



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**Σχολή Εφαρμοσμένων Μαθηματικών
και Φυσικών Επιστημών**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**“Διερεύνηση μεθόδων τεχνικών και τεχνολογιών του
Σημασιολογικού Ιστού”**



Αθήνα, Μάρτιος 2012

Μαρία-Σοφία Λιανού

Επιβλέπων Καθηγητής

Πέτρος Στεφανέας

Πίνακας περιεχομένων

Πίνακας περιεχομένων	ii
Κατάλογος Εικόνων	vi
Κατάλογος Πινάκων.....	viii
0. Περίληψη.....	1
1. Εισαγωγή.....	5
1.1. Ιστορικά στοιχεία	5
1.2. Σημασιολογικός Ιστός	12
1.3. Web 1.0, Web 2.0 και Web 3.0	15
1.4. Τεχνολογίες του Σημασιολογικού Ιστού	17
1.5. Αρχιτεκτονική του Σημασιολογικού Ιστού	20
1.5.1. 1ο Επίπεδο – Κωδικοποίηση και Διευθυνσιοδότηση Εγγράφων	22
1.5.2. 2ο Επίπεδο - Αναπαράσταση Περιεχομένου Εγγράφων	22
1.5.3. 3ο Επίπεδο - Αναπαράσταση νοήματος	23
1.5.4. 4ο Επίπεδο - Καθορισμός λεξιλογίου.....	24
1.5.5. 5ο Επίπεδο - Λογική.....	24
1.5.6. 6ο Επίπεδο - Αποδείξεις.....	25
1.5.7. 7ο Επίπεδο - Έλεγχος Αξιοπιστίας – Ψηφιακές υπογραφές.....	25
1.5.8. Διεπαφή Χρηστών	26
1.6. Νεότερες Τεχνολογίες στη στοίβα του Σημασιολογικού Ιστού	26
1.6.1. SPARQL.....	26
1.6.2. Rif.....	27
1.6.3. SWRL	28
1.6.4. Προγράμματα Περιγραφικής Λογικής	29
1.7. Εναλλακτική Στοίβα Σημασιολογικού Ιστού	30
1.8. Στόχοι Σημασιολογικού Ιστού.....	33
2. Αναπαράσταση γνώσης.....	36
2.1. Δημοφιλέστερες μέθοδοι Αναπαράστασης Γνώσης.....	37
2.1.1. Λογική	38
2.1.2. Δομημένες αναπαραστάσεις γνώσης.....	39
2.1.3. Κανόνες (if-then rules)	41
2.1.4. Διαδικαστική Αναπαράσταση	43
3. Η Λογική για το Σημασιολογικό Ιστό	44
3.1. Λογική	44

3.1.1.	Προτασιακή Λογική	45
3.1.2.	Λογική Πρώτης Τάξης	47
3.1.3.	Περιγραφικές Λογικές	50
3.2.	Οντολογίες	52
3.2.1.	OWL 1	52
3.2.2.	Από την OWL 1 στην OWL 2	57
3.2.3.	OWL 2	58
3.2.4.	Πιθανή μελλοντική ανάπτυξη της OWL	61
3.3.	RuleML	62
3.3.1.	Κανόνες	62
3.3.2.	RuleML	64
3.3.3.	Πρότυπα σήμανσης και πρωτοβουλίες σχετικά με τη RuleML	67
3.3.3.1.	Mathematical Markup Language (MathML)	68
3.3.3.2.	DARPA Agent Markup Language (DAML)	68
3.3.3.3.	Predictive Model Markup Language (PMML)	69
3.3.3.4.	Attribute Grammars in XML (AG-markup)	70
3.3.3.5.	Extensible Stylesheet Language Transformations (XSLT)	71
3.3.4.	Το Σχέδιο της RuleML	72
4.	Εφαρμογές	75
4.1.	DBpedia	75
4.1.1.	DBpedia Live	77
4.1.2.	Ελληνική DBpedia	79
4.2.	Linking Open Data	80
4.3.	FOAF	85
4.3.1.	FOAF-a-matic	88
4.4.	SIOC	89
4.4.1.	Υλοποιήσεις και Εφαρμογές	94
4.5.	NextBio	95
4.6.	Cyc	97
4.6.1.	Ο Cyc Knowledge Server	98
4.6.2.	OpenCyc	102
4.7.	Μεταδεδομένα σε Ψηφιακές Βιβλιοθήκες	105
4.7.1.	Πρωτόκολλο Μεταδεδομένων Dublin Core	107
4.8.	Σημασιολογικό Δίκτυο Αισθητήρων	108
4.8.1.	Sensor Wiki	109

4.8.2.	Έργο ΠΛΑΤΩΝΑΣ	109
4.9.	Μεταδεδομένα σε Video	111
4.9.1.	Αρχειοθετημένα λεξιλόγια για μεταδεδομένα σε Video	113
4.9.2.	Προτεινόμενα βασικά στοιχεία	115
4.10.	Inspire.....	116
4.10.1.	Inspire Geoportal	119
4.10.1.1.	Discovery / Viewer.....	119
4.10.1.2.	Validator	120
4.10.1.3.	Metadata Editor	120
4.10.2.	GIMED: Ελληνικός Metadata Editor	120
5.	Κατασκευή οντολογίας με τη βοήθεια βάσης δεδομένων.....	122
5.1.	Protégé.....	122
5.2.	Υλοποίηση οντολογίας.....	122
5.3.	Συμπεράσματα.....	128
6.	Παράρτημα	131
6.1.	Πολυπλοκότητα των προφίλ της OWL 2	131
6.2.	Στοιχεία μεταδεδομένων Dublin Core.....	134
6.3.	Τροποποιημένο σχήμα βάσης Εργαλείου semfe.gr.....	135
	Βιβλιογραφία.....	142

Κατάλογος Εικόνων

Εικόνα 1: Tim Berners Lee	5
Εικόνα 2: Ο πρώτος Web Server	8
Εικόνα 3: Μιχάλης Δερτούζος (1935-2001).....	10
Εικόνα 4: Ιστορικό λογότυπο του www σχεδιασμένο από τον Ρόμπερτ Κάιλιου	13
Εικόνα 5: Συνοπτική περιγραφή Web 1.0, 2.0 και 3.0	15
Εικόνα 6: Κουμπιά διαμοίρασης (Web 2.0).	15
Εικόνα 7: Web 1.0, 2.0 και 3.0	16
Εικόνα 8: Ευφυείς Προσωπικοί Πράκτορες	19
Εικόνα 9: Ο πύργος των επιπέδων του Σημασιολογικού Ιστού.....	21
Εικόνα 10: «Πλήρης» γενικευμένη στοίβα Σημασιολογικού Ιστού.....	30
Εικόνα 11: Εναλλακτική στοίβα Σημασιολογικού Ιστού	31
Εικόνα 12: Λογότυπο των Περιγραφικών Λογικών (Enrico Franconi).....	50
Εικόνα 13: Βάση γνώσης σε Περιγραφικές Λογικές.....	50
Εικόνα 14: Λογότυπο OWL.....	53
Εικόνα 15: Υπογλώσσες της OWL 1.....	55
Εικόνα 16: OWL 1 και OWL 2 profiles	60
Εικόνα 17: Λογότυπο RuleML	64
Εικόνα 18: Λογότυπο MathML	68
Εικόνα 19: Λογότυπο DAML	69
Εικόνα 20: Λογότυπο PMML	69
Εικόνα 21: Εικόνα XSLT.....	71
Εικόνα 22: Υπογλώσσες της οικογένειας της RuleML	73
Εικόνα 23: Λογότυπο DBpedia	75
Εικόνα 24: Τα κύρια συστατικά του συστήματος της DBpedia-Live	78
Εικόνα 25: Λογότυπο Ελληνικής DBpedia	79
Εικόνα 26: Λογότυπο Linking Open Data.....	81
Εικόνα 27: LOD Cloud, Μάιος 2007.....	83
Εικόνα 28: LOD Cloud, Σεπτέμβριος 2011	84
Εικόνα 29: Λογότυπο FOAF	86
Εικόνα 30: Λογότυπο SIOC.....	89
Εικόνα 31: Διασύνδεση Κοινοτικών Ιστοσελίδων.	92

Εικόνα 32: Λογότυπο Nextbio.....	95
Εικόνα 33: Πλατφόρμα NextBio	97
Εικόνα 34: Ο Cys knowledge server	99
Εικόνα 35: Στιγμιότυπο της βάσης γνώσης της Cys	101
Εικόνα 36: Αρχιτεκτονική προτεινόμενης πλατφόρμας "ΠΛΑΤΩΝΑΣ"	110
Εικόνα 37: Λογότυπο Inspire.....	117
Εικόνα 38: Σχέσεις μεταξύ των κύριων κλάσεων.....	124
Εικόνα 39: Στιγμιότυπο protégé [1].....	125
Εικόνα 40: Στιγμιότυπο protégé [2].....	126
Εικόνα 41: Στιγμιότυπο protégé [3].....	127
Εικόνα 42: Στιγμιότυπο protégé [4].....	128
Εικόνα 43: Πολυπλοκότητα των προφίλ της OWL 2	133
Εικόνα 44: Στοιχεία μεταδεδομένων Dublin Core	134
Εικόνα 45: Σχήμα βάσης με foreign keys.....	137
Εικόνα 46: Σχέσεις κλάσεων/υποκλάσεων των ροών	138
Εικόνα 47: Σχέσεις καθηγητών με τομείς και με βαθμίδες	139
Εικόνα 48: Σχέσεις κατευθύνσεων/ροών με κατηγορίες μαθημάτων ανά εξάμηνο ..	140
Εικόνα 49: Σχέσεις κατηγοριών μαθημάτων και των υποκλάσεων τους	141

Κατάλογος Πινάκων

Πίνακας 1: Μορφές κανόνων και η σημασιολογία τους	42
---	----

0. Περίληψη

Ο Παγκόσμιος Ιστός είναι ένα ταχέως αναπτυσσόμενο μέσο ενημέρωσης, επικοινωνίας και ηλεκτρονικών αγορών. Από την απαρχή του, ο Tim Berners-Lee είχε ως όραμα έναν Ιστό, στον οποίο η πληροφορία θα έπαιζε πολύ πιο σημαντικό ρόλο από αυτόν που έχει στο σήμερα. Ο Σημασιολογικός Ιστός αποτελεί μια επέκταση του υπάρχοντος Διαδικτύου, με σκοπό οι πληροφορίες που παρουσιάζονται στους χρήστες να ορίζονται με τέτοιο τρόπο, ώστε να είναι «κατανοητές» τόσο από τους ανθρώπους όσο και από τα προγράμματα-μηχανές. Κάποια ουσιώδη κομμάτια του Σημασιολογικού Ιστού, είναι η λογική, οι οντολογίες και οι κανόνες. Οι τεχνολογίες του Σημασιολογικού Ιστού μπορούν να εφαρμοστούν σε μεγάλο αριθμό πεδίων και να δημιουργήσουν ενδιαφέρουσες εφαρμογές. Στο πλαίσιο αυτών των εφαρμογών, υλοποιούμε μια οντολογία σχετική με τη δομή και το αντικείμενο αυτής της Σχολής, με τη βοήθεια μιας βάσης δεδομένων.

