» του RDF Schema). Αυτό σημαίνει ότι ακόμη και οι απλοί RDF Schema πράκτορες έχουν τη δυνατότητα να επεξεργάζονται τις οντολογίες OIL, και να συλλαμβάνουν όσο το δυνατόν περισσότερο από το νόημά τους με τις περιορισμένες δυνατότητές τους.

- Η Standard Oil είναι μια γλώσσα που προορίζεται να συλλαμβάνει τις απαραίτητες τρέχουσες αρχές μοντελοποίησης, οι οποίες και παρέχουν επαρκή εκφραστική ισχύ και είναι απόλυτα κατανοητές, επιτρέποντας, έτσι, η σημασιολογία να είναι ακριβώς καθορισμένη και η πλήρης συναγωγή να είναι βιώσιμη.

- Η Instance OIL περιλαμβάνει διεξοδική, ατομική ολοκλήρωση. Ενώ το προηγούμενο στρώμα - η Standard Oil - περιλαμβάνει κατασκευές μοντελοποίησης, Η Instance OIL περιλαμβάνει πλήρεις δυνατότητες βάσεων δεδομένων.

- Η Heavy OIL μπορεί να περιλαμβάνει πρόσθετες δυνατότητες αναπαράστασης (και συλλογιστικής). Ειδικά μια πιο εκφραστική γλώσσα κανόνων και οι παροχές μετά-κλάσεων φαίνονται ιδιαίτερα επιθυμητές. Αυτές οι επεκτάσεις της OIL ορίζονται σε συνεργασία με την πρωτοβουλία DAML για μια γλώσσα κανόνων για τον Ιστό.

Η αρχιτεκτονική σε στρώματα της OIL έχει τρία βασικά πλεονεκτήματα⁸:

- Μια εφαρμογή δεν είναι ανάγκη να εργάζεται με μια γλώσσα που προσφέρει σημαντικά μεγαλύτερη εκφραστικότητα και πολυπλοκότητα από αυτή που πραγματικά χρειάζεται.

- Εφαρμογές που επιτρέπεται να επεξεργαστούν ένα χαμηλότερο επίπεδο πολυπλοκότητας είναι ακόμη σε θέση να συλλάβουν τα ίδια στοιχεία της οντολογίας.

- Μια εφαρμογή που έχει κατανόηση ενός υψηλότερου επιπέδου πολυπλοκότητας μπορεί επίσης να κατανοήσει οντολογίες που εκφράζονται σε πιο απλή γλώσσα οντολογίας.

Ο ορισμός μιας γλώσσας οντολογίας ως επέκταση του RDF Schema σημαίνει ότι κάθε οντολογία RDF Schema είναι μια έγκυρη οντολογία στη νέα γλώσσα (δηλαδή, ένας επεξεργαστής OIL θα κατανοεί, επίσης, RDF Schema). Ωστόσο, η άλλη κατεύθυνση είναι επίσης διαθέσιμη: ο ορισμός μιας επέκτασης OIL όσο το δυνατόν πλησιέστερης στο RDF Schema επιτρέπει τη μέγιστη επαναχρησιμοποίηση των υφιστάμενων εφαρμογών που είναι βασισμένες σε RDF Schema και των εργαλείων. Ωστόσο, δεδομένου ότι η γλώσσα οντολογίας περιέχει συνήθως νέες πτυχές (και ως εκ τούτου νέο λεξιλόγιο, που ένας επεξεργαστής RDF Schema δεν κατανοεί), η 100% συμβατότητα δεν είναι δυνατή. Ακολουθεί ένα παράδειγμα. Η ακόλουθη έκφραση OIL ορίζει ως κλάση την herbivore (φυτοφάγα), η οποία είναι μια υπο-κλάση της animal και διαζευκτικό του carnivores (σαρκοφάγα).

```
<rdfs: Class rdf: ID="herbivore">
<rdf: type
rdf: resource="http://www.ontoknowledge.org/oil/RDFS-
schema/#DefinedClass"/>
<rdfs: subclassOf rdf: resource="#animal"/>
<rdfs: subclassOf>
<oil: NOT>
```

⁸ Olivier Mehani, Guillaume Jourjon, Thierry Rakotoarivelo, and Max Ott, "An instrumentation framework for the critical task of measurement collection in the future Internet," *Computer Networks*, vol. 63, pp. 68-83, Apr. 2014. doi: 10.1016/j.bjp.2014.01.007 [Online]. Available: <http://www.nicta.com.au/pub?id=6718>


```

        <oil: hasOperand rdf: resource="#carnivore"/>
    </oil: NOT>
</rdfs: subClassOf>
</rdfs: Class>

```

Μια εφαρμογή περιορισμένη σε απλό RDFS είναι ακόμη σε θέση να συλλάβει μερικές πτυχές του ορισμού αυτού.

```

<rdfs: Class rdf: ID="herbivore">
    <rdfs: subClassOf rdf: resource="#animal"/>
    <rdfs: subClassOf>
        ...
    </rdfs: subClassOf>
</rdfs: Class>

```

Κατανοεί ότι το herbivore είναι μια υπό-κλάση της animal και υπό-κλάση μιας δεύτερης κλάσης που δε μπορεί να κατανοήσει σωστά. Αυτός φαίνεται να είναι ένας χρήσιμος τρόπος να διατηρηθεί η πολυπλοκότητα της σημασιολογίας για απλούστερες εφαρμογές.

Η OIL έχει πολύ ισχυρά υποστηρικτικά εργαλεία στους ακόλουθους τομείς: Συντάκτες Οντολογιών για τη δημιουργία νέων οντολογιών, εργαλεία σχολιασμού οντολογιών για σύνδεση αδόμητων και ημι-δομημένων πηγών πληροφοριών με οντολογίες και συλλογιστική με οντολογίες με προηγμένες δυνατότητες υπηρεσιών απάντησης ερωτημάτων, υποστήριξη δημιουργίας οντολογιών και βοήθεια στις αντιστοιχίσεις μεταξύ διαφορετικών οντολογιών.

Οι συντάκτες οντολογιών βοηθούν τους μηχανικούς γνώσης να χτίζουν οντολογίες. Υποστηρίζουν τον ορισμό της ιεραρχίας των εννοιών, τον ορισμό των χαρακτηριστικών των εννοιών, και τον ορισμό των αξιωμάτων και των περιορισμών. Θα πρέπει να παρέχουν γραφικές διεπαφές και να συμμορφώνονται με τα ισχύοντα πρότυπα της δικτυοκεντρικής ανάπτυξης λογισμικού. Επιτρέπουν τον έλεγχο, την περιήγηση, την κωδικοποίηση και την τροποποίηση οντολογιών και υποστηρίζουν με αυτόν τον τρόπο την ανάπτυξη οντολογιών και το έργο της συντήρησης.

4.3.5 DARPA Agent Markup Language+OIL (DAML+OIL)

Η DAML + OIL έχει αναπτυχθεί από μια μεικτή επιτροπή από τις ΗΠΑ και την Ευρωπαϊκή Ένωση (IST) στα πλαίσια της DAML, ένα έργο DARPA που επιτρέπει την σημασιολογική διαλειτουργικότητα στην XML. Ως εκ τούτου, η DAML + OIL συμμαρίζεται τον ίδιο στόχο με την OIL.

Η DAML + OIL είναι μια γλώσσα οντολογίας, και ως τέτοια έχει σχεδιαστεί για να περιγράφει τη δομή ενός τομέα. Η DAML + OIL έχει μια αντικειμενοστραφής προσέγγιση, με τη δομή του τομέα να περιγράφεται με τη μορφή κλάσεων και ιδιοτήτων. Μια οντολογία αποτελείται από ένα σύνολο αξιωμάτων. Η υποστήριξη ότι οι πόροι (ζεύγη των πόρων) είναι στιγμιότυπα των DAML + OIL κλάσεων (ιδιότητες) αφήνεται στο RDF, μια εργασία για την οποία είναι κατάλληλο. Όταν ένας πόρος *r* είναι ένα στιγμιότυπο της κλάσης *C* λέμε ότι η *r* είναι τύπου *C*.

Από μια επίσημη άποψη, η DAML + OIL μπορεί να θεωρηθεί ότι ισοδυναμεί με μια πολύ εκφραστική Περιγραφική Λογική, με τη DAML + OIL οντολογία να αντιστοιχεί σε ορολογία Περιγραφικής Λογικής. Όπως στην Περιγραφική Λογική, οι DAML + OIL κλάσεις μπορούν να είναι ονόματα (URI) ή εκφράσεις, και μια ποικιλία κατασκευαστών παρέχεται για την

κατασκευή εκφράσεων κλάσεων. Η εκφραστική ισχύς της γλώσσας καθορίζεται από τους κατασκευαστές κλάσεων, καθώς και από τα είδη αξιωμάτων που υποστηρίζονται.

Η DAML + OIL υποστηρίζει το πλήρες εύρος των τύπων δεδομένων του XML Schema. Αυτό διευκολύνεται από τη διατήρηση ενός καθαρού διαχωρισμού μεταξύ των στιγμιότυπων των κλάσεων "αντικειμένων" (που ορίζονται με τη γλώσσα οντολογίας) και των στιγμιότυπων των τύπων δεδομένων (που ορίζονται με το σύστημα τύπων του XML Schema). Συγκεκριμένα, θεωρείται ότι ο τομέας της ερμηνείας των κλάσεων αντικειμένων είναι ασύνδετος με τον τομέα της ερμηνείας των τύπων δεδομένων, έτσι, ένα στιγμιότυπο μιας κλάσης αντικειμένων δε μπορεί ποτέ να έχει την ίδια ερμηνεία με την τιμή ενός τύπου δεδομένων, και ότι το σύνολο των ιδιοτήτων των αντικειμένων είναι ασύνδετο με το σύνολο των ιδιοτήτων του τύπου δεδομένων.

Η ασυνδετότητα των τομέων των αντικειμένων και των τύπων δεδομένων είχε ως κίνητρο τόσο φιλοσοφικούς, όσο και ρεαλιστικούς λόγους:

- Οι τύποι δεδομένων θεωρούνται ήδη επαρκώς δομημένοι από τα ενσωματωμένα κατηγορήματα, και, επομένως, δεν είναι πρόβλημα να δημιουργούνται νέες κλάσεις τιμών τύπων δεδομένων με χρήση της γλώσσας οντολογίας.

- Η απλότητα και η πυκνότητα της γλώσσας οντολογίας δεν διακυβεύονται: ακόμη και αν η απαρίθμηση όλων των τύπων δεδομένων του XML Schema που θα αύξανε σημαντικά την πολυπλοκότητά της, προσθέτοντας μια λογική θεωρία για κάθε τύπο δεδομένων, ήταν δυνατή, θα οδηγούσε σε μια γλώσσα μνημειακών διαστάσεων.

- Η σημασιολογική ακεραιότητα της γλώσσας δεν τίθεται σε κίνδυνο: ο ορισμός θεωριών για όλους τους τύπους δεδομένων του XML Schema θα ήταν δύσκολος ή αδύνατος χωρίς την επέκταση της γλώσσας σε κατευθύνσεις των οποίων σημασιολογικά θα ήταν δύσκολο να συλληφθεί εντός του υπάρχοντος πλαισίου.

Από θεωρητικής απόψεως, αυτό το σχέδιο σημαίνει ότι η γλώσσα οντολογίας μπορεί να ορίσει περιορισμούς στις τιμές των δεδομένων, αλλά δεδομένου ότι οι τιμές των δεδομένων δεν μπορούν ποτέ να είναι στιγμιότυπα κλάσεων αντικειμένων, δε μπορούν να εφαρμόσουν πρόσθετους περιορισμούς στα στοιχεία του τομέα του αντικειμένου. Αυτό επιτρέπει στο σύστημα τύπων να επεκτείνεται χωρίς τυχόν επιπτώσεις στη γλώσσα οντολογίας, και αντίστροφα.

Από πρακτική άποψη, οι DAML + OIL εφαρμογές μπορούν να επιλέξουν την υποστήριξη ορισμένων ή όλων των τύπων δεδομένων του XML Schema. Για τους υποστηριζόμενους τύπους δεδομένων, μπορούν να εφαρμόσουν είτε το δικό τους ελεγκτή τύπων / επικυρωτή ή να βασιστούν σε μια εξωτερική συνιστώσα. Η δουλειά του ελεγκτή / επικυρωτή τύπων είναι να παίρνει μηδέν ή περισσότερες τιμές δεδομένων και έναν ή περισσότερους τύπους δεδομένων, και να καθορίζει εάν υπάρχει οποιαδήποτε τιμή δεδομένων που είναι ίση με κάθε μία από τις καθορισμένες τιμές δεδομένων και αν είναι στιγμιότυπο σε κάθε έναν από τους καθορισμένους τύπους δεδομένων.

Η DAML + OIL είναι χτισμένη σε RDF (S). Το όνομά της εμμέσως υποδηλώνει ότι υπάρχει στενή σχέση με την OIL. Αντικαθιστά την αρχική προδιαγραφή, η οποία ονομαζόταν DAML-ONT, και βασιζόταν επίσης στη γλώσσα OIL. Τα OIEd, OntoEdit, Protégé2000, και WebODE είναι εργαλεία συγγραφής DAML + OIL οντολογιών.

4.4 Προβλήματα στην Πορεία

Οι πολλαπλές επιδράσεις στην OWL έχουν οδηγήσει σε μια σειρά προβλημάτων. Μερικά από αυτά τα προβλήματα είναι αποτέλεσμα αντικρουόμενων απαιτήσεων, όπως με την σύγκρουση ανάμεσα στη χρήση RDF / XML ως την επίσημη σύνταξη της OWL και σε μια εύκολα αναγνώσιμη σύνταξη. Μερικά από αυτά τα προβλήματα προκύπτουν από την ανάγκη επέκτασης προηγούμενων λύσεων, όπως με τα προβλήματα που προκύπτουν στη διαμόρφωση μια επέκτασης του RDF που ενσωματώνει πληροφορίες που δεν ταιριάζουν καλά στην άποψη του RDF.

4.4.1. Συντακτικά Προβλήματα

Για διάφορους λόγους, συμπεριλαμβανομένης της διατήρησης συνδέσεων μεταξύ των πλαισίων και των Περιγραφικών Λογικών, η OWL θα πρέπει να έχει μια εύκολα αναγνώσιμη σύνταξη που να είναι εύκολα κατανοητή και εύκολη στη δημιουργία. Ωστόσο, ήταν μια απαίτηση της OWL να χρησιμοποιεί XML ως κανονιστική σύνταξη και, επιπλέον, να χρησιμοποιεί την XML με τον ίδιο τρόπο που χρησιμοποιείται στο RDF. Η απαίτηση αυτή είχε ήδη αντιμετωπιστεί από την OIL και, αργότερα, από τη DAML + OIL. Η OIL έχει τόσο RDF / XML, όσο και XML συντακτικό, ενώ η DAML + OIL έχει μόνο RDF / XML σύνταξη.