Abstract

The World Wide Web is a rapidly growing medium of information, communication and financial transactions. Since its very beginning, Tim Berners-Lee had a vision of a Web in which information would hold a more significant role than it holds today. Semantic Web is an extension of the existing Internet, with a purpose that all the information presented to users is defined in such a way so that it can be "understood" by both people and machine programs. Some substantial parts of the Semantic Web are logic, ontologies and rules. Semantic Web technologies can be applied in many fields and are capable to create interesting applications. In light of these applications, we implement an ontology related to the structure and purpose of this School, with the aid of a database.

1. Εισαγωγή

1.1. Ιστορικά στοιχεία

“ Έχω ένα όνειρο για τον Ιστό, στον οποίο οι υπολογιστές θα είναι σε θέση να αναλύουν όλα τα δεδομένα στον Ιστό – το υλικό, τους συνδέσμους και τις συναλλαγές μεταξύ ανθρώπων και υπολογιστών. Ένας ‘Σημασιολογικός Ιστός’, ο οποίος θα καθιστούσε κάτι τέτοιο δυνατό, έχει ακόμα καιρό για να αναδυθεί, αλλά όταν το κάνει, οι μηχανισμοί του εμπορίου, της γραφειοκρατίας και της καθημερινής ζωής μας θα διαχειρίζονται από ομιλούσες μηχανές προς μηχανές. Οι ‘Ευφυείς Πράκτορες’, που οι άνθρωποι οραματίζονταν για ολόκληρα χρόνια, επιτέλους θα υλοποιηθούν.

– Tim Berners-Lee, 1999

Ένας σοφός του καιρού μας, ένας οραματιστής, ένας απλός άνθρωπος, ο εφευρέτης του Παγκόσμιου Ιστού, έδωσε μέσα σε λίγες γραμμές το στίγμα για την εξέλιξη που πίστευε πως θα ακολουθούσε το δημιούργημά του. Αρχικά, λίγοι πίστεψαν στην ιδέα του. Με το πέρασμα του χρόνου και μετά από πολλές επίμονες προσπάθειες, όλο και περισσότεροι άρχισαν να ασχολούνται με κάτι που φαινόταν να έχει εξαιρετικά πολλές προοπτικές θετικής εξέλιξης. Η δικαίωση δεν άργησε να έρθει.

Το 1991, ένας Άγγλος Φυσικός, εργαζόμενος εκείνο τον καιρό στον Ευρωπαϊκό Οργανισμό Πυρηνικών Ερευνών (Organisation Européenne pour la Recherche Nucléaire, CERN), ο Tim Berners Lee¹, χρησιμοποίησε τον προσωπικό του υπολογιστή, έναν Apple NeXT, ως τον πρώτο web server στην ιστορία. Η λογική του Tim Berners Lee ήταν απλή, αλλά υπέρμετρα φιλόδοξη για την εποχή του· έστω ότι όλες οι αποθηκευμένες πληροφορίες στους



Εικόνα 1: Tim Berners Lee

¹ Tim Berners-Lee, γεννημένος στις 8 Ιουνίου το 1955 στο Λονδίνο της Μεγάλης Βρετανίας. (129)

ηλεκτρονικούς υπολογιστές όλου του κόσμου ήταν συνδεδεμένες μεταξύ τους και έστω ότι θα μπορούσε να δημιουργηθεί ένας χώρος όπου τα πάντα θα μπορούσαν να συνδέονται με τα πάντα. Το γεγονός αυτό θα έδινε τη δυνατότητα σε όλους τους χρήστες υπολογιστών να έχουν πρόσβαση σε όλα τα bits πληροφορίας, όλων των υπολογιστών του κόσμου. Σκοπός αυτού του εγχειρήματος θα ήταν η διευκόλυνση της ανθρώπινης διάνοιας, με την εκπληκτική διαίσθησή της, να κινηθεί μέσα στο δαίδαλο των πληροφοριών.

Προκειμένου να γίνει ένα ξεκίνημα σε αυτό το όραμα, ο Berners Lee έγραψε ένα πρόγραμμα για να εφαρμοστεί η δομή ενός ιστού. Το Enquire, όπως ήταν το όνομα αυτού του προγράμματος ήταν γραμμένο σε γλώσσα Pascal και διέθετε δυο είδη συνδέσμων: τους «εσωτερικούς» και τους «εξωτερικούς». Οι «εσωτερικοί» σύνδεσμοι ήταν αμφίδρομοι και παρέπεμπαν από μια σελίδα ενός αρχείου σε μια άλλη σελίδα του ίδιου αρχείου. Οι «εξωτερικοί» σύνδεσμοι λειτουργούσαν μόνο προς μια κατεύθυνση και συνέδεαν σελίδες διαφορετικών αρχείων. Στην ουσία, όλες οι σελίδες ήταν γραμμένες σε μια «εμπλουτισμένη» γλώσσα ή, αλλιώς, οι σελίδες αποτελούσαν «υπερκείμενα» (hypertexts), ώστε να είναι δυνατή η παρουσία συνδέσμων.

Το Διαδίκτυο (Internet) λειτουργούσε ήδη ευρέως από τη δεκαετία του 70, όμως η μεταφορά πληροφοριών ήταν σημαντικά περίπλοκη διαδικασία για κάποιον μη-ειδικό. Ήταν μείζονος σημασίας ανάγκη να εφαρμοστούν κάποιες τυποποιήσεις (standards), για να μπορέσει να μπει μια τάξη στη μεταφορά δεδομένων μεταξύ υπολογιστών διαφορετικής αρχιτεκτονικής, με διαφορετικά λειτουργικά συστήματα σε διαφορετικά μέρη του πλανήτη. Με αυτό τον τρόπο, το Διαδίκτυο και οι υπηρεσίες του θα γίνονταν πολύ πιο προσιτές στον απλό κόσμο. Στην ουσία, το Διαδίκτυο είναι ένα δίκτυο δικτύων, ο αριθμός των οποίων συνεχώς μεταβάλλεται. Τη βάση του, ωστόσο, αποτελεί ένα σύνολο τυποποιημένων πρωτοκόλλων, σύμφωνα με τα οποία οι υπολογιστές-πομποί στέλνουν δεδομένα σε κωδικοποιημένη μορφή, μέσω γραμμών τηλεφώνου συνήθως, και οι υπολογιστές-δέκτες τα αποκωδικοποιούν και τα παρουσιάζουν στον τελικό χρήστη.

Στα τέλη της δεκαετίας του 80, ο Berners Lee παρουσιάζει το CERN σαν ένα τεχνολογικό χωνευτήρι. Κάποιοι επιστήμονες χρησιμοποιούσαν το λειτουργικό

σύστημα VAX/VMS της Digital και τα Πρωτόκολλα επικοινωνίας DECnet², ενώ άλλοι χρησιμοποιούσαν το λειτουργικό σύστημα Unix και το Πρωτόκολλο επικοινωνίας TCP/IP³. Με τη λογική ότι το Πρωτόκολλο TCP είχε αρχίσει να υποστηρίζεται από το VMS και ότι το TCP/IP είχε αναπτυχθεί από ένα Πανεπιστήμιο, αυτό του Wollongong στην Αυστραλία, και όχι από κάποια εταιρεία, ο Berners Lee αποφάσισε να το εισαγάγει δοκιμαστικά σε ένα πρόγραμμα που είχε αναπτύξει, το RPC (Remote Procedure Call). Με το RPC, όπως φανερώνει και το όνομά του, ένας προγραμματιστής μπορούσε να γράψει ένα πρόγραμμα σε ένα είδος υπολογιστή, αλλά ταυτόχρονα να καλεί διαδικασίες και σε άλλους υπολογιστές, ακόμα και στην περίπτωση που εκείνοι χρησιμοποιούσαν διαφορετικά λειτουργικά συστήματα ή και γλώσσες.

Ο Berners Lee ήταν πια πεπεισμένος ότι το Πρωτόκολλο TCP/IP ήταν ιδανικό για την εύκολη επικοινωνία μεταξύ διαφορετικών υπολογιστών. Το όραμά του, για μια υπηρεσία ανταλλαγής πληροφοριών μεταξύ απλών χρηστών, για έναν Παγκόσμιο Ιστό (World Wide Web) όπως το είχε ονομάσει, σιγά-σιγά έπαιρνε σάρκα και οστά. Επόμενο βήμα ήταν να πείσει ορισμένα στελέχη του CERN να χρηματοδοτηθεί η προσπάθειά του. Τον Μάρτιο του 1989, πήρε την απόφαση να γράψει μια πρόταση σχετική με το έργο του. Στις αρχές του 1990 δεν είχε πάρει ακόμα κάποια απάντηση, οπότε την επαναδιατύπωσε το Μάιο του ίδιου χρόνου, χωρίς όμως πάλι κάποια θετική εξέλιξη.

Στο CERN επικρατούσε το δόγμα «Μη φτιάχνεις κάτι μόνος σου εάν αυτό υπάρχει ήδη και μπορείς να το αγοράσεις». Παρόλ' αυτά, ο Berners Lee ήταν αποφασισμένος να δημιουργήσει μόνος του τον Παγκόσμιο Ιστό. Με τη βοήθεια του υπολογιστή NeXT και χρησιμοποιώντας τη γλώσσα HTML⁴ για τη συγγραφή υπερκειμένων, πέρασε στο επόμενο στάδιο που ήταν η κατασκευή ενός πρωτοκόλλου μεταφοράς υπερκειμένων μεταξύ υπολογιστών, του HTTP (HyperText Transfer

²DECnet: Αποτελεί μια συλλογή Πρωτοκόλλων Επικοινωνίας η οποία δημιουργήθηκε από τη Digital Equipment Corporation και αρχικά κυκλοφόρησε το 1975. (130)

³TCP/IP (Transmission Control Program/Internet Protocol, Πρόγραμμα Ελέγχου Μετάδοσης και Πρωτόκολλο του Internet): αποτελεί μια συλλογή Πρωτοκόλλων Επικοινωνίας στα οποία βασίζεται το Διαδίκτυο. (128)

⁴HTML (HyperText Markup Language, Γλώσσα Σήμανσης Υπερκειμένου): αποτελεί την κύρια γλώσσα σήμανσης για τις Ιστοσελίδες. (131)

Protocol), καθώς και το σύστημα διευθυνσιοδότησης των εγγράφων αυτών, το URI⁵ (Universal Resource Identifier).

Στις αρχές του 1991, ο πρώτος Web server ήταν γεγονός. Επισήμως, η διεύθυνσή του ήταν η nxoc01.cern.ch (NeXT, Online Controls, 1), η οποία παρέπεμπε στον συγκεκριμένο υπολογιστή NeXT που χρησιμοποιούσε. Θέλησε όμως να κατοχυρώσει και τη διεύθυνση info.cern.ch (η οποία λειτουργεί ακόμα και σήμερα), με τη λογική ότι ο server θα μπορούσε μελλοντικά να μεταφερθεί αλλού και άρα να εξακολουθεί να είναι προσπελάσιμος μέσω Διαδικτύου.



Εικόνα 2: Ο πρώτος Web Server

Ο Παγκόσμιος Ιστός, αν και σε πρωτόγονη μορφή ήταν πλέον γεγονός. Έπρεπε πλέον να γίνουν δυο ακόμα βήματα. Πρωταρχικά, να προστεθεί κάποιου είδους χρήσιμο υλικό στον Ιστό, για να ενδιαφερθεί ένας χρήστης να χρησιμοποιήσει την υπηρεσία. Κατά δεύτερο λόγο, να κατασκευαζόταν ένα πρόγραμμα πλοήγησης, φιλικό προς τον χρήστη και προς υπολογιστές διαφορετικής αρχιτεκτονικής, ως εργαλείο ανεύρεσης πληροφοριών. Η ασυμβατότητα μεταξύ υπολογιστών, που χρησιμοποιούσαν διαφορετικά λειτουργικά συστήματα ήταν το σημαντικότερο εμπόδιο στην επικοινωνία μεταξύ τους. Το World Wide Web ήρθε να βάλει τέλος σε αυτό το φαινόμενο, ενώ παράλληλα θα παρείχε ένα αποτελεσματικό εργαλείο συντονισμού της ανθρώπινης δραστηριότητας. Ήταν ασφαλώς δυσνόητο για εκείνα τα χρόνια, το γεγονός ότι ο Παγκόσμιος Ιστός στηριζόταν πάνω σε τρεις έννοιες: το URI, το HTTP⁶ και την HTML. Δεν υπήρχε κάποιος κεντρικός υπολογιστής να «ελέγχει» την ποσότητα και το είδος των πληροφοριών που διακινούνταν μεταξύ των υπολογιστών, ούτε καν ένας οργανισμός σε κάποιο σημείο του πλανήτη που να συντονίζει την εύρυθμη λειτουργία του. Το Web δεν ήταν ένα φυσικό αντικείμενο μέσα σε έναν χώρο. Ήταν ό ίδιος ο χώρος, μέσα στον οποίο μπορούσε να υπάρξει η πληροφορία.

⁵ URI: (Uniform Resource Identifier, Ενιαίος Εντοπιστής Πόρων): είναι ένα αλφαριθμητικό χαρακτήρων το οποίο προσδιορίζει ένα όνομα ή μια πηγή στο Διαδίκτυο. (134)

⁶ HTTP (HyperText Transfer Protocol, Πρωτόκολλο Μεταφοράς Υπερκειμένου): είναι το Πρωτόκολλο του Παγκόσμιου Ιστού με το οποίο γίνεται η μεταφορά δεδομένων ανάμεσα σε έναν διακομιστή (server) και έναν πελάτη (client). (132)

Η σημαντικότερη καινοτομία του Παγκόσμιου Ιστού ήταν το URI. Από τη στιγμή που κάποιο έγγραφο, σε έναν server, διέθετε μια διεύθυνση, τότε αυτό θα ήταν προσπελάσιμο από οποιονδήποτε υπολογιστή συνδεδεμένο στο Διαδίκτυο, εύκολα και γρήγορα (εύκολα λόγω του URI και γρήγορα λόγω του πρωτοκόλλου HTTP). Αργότερα, το URI θα μετονομαζόταν σε URL⁷ (Uniform Resource Locator), για να τονιστεί, όχι τόσο η σημασία της καθολικότητας και της διάρκειας της πληροφορίας, όσο η σημασία της σταθερότητας στις διευθύνσεις.

Η γλώσσα HTML κατασκευάστηκε με βάση μια οικογένεια γλωσσών μορφοποίησης, την SGML⁸ (Standard Generalized Markup Language), που εκείνη την εποχή είχε πολλούς οπαδούς, καθώς θεωρούνταν το μόνο αποτελεσματικό πρότυπο εγγραφών στην κοινότητα του υπερκειμένου. Στην ουσία, δημιουργήθηκε με βάση κάποιους στοιχειώσεις κανόνες σχεδιασμού και ορισμένες πρακτικές, ακόμη και πολιτικές επιλογές. Μια βασική αρχή της φιλοσοφίας με την οποία κατασκευάστηκε ήταν ότι η HTML έπρεπε να μεταφέρει τη δομή του υπερκειμενικού εγγράφου, χωρίς όμως να μεταφέρει και τις λεπτομέρειες της παρουσιάσής του. Μόνο κατ' αυτό τον τρόπο θα μπορούσαν να παρουσιαστούν ικανοποιητικά τα έγγραφα σε οποιονδήποτε από τους διάφορους τύπους οθόνης που κυκλοφορούσαν και να τυπωθούν σε οποιοδήποτε μέγεθος χαρτιού.

Το ενδιαφέρον του κόσμου άρχισε να στρέφεται προς τον Παγκόσμιο Ιστό, αφού πρώτα έκαναν την εμφάνισή τους ορισμένα -δύσχρηστα μεν, αποτελεσματικά δε- προγράμματα πλοήγησης point-and-click για υπολογιστές εκτός των NeXT. Αυτά ήταν τα Erwise, Viola και Midas για τους υπολογιστές που χρησιμοποιούσαν περιβάλλον X-Windows και το Samba για τους υπολογιστές Macintosh. Ο αριθμός των servers αυξανόταν ραγδαία και το 1993 είχαν φτάσει τους 50 παγκοσμίως. Εκείνη την εποχή έκανε την εμφάνισή του ένας καινούριος πλοηγός, από το Πανεπιστήμιο του Illinois και συγκεκριμένα από το τμήμα NCSA (National Center for Supercomputing Applications), με το όνομα Mosaic. Ο NCSA Mosaic έγινε τόσο δημοφιλής, που πολλά μέσα μαζικής ενημέρωσης τον παρουσίαζαν σαν τον ίδιο τον

⁷URL (Uniform Resource Locator, Ενιαίος Εντοπιστής Πόρων): είναι ένα συγκεκριμένο σύνολο αλφαριθμητικών, το οποίο αποτελεί μια αναφορά σε ένα πόρο του Διαδικτύου. (133)

⁸SGML (Standard Generalized Markup Language, Πρότυπη Γενικευμένη Γλώσσα Σήμανσης): αποτελεί μια γλώσσα προσδιορισμού τύπου εγγράφων. Οι γλώσσες που συμμορφώνονται με την SGML ονομάζονται εφαρμογές SGML. (135)

Παγκόσμιο Ιστό. Παράλληλα, το Πανεπιστήμιο της Μινεσότα, που είχε αναπτύξει το Πρωτόκολλο Gopher, χρήσιμο για την εύρεση διαφόρων εγγράφων στο Διαδίκτυο, αποφάσισε να επιτρέπει τη χρήση του συγκεκριμένου Πρωτόκολλου μόνο επί πληρωμή.

Τέτοιου είδους κινήσεις έκαναν τον κόσμο να πιστεύει ότι και το CERN θα έκανε κάτι ανάλογο με τον Παγκόσμιο Ιστό, προσφέροντας δηλαδή τη συγκεκριμένη υπηρεσία με συνδρομή. Το CERN δεν είχε ακόμα αποφασίσει τι θα πράξει με τον Ιστό. Ο Berners Lee κινήθηκε άμεσα, προκειμένου να προλάβει καταστάσεις. Έκανε αίτηση στο CERN ζητώντας να χαρακτηριστεί το λογισμικό του Παγκόσμιου Ιστού ως ελεύθερο, χωρίς κανέναν απολύτως περιορισμό. Στις 30 Απριλίου 1993, το CERN συμφώνησε να επιτρέψει σε όποιον ήθελε να χρησιμοποιεί ελεύθερα το Πρωτόκολλο και τον κώδικα του Ιστού, να δημιουργήσει έναν server ή ένα πρόγραμμα πλοήγησης, να τα διαθέτει ελεύθερα ή να τα πουλάει, χωρίς το CERN να προβάλει κανενός είδους αξιώσεις ή να επιβάλει οποιουδήποτε περιορισμούς.

Μετά από αυτά τα γεγονότα, ο Berners Lee έθεσε ως προτεραιότητα τη δημιουργία ενός οργανισμού, σκοπός του οποίου θα ήταν η οργάνωση, ο συντονισμός, η τυποποίηση διαδικασιών και η διευθέτηση ζητημάτων που θα απασχολούσαν τον Παγκόσμιο Ιστό. Σε συνεργασία με τον Έλληνα καθηγητή και Διευθυντή του Εργαστηρίου Πληροφορικής Επιστήμης στο Πανεπιστήμιο της Μασαχουσέτης (MIT) Μιχάλη Δερτούζο⁹, πάρθηκε η απόφαση να ιδρυθεί μια κοινοπραξία (consortium), έδρα της οποίας θα ήταν τόσο η Ευρώπη όσο και η Αμερική. Συγκεκριμένα, η μία έδρα θα ήταν στο Πανεπιστήμιο της Μασαχουσέτης (Η.Π.Α.) και η άλλη στο CERN (σύνορα Ελβετίας και Γαλλίας). Το World Wide Web Consortium (W3C) ιδρύθηκε επισήμως το φθινόπωρο του 1994 και ο Tim Berners Lee θα αναλάμβανε τη θέση του Γενικού Διευθυντή της κοινοπραξίας, για τη διευκόλυνση της αρμονικής εξάπλωσης του Παγκόσμιου Ιστού.