Δεδομένης απλά ως μιας σύνταξης για την OWL, το RDF στη μορφή RDF / XML έχει μια σειρά από προβλήματα. Αυτά τα προβλήματα μπορούν να ξεπεραστούν, αλλά κάνουν πιο πολύπλοκη την OWL από όσο θα ήταν διαφορετικά.

Ένα πρόβλημα είναι ότι το RDF / XML είναι εξαιρετικά φλύαρο. Συγκρίνετε, για παράδειγμα, τις πληροφορίες σχετικά με μια κλάση, όπως θα ορίζονταν σε μια σύνταξη Περιγραφικής Λογικής

Student = Person $\Pi \geq 1$ enrolledIn

(ένας μαθητής είναι ένα άτομο που συμμετέχει σε ένα πράγμα), με το πώς θα ορίζονταν με σύνταξη RDF/XML της OWL.

```
<owl: Class rdf: ID="Student">
<owl: intersectionOf rdf: parsetype="Collection">
<owl: Class rdfs: about="Person" />
<owl: Restriction>
<owl: onProperty rdf: resource="enrolledIn" />
<owl: minCardinality rdfs: datatype="xsd:Integer">
</owl: minCardinality>
</owl: Restriction>
<owl: intersectionOf>
</owl: Class>
```

Ένα άλλο πρόβλημα είναι ότι το RDF σπάει τα πάντα σε RDF τριάδες. Αυτό σημαίνει ότι πολλές δομές της OWL, όπως οι περιορισμοί ιδιοτήτων, πρέπει να κωδικοποιηθούν σε αρκετές τριάδες. Η OWL χρησιμοποιεί συνήθως μια κωδικοποίηση παρόμοια με αυτή που χρησιμοποιείται από τη DAML + OIL. Για παράδειγμα, ένας περιορισμός τιμής OWL που θα γραφόταν στο συντακτικό Περιγραφικής Λογικής ως `child.person` (η κλάση της οποίας τα στιγμιότυπα έχουν κάποιο παιδί που είναι ένα άτομο) κωδικοποιείται ως δύο τριάδες RDF κάπως έτσι

: x owl: onProperty ex: child .
: x owl: someValuesFrom ex: Person .
, όπου το x είναι το μέρος όπου γράφεται ο περιορισμός.

Ένα τρίτο πρόβλημα είναι ότι όλες οι RDF τριάδες είναι ανεξάρτητες. Αυτό σημαίνει, για παράδειγμα, ότι όσον αφορά το RDF, δεν υπάρχει απαίτηση οι δύο παραπάνω τριάδες να εμφανίζονται πάντα μαζί. Ομοίως, δεν υπάρχει απαίτηση να μην υπάρχουν επιπλέον τριάδες, οπότε αν προσθέσουμε

: x owl: onProperty ex: friend .
: x owl: allValuesFrom ex: Doctor .
στις παραπάνω δυο τριάδες δε μπορούν να αποκλειστούν στο RDF.

Ένα τέταρτο πρόβλημα είναι ότι οι RDF τριάδες είναι όλες εύκολα προσβάσιμες. Αυτό σημαίνει ότι η κυκλική και άλλες ασυνήθιστες δομές δε μπορούν να αποκλειστούν. Για παράδειγμα, δεν υπάρχει κανένα πρόβλημα στο RDF με συλλογές τριάδων όπως

: x owl: onProperty ex: child .
: x owl: allValuesFrom: x .

Αυτά τα ζητήματα δεν αντιμετωπίζονται στην OIL, η οποία δεν παρέχει καμία ένδειξη ως προς το τι πρέπει να συμβεί για τα σύνολα τριάδων που δεν ταιριάζουν με τις παραγωγές συντακτικού της γλώσσας. Η DAML + OIL έχει διαφορετική προσέγγιση, επιτρέποντας τις ασυνήθιστες κατασκευές, αλλά αρνούμενη να τους δώσει μια DAML + OIL έννοια. Η OWL έχει ακολουθήσει περίπου τη λύση της DAML + OIL, αλλά με αρκετές τροποποιήσεις.

4.4.2 Σημασιολογικά Προβλήματα

Ακόμα κι αν τα θέματα της σύνταξης αντιμετωπιστούν, παραμένουν ακόμα τα ζητήματα που σχετίζονται με την έννοια. Το RDF παρέχει ένα νόημα για κάθε τριάδα, οπότε αν η OWL θεωρηθεί ότι αποτελεί μια επέκταση του RDF, η έννοια που παρέχει η OWL για τις τριάδες πρέπει να είναι μια επέκταση της έννοιας του RDF.

Αυτό δεν ήταν μεγάλο ζήτημα όταν σχεδιάστηκαν η OIL και η DAML + OIL, καθώς η έννοια του RDF δεν ήταν πολύ καλά προσδιορισμένη. Η OIL, συγκεκριμένα, δεν συσχετίζει την έννοια του RDF της RDF / XML σύνταξης της με την έννοια της OIL σε αυτή τη σύνταξη. Η RDF / XML σύνταξη για ορισμένες δομές OIL συμφωνεί πάνω κάτω με την έννοια του RDF σε αυτές τις δομές, αλλά αυτό δεν αποτελεί κανόνα για όλες αυτές τις δομές.

Η DAML + OIL συμμορφώνεται καλύτερα με την έννοια του RDF στην σύνταξή του. Το μοντέλο θεωρίας της DAML + OIL περιλαμβάνει μια σημασιολογική συνθήκη για τις τριάδες που είναι κοντά στην έννοια των RDF (όπως ορίζονται εκείνη τη στιγμή) τριάδων. Περαιτέρω, η DAML + OIL χρησιμοποιεί το ενσωματωμένο RDF και RDFS λεξιλόγιο σε μεγαλύτερο βαθμό από ό, τι κάνει η OIL, και το χρησιμοποιεί με τρόπο γενικά συμβατό με την έννοια του RDF ή του RDFS (όπως ορίζεται εκείνη τη στιγμή) για αυτό το λεξιλόγιο. Για παράδειγμα, η DAML + OIL χρησιμοποιεί το rdfs: subclassOf για να συσχετίσει κλάσεις με υπερκλάσεις, συμπεριλαμβανομένων των περιορισμών DAML + OIL περιορισμούς, ενώ η OIL χρησιμοποιεί το oil: hasSlotConstraint σε ορισμένες από αυτές τις καταστάσεις.

Ακόμα και όταν αναπτυσσόταν η DAML + OIL, ωστόσο, υπήρχαν ορισμένες πτυχές της σημασίας του RDFS που δεν μπορούσαν να συμβιβαστούν με την κατάλληλη έννοια στη DAML + OIL. Ειδικότερα, το RDFS τότε είχε μια ασυνήθιστη έννοια για τομείς και εύρη

ιδιοτήτων. Μόνο ένα εύρος επιτρεπόταν για τις ιδιότητες και πολλοί τομείς αντιμετωπίζονταν διαζευκτικά. Για παράδειγμα, τα

```
ex: foo rdfs: domain ex: Person .
ex: foo rdfs: domain ex: Rock .
```

θα επέτρεπαν και το people και το rocks να συμμετέχουν στην ιδιότητα foo.

Αυτή η διαζευκτική ανάγνωση των τομέων προκάλεσε προβλήματα στην σημασιολογία της DAML + OIL, όποτε πάρθηκε μια επιλογή για να αλλάξει αυτό σε συνδετική ανάγνωση και έγινε αίτηση στη νεοσύστατη Ομάδα Εργασίας του RDF Core να αλλάξει το RDFS έτσι, ώστε να επιτρέπει πολλούς τομείς και εύρη, και τα δύο με συνδετική ανάγνωση. Ως μέρος του σχεδίου του ξεκαθαρίσματος της σημασιολογίας του RDF και του RDFS, η Ομάδα Εργασίας του RDF Core αποφάσισε να κάνει αυτή την αλλαγή, εξαλείφοντας ένα πρόβλημα για την OWL.

Παράλληλα με το ξεκαθάρισμα των RDF και RDFS, η ομάδα εργασίας του Core RDF αποφάσισε επίσης να θέσει το RDF σε ένα πιο σταθερό σημασιολογικό έδαφος. Το έκανε αυτό παρέχοντας ένα μοντέλο θεωρίας για το RDF και το RDFS, μαζί με μια τυπική αντιμετώπιση της συμπεραματολογίας για τα RDF και RDFS. Αυτό σημαίνει ότι πλέον παρέχεται μεγαλύτερη έννοια στα RDF και RDFS με την οποία πρέπει να είναι συμβατή η OWL. Ειδικότερα, όλες οι τριάδες που χρησιμοποιούνται για την κωδικοποίηση της σύνταξης της OWL έχουν πλέον νόημα RDF, και αυτή η έννοια RDF πρέπει να ληφθεί υπόψη από την OWL, για να είναι η σημασιολογία της OWL πλήρως συμβατή με εκείνη των RDF και RDFS.

Ούτε η OIL ούτε η DAML + OIL προέβλεπαν μια τυποποιημένη θεωρία συμπεράσματος. Αυτό ήταν κοινό στους φορμαλισμούς που επηρεάστηκαν από την OIL και τη DAML + OIL. Τα πλαίσια παρείχαν γενικώς μια διεπαφή με τις εσωτερικές δομές δεδομένων, αντί οποιωνδήποτε άλλων συμπερασμάτων ή ακόμη και ερωτημάτων. Οι Περιγραφικές Λογικές παρέχουν μια τυπική θεωρία ερωτημάτων, αλλά αυτό είναι κάπως διαφορετικό από μια τυπική θεωρία συμπεράσματος.

Η διαφορά είναι ότι τα ερωτήματα της Περιγραφικής Λογικής θα μπορούσαν να είχαν οριστεί για την DAML + OIL με έναν τρόπο που θα είχε βοηθήσει να κρυφτεί την έννοια των RDF τριάδων. Για παράδειγμα, η ερώτηση αν ένα αντικείμενο ανήκει σε μια κλάση θα μπορούσε να προσθέσει την σύνταξη που χρησιμοποιείται για να οριστεί η κλάση στα ερωτήματα. Ωστόσο, μία τυπική θεωρία λογικών συμπερασμάτων δεν μπορεί να το κάνει αυτό.

Οι επιπτώσεις αυτής της αλλαγής μπορούν να φανούν σε ένα απλό παράδειγμα. Δεδομένων των ακόλουθων πληροφοριών

```
ex: John rdf: type ex: Student .
ex: John rdf: type ex: Employee .
```

θα ήταν αρκετά εύκολο να το διαμορφώσει έτσι ώστε η ερώτηση αν ο John ανήκε στην τομή των Student και Employee να εξασφαλίζει πρώτα ότι υπάρχει αυτή η τομή και στην συνέχεια να θέτει το ερώτημα αν ο John ανήκει σε αυτήν. Ωστόσο, η μετατροπή αυτού σε μια συνεπαγωγή απαιτεί οι παραπάνω πληροφορίες να συνεπάγονται

```
: c owl: intersectionOf: I1 .
: I1 rdf: first ex: Student .
: I1 rdf: rest: I2 .
: I2 rdf: first ex: Employee .
```


: l2 rdf: rest rdf: nil .
ex: John rdf: type: c .

τα οποία, εξαιτίας της έννοιας που αποδίδεται σε όλες τις τριάδες του RDF, απαιτούν την ύπαρξη των τριάδων που κωδικοποιούν την σύνταξη.

Η OWL πρέπει ως εκ τούτου να αναπτύξει μια μέθοδο που αυξάνει τη νέα σημασιολογία RDF τόσο ώστε να στηρίζει τα παραπάνω συμπεράσματα, χωρίς να είναι πολύ ισχυρή, και καταλήξει, έτσι, σε σημασιολογικά παράδοξα.

4.4.3 Εκφραστική Ισχύς

Επειδή πολλά πράγματα αναμένονταν από την OWL, υπήρξαν πολλά αιτήματα για εκφραστική ισχύ πέρα από αυτή που συνήθως παρέχεται από τις Περιγραφικές Λογικές. Για παράδειγμα, πολλοί χρήστες ήθελαν να έχουν την ικανότητα να συσχετίσουν πληροφορίες με κλάσεις και ιδιότητες και να κάνουν τις κλάσεις να ανήκουν σε άλλες κλάσεις, όπως είναι δυνατό στο RDF. Ομοίως, υπήρξαν πολλά αιτήματα για εκφραστική ισχύ πέρα από των RDF και RDFS. Για παράδειγμα, πολλοί χρήστες ήθελαν να μπορούν να παράσχουν τοπικά τύπους για τιμές ιδιοτήτων, όπως είναι δυνατό στις Περιγραφικές Λογικές. Η OWL έπρεπε να σχεδιαστεί για να επιτρέπει με κάποιον τρόπο αυτά τα είδη εκφραστικότητας διατηρώντας τις συνδέσεις με τις ρίζες της.

Όταν αναπτύχθηκε η DAML + OIL, ο μόνος τύπος δεδομένων που υποστηριζόταν από το RDF ήταν οι λεκτικές σταθερές: σχεδόν αδιαφοροποίητες τιμές με τη μορφή συμβολοσειρών. Έτσι, η DAML + OIL έπρεπε να παράσχει τη δική της λύση για τους τύπους δεδομένων, και έτσι έκανε, επιτρέποντας τη χρήση των τύπων δεδομένων της XML Schema. Ωστόσο, οποιαδήποτε λογική λύση για τους τύπους δεδομένων που χρησιμοποιεί μόνο την σύνταξη RDF χρειάζεται βοήθεια από το RDF (δηλαδή, μια επέκταση στην σύνταξη RDF), και έτσι η λύση της DAML + OIL παρέμεινε ελλιπής.

Αργότερα, το RDF απέκτησε δική του έκδοση τύπων δεδομένων, όμοια αλλά διαφορετική από τη λύση της DAML + OIL. Έτσι, η OWL υιοθέτησε την έκδοση του RDF.

4.4.4 Υπολογιστικά Προβλήματα

Μια πτυχή της OWL που τη διαχωρίζει από τα RDF και RDFS είναι ότι υποστηρίζει ένα πλούσιο σύνολο εξαγωγών συμπερασμάτων. Μερικές από αυτές τις εξαγωγές συμπερασμάτων είναι αρκετά προφανείς, όπως το παράδειγμα που δόθηκε παραπάνω σχετικά με τους φοιτητές και τους εργαζόμενους, και έτσι φαίνονται εύκολα στον υπολογισμό. Άλλες εξαγωγές συμπερασμάτων που υποστηρίζονται από την OWL, όμως, είναι αρκετά πολύπλοκες.

Συγκεντρώνοντας όλες τις επιθυμίες για την αναπαράσταση της OWL θα είχε δημιουργηθεί ένας φορμαλισμός, όπου το βασικό πρόβλημα εξαγωγής συμπερασμάτων θα ήταν η δυσκολία λήψης αποφάσεων. Για παράδειγμα, επιτρέποντας τον ισχυρισμό σχέσεων μεταξύ των αλυσίδων ιδιοτήτων (όπως λέγοντας ότι ένας θείος είναι συγκεκριμένα ο αδελφός ενός γονέα) θα έκανε την συνεπαγωγή της OWL ανίκανη για λήψη αποφάσεων. Επιπλέον, ορισμένες πτυχές του RDF, όπως η χρήση των κλάσεων ως στιγμιοτύπων, αλληλεπιδρούν με την υπόλοιπη OWL και δημιουργούν υπολογιστικές δυσκολίες, οδηγώντας την OWL πέρα από το εύρος των πρακτικών αλγορίθμων και των εφαρμοσμένων συστημάτων συλλογιστικής.

Η OWL έπρεπε ως εκ τούτου να παράσχει μια λύση για αυτά τα υπολογιστικά ζητήματα εξακολουθώντας να διατηρεί την προς τα πάνω συμβατότητα με τα RDF και RDFS.

4.5 Λύσεις

Η ομάδα εργασίας Οντολογιών Ιστού πέρασε το μεγαλύτερο μέρος ενός έτους για να ξεπεράσει τις βασικές τάσεις που διέπουν τα παραπάνω προβλήματα. Η δυσκολία δεν έγκειται στο κάθε πρόβλημα ξεχωριστά, αλλά στον συνδυασμό όλων των παραπάνω προβλημάτων και των περιορισμών που έχουν τεθεί σχετικά με το σχεδιασμό της OWL. Θα ήταν πολύ πιο εύκολο, για παράδειγμα, να πληρούνται όλες οι παραπάνω απαιτήσεις, εάν η OWL μπορούσε να έχει χρησιμοποιήσει μια προέκταση της σύνταξης RDF. Εάν αυτό είχε επιτραπεί, η OWL θα μπορούσε να είχε προσθέσει νέα, φυσική σύνταξη για τις δομές της, των οποίων η σημασιολογία δεν θα απαιτούνταν να ακολουθήσουν την έννοια των τριάδων του RDF.

Παρ' όλα αυτά βρέθηκε μια βιώσιμη λύση που πληρεί όλες τις παραπάνω απαιτήσεις. Ή, στην πραγματικότητα, είναι πιο ακριβές να πούμε ότι τρεις λύσεις έχουν βρεθεί, καθεμία από τις οποίες ικανοποιεί σχεδόν όλες από τις παραπάνω απαιτήσεις.

OWL DL: Αν η φιλική σύνταξη ή η ικανότητα λήψης αποφάσεων για εξαγωγή συμπερασμάτων θεωρούνται πρωταρχικής σημασίας, τότε, η OWL DL, μια έκδοση της OWL με ικανότητα λήψης αποφάσεων για εξαγωγή συμπερασμάτων που μπορεί να γραφτεί σε μια μορφή πλαισίου ή Περιγραφικής Λογικής, είναι η ενδεδειγμένη.

OWL Lite: Εάν μια ακόμη πιο εύκολη σύνταξη και η τρωτή εξαγωγή συμπερασμάτων θεωρούνται πρωταρχικής σημασίας, τότε η OWL Lite, ένα συντακτικό υποσύνολο της OWL DL, είναι η ενδεδειγμένη.

OWL Full: Αν η συμβατότητα με τα RDF και RDFS θεωρείται πρωταρχικής σημασίας, τότε η OWL Full, μια συντακτική και σημασιολογική επέκταση του RDFS, είναι η κατάλληλη.

4.5.1 Αναγνωσιμότητα

Όπως φαίνεται από τα παραπάνω παραδείγματα, η OWL δεν είναι πολύ ευανάγνωστη όταν γράφεται ως RDF / XML ή ακόμα και ως RDF τριάδες. Μέρος αυτού του προβλήματος είναι ότι η RDF / XML σύνταξη είναι εξαιρετικά φλύαρη, αλλά το μεγαλύτερο μέρος του προβλήματος αναγνωσιμότητας είναι η κωδικοποίηση των δομών της OWL σε RDF / XML ή σε RDF τριάδες.

Για να αντιμετωπιστεί εν μέρει αυτό το πρόβλημα, δημιουργήθηκε μια αφηρημένη σύνταξη για την OWL, μαζί με την αντιστοίχιση από την αφηρημένη σύνταξη προς τα γραφήματα RDF. Η αφηρημένη σύνταξη είναι πιο κοντά στην σύνταξη της OIL από ό, τι στην DAML + OIL, αλλά χωρίς την ακραία έμφαση της OIL για αναγνωσιμότητα. Σε αυτήν την αφηρημένη σύνταξη το παραπάνω παράδειγμα θα γραφτεί

```
Class(Student complete
Person
restriction(enrolledIn minCardinality(1))).
```

Τότε ορίστηκε η OWL DL ως το συντακτικό υποσύνολο της OWL που προκύπτει από τη μετάφραση από την αφηρημένη σύνταξη σε γραφήματα RDF. Δηλαδή, ένα γράφημα RDF είναι μια οντολογία OWL DL μόνο όταν είναι η μετάφραση κάποιας οντολογίας στην αφηρημένη σύνταξη. Οι χρήστες και τα εργαλεία που ενδιαφέρονται περισσότερο για την αναγνωσιμότητα από ό, τι για το RDF / XML μπορούν να χρησιμοποιήσουν αυτή την αφηρημένη σύνταξη εσωτερικά, ή ακόμα και εξωτερικά για παρουσίαση στους χρήστες,

διατηρώντας την RDF / XML σύνταξη για τους σκοπούς της ανταλλαγής μεταξύ των εφαρμογών.

4.5.2 Χειρισμός Λανθασμένων Γραφημάτων

Επειδή η OWL Full επιτρέπει αυθαίρετα γραφήματα RDF, πρέπει να είναι σε θέση να χειρίζεται ακατάλληλη σύνταξη OWL. Η OWL χρησιμοποιεί μια επέκταση της λύσης της DAML + OIL: μόνο στις τριάδες που συνθέτουν καλοσχηματισμένες δομές OWL δίνεται ένα επιπλέον νόημα, έτσι το

: x owl: onProperty ex: child .

από μόνο του δεν έχει καμία ιδιαίτερη έννοια OWL.

Για να χειριστεί τις υποθέσεις πάρα πολλών τριάδων, η OWL χρησιμοποιεί και πάλι τη λύση της DAML + OIL να επιλέξει όλα τα καλοσχηματισμένα υποσύνολα και να τους δώσει OWL έννοια. Αυτό έχει ασυνήθιστες συνέπειες, για παράδειγμα, το

: x owl: onProperty ex: child .

: x owl: someValuesFrom ex: Person .

: x owl: onProperty ex: friend .

: x owl: allValuesFrom ex: Doctor .

καταλήγει να εξισώνει την επέκταση τεσσάρων διαφορετικών περιορισμών OWL (όλοι οι πιθανοί συνδυασμοί των δύο ιδιοτήτων και των δύο κλάσεων), το οποίο είναι σχεδόν βέβαιο ότι δεν είναι αυτό που ήθελε ο χρήστης. Η λύση αυτή, ωστόσο, διατηρεί τη μονοτονία, και το (ενδεχομένως) μη-διαισθητικό νόημα είναι ένα μικρό πρόβλημα, δεδομένου ότι τέτοιες εν λόγω ακατάλληλες κατασκευές μπορούν εύκολα να αποφευχθούν. Η έλλειψη δομής στα γραφήματα RDF πρέπει να αντιμετωπιστεί με σημασιολογικά μέσα, τα οποία περιγράφονται στην συνέχεια.

4.5.3 Παροχή μιας βιώσιμης Σημασιολογικής Θεωρίας για την OWL

Καθώς το RDF έχει τώρα μια θεωρία μοντέλου, με μια ολοκληρωμένη έννοια συνεπαγωγής, OWL πρέπει να παρέχει ένα προς τα πάνω συμβατό μοντέλο θεωρίας που να χειρίζεται κατάλληλα συνεπαγωγές στις δομές της OWL. Αυτό θα ήταν εύκολο αν η OWL μπορούσε να επεκτείνει την σύνταξη του RDF, καθώς τότε αυτά τα νέα κομμάτια της σύνταξης θα μπορούσαν να έχουν μια έννοια μόνο της OWL. Ωστόσο, ήταν μια απαίτηση η OWL να έχει σύνταξη RDF, και ότι αυτή η σύνταξη θα κρατούσε όλο το RDF νόημα. Αυτή η αμφίδρομη απαίτηση συμβατότητας είναι πολύ ισχυρότερη από ό, τι επιβάλλεται συνήθως μεταξύ ασθενέστερων φορμαλισμών (όπως η προτασιακή λογική) και ισχυρότερων φορμαλισμών (όπως η λογική firstorder), όπου ο ισχυρότερος φορμαλισμός επιτρέπεται να επεκτείνει την σύνταξη του ασθενέστερου φορμαλισμού.

Η σοβαρότερη πτυχή αυτού του προβλήματος είναι ότι οι συντακτικές δομές της OWL που κωδικοποιούνται ως πολλαπλές τριάδες πρέπει να διατηρήσουν την έννοια του RDF για αυτές τις τριάδες. Καθώς όλες οι τριάδες RDF φέρουν σημασιολογικές συνθήκες, δε μπορούν απλώς να συναχθούν από το τίποτα. Αντίθετα, η σημασιολογία της OWL έπρεπε να προσθέσει ειδικούς περιορισμούς που ουσιαστικά ορίζουν ότι κάθε ερμηνεία OWL πρέπει να περιλαμβάνει ορισμένες δομές. (Τέτοιοι περιορισμοί συνήθως ονομάζονται αρχές κατανόησης.) Για παράδειγμα, μία αρχή κατανόησης OWL δηλώνει ότι κάθε μοντέλο πρέπει

να περιλαμβάνει όλες τις πεπερασμένες λίστες των κλάσεων. Άλλη δηλώνει ότι κάθε τέτοια λίστα πρέπει να έχει μια αντίστοιχη τετμημένη κλάση, προϋποθέτοντας ότι υπάρχει κάποια κλάση που συνδέεται με τη λίστα με μια ιδιότητα owl: intersectionOf.

Αυτές οι αρχές κατανόησης υποστηρίζουν την συνεπαγωγή από το

ex: John rdf: type ex: Student .
 ex: John rdf: type ex: Employee .

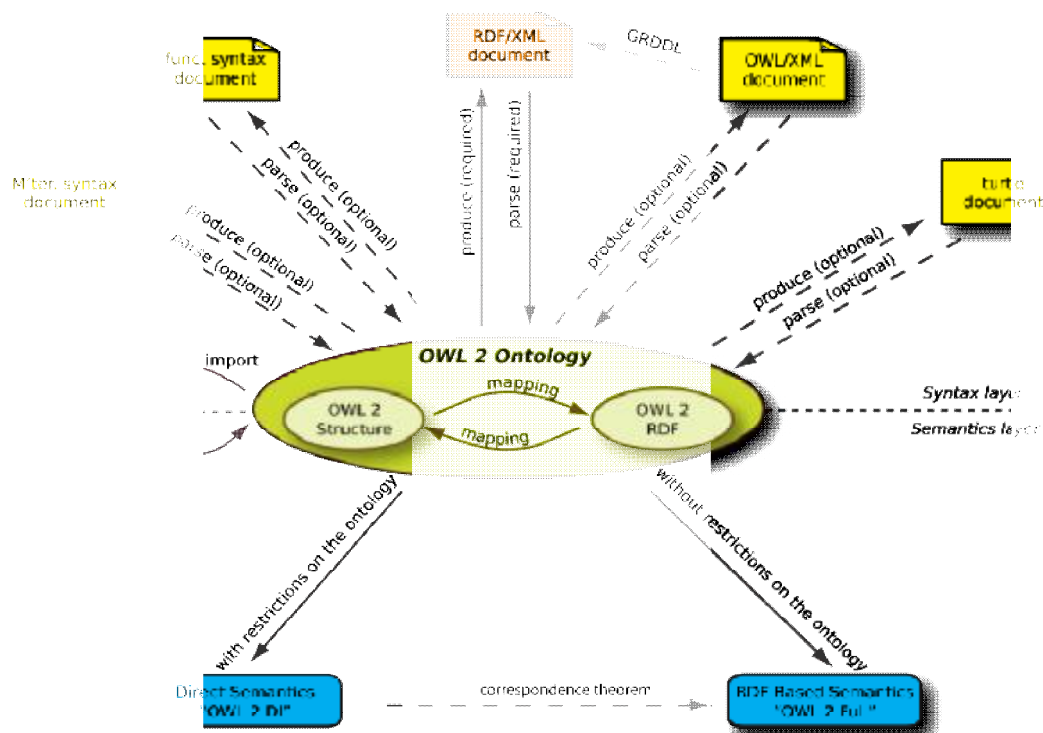
στο

: c owl: intersectionOf: I1 .
 : I1 rdf: first ex: Student .
 : I1 rdf: rest: I2 .
 : I2 rdf: first ex: Employee .
 : I2 rdf: rest rdf: nil .
 ex: John rdf: type: c .

διότι το

: I1 μπορεί να είναι η απαιτούμενη λίστα των Student και Employee,
 το: I2 μπορεί να είναι η απαιτούμενη ουρά της λίστας αυτής, καθώς και
 το: c μπορεί να είναι η απαιτούμενη τομή της λίστας.

Επιπλέον (και συνήθως) οι σημασιολογικές συνθήκες προϋποθέτουν ότι το ex: John ανήκει στο: c, ολοκληρώνοντας όλα όσα απαιτούνται για την συνεπαγωγή.



Εικόνα 16: Δομή της OWL

<http://www.w3.org/TR/owl2-overview/>

4.5.4 Αποφυγή Παράδοξων

Οι αρχές κατανόησης είναι πολύ ισχυρές, δεδομένου ότι δημιουργούν κάτι από το τίποτα (ή, τουλάχιστον, κάτι από πολύ λίγα). Αυτή η δύναμη μπορεί εύκολα να οδηγήσει σε σοβαρά προβλήματα.