Εικόνα 3: Μιχάλης Δερτούζος (1935-2001)

⁹Μιχάλης Δερτούζος (5 Νοεμβρίου 1936 - 27 Αυγούστου 2001): διακεκριμένος καθηγητής και διευθυντής του Εργαστηρίου Επιστήμης Υπολογιστών (Laboratory for Computer Science, LCS) του Ινστιτούτου Τεχνολογίας της Μασαχουσέτης (Massachusetts Institute of Technology, MIT) από το 1974 ως το 2001. (136)

Κύριος σκοπός του W3C θα ήταν να βοηθήσει το Web να εκμεταλλευθεί πλήρως τις δυνατότητές του. Αυτό θα γινόταν κυρίως με την επεξεργασία κοινών πρωτοκόλλων που θα επέτρεπαν τη συνεργασία μεταξύ διαφορετικών υπολογιστών και προγραμμάτων λογισμικού, συμβάλλοντας έτσι στην εξέλιξη του Παγκόσμιου Ιστού. Βασική αρχή για την επίτευξη αυτού του στόχου θα ήταν η δημιουργία ανοικτών τεχνικών προδιαγραφών: Ένας μικρός πυρήνας ανθρώπων, που θα είχαν τη βάση τους στο Εργαστήριο Πληροφορικής Επιστήμης του MIT και σε Ιδρύματα στην Ευρώπη και την Ασία, θα δημιουργούσαν τις προδιαγραφές και δείγματα του κώδικα. Κατόπιν, τα Μέλη του W3C -αλλά και οποιοσδήποτε άλλος ενδιαφερόμενος- θα μπορούσαν να πάρουν τις προδιαγραφές και να επεκτείνουν κώδικα και να τα χρησιμοποιήσουν σε διάφορες εφαρμογές, είτε δωρεάν είτε εμπορικές.

Από τα πρώτα Πρωτόκολλα που βασίστηκαν σε αυτή τη λογική ήταν το PICS (Platform for Internet Content Selection), το οποίο δημοσιοποιήθηκε τον Μάρτιο του 1996 και σκοπός του ήταν να ρυθμίζει το επίπεδο καταλληλότητας του περιεχομένου του Παγκόσμιου Ιστού σε κάποιον υπολογιστή. Το συγκεκριμένο Πρωτόκολλο προήλθε έπειτα από μια ανησυχία της IBM ότι θα υπήρχε μεγάλη αντίδραση, τόσο σε τοπικό όσο και σε εθνικό επίπεδο, στο θέμα που αφορούσε στο ακατάλληλο υλικό για ανήλικους, κυρίως πορνογραφικό, μέσω του Παγκόσμιου Ιστού. Η αντίδραση αυτή δεν άργησε να έρθει, όταν εμφανίστηκε το νομοσχέδιο περί Αξιοπρέπειας των Επικοινωνιών, που στην ουσία επιχειρούσε να χειραγωγήσει το υλικό που διακινούνταν στο Διαδίκτυο. Με την παρουσίαση του PICS, το κλίμα αναστράφηκε και σύντομα ο νόμος κρίθηκε αντισυνταγματικός και αποσύρθηκε. Με το PICS, το W3C απέδειξε ότι μπορούσε να δουλέψει γρήγορα και αποτελεσματικά και μάλιστα σε μια νέα περιοχή –στην περιοχή όπου η τεχνολογία, η κοινωνία και η πολιτική αλληλεπικαλύπτονται.

Στο καταστατικό του W3C αναφέρεται ο όρος ότι όλο το λογισμικό που δημιουργεί για να υποστηρίξει το έργο του, μπορεί να χρησιμοποιηθεί ελεύθερα από το κοινό. Αυτός είναι ένας τρόπος να προωθούνται οι συστάσεις, ο διάλογος και ο πειραματισμός. Επιτρέπεται σε όλους να συμμετέχουν στον έλεγχο των νέων πρωτοκόλλων και σε νέες εταιρείες να μουν αμέσως στο παιχνίδι της δημιουργίας λογισμικού για τον Παγκόσμιο Ιστό. Το μόνο που χρειάζεται να κάνει κανείς είναι να επισκεφθεί τον διαδικτυακό τόπο του W3C και να κατεβάσει αυτά τα εργαλεία, χωρίς καμία απολύτως χρέωση.

Για να μπορούμε πραγματικά να χρησιμοποιούμε τον Παγκόσμιο Ιστό για τη συνεργασία μας, χρειαζόμαστε εργαλεία υψηλής ποιότητας και χρηστικότητα. Για να μοιραστούν οι άνθρωποι τη γνώση, ο Ιστός πρέπει να αποτελεί έναν καθολικό χώρο, όπου θα μπορούν να κινούνται ανεμπόδιστα όλοι οι υπερκειμενικοί σύνδεσμοι. Ο πιο απλός παράγοντας, που ελέγχει τον Ιστό ως μέσο ανθρώπινης επικοινωνίας, είναι η δύναμη των διαφόρων μορφών (formats) των δεδομένων που χρησιμοποιούνται για να αναπαραστήσουν το υπερκείμενο, τα γραφικά και άλλες μορφές πληροφορίας. Υπό την πίεση του ότι τα αποτελέσματά τους ήταν άμεσα ορατά στο χρήστη, οι μορφές αυτές εξελίχθηκαν με ταχείς ρυθμούς, γιατί το κάθε μέσο ήταν ουσιαστικά ανεξάρτητο από τα υπόλοιπα. Οι προσπάθειες που έγιναν τα τελευταία χρόνια για να εισέλθει ο Παγκόσμιος Ιστός στα πάντα, από τις τηλεοράσεις έως τις οθόνες των κινητών τηλεφώνων, έδειξαν πολύ καθαρά πόσο αναγκαία είναι η ανεξαρτησία από τη συσκευή. Αυτό είχε ως αποτέλεσμα τη δημιουργία ακόμα πιο προωθημένων μορφών γραφικών, που είναι πιο εύκολο να παρουσιάσουν μια εικόνα σε οθόνες διαφορετικού μεγέθους και τεχνολογίας. Ένα από τα πιο πρόσφατα παραδείγματα είναι οι εικόνες της μορφής Scalable Vector Graphics (SVG), η οποία είναι βασισμένη στην XML. Οι συγκεκριμένες εικόνες περιγράφονται ως αφηρημένα σχήματα, αφήνοντας το πρόγραμμα πλοήγησης ελεύθερο να γεμίσει τα pixels με τρόπο που η εικόνα να μπορεί να εμφανίζεται με την ίδια ευκρίνεια στην οθόνη ενός ρολογιού χειρός όσο και σε στην γιγαντοοθόνη ενός κινηματογράφου. (1)

1.2. Σημασιολογικός Ιστός

“ Ο Σημασιολογικός Ιστός δεν είναι ένας ξεχωριστός Ιστός αλλά επέκταση του υπάρχοντος, όπου ένα σαφές νόημα δίνεται στις πληροφορίες και που, ακόμα, επιτρέπει στους ανθρώπους να συνεργάζονται με τους υπολογιστές.

– Tim Berners-Lee, 2001

“ Η ανθρωπότητα έκανε ένα τεράστιο λάθος όταν πριν από 300 χρόνια διαχώρισε την τεχνολογική εξέλιξη από τη φιλοσοφία και τον ανθρωπισμό...

Πρέπει να ζυπνήσουμε, να ακούσουμε μέσα μας και να αφουγκραστούμε ιδιαίτερα τα πιο απόμακρα κομμάτια του διασπασμένου είναι μας... Η επανένωση δεν είναι εύκολη. Αξίζει όμως να γίνει ο διηνεκής στόχος μας. Ας διορθώσουμε το σφάλμα του Διαφωτισμού και ας κοιτάζουμε ξανά εκστατικοί το ηλιοβασίλεμα, τον τροχό και ό,τι μπορεί να βρίσκεται πίσω τους.

– Μιχάλης Δερτούζος, «Η ανολοκλήρωτη επανάσταση», 2001



Εικόνα 4: Ιστορικό λογότυπο του www
σχεδιασμένο από τον Ρόμπερτ Κάιλιου

Γενικά θεωρούμε πως με την ονομασία «Παγκόσμιος Ιστός» (World Wide Web, WWW) αναφερόμαστε στο Internet. Αυτό αποτελεί ένα λάθος, καθώς το Internet είναι μόνο μια (η δημοφιλέστερη) από τις εφαρμογές του Παγκόσμιου Ιστού. (2)

Ο Παγκόσμιος Ιστός είναι ένα ταχέως αναπτυσσόμενο μέσο ενημέρωσης, επικοινωνίας και ηλεκτρονικών αγορών. (3) Πιο παλιά, ο ρόλος των υπολογιστών και των δικτύων στην ανθρώπινη επικοινωνία μέσω του Παγκόσμιου Ιστού ήταν να δημιουργούν το χώρο, όπου θα διακινούνταν η πληροφορία και, από κει και πέρα, να «συμπεριφέρονταν» σαν να μην υπάρχουν. Θα ήταν όμως χρήσιμο να δοθεί ένας πιο ενεργητικός ρόλος στους υπολογιστές και οι αναλυτικές τους ικανότητες θα μπορούσαν να μας βοηθήσουν να δώσουμε ένα νόημα στον τεράστιο όγκο των πληροφοριών και του ανθρώπινου λόγου που κυκλοφορεί στον Ιστό.