Για παράδειγμα, το παράδοξο του Ράσελ είναι παράδοξο ακριβώς λόγω των αρχών κατανόησης που ενσωματώθηκαν σε μια πρώιμη έκδοση της θεωρίας συνόλων. Αυτή η πρώιμη έκδοση της θεωρίας συνόλων είχε μια αρχή κατανόησης που δήλωνε ότι ένα σύνολο θα μπορούσε να κατασκευαστεί από τα πράγματα που ικανοποιούν μία φόρμουλα με μια ελεύθερη μεταβλητή, π.χ., τον τύπο της ύπαρξης ενός ανθρώπου, $x \hat{=} \text{human}$, θα μπορούσε να χρησιμοποιηθεί για την κατασκευή του συνόλου των ανθρώπων, $\{x \mid x \hat{=} \text{human}\}$. Δυστυχώς, από τον τύπο που δεν ανήκει κάτι στον εαυτό του, $x \hat{=} x$, προκύπτει το σύνολο $\{x \mid x \hat{=} x\}$. Αυτό το σύνολο προκαλεί προβλήματα όπου υπάρχει, επειδή είναι αδύνατο να προσδιορίσει αν κάτι ανήκει στον εαυτό του. Οι αρχές κατανόησής του, απαγορεύουν ότι υπάρχει παντού, οδηγώντας έτσι αυτήν την πρώιμη έκδοση της θεωρίας συνόλων σε κατάρρευση.

Μια παρόμοια κατάσταση μπορεί να προκύψει με την OWL. Φαίνεται φυσικό να θέλουμε να έχουμε κυκλικές δομές σαν της OWL, για παράδειγμα, κλάσεις των οποίων τα στιγμιότυπα σχετίζονται μόνο με άλλα στιγμιότυπα της κλάσης, όπως

```
: c owl: onProperty ex: child .  
: c owl: allValuesFrom: c .
```

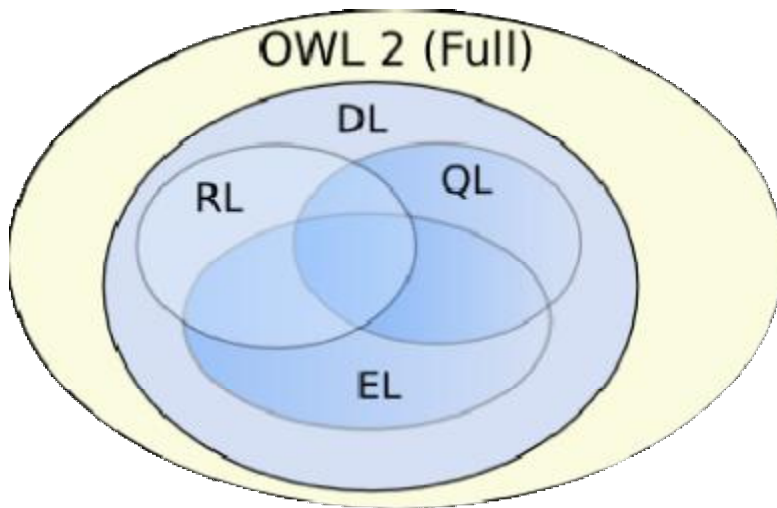
που θα μπορούσε να χρησιμοποιηθεί σε μια απλοϊκή αναπαράσταση ορισμένων πτυχών της βιολογίας.

Ωστόσο, η ύπαρξη αρχών κατανόησης για τέτοιες κυκλικές κλάσεις μπορεί εύκολα να οδηγήσει σε απαίτηση για την ύπαρξη κλάσεων, όπως

```
: c owl: onProperty rdf: type .  
: c owl: allValuesFrom: d .  
: d owl: complementOf: l .  
: l rdf: first: c .  
: l rdf: rest rdf: nil .
```

η οποία είναι η κλάση των πραγμάτων που δεν έχουν κανενός τύπου σχέση με την ίδια την κλάση. Τα αντικείμενα που ανήκουν σε αυτή την κλάση δεν μπορούν να ανήκουν σε αυτήν, και αντίστροφα, έτσι ώστε αν οι αρχές κατανόησης απαιτούσαν την ύπαρξη αυτής της κλάσης, τότε κάθε οντολογία OWL θα ήταν παράδοξη.

Για την αποφυγή αυτών των παράδοξων, οι αρχές κατανόησης της OWL δεν απαιτούν ποτέ την ύπαρξη κυκλικών αλυσίδων αναφοράς όπως παραπάνω. Ωστόσο, αυτό δεν σημαίνει ότι υπάρχουν συνεπαγωγές όπως θα περίμενε κανείς, όπως αυτές που αφορούν τη δομή με το `ex: child` παραπάνω, για παράδειγμα η ύπαρξη ενός ατόμου χωρίς παιδιά ανήκει σε μια τέτοια δομή, που δεν είναι έγκυρη στην OWL. Ο σχεδιασμός αυτών των αρχών κατανόησης χρειάστηκε μεγάλη προσπάθεια (μεγάλο ποσοστό εμπλέκεται με τον καθορισμό των βασικών κανόνων για τις αρχές).



Εικόνα 17: Διάγραμμα Venn

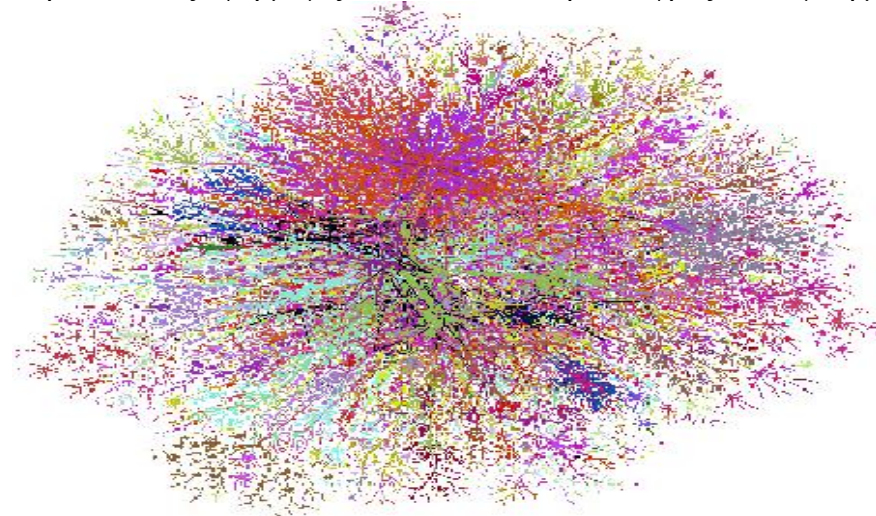
http://wiki.opensemanticframework.org/index.php/Metamodeling_in_Domain_Ontologies

4.5.5 Διατήρηση Ικανότητας Λήψης Αποφάσεων

Η OWL Full δεν είναι ικανή για λήψη αποφάσεων (για διάφορους λόγους), και ακόμη και η OWL DL θα μπορούσε εύκολα να μην είναι ικανή για λήψη αποφάσεων αν περιελάμβανε ορισμένες δομές που είναι γνωστό ότι προκαλούν ανικανότητα λήψης αποφάσεων στις Περιγραφικές Λογικές. Ως εκ τούτου, η OWL DL σχεδιάστηκε προσεκτικά για να παραμείνει ικανή για λήψη αποφάσεων, και δεν περιλαμβάνει, για παράδειγμα, σχέσεις μεταξύ αλυσίδων ρόλων, που θα προκαλούσαν ανικανότητα λήψης αποφάσεων, προκαλώντας πρόβλημα στην OWL DL.

Αυτό δεν σημαίνει ότι η εξαγωγή συμπερασμάτων στην OWL DL δεν είναι δύσκολη. Η OWL DL έχει ένα δύσκολο πρόβλημα συνεπαγωγής, καθώς η συμπερασματολογία στη SHOIN(D) είναι η χειρότερη περίπτωση εκθετικού χρόνου (NExpTime) πολυπλοκότητας, και η OWL DL πρέπει να έχει την ίδια πολυπλοκότητα.

Η OWL Lite είναι καλύτερη από αυτή την άποψη. Ο συμπερασμός στη SHIF (D) είναι στη χειρότερη περίπτωση εκθετικού χρόνου (ExpTime) πολυπλοκότητας, και η OWL Lite έχει την ίδια πολυπλοκότητα. Επιπλέον, υπάρχουν πρακτικοί αλγόριθμοι βελτιστοποίησης για συμπερασματολογία στην OWL Lite, όπως ο αλγόριθμος που διέπει τη λογική των συστημάτων FACT και RACER. Τα συστήματα αυτά έχουν αποδειχθεί ότι λειτουργούν καλά σε ρεαλιστικές εφαρμογές και είναι σε θέση να εκφράζουν λογική με μεγάλες οντολογίες.



ΚΕΦΑΛΑΙΟ 5^ο

5.1 Συστήματα βασισμένα σε οντολογίες

Οι οντολογίες είναι μια έννοια που έχει εισαχθεί στην πληροφορική τα τελευταία δεκαπέντε χρόνια όμως πολλά συστήματα έχουν βασιστεί σε αυτές. Στην παράγραφο αυτή θα γίνει μια σύντομη αναφορά σε μερικά από τα συστήματα που χρησιμοποιούν σημασιολογική προσέγγιση μέσω οντολογιών. Τα συστήματα αυτά χρησιμοποιούνται κυρίως για κατηγοριοποίηση αλλά και για άλλες εφαρμογές.

5.1.1 OntoSeek

Το OntoSeek είναι ένα πρωτότυπο αποτέλεσμα συνεργασίας του Corinto (τμήμα της IBM Semea, Apple στην Ιταλία) και του Ladseb-CNR (ιστιτούτο συστημάτων επιστήμης και βιοϊατρικής τεχνολογίας) για τμήμα του έργου για ανάκτηση και επαναχρησιμοποίηση αντικειμενοστραφών τμημάτων λογισμικού.

Αποτελεί ένα σύστημα που σχεδιάστηκε για ανάκτηση πληροφορίας από χρυσό οδηγό και καταλόγους προϊόντων βασισμένη στο περιεχόμενο. Συνδυάζει έναν μηχανισμό αντιστοίχισης περιεχομένου οδηγούμενο από οντολογίες με έναν εκφραστικό φορμαλισμό αναπαράστασης. Τα παρακάτω αποτελούν τις βασικές σχεδιαστικές επιλογές για το OntoSeek:

1. Η επιλογή να χρησιμοποιηθούν αφηρημένοι όροι φυσικής γλώσσας για ακριβείς περιγραφές πόρων στη φάση της κωδικοποίησης .
2. Πλήρης εννοιολογική ευελιξία για τα ερωτήματα, μέσα από μια διαδικασία σημασιολογικού ταιριάσματος οδηγούμενο από οντολογίες μεταξύ ερωτημάτων και περιγραφών πόρων
3. Αλληλεπιδραστική βοήθεια στο σχηματισμό, τη δημιουργία και την εξειδίκευση μορφοποίησης ερωτημάτων
4. Μια αρχιτεκτονική τελευταίας γενιάς
5. Καλούς παράγοντες ανάκλησης και ακρίβειας, και δικαιολογημένη αποδοτικότητα σε μαζικά δεδομένα
6. Καλή κλιμάκωση και φορητότητα

Το σύστημα σχεδιάστηκε για να διαχειρίζεται ομογενή και ετερογενή δεδομένα καταλόγων προϊόντων. Οι ετερογενείς κατάλογοι προϊόντων είναι περισσότεροι δύσκολοι στη διαχείριση επειδή έχουν μεγαλύτερη πολυπλοκότητα στην περιγραφή χρησιμοποιήθηκαν απλοί γράφοι σημασιολογίας για να αναπαρασταθούν ερωτήματα και περιγραφές πόρων. Το σύστημα δε βασίστηκε σε οντολογίες που φτιάχτηκαν από την αρχή αλλά χρησιμοποίησε έτοιμες οντολογίες. Συγκεκριμένα, χρησιμοποιήθηκε η οντολογία Sensus που συμπληρώνει για απλή ταξινομική προσέγγιση με 50000 κόμβους που λήφθηκαν από το WordNet.

5.1.2 Telltale

Το Telltale αποτελεί ένα σύστημα κατηγοριοποίησης εγγράφων βασισμένο στους καταλόγους τους Yahoo. Αυτοί οι κατάλογοι αποτελούν κατηγορίες που προσφέρουν ένα κανονικοποιημένο και καθολικό τρόπο για αναφορά ή περιγραφή αντικειμένων του

πραγματικού κόσμου και κατ' επέκταση σημασιολογική κατηγοριοποίηση του περιεχομένου εγγράφων. Οι κατάλογοι του Παγκόσμιου Ιστού όπως το Yahoo! Προσφέρουν μια τεράστια ιεραρχία κατηγοριών που καλύπτουν κάθε πτυχή ανθρώπινης προσπάθειας. Αυτά τα θέματα μπορούν να χρησιμοποιηθούν σαν περιγραφείς, ομοίως με τον τρόπο εύρεσης ομοιότητας που χρησιμοποιούν οι βιβλιογράφοι για παράδειγμα.

Το Telltale χρησιμοποιεί τη μέθοδο n-grams για να υπολογίσει την ομοιότητα μεταξύ εγγράφων. Κατά τη δημιουργία του έγιναν πειραματικές μετρήσεις με διάφορες περιγραφές για τους καταλόγους του Yahoo! ώστε να κατηγοριοποιηθούν ιστοσελίδες. Τα αποτελέσματα των πειραμάτων έδειξαν ότι η κατηγοριοποίηση ήταν καλύτερη στην περίπτωση που χρησιμοποιούνταν σύντομες περιγραφές για τις διάφορες κατηγορίες.

Για το σύστημα αυτό με βάση τις κατηγορίες των καταλόγων δημιουργήθηκε μια ταξινόμια – οντολογία πάνω στην οποία θα βασιστεί η κατηγοριοποίηση. Κατά την κατηγοριοποίηση γινόταν υπολογισμός των βαρών για κάθε όρο. Το βάρος κάθε όρου υπολογιζόταν ως η διαφορά μεταξύ του μετρήματος ενός δοσμένου n-gram για ένα έγγραφο, κανονικοποιημένο στο μέγεθός του, και το κανονικοποιημένο μέσο όρο του μετρήματος σε όλα τα έγγραφα για αυτό το n-gram. Αυτό παρέχει το βάρος για κάθε n-gram σε ένα έγγραφο σχετικό με το μέσο όρο της συλλογής. Η ομοιότητα μεταξύ εγγράφων μετά υπολογίζεται από το συνημίτονο των δύο διανυσμάτων αναπαράστασης.

5.1.3 Ontology-Based Semantic Online Classification of Documents

Το σύστημα αυτό περιγράφει μια μέθοδο κατηγοριοποίησης κειμένων με στόχο την αποσαφήνιση ερωτημάτων χρηστών. Οι τεχνικές που εφαρμόζονται βασίζονται σε συνδυασμό τεχνικών δεικτοδότησης και σε τεχνικές βασισμένες σε οντολογίες ανάκτησης πληροφορίας σε ένα αλληλεπιδραστικό σύστημα ανάκτησης πληροφορίας. Το σύστημα προτείνει μια προσέγγιση για κατηγοριοποίηση αποτελεσμάτων αναζήτησης αντιστοιχίζοντάς τα σε σημασιολογικές κλάσεις που ορίζονται από τη λογική ενός ερωτήματος.

Τα κριτήρια που ορίζουν κάθε κλάση ή «λογικό κατάλογο» αντλούνται από τις έννοιες μιας προσδιορισμένης οντολογίας, εδώ το MultiWordNet. Σχολιάζονται κάθε στοιχείο του συνόλου αποτελεσμάτων με τον λογικό κατάλογο στον οποίο έχει κατηγοριοποιηθεί, ο χρήστης παίρνει πρόσθετη πληροφορία για κάθε αντικείμενο. Ο συγκεκριμένος όρος αποτελέσματος αποσαφηνίζεται σε αντιστοιχία με το υποκείμενο έγγραφο και μπορεί έτσι να αποφασιστεί πιο εύκολα αν το έγγραφο είναι σχετικό ή όχι για το συγκεκριμένο ερώτημα .

5.2 Οντολογίες αναφοράς - προφίλ

Τα συστήματα προσωποποίησης του Παγκόσμιου Ιστού που χρησιμοποιούν σημασιολογία βασίζονται πάντα σε μια οντολογία αναφοράς. Η οντολογία αναφοράς χρησιμοποιείται για κατηγοριοποίηση των επιλογών των χρηστών ώστε να καθοριστεί το προφίλ του και να εξατομικευτούν οι μελλοντικές επιλογές του βάσει του προφίλ που έχει δημιουργηθεί για αυτόν.

Στην ενότητα αυτή γίνεται μια αναφορά στις βασικότερες οντολογίες αναφοράς που έχουν χρησιμοποιηθεί σε συστήματα προσωποποίησης καθώς και κατάλογοι του Παγκόσμιου Ιστού πάνω στους οποίους έχουν βασιστεί οντολογίες που περιγράφουν προφίλ.

5.2.1 SUMO — the Suggested Upper Merged Ontology

Η Sumo είναι μια συλλογή από περίπου 1000 καλά ορισμένες έννοιες, διασυνδεδεμένες σε ένα σημασιολογικό δίκτυο και συνοδευόμενων από ένα σύνολο αξιωμάτων. Οι έννοιες έχουν εύρος από πολύ γενικές, όπως «Ποσότητα», σε πολύ συγκεκριμένες, όπως «Πουλί». Τα αξιώματα αντανακλούν κυρίως την αντίληψη της κοινής λογικής που αναγνωρίζονται μεταξύ των εννοιών. Η SUMO σχεδιάστηκε ως ένα υπόστρωμα ανεξάρτητο από πεδίο για σχεδίαση οντολογιών πεδίων. Τα αξιώματα βοηθούν στο να περιοριστούν οι διερμηνείες για τις έννοιες και να παρέχουν οδηγίες για συστήματα αυτοματοποιημένης αιτιολόγησης που επεξεργάζονται βάσεις, γνώσεις που συνάδουν με την οντολογία SUMO. Ένα παράδειγμα ενός τέτοιου αξιώματος είναι: «Αν το *c* είναι στιγμιότυπο της «Καύσης», τότε υπάρχει η θέρμανση *h* και το ακτινοβόλο φως *L* έτσι ώστε και το *h* και το *L* είναι παράγωγα του *c*».

Αυτό η περίπλοκη, αλλά λογική, πρόταση λέει ότι η διαδικασία της θέρμανσης και η διαδικασία εκπομπής φωτός συνοδεύουν κάθε διαδικασία καύσης. Επιπλέον, αυτό το αξίωμα κωδικοποιείται στο SUMO σε μια επίσημη λογική γλώσσα.

Οι έννοιες στο SUMO οργανώνονται σε μια απλή ιεραρχία που έχει σαν ρίζα την έννοια «Entity», που αναπαριστά την πιο γενική έννοια.

Οι υποκλάσεις της κλάσης είναι συνήθως αμοιβαία αποκλειόμενες δηλαδή δεν μοιράζονται κοινά στιγμιότυπα. Για παράδειγμα, τίποτα δεν μπορεί να είναι και αφηρημένο και φυσικό, ούτε και αντικείμενο και διαδικασία. Αυτή η ιδιότητα ορίζεται αποκλειστικά στην SUMO. Παρόλα αυτά, κάποιες κλάσεις μπορούν να έχουν πολλαπλές υπερκλάσεις. Για παράδειγμα, η κλάση Human (άνθρωπος) μπορεί να είναι Υπόκλιση της κλάσης Hominid (ανθρωποειδές - είναι μέλος της κλάσης των ζώων) και της κλάσης Cognitive Agent (μια οντότητα με την ικανότητα να σκέφτεται λογικά).

Ένα από τα μειονεκτήματα της SUMO είναι η σχετικά μικρή κάλυψή της που δεν της επιτρέπει να είναι αποδοτική για ανοιχτού πεδίου εφαρμογές. Επίσης, έχει έλλειψη μιας σύνδεσης μεταξύ των εννοιών της και των λέξεων της φυσικής γλώσσας. Αυτοί οι περιορισμοί έχουν παρακαμφθεί μερικώς συνδέοντας την οντολογία SUMO με το λεξικό WordNet.

5.2.2 WordNet — Μια Online λεξικογραφική βάση δεδομένων

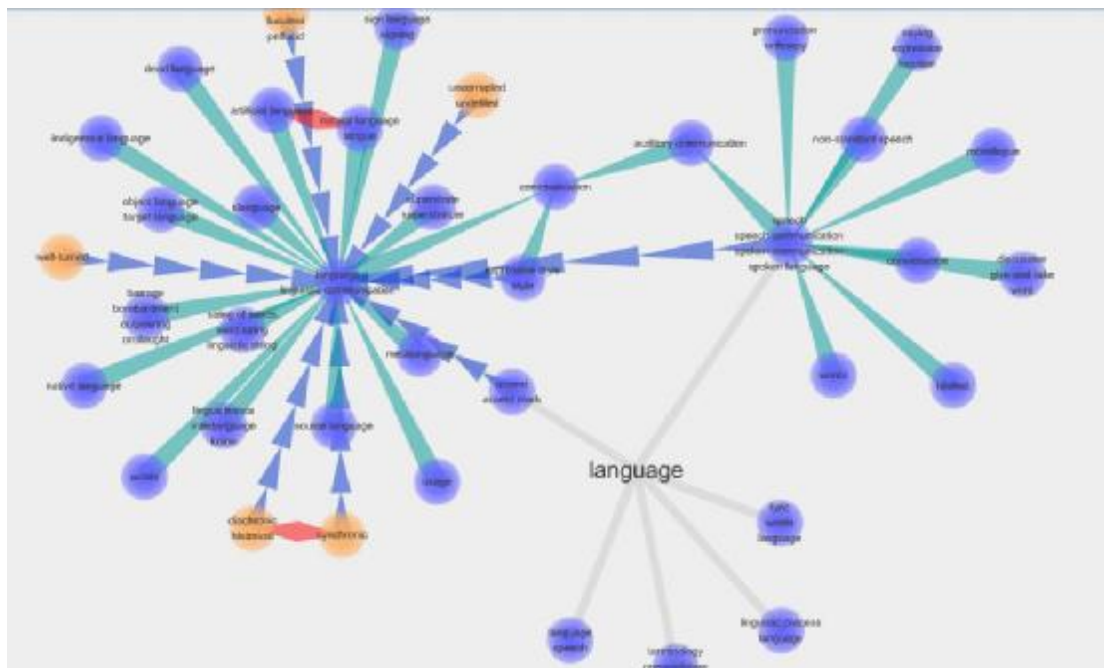
Το WordNet είναι μια ελεύθερα διαθέσιμη on-line λεξικογραφική βάση δεδομένων. Το τμήμα γλωσσολογίας του Πανεπιστημίου Πρίνστον τη δημιούργησε σαν αποτέλεσμα της ψυχολογολογικής έρευνας. Παρόλα αυτά την τελευταία δεκαετία το WordNet αποδείχτηκε πολύ χρήσιμη πηγή για αυτοματοποιημένη επεξεργασία της φυσικής γλώσσας.

Τεχνικά, το WordNet είναι ένας ηλεκτρονικός θησαυρός, ορίζοντας μεγάλα σύνολα εννοιών λέξεων, διασυνδεδεμένα με σημασιολογικούς δείκτες.

Οι έννοιες λέξεων διασυνδέονται με φόρμες λέξεων που μπορούν να τις εκφράσουν. Μπορούμε να δούμε στην εικόνα ότι η σχέση μεταξύ φορμών λέξεων και οι έννοιες λέξεων είναι m-n- οι φόρμες λέξεων μπορούν να έχουν πολλές έννοιες, και πολλές φόρμες λέξεων μπορούν να αναφέρονται σε πολλές έννοιες. Το πρώτο φαινόμενο ονομάζεται πολυσημία και το δεύτερο συνωνυμία. Η αντιμετώπιση μιας τέτοιας αμφισημίας της φυσικής γλώσσας είναι η πρόκληση κλειδί στην αυτοματοποιημένη επεξεργασία της φυσικής γλώσσας.

Κάθε είσοδος λέξεων εννοιών (που καλείται επίσης σύνολο συνωνυμίας, ή synset), συνοδεύεται με σύντομους άτυπους ορισμούς (που καλούνται gloss), και λίστες φορμών λέξεων που μπορούν να αναπαραστήσουν το synset στην προφορική ή στην γραπτή γλώσσα. Τα synsets κρατούνται ξεχωριστά για διαφορετικά μέρη του λόγου: υπάρχουν βάσεις δεδομένων για ουσιαστικά, ρήματα, επίθετα και επιρρήματα. Θα πρέπει να σημειωθεί ότι οι σημασιολογικές συσχετίσεις μεταξύ synsets είναι διαφορετικές για διαφορετικά μέρη του λόγου. Για παράδειγμα, για ουσιαστικά μια βασική συσχέτιση μεταξύ synsets είναι μια is-a συσχέτιση, που είναι γνωστή από τη μοντελοποίηση δεδομένων. Στο WordNet, αυτή η συσχέτιση καλείται υπερωνυμία / υπωνυμία.

Μπορεί να φαίνεται με την πρώτη ματιά ότι τα synsets στο WordNet χτίζουν ένα μεγάλο σημασιολογικό δίκτυο, όπως ξέρουμε είναι ένα παράδειγμα αναπαράστασης γνώσης της τεχνητής νοημοσύνης. Παρόλα αυτά, ένας πιο επιμελής έλεγχος αποκαλύπτει ότι οι σημασιολογικές συσχετίσεις στο WordNet είναι μερικές φορές πολύ ασαφείς και μη λογικές, και δεν μπορούν να χρησιμοποιηθούν για εξαγωγή συμπεράσματος. Οι συσχετίσεις κωδικοποιούνται από λεξικογράφους, και σηματοδοτούνταν σαν ομοιότητα που κατανοούν οι άνθρωποι από συσχετίσεις μεταξύ λέξεων εννοιών. Επιπλέον, εξαιτίας του υπερβολικού μεγέθους του σημασιολογικού δικτύου, η αιτιολογία σχεδιασμού των σημασιολογικών συσχετίσεων είναι μάλλον λογική, χωρίς να δίνεται έμφαση στην γενικότερη δομή του συνολικού δικτύου.



Εικόνα 18: Παράδειγμα WordNet

<https://www.language-trainers.co.uk/blog/2009/05/04/visualising-the-wordnet/>

Θεωρείστε ένα παράδειγμα ενός synset που αντιστοιχεί στη λέξη blues. Το WordNet ορίζει τη λέξη blues σαν «ένα είδος παραδοσιακού τραγουδιού που ξεκίνησε από τους Μαύρους Αμερικανούς στην αρχή του 20ού αιώνα, έχει έναν μελαγχολικό ήχο από επαναλαμβανόμενη χρήση από blues νότες».

Μπορεί να φανεί από την εικόνα ότι αυτό το μικρό σημασιολογικό δίκτυο είναι αρκετά λάθος κατασκευασμένο, εξαιτίας της χαλαρής μετάφρασης των ορισμών των εννοιών. Για παράδειγμα, το Blues είναι και αφηρημένη έννοια (πνευματικό, μη υπαρκτό) και αντικείμενο (φυσική έννοια). Ομοίως, η έννοια «folksong» έχει δύο έννοιες: μια έννοια σημαίνει μια κλάση από τραγούδια που είναι παραδοσιακά, και είναι και ένα είδος της έννοιας «song». Παρόλα αυτά, η έννοια μπορεί επίσης να κατανοηθεί σαν ένα χαρακτηριστικό τραγουδιών, που είναι είδος της κλάσης «music genre». Στην κοινή γλώσσα αυτές οι δύο έννοιες δεν διακρίνονται αποκλειστικά, και έτσι αυτός ο διαχωρισμός δεν διαχειρίζεται στο WordNet.

Παρόμοια κατάσταση συμβαίνει με πολλαπλές υπερωνυμίες της έννοιας «music genre», όπου παραλείπεται η διάκριση μεταξύ μιας διαδικασίας (music), και του ρόλου του (social relation). Φυσικά, υπάρχουν πολλά παρόμοια προβλήματα που περιπλέκουν την εξερεύνηση του WordNet σαν πηγή για αυτοματοποιημένη επεξεργασία της φυσικής γλώσσας.

Είναι ενδιαφέρον να ερωτηθεί αν αυτές οι ασυμφωνίες είναι εσωτερικές σε πηγές όπως το WordNet, ή αν μπορούν να αποφευχθούν από περισσότερο προσεκτικό σχεδιασμό. Το πρόβλημα είναι ότι λογικές και γλωσσολογικές αιτιολογίες για οργάνωση εννοιών λέξεων είναι αρκετά διαφορετικές, ειδικά για πολύ γενικές έννοιες, που είναι κοντά στη ρίζα της ιεραρχίας.