Η αρχική σχεδίαση του Διαδικτύου είχε ως στόχο να αποτελέσει αυτό ένα χώρο ελεύθερης διακίνησης πληροφοριών. Χάρη σε αυτό θα γινόταν πιο εύκολη η επικοινωνία των ανθρώπων καταρχήν μέσω των υπολογιστών και ενδεχομένως μέσω αυτόβουλων ενεργειών τους. Ένα από τα μεγαλύτερα εμπόδια προς αυτή τη κατεύθυνση είναι ότι η συντριπτική πλειοψηφία των πληροφοριών του διαδικτύου είναι προσανατολισμένη στον άνθρωπο-χρήστη. Ο Tim Berners-Lee, ως αρχικό όραμα είχε ένα Ιστό στον οποίο η πληροφορία θα έπαιζε πολύ πιο σημαντικό ρόλο από αυτόν που έχει στο σημερινό Ιστό. Για να συμβεί κάτι τέτοιο, θα έπρεπε να γίνει εισαγωγή των δεδομένων σε μορφή που να μπορούν να κατανοήσουν οι μηχανές, είτε

άμεσα είτε έμμεσα. Κατ' αυτόν τον τρόπο δημιουργείται ο Σημασιολογικός Ιστός – ένας ιστός δεδομένων, που μπορούν να τα επεξεργαστούν οι μηχανές. (4)

Ως τώρα τα έγγραφα που αποτελούν τον Παγκόσμιο Ιστό στηρίζονται σε τυποποιημένες γλώσσες, όπως η HTML και χρησιμοποιούν Πρωτόκολλα τα οποία καθιστούν το περιεχόμενό τους αναγνώσιμο (κυρίως μέσω της φυσικής γλώσσας) μόνο από τους ανθρώπους, οπότε το περιεχόμενο δεν είναι κατανοητό ως προς το νόημά του από τις μηχανές. Συνεπώς, σήμερα, ο όρος «ανάκτηση πληροφοριών» που χρησιμοποιείται σε σχέση με τις μηχανές αναζήτησης, αντιστοιχεί πιο πολύ σε «εντοπισμό θέσης» και ο χρήστης είναι αυτός που καλείται να εξάγει τις πληροφορίες που ψάχνει. Γενικά, τα προβλήματα που παρουσιάζονται ανήκουν στις ακόλουθες κατηγορίες: (α) υψηλή ανάκληση – χαμηλή ακρίβεια, (β) χαμηλή ή καθόλου ανάκληση, (γ) ευαίσθητα στο λεξιλόγιο αποτελέσματα και (δ) τα αποτελέσματά τους είναι μεμονωμένες σελίδες. Τα αποτελέσματα των αναζητήσεων στον ιστό δεν είναι άμεσα προσπελάσιμα από άλλα εργαλεία λογισμικού και συνεπώς οι μηχανές αναζήτησης είναι συχνά απομονωμένες εφαρμογές. (3)

Ο Σημασιολογικός Ιστός αποτελεί μια επέκταση του υπάρχοντος Διαδικτύου. Σκοπός του είναι οι πληροφορίες που παρουσιάζονται στον χρήστη να ορίζονται με τρόπο που να τις κάνει «κατανοητές» τόσο από τους ανθρώπους όσο και από τα προγράμματα-μηχανές. Τελικός σκοπός είναι να επιτρέπει σε αυτοματοποιημένα προγράμματα (πράκτορες) να διαμοιράζονται και να επεξεργάζονται πληροφορίες και δεδομένα για τους ανθρώπους-χρήστες τους. Προωθείται από την Κοινοπραξία Παγκοσμίου Ιστού. Αποτελεί τη μεγαλύτερη προσπάθεια αυτόματης ενοποίησης συστημάτων ώστε να συνεργάζονται διαλειτουργικά σε Παγκόσμιο επίπεδο. Η σχετική δραστηριότητα (Semantic Web Activity) έχει δημιουργηθεί για να διαδραματίσει έναν ηγετικό ρόλο, τόσο στο σχεδιασμό προδιαγραφών, όσο και στην ανοικτή ανάπτυξη της τεχνολογίας μέσω συνεργασίας. (4)

Έχουν γίνει αρκετά βήματα προς την υλοποίηση αυτού του σχεδιασμού. Πρόκληση αποτελεί η υιοθέτηση του σχεδιασμού και της τεχνολογίας και όχι η επιστημονική υιοθέτηση, καθώς υπάρχουν επιμέρους λύσεις για τα επιστημονικά προβλήματα. Προς το παρόν οι μεγαλύτερες ανάγκες βρίσκονται στους τομείς της ενοποίησης (integration), της προτυποποίησης (standardization), της ανάπτυξης εργαλείων και της αποδοχής τους από τους χρήστες. (3)

1.3. Web 1.0, Web 2.0 και Web 3.0

Web 1.0	Web 2.0	Web 3.0
"the mostly read only web"	"the wildly read-write web"	"the portable personal web"
45 million global users (1996)	1 billion+ global users (2006)	focused on the individual
focused on companies	focused on communities	lifestream
home pages	blogs	consolidating dynamic content
owning content	sharing content	the semantic web
Britannica Online	Wikipedia	widgets, drag & drop mashups
HTML, portals	XML, RSS	user behavior ("me-onomy")
web forms	web applications	iGoogle, NetVibes
directories (taxonomy)	tagging ("folksonomy")	user engagement
Netscape	Google	advertainment
pages views	cost per click	
advertising	word of mouth	

Εικόνα 5: Συνοπτική περιγραφή Web 1.0, 2.0 και 3.0

Την περίοδο που διανύουμε, ο Ιστός περνάει σε ένα νέο στάδιο εξέλιξης. Πολλοί δε θέλουν να το ξεχωρίσουν από το ήδη υπάρχον κύμα του Ιστού (το κοινώς αποκαλούμενο Web 2.0), όμως αυτή η εξέλιξη έχει διαφορετικό στόχο. (5)

Ο όρος Web 1.0 αναφέρεται στο πρώτο στάδιο του Παγκόσμιου Ιστού. Χαρακτηρίζεται και ως «ο παλιός Ιστός». Στο στάδιο αυτό, το κυρίαρχο χαρακτηριστικό ήταν ότι αποτελούνταν από Ιστοσελίδες στις οποίες οι χρήστες περιορίζονταν στη λήψη πληροφοριών. Από το 2004, χρησιμοποιείται ο όρος Web 2.0 για να περιγράψει το Κοινωνικό Διαδίκτυο (ένα σύνολο κοινωνικών σχέσεων που συνδέει τους ανθρώπους μέσα στον Παγκόσμιο Ιστό (6), την αποκαλούμενη "δεύτερη γενιά του Web". Είναι δύσκολο να αποδώσουμε αποκλειστικά έναν από τους δύο όρους σε πολλούς ιστότοπους. Αλλά υπάρχουν και άλλοι, όπως για παράδειγμα το Facebook και το Youtube, για τους οποίους είναι σαφής αυτός ο διαχωρισμός.



Εικόνα 6: Κουμπιά διαμοίρασης (Web 2.0).

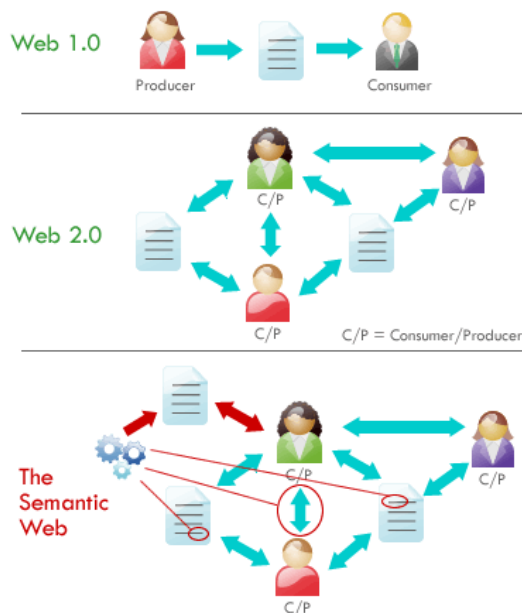
Μια συγκεκριμένη Ιστοσελίδα μπορούμε να πούμε ότι ανήκει στο Web 2.0, εάν διαθέτει τα παρακάτω στοιχεία:

- τεχνολογία: ο κώδικας που χρησιμοποιείται για τη κατασκευή της Ιστοσελίδας, καθώς και οι τεχνολογίες, τα οποία επιτρέπουν την αλληλεπίδραση με τον χρήστη,
- δομή: ο σκοπός ύπαρξης της Ιστοσελίδας και η εμφάνισή της,
- κοινωνιολογική προέκταση: αν υπάρχουν οι έννοιες των «φίλων» και των «ομάδων».

Θέλουμε Ιστοσελίδες οι οποίες να ενσωματώνουν μια ισχυρή κοινωνική συνιστώσα, η οποία να περιλαμβάνει προφίλ χρηστών, συνδέσμους φίλων, Ιστοσελίδες οι οποίες ενθαρρύνουν τους χρήστες να δημιουργήσουν περιεχόμενο στη μορφή κειμένου ή εικόνας ή βίντεο, ακολουθούμενα από σχόλια, ετικέτες και βαθμολογία.

Υπάρχει μια μεγάλη μεταβολή της κίνησης στο Διαδίκτυο, η οποία προήλθε από την δραματική αύξηση στη χρήση των Ιστοσελίδων του Web 2.0. Η βασική διαφορά μεταξύ του Web 1.0 και του Web 2.0 συνοψίζεται στο ότι οι δημιουργοί περιεχομένου ήταν λίγοι στο Web 1.0, με τη συντριπτική πλειοψηφία των χρηστών να συμπεριφέρεται σαν «καταναλωτές περιεχομένου», ενώ στο Web 2.0 ο κάθε ένας που λαμβάνει μέρος μπορεί να είναι δημιουργός περιεχομένου. Για να μεγιστοποιηθεί η πιθανότητα για συμμετοχή όσο περισσότερων στη δημιουργία περιεχομένου, έχουν δημιουργηθεί πολλά τεχνολογικά βοηθήματα¹⁰. (7)

Αν και το Web 2.0 είναι πρόσφατο και οι ιδέες πίσω από αυτό τώρα τελευταία έχουν αρχίσει να βρίσκουν ευρεία αποδοχή, έχει ήδη αρχίσει η



Εικόνα 7: Web 1.0, 2.0 και 3.0

¹⁰ Δε μελετάμε το κοινωνικό αντίκτυπο του Web 2.0 αλλά την τεχνολογική και εξελικτική του πορεία.

εισαγωγή σε αυτό τμημάτων που εν τέλει θα αποτελέσουν το αποκαλούμενο Web 3.0. Είναι ένα στάδιο του Ιστού στο οποίο οι Ιστοσελίδες, η ενημέρωση και οι βάσεις δεδομένων θα είναι «εξυπνότερα» και θα μπορούν αυτόματα να εξάγουν νόημα και όχι μόνο περιεχόμενο από τα δεδομένα.