5.3 Το Protégé - 2004

Η πλατφόρμα πρωτοεμφανίστηκε το 1988 και αποτελούσε απλώς ένα μέσο για τη δημιουργία εργαλείων ανάκτησης γνώσης για έμπειρα συστήματα. Σήμερα, το Protégé έχει εξελιχθεί σε ένα σύγχρονο εργαλείο μοντελοποίησης γνώσης, χάρη στη συστηματική εξέλιξή του από το Τμήμα Ιατρικής Πληροφορικής του Πανεπιστημίου του Stanford (Stanford Medical Informatics – SMI). Αυτή τη στιγμή χρησιμοποιείται από 4.500 χρήστες σε όλο τον κόσμο. Οι χρήστες του προσφέρουν τις γνώσεις τους σε μια on-line κοινότητα, που έχει δημιουργηθεί για να βοηθήσει στην εξάπλωση της γνώσης και την επίλυση προβλημάτων που οι χρήστες αντιμετωπίζουν, γεγονός το οποίο έχει συντελέσει σημαντικά στην περαιτέρω ανάπτυξη της εφαρμογής από την προγραμματιστική ομάδα.

Το Protégé - 2004 είναι μία πλατφόρμα ανοιχτού κώδικα (open source ontology editor), η οποία παρέχει ένα σύνολο εργαλείων για κατασκευή και διαχείριση οντολογιών και βάσεων γνώσης. Εκτός από τον πυρήνα του, που αποτελεί μία πλήρη εφαρμογή ανάπτυξης και διαχείρισης οντολογιών, προσφέρει και ένα περιβάλλον για ανάπτυξη οντολογικών εφαρμογών (Application Programming Interface - API), που βασίζεται στη γλώσσα JAVA. Μέσω του Protégé - 2004 είναι δυνατή η δημιουργία μιας OWL οντολογίας, η επεξεργασία των στοιχείων της (κλάσεων, ιδιοτήτων και στιγμιότυπων), καθώς και η εκτέλεση ερωτήσεων (queries) πάνω στην οντολογία.

Σκοπός του είναι να καθοδηγήσει την ανάπτυξη συστημάτων καθιστώντας δυνατή την επαναχρησιμοποίηση οντολογιών και μεθόδων επίλυσης προβλημάτων, μειώνοντας έτσι το χρόνο που απαιτείται για την ανάπτυξη και τη συντήρηση προγραμμάτων.

5.3.1 Γραφικό Περιβάλλον Εργασίας – Βασικές Λειτουργίες

Το Γραφικό Περιβάλλον Εργασίας (Graphical User Interface - GUI) του Protégé-2004 καθιστά δυνατά τα εξής:

- τη μοντελοποίηση μίας οντολογίας που περιγράφει κάποιο πεδίο,
- την υλοποίηση ενός εργαλείου δημιουργίας μίας βάσης γνώσης για τη συγκέντρωση πληροφοριών,
- την εισαγωγή συγκεκριμένων στιγμιότυπων δεδομένων και τη δημιουργία μίας βάσης γνώσης, και
- την εκτέλεση εφαρμογών.

Για το σχεδιασμό και ανάπτυξη της οντολογίας χρησιμοποιήσαμε το λογισμικό PROTÉGÉ - 2004.

Καθορισμός του τομέα και του πεδίου της οντολογίας

Κατά το αρχικό στάδιο της ανάπτυξης της οντολογίας μας καθορίσαμε τον τομέα και το πεδίο της οντολογίας. Ο καθορισμός αυτός προέκυψε απαντώντας διάφορες βασικές ερωτήσεις όπως:

- Ποιος είναι ο τομέας που θα καλύψει η οντολογία;
- Που θα χρησιμοποιηθεί η οντολογία;
- Για ποιους τύπους ερωτήσεων θα πρέπει οι πληροφορίες στην οντολογία να δώσουν τις απαντήσεις;
- Ποιος θα χρησιμοποιήσει και θα διατηρήσει την οντολογία;

Ο τομέας που θα καλύψει η οντολογία που αναπτύξαμε είναι ένα *περιβάλλον φωτογραφικής μηχανής*. Το πεδίο γνώσης της οντολογίας είναι η αναπαράσταση όλου του

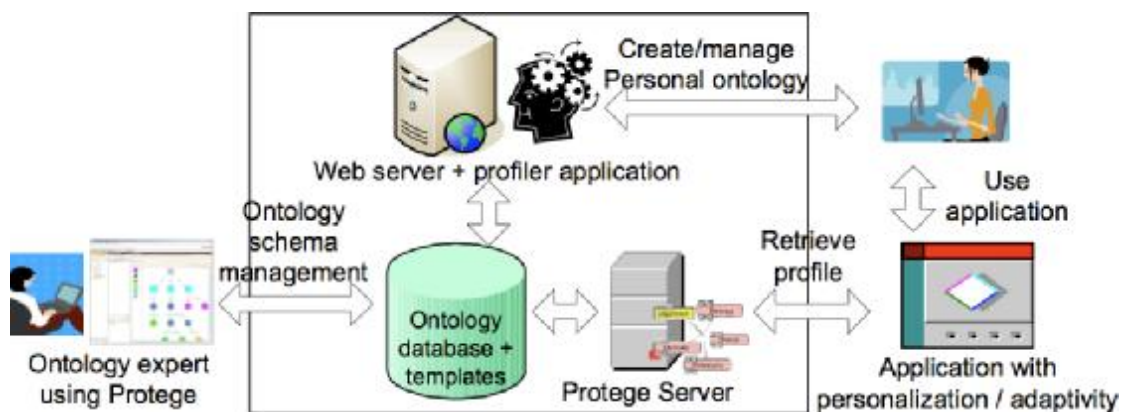
εξοπλισμού που απαιτείται για την δημιουργία μιάς σύγχρονης επαγγελματικής ή απλής χρήσης φωτογραφικής μηχανής.

Ο σκοπός αυτής της οντολογίας θα μπορούσε να είναι η χρήση της σε ένα ηλεκτρονικό σύστημα οδηγού αγορών μιας επιχείρησης που δραστηριοποιείται στο χώρο της τεχνολογίας και παρέχει ολοκληρωμένες λύσεις στους πελάτες της. Ο οδηγός αυτός θα πρέπει να παρέχει όλες τις πληροφορίες που χρειάζεται κάποιος που θέλει ένα είδος φωτογραφικής μηχανής. Ο ηλεκτρονικός οδηγός, που πιθανώς να έχει την μορφή ενός *wizard*, θα πρέπει να έχει τη δυνατότητα να προτείνει πιθανούς συνδυασμούς ανάλογα με τις ανάγκες του ενδιαφερόμενου αλλά και ταυτόχρονα να καθοδηγεί τον χρήστη σε σωστές επιλογές αποτρέποντας λάθη επιλογής, διενέξεις και στοιχεία ασυμβατότητας.

Η οντολογία μας θα περιέχει έννοιες και ορολογίες της πληροφορικής, φωτογραφικής μηχανής για τα προϊόντα που θα περιλαμβάνει. Δεν θα είναι εξειδικευμένη γνώση καθώς ο χρήστης που θα κληθεί να την χρησιμοποιήσει θα είναι ένας πελάτης που επιθυμεί να προμηθευτεί κάποια προτεινόμενη λύση. Συνεπώς θα περιορίζεται σε κοινές ορολογίες και σε προδιαγραφές που περιγράφουν ένα προϊόν όπως σε ένα οποιοδήποτε κατάστημα.

Η οντολογία μας θα έχει καθαρά συμβουλευτικό χαρακτήρα και θα περιέχει τιμές και λογιστικά δεδομένα καθώς η πιο πιθανή χρήση της θα μπορούσε να είναι η παροχή της από το internet ώστε κάποιος πελάτης, πριν επισκεφτεί το κατάστημα, θα μπορούσε να ετοιμάσει την λύση του, ώστε επισκεπτόμενος το κατάστημα να πάρει την οικονομική προσφορά, χωρίς να απαιτείται η ύπαρξη κάποιου ειδικού συμβούλου στο κατάστημα. Ως εκ τούτου ο ιδιοκτήτης του καταστήματος, με αρκετά λιγότερες ανάγκες σε προσωπικό θα μπορούσε να παρέχει τα προϊόντα και τις λύσεις του το ίδιο ποιοτικά και ίσως με μεγαλύτερη παραγωγικότητα.

Βέβαια θα πρέπει να τονίσουμε πως η οντολογία μας απαιτεί τη συνεχή συντήρηση και εισαγωγή δεδομένων ανάλογα με την εξέλιξη της τεχνολογίας και των προϊόντων που προσφέρει το κατάστημα.



Εικόνα 19: Overall system architecture

<http://alandix.com/academic/papers/PIM2008-personal-ontology/>

ΚΕΦΑΛΑΙΟ 6^ο

Συμπεράσματα και παρατηρήσεις

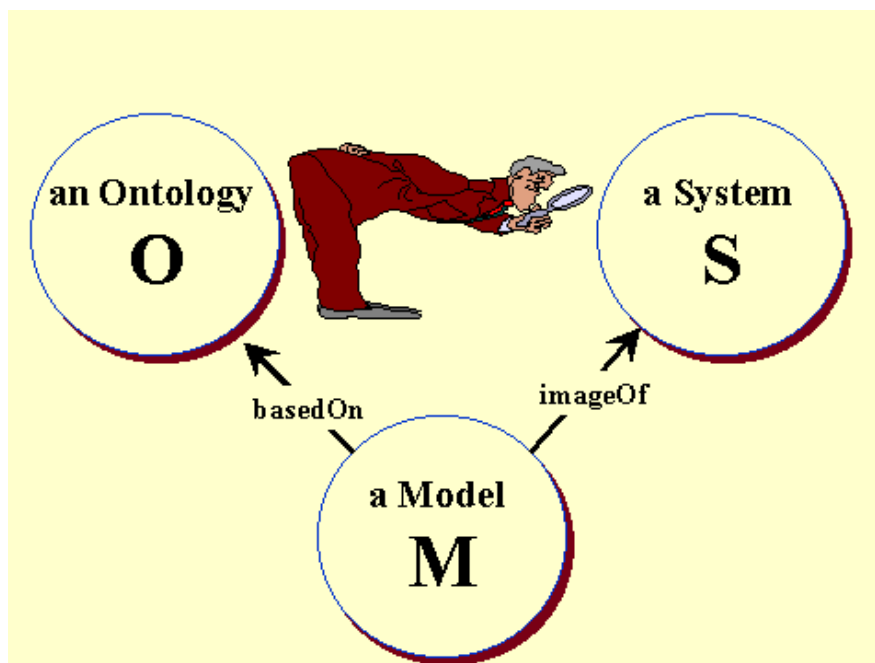
6.1 Επίλογος - Συμπεράσματα

Ο Σημασιολογικός Ιστός αποτελεί αδιαμφισβήτητα το νέο ορόσημο για την εξέλιξη των πληροφοριακών συστημάτων.

Συμπερασματικά, θα μπορούσαμε να πούμε ότι το περιβάλλον του θα μπορούσε να αποτελέσει μελλοντικά ένα ιδανικό περιβάλλον για ανταλλαγή στοχευόμενων δεδομένων. Σημαντικό ρόλο, στον Σημασιολογικό Ιστό μπορούν να διαδραματίσουν οι έξυπνοι πράκτορες, οι οποίοι είναι ικανοί να μεταβάλλουν τον τρόπο πλοήγησης, ανεύρεσης και συλλογής πληροφοριών των χρηστών, γιατί πλέον, θα είναι σε θέση να κατανοήσουν και να αξιολογήσουν την πληροφορία με τη βοήθεια οντολογιών. Επίσης, οι πράκτορες θα μπορούν να επικοινωνούν και να συνεργάζονται καλύτερα, διότι ο Σημασιολογικός Ιστός θα είναι στην ουσία μια συλλογή από υπηρεσίες, οι οποίες θα περιγράφονται από οντολογίες.

Η συνεχής και αυξανόμενη χρήση του Σημασιολογικού Ιστού μπορεί να δημιουργήσει ένα κοινά αποδεκτό δίκτυο, όπου οι καλές πρακτικές θα ανευρίσκονται και θα επαναχρησιμοποιούνται εύκολα και γρήγορα. Αξίζει εδώ να αναφερθεί ότι η δημιουργία μιας αποθήκης κοινά αποδεκτών εννοιών, αποτελεί πρόκληση για την επίτευξη εύκολης επαναχρησιμοποίησης, καθώς η δημιουργία μιας παγκοσμίως αποδεκτής γλώσσας, είναι δύσκολο να επιτευχθεί.

Πολλοί είναι αυτοί που υποστηρίζουν πως ο Σημασιολογικός Ιστός είναι το επόμενο βήμα ανάπτυξης του Διαδικτύου, στο οποίο θα προσδώσει ένα πλήθος χαρακτηριστικών που ακόμη απουσιάζουν, όπως η δυνατότητα αναζήτησης με σημασιολογικά κριτήρια, η δυνατότητα εύκολης επικοινωνίας μεταξύ των εφαρμογών και η ανταλλαγή δεδομένων μεταξύ των εφαρμογών χωρίς τη διαμεσολάβηση του χρήστη.



Εικόνα 20: An ontology as a filter

http://www.flaterco.com/purls/oopsla98_bezivin/

Πράγματι ο Σημασιολογικός Ιστός παρέχει λύσεις για όλα τα προαναφερόμενα, παρόλα αυτά χρειάζεται να λυθούν αρκετά προβλήματα που υπάρχουν, όπως η διαχείριση και προβολή μεγάλων οντολογιών, η δυνατότητα χρήσης των υπάρχουσών Βάσεων Δεδομένων και η δημιουργία απλών και σταθερών εργαλείων που να αξιοποιούν τις δυνατότητές του. Για αυτόν το λόγο το ενδιαφέρον της ακαδημαϊκής κοινότητας είναι μεγάλο. Καθημερινά γίνονται συζητήσεις για βελτίωση των ήδη υπάρχοντων προτύπων και εργαλείων, χαρακτηριστικό παράδειγμα αποτελεί πως τον έναν σχεδόν χρόνο ενασχόλησης του συγγραφέα με τον Σημασιολογικό Ιστό, έχουν παρουσιαστεί καινούργιες εκδόσεις για το σύνολο σχεδόν των εργαλείων που μελέτησε.

Τέλος, χρησιμοποιώντας το λογισμικό PROTÉGÉ 2004 αναπτύξαμε μια δική μας οντολογία που αφορά στο πεδίο μιας φωτογραφικής μηχανής. Η οντολογία που αναπτύξαμε βασίστηκε σε έτοιμη οντολογία που ανακτήσαμε από το διαδίκτυο και περιείχε σημαντικό πλήθος όρων που θα χρησιμοποιούσαμε στην δική μας οντολογία. Βασιζόμενη λοιπόν στην αρχική οντολογία, προβήκαμε στις απαραίτητες προσθήκες, διαγραφές και μεταποιήσεις προκειμένου η τελική οντολογία να ανταποκρίνεται στις απαιτήσεις μας και στις ερωτήσεις επάρκειας που θέσαμε για την εφαρμογή μας κατά την εκκίνηση της διαδικασίας.

Εν κατακλείδι, η αξιοποίηση του Σημασιολογικού Ιστού στη μάθηση και τη διδασκαλία θα επιφέρει μια επαναστατική εποχή στο χώρο της εκπαίδευσης, αφού η αλληλεπίδραση ενός χρήστη με τον υπολογιστή, θα επιτυγχάνεται στα πλαίσια της τεχνητής νοημοσύνης και θα επιτρέπει ουσιαστική και εννοιολογική επικοινωνία μεταξύ τους.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Aktas M. Nacar N. and Menczer F. Personalizing PageRank based on domain profiles. In Proceedings of the KDD Workshop on Web Mining and Web Usage, pp. 83-90, 2004
2. Androutsopoulos I., Galanis D. 2005. A practically unsupervised learning method to identify single-snippet answers to definition questions on the web. In the HLT/EMNLP Conference, pp. 323-330.
3. Anick, P. (2004). Using terminological feedback for Web search refinement: a log-based study. In Proceedings of WWW '04, 89-95.
4. Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A. & Raghavan, S. (2001), 'Searching the web', ACM Transactions on Internet Technology (TOIT) 1(1), 2{43.
5. Armstrong R., Freitag D., Joachims T., Mitchell T. 1995. WebWatcher: A Learning Apprentice For The World Wide Web. In Proceedings of the AAAI Spring Symposium On Information Gathering, pp. 6-12
6. Baeza-Yates R. A. and Ribeiro-Neto. Modern Information Retrieval. Addison-Wesley Longman Publishing Co., Inc., 1999
7. Beitzel S. Understanding and Classifying Web Queries. Ph.D. Dissertation, Illinois Institute of Technology May 2006
8. Belkin N.J., Cool C., Head J., Jeng J., Kelly D., Lin S., Lobash L., Park S.Y., Savage-Knepshield P. and Sikora C. Relevance Feedback versus Local Context Analysis as Term Suggestion Devices: Rutgers' TREC-8 Interactive Track Experience. TREC 2000
9. Berendt B., Using Site Semantics to Analyze, Visualize and Support Navigation, in Data Mining and Knowledge Discovery Journal (2002), 6: 37-39.
10. Bharat K., T. Kamba, and M. Albers. Personalized Interactive News on the Web. Multimedia Systems, 6(5): 349-358,1998
11. Borko, H. and Bernier, L. (1975). Abstracting Concepts and Methods. San Diego, CA: Academic Press.
12. Brin S., Page L. 1998 "The Anatomy of a Large-Scale Hypertextual Web Search Engine" In Proceedings of the 7th International World Wide Web Conference, Australia
13. Bruza P., McArthur R., Dennis S. 2000 "Interactive Internet Search: Keyword, Directory and Query Reformulation Mechanisms Compared" In Proceedings of the 23rd Annual ACM SIGIR Conference, Athens, Greece
14. Buchholz S. 2002. Using grammatical relations, answer frequencies and the World Wide Web for TREC question answering. In the 10th TREC Conference.

15. Efthimiadis E. 2000 “ Interactive Query Expansion: a User-Based Evaluation in a Relevance Feedback Environment” In Journal of the American Society for Information Science
16. Eirinaki M., Vazirgiannis M., M.SEWeP: Using Site Semantics and a Taxonomy to Enhance the Web Personalization Process, in Proceedings of the 9th ACM International Conference of Knowledge Discovery and Data Mining (SIGKDD’03), Washington DC, August 2003.
17. Einat Amitay, Cecile Paris, “Automatically Summarizing Web Sites - Is There A Way Around It?”
18. Ferragina, P., Gulli, A. 2005. A personalized search engine based on web-snippet hierarchical clustering. In Special Interest Tracks & Posters, 14th Intl. WWW Conference.
19. Field A., P. Hartel, and W. Mooij. Personal DJ, an Architecture for Personalized Content Delivery. In Proc. Of the 10th International World Wide Web Conference (WWW10), pages 1-8, 2001
20. Foltz P., Kintsch W., Landauer K. 1998. Textual coherence using latent semantic analysis. In Discourse Processes, 25(2&3): 285-307.
21. Gauch, S., Chafee, J. and Pretschner, A. (2004). Ontologybased personalized search and browsing. Web Intelligence and Agent Systems, 1(3-4): 219-234.
22. Gliozzo, A., Strapparava, C., Dagan, I., 2004. Unsuper-vised and Supervised Exploitation of Semantic Do-mains in Lexical Disambiguation. In Computer Speech and Language, 18(3) pp. 275-299
23. Gong Y.H and Liu X. “Generic Text Summarization using Relevance Measure and Latent Semantic Analysis”. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, Louisiana, USA, 2001, pp. 19-25.
24. Google patent 2003. Detecting query-specific duplicate documents. US patent No. 6615209.
25. Google Personal: <http://labs.google.com/personalized>
26. Grefenstette G. 1992 “Use of Syntactic Context to Produce Term Association Lists for text Retrieval” In Proceedings of the 15th Annual ACM SIGIR Conference, pp. 89-97
27. Grosz B., Joshi A., Weinstein S. 1995. Centering: A framework for modeling the local coherence of discourse. In Computational Linguistics, 21(2): 203-225.
28. Hahn, U. and Mani, I. (2000). The Challenges of Automatic Summarization. IEEE Computer, 33(11), 29-36.
29. Harman D. 1991 “How Effective is suffixing?” In Journal of the American Society for Information Science, Vol. 42, No. 1, pp. 7-15

30. Harman D. Towards Interactive Query Expansion. In Proceedings of the SIGIR Conference, pp. 321-331, 1988
31. Haveliwala T. H. Topic-Sensitive PageRank. In Proc. of the 11th International World Wide Web Conference (WWW2002), pages 517–526, 2002.
32. Higgins D., Burstein J. Marcu D., Gentile C. 2004. Evaluating multiple aspects of coherence in student essays. In Proceedings of the NAACL Conference, pp. 185-192.
33. Hill W., L. Stead, M. Rosenstein, and G.W. Furnas. Recommending and Evaluating Choice in a Virtual Community of Use, in Proc. Of the 22nd Annual International ACM SIGIR Conference on Research and Development on Human Factors in Computing Systems (CHI '95), pages 194-139, 2001
34. Hjelsvold R., S. Vdaygiri, and Y. Leaute. Web-Based Personalization and Management of Interactive Video in Proc. Of the 10th International World Wide Web Conference (WWW10), pages 129-139, 2001
35. Hovy E., Hermjakob U., Lin C.Y. 2001. The user of external knowledge on factoid QA. In the 10th TREC Conference.
36. Hull D. 1996 “Stemming Algorithms: A Case Study for Detailed Evaluation” Journal of the American Society for Infromation Science, 47 (1): 70-84
37. Ikeji A. Fotouhi F. 1999 “An Adaptive Real-Time Web Search Engine” In Proceedings of WIDM '99 Kansas City, USA
38. Ittycheriah A., Franz M., Roukos S. 2001. IBM’s statistical question answering system-TREC10. 10th TREC Conf.
39. J. Goecks and J. Shavlick. Learning users' interests by unobtrusively observing their normal behavior. In Proceedings of the IJCAI Workshop on Machine Learning for Information Filtering, 1999.
40. Jansen B.J., Spink A. Saracevic T. 2000. Real life, real users, and real needs: A study and analysis of user queries on the Web. Information Processing & Management, 36(2): 207-227.
41. Jarvelin K. Kristensen J., Niemi T., Sormunen E., Keskustalo H. 1996 “A Deductive Data Model for Query Expansion” In Proceedings of the 19th Annual ACM SIGIR Conference, New York
42. Jeh G., Widom J. Scaling personalized web search. In Proceedings of the 12th Intl. World Wide Web Conference, pp. 271-279, 2003
43. Jing H., Tzoukermann E. 1999 “Information Retrieval Based on Context Distance and Morphology” In Proceedings of the 22nd Annual ACM SIGIR Conference, Berkley, USA, pp. 90-96

44. Jing J. and W.B Croft. An association theaurus for information retrieval. In Proc. RIAO'94, pages 146-160,1994
45. Joachims T., Freitag D., Mitchell T. August 1997. WebWatcher: A Tour Guide for the World Wide Web. In Proceedings of IJCAI'97
46. Jones S., Gatford S., Robertson M., Hancock B., Secker J. 1995 "Interactive Thesaurus Navigation: Intelligence Rules ok?" Journal of the American Society for Information Science 46(1): 52-59
47. Joseph A. Konstan , Bradley N. Miller , David Maltz , Jonathan L. Herlocker , Lee R. Gordon , John Riedl, GroupLens: applying collaborative filtering to Usenet news, Communications of the ACM, v.40 n.3, p.77-87, March 1997
48. Kelly D. and Teevan J. Implicit feedback for inferring user preference: a bibliography. In SIGIR Forum 32(2), pp. 18-28, 2003
49. Kilgarriff A. 1992 Polysemy Ph.D Thesis, University of Sussex
50. Koenamann J. and Belkin N. A case for interaction: A study of interactive information retrieval behavior and effectiveness. In Proceedings of CHI Conference, pp. 205-212, 1996
51. Koutrika G. and Ioannidis Y. Personalized queries under a generalized preference model. In Proceedings of the ICDE Conference, 2005
52. Kozanidis L., Tzekou P., Zotos N. Stamou S., Christodoulakis D. Ontology-based adaptive query refinement. In Proceedings of the 3rd WebIST Conference, 2007.
53. Krishna Bharat and George A. Mihaila. When experts agree: Using non-a_liated experts to rank popular topics. In Proceedings of the Tenth International World Wide Web Conference, 2001
54. Krishnan V. and S.G. Chang. Customized Internet Radio. In Proc. Of the 9th International World Wide Web Conference (WWW9), pages 609-618, 2000
55. Kristensen J. 1993 "Expanding End-Users' Query Statements for Free Text Searching with a Search-Aid thesaurus" In Information Processing and Management, 29(6): 733-744
56. Kritikopoulos A., Sideri M., Varlamis I. 2007. Success Index: Measuring the efficiency of search engines using implicit user feedback. In the 11th Pan-Hellenic Conference on Informatics, Special Session on Web Search and Mining.
57. Krovetz R. 1993 "Viewing Morphology as an Inference Process" In Proceedings of the 16th Annual ACM SIGIR Conference, Pittsburgh, USA, pp. 40-48
58. Krovetz R. 1997 "Homonymy and Polysemy in Infromation Rerieval" In Proceedings of the ACL/EACL Conference

59. Krovetz R. Croft B. 1992 "Lexical Ambiguity and Information Retrieval" In ACM Transactions on Information Systems, Vol. 10(2), pp. 115-141
60. Kurki T., Jokela S., Sulonen R. and Turpeinen M. 1999. Agents in Delivering Personalized Content Based on Semantic Metadata. In Proceedings of the 1999 AAAI Spring Symposium Workshop on Intelligent Agents in Cyberspace, pp. 84-93.
61. Kwok C., Etzioni O., Weld D. 2001. Scaling question answering to the web. In the 10th WWW Conference, pp. 150-161.
62. Lancaster F., Fayen E. 1973 Information Retrieval On-Line Melville Publishing Co. Chapter 6, Los Angeles, C.A
63. Lapata M., Barzilay R. 2005. Automatic evaluation of text coherence: models and representations. In the Intl. Joint Conferences on Artificial Intelligence.
64. Lawrence S. and C. L. Giles. "Accessibility of information on the web". Nature, 400: 107-109, 1999.
65. Lee G.G., Seo J., Lee S., Jung H., Cho B.H., Lee C., Kwak B.K., Cha J., Kim D., An J., Kim H., Kim K. 2001. SiteQ: Engineering high performance QA system
66. using lexico-semantic pattern matching and shallow NLP. In the 10th TREC Conference.
67. Lieberman H. 1995. Letizia: An Agent That Assists Web Browsing. In Proceedings of the 14th International Joint Conference On Artificial Intelligence, pp. 924-929.
68. Lieberman H. May 1997. Autonomous Interface Agents. In Proceedings of the ACM Conference on Computers and Human Interaction (CHI'97).
69. Light M., Mann G.S., Riloff E., Breck E. 2001. Analyses for elucidating current question answering technology. Journal of Natural Language Engineering, Special Issue in Question Answering.
70. Lin D., Pantel P. 2001. Discovery of inference rules for question answering. Natural Language Engineering 7(4): 343-360.
71. Liu F., Yu C., and Meng W. Personalized web search by mapping user queries to categories. In Proceedings of the CIKM Conference, pp. 558-565, 2002
72. Lovins J. 1968 "Development of a Stemming Algorithm" Mechanical Translation and Computational Linguistics, Vol. 11, pp. 22-31
73. Luhn H.P. "The Automatic Creation of Literature Abstracts" In IBM Journal of Research and Development, Vol. 2, No.2, April 1968, pp. 159-165.
74. Lyons J. 1977. Semantics, Vol. 1 & 2. Cambridge University Press

75. Magdalini Eirinaki and Michalis Vazirgiannis, "Web Mining for Web Personalization". In ACM Transactions on Internet Technology, Vol. 3, No. 1, February 2003, Pages 1-27.
76. Manber U., A. Patel, and J. Robinson. Experience with Personalization on Yahoo! Communications of the ACM, 43(8): 35-39, 2000
77. Mandala R., Tokunaga T., Tanaka H. 1999 "Combining Multiple Evidence form Different Types of Thesaurus for Query Expansion" In Proceedings of the 22nd Annual ACM SIGIR Conference, Berkley, USA, pp. 191-197
78. Mani, I. (2001). Automatic Summarization. Amsterdam: John Benjamin's Publishing Company
79. Mark Claypool , David Brown , Phong Le , Makoto Waseda, Inferring User Interest, IEEE Internet Computing, v.5 n.6, p.32-39, November 2001
80. Masahiro Morita , Yoichi Shinoda, Information filtering based on user behavior analysis and best match text retrieval, Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, p.272-281, July 03-06, 1994, Dublin, Ireland
81. Merialdo B., K. T. Lee, D. Luparello, and J. Roudaire. Automatic Construction of Personalized TV News Programs. In Proc. Of the 7th ACM International Conference on Multimedia (Multimedia '99), pages 323-331, 1999
82. Miller G. A., Charles W. G. 1991 "Contextual Correlates of Semantic Similarity" In Language and Cognitive Processes, Vol. 6, pp. 1-28
83. Mishne G., de Rijke M., Jijkoun V. 2005. Using a reference corpus as a user model for focused information retrieval, In Journal of Digital Information Management, 3(1): 47-52.
84. Mladenic D. 1998. Personal WebWatcher: Design and Implementation. ! Technical Report IJS-DP-7472, J. Stefan Institute, Department for Intelligent Systems, Ljubljana, Slovenia
85. Montebello M., Gray W. and Hurley S. July 1998. A Personable Evolvable Advisor for WWW Knowledge-Based Systems. In Proceedings of the 1998 International Database Engineering and Application Symposium (IDEAS'98), pp. 224-233.
86. Morita, M. and Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. In *Proceedings of SIGIR '94*, 272-281.
87. Prager J., Chu-Carroll J., Czuba K. 2001. Use of WordNet hypernyms for answering what-is questions. In Proceedings of the TREC-2002 Conference, NIST.
88. Prager J.M., Radev D.R., Czuba K. 2001. Answering what-is questions by virtual annotation. In Proceedings of Human Language Technologies Conference, pp. 26-30.