Με τη χρήση γλώσσας σήμανσης (mark-up language) πίσω από κάθε σελίδα, όταν οι προγραμματιστές συμφωνήσουν σε ένα κοινό σύνολο ορισμών θα μας δίνεται η δυνατότητα διασύνδεσης όλων των στοιχείων σε αναρίθμητες βάσεις δεδομένων. (8) Χάρη στις διαδικτυακές υπηρεσίες όπως ο Σημασιολογικός Ιστός, τα microformats, η αναζήτηση μέσω της φυσικής γλώσσας, η εξόρυξη δεδομένων, οι συστάσεις των Πρακτόρων και οι τεχνολογίες της Τεχνητής Νοημοσύνης, δίνεται έμφαση στη διευκόλυνση της κατανόησης των πληροφοριών από τις μηχανές με απώτερο σκοπό η χρήση του Ιστού, για τον εκάστοτε χρήστη, να αποκτήσει ένα πιο παραγωγικό και διαισθητικό τόνο.

Το Web 3.0 θα είναι πιο συνδεδεμένο, ανοιχτό, και «εξυπνο» με τη χρήση των τεχνολογιών του Σημασιολογικού Ιστού, τις κατανεμημένες βάσεις δεδομένων, την επεξεργασία της φυσικής γλώσσας, τον συμπερασμό των μηχανών, τη μηχανική μάθηση¹¹ και τους αυτόνομους Πράκτορες. (5) Γενικά, το Web 3.0 είναι ο Ιστός των δεδομένων. Ο Tim Berners-Lee χαρακτήρισε το Σημασιολογικό Ιστό ως ένα από τα συστατικά του Web 3.0, αν και πολλές φορές θα δούμε να παραθέτουν αυτά τα δύο ως συνώνυμες έννοιες. (9) Κάποια προφανή πεδία εφαρμογών αποτελούν οι επιστήμες της ζωής, η ιατρική, και η βιομηχανική έρευνα. (8)

1.4. Τεχνολογίες του Σημασιολογικού Ιστού

Ρητά Μεταδεδομένα: Το περιεχόμενο του Ιστού είναι μορφοποιημένο για ανθρώπους-αναγνώστες και όχι για προγράμματα. Κυρίαρχη γλώσσα στην οποία είναι γραμμένες οι Ιστοσελίδες είναι η HTML, η οποία παρουσιάζει τις πληροφορίες σε μια κατανοητή μορφή για τους ανθρώπους, άλλα οι υπολογιστές έχουν πρόβλημα

¹¹ Μηχανική μάθηση (machine learning): Η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα. (4)

ως προς την κατανόηση των πληροφοριών. Ασχολούνται σχεδόν αποκλειστικά με την προσπέλαση και την αναπαράσταση αυτών των πληροφοριών. Οι Ιστοσελίδες θα μπορούσαν εκτός από το να περιλαμβάνουν πληροφορίες μορφοποίησης, να περιέχουν και πληροφορίες σχετικά με το περιεχόμενό τους, κάνοντας τις αναπαραστάσεις πολύ πιο εύκολα επεξεργάσιμες από υπολογιστές.

Ο όρος μεταδεδομένα (metadata) αναφέρεται σε τέτοιου είδους πληροφορίες: δεδομένα σχετικά με δεδομένα. Τα μεταδεδομένα συλλαμβάνουν μέρος του νοήματος των δεδομένων. (3)

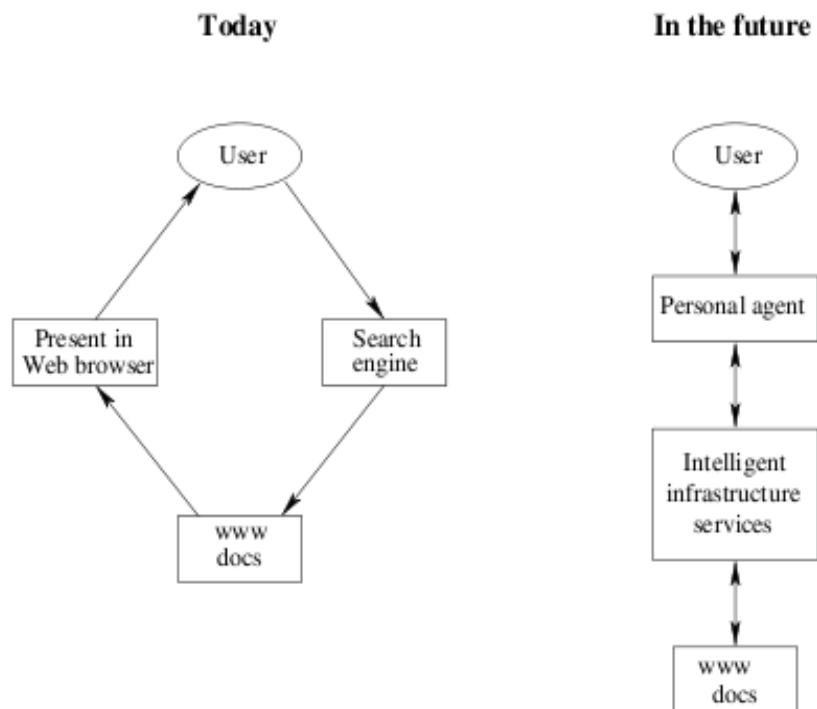
Οντολογίες: «Οντολογία είναι ένας τυπικός και σαφής ορισμός μιας κοινής και συμφωνημένης εννοιολογικής μορφοποίησης που αφορά σε ένα πεδίο ενδιαφέροντος». (10) Σε γενικές γραμμές, μια οντολογία περιγράφει ένα πεδίο ενδιαφέροντος. Συνήθως αποτελείται από μια λίστα όρων (περιγράφουν κλάσεις αντικειμένων) και τις σχέσεις μεταξύ αυτών (περιλαμβάνουν συνήθως ιεραρχίες κλάσεων). Εκτός από σχέσεις κλάσεων, οι οντολογίες περιέχουν πληροφορίες όπως ιδιότητες, περιορισμούς τιμών, προτάσεις μη επικάλυψης και προδιαγραφές λογικών σχέσεων μεταξύ αντικειμένων.

Στο πλαίσιο του Ιστού οι οντολογίες παρέχουν μια κοινή κατανόηση ενός πεδίου. Μια τέτοια κατανόηση είναι απαραίτητη για να ξεπεραστούν οι διαφορές στην ορολογία. Είναι χρήσιμες για την οργάνωση και πλοήγηση σε Ιστότοπους, για τη βελτίωση της ακρίβειας των αναζητήσεων στον Ιστό και την αξιοποίηση πληροφοριών γενίκευσης/εξειδίκευσης.

Λογική: Η Λογική (logic) είναι το επιστημονικό πεδίο που μελετάει τις αρχές της συλλογιστικής (reasoning). Προσφέρει τυπικές γλώσσες για την έκφραση γνώσης και παρέχει μια κατανοητή τυπική σημασιολογία (formal semantics). Τα προγράμματα αυτοματοποιημένης συλλογιστικής (automated reasoners) μπορούν να συνάγουν συμπεράσματα από την υπάρχουσα γνώση, καθιστώντας έτσι γνωστή την υπονοούμενη γνώση. Επίσης, σε ένα από τα πλεονεκτήματά της συγκαταλέγεται το ότι μπορεί να παρέχει αιτιολογήσεις των συμπερασμάτων, καθώς η ακολουθία των βημάτων συμπερασμού (inference steps) είναι δυνατό να ανιχνευθεί εκ νέου. (3)

Πράκτορες: Οι πράκτορες είναι προγράμματα λογισμικού που λειτουργούν αυτόνομα και προνοητικά. Εξελίχθηκαν από τις έννοιες του αντικειμενοστρεφούς προγραμματισμού και της ανάπτυξης του λογισμικού που βασίζεται σε συστατικά μέρη (component-based software development).

Οι Προσωπικοί Πράκτορες είναι δυναμικοί και ευπροσάρμοστοι ευφυείς Πράκτορες που καλύπτουν τις απαιτήσεις μεμονωμένων ατόμων. (11) Ο ρόλος τους στο Σημασιολογικό Ιστό θα είναι να δέχονται κάποιες εργασίες και προτιμήσεις από το χρήστη, να αναζητούν πληροφορίες από πηγές του Ιστού, να επικοινωνούν με άλλους πράκτορες και να συγκρίνουν πληροφορίες για τις απαιτήσεις και προτιμήσεις του χρήστη. Όλα αυτά χωρίς να έχουν ως σκοπό να αντικαταστήσουν τους ανθρώπους ως χρήστες στο Σημασιολογικό Ιστό ή να λαμβάνουν απαραίτητα αποφάσεις. Πιο συνοπτικά θα μπορούσαμε να πούμε πως ο ρόλος τους θα είναι η συλλογή και οργάνωση πληροφοριών και η παρουσίαση επιλογών προς τους χρήστες, από τις οποίες θα διαλέξουν τελικά οι ίδιοι οι χρήστες.



Εικόνα 8: Ευφυείς Προσωπικοί Πράκτορες

Οι πράκτορες του Σημασιολογικού Ιστού θα χρησιμοποιούν όλες τις τεχνολογίες που έχουμε περιγράψει:

- ❖ Μεταδεδομένα: προσδιορισμός και εξαγωγή πληροφοριών από πηγές του Ιστού.
- ❖ Οντολογίες: υποστήριξη των διαδικτυακών αναζητήσεων, ερμηνεία των πληροφοριών που ανακτώνται, και επικοινωνία με άλλους πράκτορες.

❖ Λογική: επεξεργασία των πληροφοριών που ανακτώνται και εξαγωγή συμπερασμάτων.

Θα υπάρξει πιθανότατα απαίτηση και επιπλέον τεχνολογιών, όπως για παράδειγμα οι γλώσσες επικοινωνίας πρακτόρων. Οι περισσότερες τεχνολογίες που είναι απαραίτητες για την υλοποίηση του Σημασιολογικού Ιστού αναπτύσσονται με βάση εργασίες που έχουν γίνει στη περιοχή της Τεχνητής Νοημοσύνης. Συνεπώς, η πρόοδος αυτής της Τεχνολογίας θα συντελέσει σε έναν καλύτερο Σημασιολογικό Ιστό. Η τρέχουσα τεχνολογία της Τεχνητής Νοημοσύνης είναι ήδη επαρκής για να συμβάλει ουσιαστικά στην υλοποίηση του οράματος του Σημασιολογικού Ιστού, δηλαδή δε χρειάζεται να περιμένουμε μέχρι να φτάσει σε υψηλά επίπεδα επιτυχίας.