89. Pretschner A., Gauch S. Ontology-based personalized search. In Proceedings of the 11th IEEE Intl. Conference on Tools with Artificial Intelligence, pp. 391-398, 1999
90. Qiu and H.P. Frei 1993 "Concept Based Query Expansion" In Proceedings of the 16th Annual ACM SIGIR Conference, pp. 160-169
91. Proceedings of the 15th Intl. World Wide Web Conference, pp. 727-236, 2006
92. Retrieval Systems" In SIGIR Forum, Vol. 15, No. 3, pp. 6-36
93. Robertson S. E., K. Sparck Jones. Relevance weighting of search terms. Journal of the American Society for Information Sciences, 27(3): 129-146, 1976
94. Robertson S.E., Walker S., Hancock-Beaulieu M., Gatford M., Payne A. 1995. Okapi at TREC-4. 4th TREC Conf.
95. Rucker J. and Polanco M. March 1997. Siter: Personalized Navigation For The Web. Communications of the ACM, 40(3), pp. 73-75.
96. Salton G. The SMART Retrieval System – Experiments in Automatic Document Processing. Prentice Hall Inc., 1971
97. Salton G., C. Buckley. Term-weighting approaches in automatic retrieval. Information Processing & Management, 24(5): 513-523, 1988
98. Salton G., M. E. Lesk. Computer evaluation of indexing and text processing. Journal of the ACM, 15(1): 8-36, January 1968
99. Salton G., McGill M. 1983 Introduction to Modern Information Retrieval Singapore: McGraw-Hill
100. Shavlik J. and Eliassi-Rad T. 1998. Intelligent Agents for Web-Based Tasks: An Advice-Taking Approach. In Working Notes of the AAAI/ICML-98 Workshop on Learning for text categorization, Madison, WI.
101. Shavlik J., Calcari S., Eliassi-Rad T., and Solock J. 1999. An Instructable, Adaptive Interface for Discovering and Monitoring Information on the World Wide Web. In Proceedings of the 1999 International Conference on Intelligent User Interfaces, Redondo Beach, CA
102. Shen, X. and Zhai, C. X. (2003). Exploiting query history for document ranking in interactive information retrieval. In *Proceedings of SIGIR '03 (Poster)*, 377-378.
103. Si L. and Callan P. "Relevant Document Distribution for Resource Selection". In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003, pp. 298-305.

104. Smyth B. and P. Cotter. A Personalized Television Listings Service. *Communications of the ACM*, 43(8): 107-111,2000.
105. Song Y.I., Han K.S., Rim H.C. 2004. A term weighting method based on lexical chain for automatic summarization. In the 5th CICLing Conference, pp. 636-639.
106. Sorensen H. and McElligott M. December 1995. PSUN: A Profiling System for Usenet News. In *Proceedings of CIKM'95 Workshop on Intelligent Information Agents*.
107. Teevan J., Dumais S., Horvitz E. Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th Intl. Conference on Research and Development in Information Retrieval*, pp. 449-456, 2005.
108. Teevan, J., Alvarado, C., Ackerman, M. S. and Karger, D. R. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of CHI '04*, 415-422.
109. Tellex S., Katz B., Lin J., Fernandes A., Marton G. 2003. Quantitative evaluation of passage retrieval algorithms for question answering. In the 26th SIGIR Conference.
110. Thorsten Joachims, Optimizing search engines using clickthrough data, *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, July 23-26, 2002, Edmonton, Alberta, Canada
111. Tomlin J.A. A new paradigm for ranking pages on the Web. In *Proceedings of the 12th WWW Conference 2003*, 350-355
112. Tsandilas T. and M.C. Schraefel. User-Controlled Link Adaptation . In *Proc. Of the 14th ACM Conference on Hypertext and Hypermedia (HT '03)*, pages 152-160, 2003
113. Van Hage W.R, de Rijke M., Marx M. 2004. Information retrieval support for ontology construction and use. In *Intl. Semantic Web Conference*, pp. 518-533.
114. Van Rijsbergen, C.J. *Information retrieval*, 2nd Edition. London, 1979: Butterworths. ISBN 0408709294.
115. Vaswani P., Cameron, J., *The NPL Experiments in Statistical Word Associations and their Use in Document Indexing and Retrieval*, No. 42. National Physical Laboratory, 1970.
116. Vicedo J.L., Ferrandez A. 2001. University of Alicante at TREC-10. In the 10th Text Retrieval Conference.
117. Voorhees E. 1994 "Query Expansion Using Lexical-Semantic Relations" In *Proceedings of the 17th Annual ACM SIGIR Conference*, New York, USA, pp. 61-69
118. Voorhees E. 1998 "Variations in Relevance Judgements and the Measurement of Retrieval Effectiveness" In *Proceedings of the 21st Annual ACM SIGIR Conference*, Melbourne, Australia

119. Voorhees E. 2003. Evaluating answers to definition questions. In Proceedings of the HLT-NAACL Conference.
120. Voorhees E. 2005. Using question series to evaluate question answering system effectiveness. HLT/EMNLP Conf.
121. Voorhees E., Harman D. 1999 “Overview of the seventh Text REtrieval Conference (TREC-7)” In Proceedings of the 7th TREC Conference, NIST Special Publication
122. Wesley W. Chu, Victor Zhenyu Liu, Wenlei Mao Techniques for Textual Document Indexing and Retrieavl via Knowledge Sources and Data Mining in INFORMATION RETRIEVAL AND CLUSTERING W.Wu Xiong andS.shekhar(Eds.)2002 Kluwer Academic Publishers
123. Widiantoro D., Ioerger T. and Yen J. 2001. Learning User Interest Dynamics with a Three-Descriptor Representation. Journal of the American Society of Information Science and Technology (JASIST) , Vol 52, No. 3, pp. 212-225.
124. Wilks Y., Stevenson M. 1977 “The Grammar of Sense: Using Part-of-Speech tags as a first step in Semantic Disambiguation”, Natural Languge Engineering, 1 (1)
125. Witten I., Moffat A., Bell T. 1994 Managing Gigabytes: Compressing and Indexing Documents and Images New York: Van Nostrand Reinhold
126. Yu C., Meng W., Wu W. and Liu K. Efficient and effective metasearch for text databases incorporating linkages among documents. In Proceedings of the ACM SIGMOD Conference, 2001.
127. Yuefeng Li, Ning Zhong “Web mining model and its applications for information gathering”
128. Zamir O., Etzioni O. 1998. Web document clustering: a feasibility demonstration. In the SIGIR Conference.
129. Zhu X., Gauch S. 2000 “Incorporating Quality metrics in Centralized-Distributed Information retrieval on the World Wide Web” In Proceedings of the 23rd Annual ACM SIGIR Conference, Athens, Greece
130. Zotos N., Tzekou V., Tsatsaronis G., Kozanidis L., Stamou S. and Varlamis I. To click or not to click? The role of contextualized and user-centric web snippets. In Proceedings of the SIGIR workshop on Focused Retrieval, 2007

ΙΣΤΟΣΕΛΙΔΕΣ - ΑΝΑΦΟΡΕΣ ΣΕ ΔΗΜΟΣΙΕΥΣΕΙΣ ΣΤΟΝ ΙΣΤΟΧΩΡΟ

1. “Scientific American: The Semantic Web”, By Tim Berners-Lee, James Hendler and Ora Lassila, <http://semanticcommunity.info/@api/deki/files/4686/=10.1.1.115.9584.pdf>
2. “Agents and the Semantic Web”, James Hendler, <http://hcs.science.uva.nl/Capita-AI/2002/papers/hendler.pdf>
3. Semantic Web Services, Sheila A. McIlraith, Tran Cao Son, and Honglei Zeng, <http://o2k.stanford.edu/people/sam/ieee01.pdf>
4. "NCSA Mosaic — September 10, 1993 Demo". Totic.org. <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
5. “Learning to match ontologies on the SemanticWeb”, AnHai Doan, Jayant Madhavan, Robin Dhamankar, Pedro Domingos, Alon Halevy, published (17/9/2003), <http://osm7.cs.byu.edu/CS652s04/DMD%2B03Learning.pdf>
6. “XML, RDF and the Relatives”, by Michel Klein, <http://ebiz.u-aizu.ac.jp/~paikic/lecture/2005-2/CFS/SemanticWeb-1/RDFXML.pdf>
7. “OIL: Ontology Infrastructure to Enable the Semantic Web”, Dieter Fensel, Ian Horrocks, Frank van Harmelen, Deborah McGuinness, and Peter F. Patel-Schneider, <http://www.informatik.uni-leipzig.de/~brewka/papers/semweb/6.OILintro.pdf>
8. “Web Ontology Language: OWL”, Grigoris Antoniou¹ and Frank van Harmelen, <http://www.cs.vu.nl/~frank.van.harmelen/postscript/OntoHandbook03OWL.pdf>
9. “XOL: An XML-Based Ontology Exchange Language”, Peter D. Karp, Vinay K. Chaudhri and Jerome Thomere, <http://www.sri.com/sites/default/files/uploads/publications/pdf/676.pdf>
10. “SHOE: A Blueprint for the Semantic Web”, Je_ Hein, James Hendler, Sean Luke, February 21, 2003, <http://www.cse.lehigh.edu/~heflin/pubs/swbook03.pdf>
11. “Ontology Languages for the Semantic Web”, Asunción Gómez-Pérez and Oscar Corcho, <http://oa.upm.es/2646/1/JCR01.pdf>
12. “Conceptual Knowledge Markup Language: An Introduction”, Robert E. Kent, <http://148.226.12.104/bitstream/123456789/10416/1/Robert%20E.%20Kent%20-%20Conceptual.pdf>
13. “Ontologies and Sematic ”, <http://www.obitko.com/tutorials/ontologies-semantic-web/description-logics.html>
14. “Resource Description Framework (RDF): Concepts and Abstract Syntax”, W3C Recommendation, 10 February 2004, <http://travesia.mcu.es/portaln/jspui/bitstream/10421/2427/1/rdf-concepts%20and%20abstract%20syntax.pdf>

15. "The Semantic Web: The roles of XML and RDF", STEFAN DECKER AND SERGEY MELNIK, FRANK VAN HARMELEN, DIETER FENSEL, MICHEL KLEIN, JEEN BROEKSTRA, MICHAEL ERDMANN, IAN HORROCKS, http://classweb.gmu.edu/kersch/infs770/Semantic_Web_16_2/Semantic%20Web.pdf
16. "OIL in a Nutshell", D. Fensel, I. Horrocks, F. Van Harmelen, S. Decker, M. Erdmann, and M. Klein, <http://www.cs.ox.ac.uk/people/ian.horrocks/Publications/download/2000/oilnutshell.pdf>
17. "OIL: Ontology Infrastructure to Enable the Semantic Web", Dieter Fensel, Ian Horrocks, Frank van Harmelen, Deborah McGuinness, and Peter F. Patel-Schneider, <http://www.informatik.uni-leipzig.de/~brewka/papers/semweb/6.OILintro.pdf>
18. <http://ontolog.cim3.net/cgi-bin/wiki.pl/>
19. <http://www.cs.utexas.edu/users/mfkb/related.html#thesauri>
20. <ftp://ftp.geneontology.org/pub/go/ontology-archive/>
21. <http://www.evoontology.org/site/Main/EvocOntologies2p9?from=Main.DownloadEvocOntologies>
http://protegewiki.stanford.edu/index.php/Protege_Ontology_Library
22. WEB 2.0: Χαρακτηριστικά και επίδρασή του σε επιχειρήσεις, κεντρική διοίκηση και χρήστες, Παρατηρητήριο για την κοινωνία της πληροφορίας, Νοέμβριος
<http://www.observe.gr/files/meletes/WEB%202.0.pdf>
23. The World Wide Web Consortium, <http://www.w3.org/>
24. Protégé, <http://protege.stanford.edu/>
25. <http://www.slideshare.net/LeeFeigenbaum/intro-to-the-semantic-web-landscape-2011>
26. Moukas A. Amalthea: Information Discovery And Filtering Using A Multiagent Evolving Ecosystem. In Proceedings of the Conference on the Practical Application of Intelligent Agents and MultiAgent Technology, 1996. <http://moux.www.media.mit.edu/people/moux/papers/PAAM96>
27. Page L., Brin S., Motwani R. and Winograd T. The pagerank citation ranking: Bringing order to the web. Technical Report, Stanford University Database Group, 1998. Available at: <http://dbpubs.stanford.edu:8090/pub/1999-66>
28. Pazzani M., Muramatsu J., Billsus D. 1996. Syskill & Webert: Identifying Interesting Web Sites. In Proceedings of the 13th National Conference On Artificial Intelligence. pp. 54-61
29. Pedersen T., Banerjee S., Patwardhan S. Maximizing semantic relatedness to perform word sense disambiguation. Available at: <http://www.d.umn.edu/~tpederse/Pubs/max-sem-relate.pdf>
30. Speretta, M. and Gauch, S. (2004). Personalizing search based on user search history. Submitted to CIKM '04. <http://www.ittc.ku.edu/keyconcept/>

31. WordNet. available at: <http://www.cogsci.princeton.edu/~wn>

32. Wu X., Palmer M. 1994. Web semantics and lexical selection. In the 32nd ACL Meeting.

33. Xu J. Croft B. 1996 “Query Expansion Using Local and Global Document Analysis” In Proceedings of the 19th Annual ACM SIGIR Conference, Zurich, Switzerland