(3)

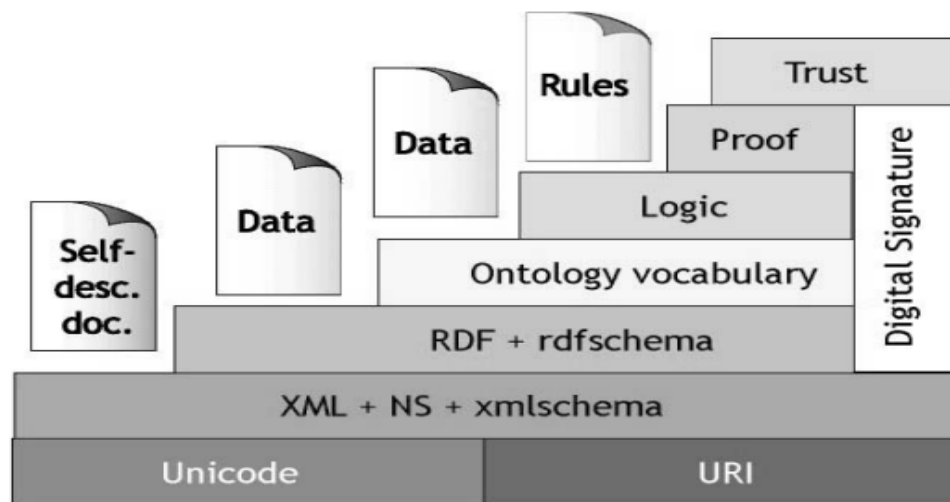
1.5. Αρχιτεκτονική του Σημασιολογικού Ιστού

Η ανάπτυξη του Σημασιολογικού Ιστού πραγματοποιείται με συγκεκριμένα βήματα ή στρώματα (layers). Κάθε στρώμα υλοποιεί ένα σύνολο τεχνολογιών οι οποίες στηρίζονται στις τεχνολογίες του προηγούμενου επιπέδου. Τα επίπεδα αυτά δεν ανταποκρίνονται αυστηρά στην έννοια της αρχιτεκτονικής λογισμικού, αλλά αποτελούν περισσότερο τεχνολογικά επίπεδα λειτουργικότητας.

Αυτό γίνεται και για πρακτικούς λόγους, όπως την εξασφάλιση της αποδοχής των τεχνολογιών κάθε στρώματος από όσο το δυνατόν περισσότερους χρήστες, εταιρίες παραγωγής λογισμικού, Οργανισμούς κτλ., κατά τη διάρκεια ανάπτυξης ενός επιπέδου και πριν να προχωρήσει η ανάπτυξη του επόμενου. Η αποδοχή των χρηστών είναι πιο εύκολη υπόθεση όταν τα βήματα είναι μικρά και εξασφαλίζουν συμβατότητα με τα προηγούμενα επίπεδα αλλά και με παλιότερες τεχνολογίες. Επιπλέον, πιστεύεται πως είναι ευκολότερο να επιτευχθεί μια γενική ομοφωνία αν τα βήματα είναι μικρά παρά αν τα βήματα είναι πιο μεγάλα. (4)

Στο ακόλουθο σχήμα αναπαρίσταται ο πύργος των επιπέδων του Σημασιολογικού Ιστού, ο οποίος οφείλεται στον Tim Berner-Lee. Για κάθε επίπεδο στο Σημασιολογικό Ιστό πρέπει να ακολουθούμε δύο αρχές:

- Συμβατότητα προς τα κάτω (downward compatability): Όσοι πράκτορες έχουν πλήρη γνώση ενός πεδίου, να μπορούν να ερμηνεύσουν και να χρησιμοποιήσουν πληροφορίες και των κατώτερων επιπέδων.
- Μερική κατανόηση προς τα πάνω (upward partial undestanding): Όσοι πράκτορες που έχουν πλήρη γνώση ενός επιπέδου, να μπορούν να εκμεταλλευτούν τουλάχιστον μερικώς, τις πληροφορίες των υψηλότερων επιπέδων. Αυτή η επιλογή θέλουμε να είναι ενεργοποιημένη και όχι απαραίτητα να παρέχεται αυτή η λειτουργικότητα από όλα τα εργαλεία.



Εικόνα 9: Ο πύργος των επιπέδων του Σημασιολογικού Ιστού

Ο «πύργος επιπέδων» του Σημασιολογικού Ιστού, όπως αυτός απεικονίζεται στη παραπάνω εικόνα, αποτελεί την κλασική στοίβα επιπέδων η οποία περιγράφει τα κύρια επίπεδα της σχεδίασης του Σημασιολογικού Ιστού. Η παραπάνω στοίβα είναι υπό συζήτηση, καθώς ο Σημασιολογικός Ιστός αποτελεί ένα ταχέως αναπτυσσόμενο Επιστημονικό Πεδίο. (3)

Τα στρώματα στο κάτω μέρος περιέχουν τις τεχνολογίες που είναι γνωστές από το υπερκείμενο. Αυτά χωρίς να χρειάζεται να υποστούν κάποια αλλαγή παρέχουν βάση για το Σημασιολογικό Ιστό. Τα μεσαία στρώματα περιέχουν τεχνολογίες που έχουν τυποποιηθεί από το W3C για να καταστήσουν δυνατή την κατασκευή εφαρμογών του Σημασιολογικού Ιστού. Τα στρώματα στην κορυφή περιέχουν τεχνολογίες που δεν έχουν ακόμη τυποποιηθεί ή περιέχουν μόνο ιδέες που πρέπει να εφαρμοστούν προκειμένου να υλοποιηθεί ο Σημασιολογικός Ιστός. (12)

1.5.1. 1ο Επίπεδο – Κωδικοποίηση και Διευθυνσιοδότηση Εγγράφων

Στο κατώτερο επίπεδο βρίσκονται οι πληροφορίες σε φυσικό και λογικό επίπεδο. Στο φυσικό επίπεδο η κωδικοποίηση των εγγράφων γίνεται με τη χρήση του διεθνώς κατοχυρωμένου προτύπου κωδικοποίησης χαρακτήρων, Unicode (Universal Code). Το πρότυπο αυτό καθορίζει το παγκόσμιο σύνολο χαρακτήρων για την κωδικοποίηση απλού κειμένου, καλύπτοντας όλες τις γλώσσες της υφηλίου.

Η διευθυνσιοδότηση των εγγράφων στο διαδίκτυο, ανεξαρτήτως περιεχομένου και Πρωτοκόλλου Επικοινωνίας, γίνεται με τη βοήθεια των Ενιαίων Προσδιοριστών Πόρων (Universal Resource Indicators, URI¹²). Οι προσδιοριστές αυτοί είναι συμβολοσειρές οι οποίες χαρακτηρίζουν μοναδικά οποιοδήποτε διαδικτυακό πόρο, για παράδειγμα μια ιστοσελίδα ή μια διεύθυνση ηλεκτρονικού ταχυδρομείου ή μια διαδικτυακή υπηρεσία κτλ, χωρίς να περιορίζονται σε μια μόνο μέθοδο ονοματολογίας και διευθυνσιοδότησης πόρων ή σε προσβάσιμες διαδικτυακές πληροφορίες. (4)

1.5.2. 2ο Επίπεδο - Αναπαράσταση Περιεχομένου Εγγράφων

Η λογική αναπαράσταση των εγγράφων στο Σημασιολογικό Ιστό βασίζεται στο πρότυπο XML (Extensible Markup Language, Επεκτάσιμη Γλώσσα Σήμανσης). Για τη δημιουργία Ιστοσελίδων χρησιμοποιούμε σήμερα την γλώσσα HTML, η οποία προέρχεται από το διεθνές πρότυπο SGML (Standard Generalized Markup Language). Η XML είναι μια μεταγλώσσα αναπαράστασης πληροφοριών. Είναι, ουσιαστικά, μια εφαρμογή SGML η οποία επιτρέπει τη συγγραφή δομημένων εγγράφων του Ιστού με τη χρήση λεξιλογίου που ορίζεται από τον χρήστη και αναπτύχθηκε για να καλύψει πολλές ελλείψεις της HTML. Έχει ως σκοπό το διαμοιρασμό εγγράφων στο διαδίκτυο μέσω μιας κοινά κατανοητής από πλευρά σύνταξης γλώσσας και είναι κατάλληλη για την αποστολή εγγράφων στον Ιστό.

¹² Υπάρχει διαφορά μεταξύ URI και URL η οποία έγκειται στο ότι οι URL διευθύνσεις αποτελούν ειδικές περιπτώσεις αναφορών URI. (4)

Ο κάθε χρήστης στην XML μπορεί, όπως αναφέραμε, να δημιουργήσει το δικό του λεξιλόγιο, κάτι που όμως κάνει πιο δύσκολη τη διαλειτουργικότητα μεταξύ των διαδικτυακών εφαρμογών. Συνεπώς, είναι απαραίτητη μια δομή για την ανταλλαγή δεδομένων μεταξύ διαδικτυακών εφαρμογών με τη χρήση κοινά κατανοητού λεξιλογίου. Το έγγραφο DTD ή XML Schema αποτελούν έγγραφα δομής XML, τα οποία ορίζουν τα επιτρεπόμενα στοιχεία του XML εγγράφου, καθώς και τον τύπο δεδομένων τους. (4)

1.5.3. 3ο Επίπεδο - Αναπαράσταση νοήματος

Η γλώσσα RDF (Resource Description Framework) αποτελεί ένα βασικό μοντέλο δεδομένων για τη συγγραφή απλών προτάσεων σχετικά με αντικείμενα του Ιστού (πόρους). Έχει ως βασικό δομικό στοιχείο μια τριάδα αντικειμένου-χαρακτηριστικού-τιμής, η οποία ονομάζεται πρόταση (statement). Αυτά αντιστοιχίζονται σε ένα υποκείμενο (subject), ένα κατηγορημα (predicate) και ένα αντικείμενο (object). Από τεχνικής άποψης, ένα σύνολο προτάσεων RDF, σχηματίζει απλώς ένα μεγάλο γράφο, συσχετίζοντας κάποια πράγματα με κάποια άλλα μέσω των ιδιοτήτων. Το μοντέλο δεδομένων της RDF δε στηρίζεται στην XML, όμως η σύνταξη της RDF είναι βασισμένη στην XML.

Η γλώσσα περιγραφής λεξιλογίου RDF, το RDF Schema (RDFS) είναι μια σημασιολογική επέκταση του RDF. Η γλώσσα RDF Schema, μπορεί να θεωρηθεί ως μια στοιχειώδης γλώσσα συγγραφής οντολογιών, η οποία βασίζεται στην RDF. Υποστηρίζει δηλώσεις που καθορίζουν αν είναι επιτρεπτή κάποια δήλωση στο βασικό μοντέλο αναπαράστασης RDF και παρέχει θεμελιώδη στοιχεία μοντελοποίησης (modeling primitives) για την οργάνωση των αντικειμένων του Ιστού σε ιεραρχίες. (3)

1.5.4. 4ο Επίπεδο - Καθορισμός λεξιλογίου

Οι οντολογίες, γενικά, περιλαμβάνουν ένα σύνολο από όρους και τις μεταξύ τους συσχετίσεις. Στο Σημασιολογικό Ιστό αποτελούν κοινά και διαμοιραζόμενα λεξικά ορολογίας με σκοπό την αμοιβαία αλληλοκατανόηση των όρων που περιλαμβάνονται σε ανταλλασσόμενα έγγραφα. (4) Κατέχουν εξέχουσα θέση, καθώς παρέχουν ένα τρόπο αναπαράστασης της σημασίας των εγγράφων και δίνουν τη δυνατότητα σε εφαρμογές του Ιστού και σε ευφυείς πράκτορες να χρησιμοποιήσουν αυτή τη σημασία. (3)

Ουσιαστικά, το συγκεκριμένο επίπεδο περιλαμβάνει γλώσσες ικανές να περιγράψουν το νόημα των δηλώσεων RDF, όπως ακριβώς η στοιχειώδης γλώσσα οντολογιών RDF Schema. Όμως, χρειαζόμαστε μια πιο ισχυρή γλώσσα οντολογιών. Επέκτασή της αποτελεί η γλώσσα OWL (Ontology Web Language) η οποία εμπλουτίζει την εκφραστική της ικανότητα με περισσότερους περιορισμούς, συσχετίσεις και τύπους ιδιοτήτων. (4) Από τις 27 Οκτωβρίου του 2009, η OWL 2 αποτελεί το πρότυπο για την περιγραφή οντολογιών στο Σημασιολογικό Ιστό. (13)

1.5.5. 5ο Επίπεδο - Λογική

Το επόμενο επίπεδο οργάνωσης του Σημασιολογικού Ιστού είναι το λογικό επίπεδο (logic layer), το οποίο χρειάζεται για να γίνονται κατανοητές διάφορες δηλώσεις, όπως οι κανόνες μετατροπής των εγγράφων από έναν τύπο σε έναν άλλο, ο έλεγχος ενός εγγράφου ως προς τη συνέπεια του, το αποτέλεσμα μιας ερώτησης με σκοπό την απόκλιση όρων ή την αναγνώριση αγνώστων όρων κτλ. (4)

Χρησιμοποιείται για την ενίσχυση της γλώσσας των οντολογιών και για τη συγγραφή δηλωτικής γνώσης, εξειδικευμένης για εφαρμογές. Παρέχει το υπόβαθρο για τη δυνατότητα αυτοματοποιημένου συλλογισμού και συμπερασμάτων βάσει των πληροφοριών που δομούνται σε μια οντολογία. (3) Με βάση τα προηγούμενα επίπεδα οργάνωσης, το λογικό επίπεδο παρέχει επιπλέον τη δυνατότητα αναπαράστασης της

Κατηγορηματικής Λογικής¹³ (δηλαδή τη χρήση λογικών τελεστών NOT, AND, OR) και τη δυνατότητα συσχέτισης των δεδομένων (για παράδειγμα, αν ισχύει το X, τότε ισχύουν όλα τα Y). (4)

Η Λογική αποτελεί τη βάση της Αναπαράστασης Γνώσης¹⁴, ειδικά με τη μορφή της Κατηγορηματικής Λογικής (γνωστή ως Λογική Πρώτης Τάξης – First-Order-Logic). Η Αναπαράσταση Γνώσης αποτέλεσε αντικείμενο μελέτης πολύ πριν την εμφάνιση του Παγκόσμιου Ιστού, ξεκινώντας από την φιλοσοφία και φτάνοντας ως τον τομέα της Τεχνητής Νοημοσύνης. (3)

1.5.6. 6ο Επίπεδο - Αποδείξεις

Το λογικό επίπεδο συνοδεύεται από το ανώτερο επίπεδο των αποδείξεων (proof) το οποίο πρακτικά περιλαμβάνει τους μηχανισμούς και τις διαδικασίες εξαγωγής συμπερασμάτων που χρησιμοποιούν την γνώση που αποτυπώνεται στο λογικό επίπεδο, την αναπαράσταση των διαδικασιών απόδειξης σε διαδικτυακές γλώσσες αναπαράστασης δεδομένων χαμηλότερου επιπέδου (όπως για παράδειγμα η XML και η RDF), καθώς και την επαλήθευση των αποδείξεων. (4) (3)

1.5.7. 7ο Επίπεδο - Έλεγχος Αξιοπιστίας – Ψηφιακές υπογραφές

Το επίπεδο της εμπιστοσύνης (trust), σε συνδυασμό με την τεχνολογία των ψηφιακών υπογραφών (digital signatures), εξασφαλίζει το βαθμό στον οποίο οι πληροφορίες που διακινούνται, επεξεργάζονται και συμπεραίνονται, είναι αξιόπιστες με αυτοματοποιημένο τρόπο. Το επίπεδο της εμπιστοσύνης προκύπτει μέσα από τη χρήση ψηφιακών υπογραφών και άλλων ειδών γνώσης, που βασίζεται σε συστάσεις από έμπιστους πράκτορες ή σε οργανισμούς αξιολόγησης και πιστοποίησης και

¹³Κατηγορηματική Λογική: Βλ. § 3.1.2

¹⁴Αναπαράσταση Γνώσης: Βλ. § 2

σωματεία καταναλωτών. Αυτό θα οδηγήσει στην εμπιστοσύνη των χρηστών στη λειτουργία του και στη ποιότητα των παρεχόμενων πληροφοριών του. (3)

1.5.8. Διεπαφή Χρηστών

Πάνω από όλα τα επίπεδα βρίσκονται η Διεπαφή των Χρηστών (user interface) και οι Εφαρμογές σε ένα ενιαίο στρώμα. Αυτό αποτελεί ένα τελικό στρώμα που θα επιτρέψει στους ανθρώπους να χρησιμοποιούν εφαρμογές του Σημασιολογικού Ιστού. (12)

1.6. Νεότερες Τεχνολογίες στη στοίβα του Σημασιολογικού Ιστού

1.6.1. SPARQL

Η γλώσσα ερωτημάτων SPARQL (SPARQL Protocol and RDF Query Language) είναι μια υποψήφια σύσταση του οργανισμού W3C για την εκτέλεση ερωτημάτων σε RDF και επομένως εξελίσσεται ταχύτατα στην πρότυπη γλώσσα ερωτημάτων για το σκοπό αυτό. Είναι μια γλώσσα που βασίζεται στην ταύτιση υποδειγμάτων γράφων¹⁵ (graph patterns), καταλαβαίνει τη σύνταξη, το μοντέλο δεδομένων και τη σημασιολογία του λεξιλογίου της RDF. Όμως, η SPARQL δεν υποστηρίζει ρητά τη σημασιολογία της γλώσσας RDFS. Επομένως, το αποτέλεσμα του ερωτήματος εξαρτάται τελικά από το αν το σύστημα που αποκρίνεται στο ερώτημα υποστηρίζει τη σημασιολογία της RDFS. (3)

¹⁵ Υπόδειγμα γράφου είναι ένας συνδυασμός υποδειγμάτων τριάδων, δηλαδή αντικειμένων που μοιάζουν με τριάδα RDF αλλά υπάρχει η δυνατότητα να χρησιμοποιηθεί μεταβλητή αντί για όρο RDF σε κάποια από τις θέσεις του υποκειμένου, του κατηγορήματος ή του αντικειμένου. (3)

1.6.2. Rif

Η ομάδα εργασίας της RIF συγκροτήθηκε στα τέλη του 2005. Η RIF (Rule Interchange Format) αποτελεί σύσταση του W3C από τις 22 Ιουνίου 2010. (14) Ο σχεδιασμός της RIF βασίζεται στην παρατήρηση ότι υπάρχουν πολλές “γλώσσες κανόνων” και εργασίες, όπως η ενσωμάτωση κανόνων ή η εισαγωγή κανόνων από ένα σύνολο κανόνων σε ένα άλλο ή η ανταλλαγή κανόνων γενικότερα, είναι δύσκολες. (15) Ο κύριος στόχος της RIF είναι να διευκολύνει την ανταλλαγή κανόνων. (16)

Η κύρια σύνταξη της RIF είναι ένα λεπτομερές XML συντακτικό που έχει σχεδιαστεί έτσι ώστε τα προγράμματα να μπορούν εύκολα να το δημιουργήσουν και να το αναλύσουν. Όμως, κατά την ανθρώπινη γραφή και ανάγνωση χρησιμοποιείται μια άλλη μικρού μεγέθους σύνταξη που έχει μια ένα-προς-ένα αντιστοιχία με την XML. Η RIF έχει σχεδιαστεί έτσι ώστε να είναι συμβατή και με άλλα Πρότυπα του Ιστού (Web Standards), όπως για παράδειγμα την RDF και την OWL. (15)

Η RIF αποτελείται από ένα σύνολο διασυνδεδεμένων διαλέκτων αντιπροσωπεύοντας γλώσσες Κανόνων με διάφορα χαρακτηριστικά. Οι τρεις πιο βασικές μορφές της RIF είναι οι ακόλουθες:

- The Basic Logic Dialect (RIF-BLD): Τεχνικά αυτή η διάλεκτος αντιστοιχεί στη Λογική Horn ¹⁶ με διάφορες σημασιολογικές και συντακτικές επεκτάσεις. Αν και δεν είναι αρκετά εκφραστική για πολλές εφαρμογές των Κανόνων, καλύπτει πολλά υπάρχοντα συστήματα Κανόνων. Η ανάπτυξη αυτής της διαλέκτου θεωρείται απαραίτητη ως σημείο έναρξης για την μελλοντική δημιουργία πιο εκφραστικών διαλέκτων.
- The Production Rule Dialect (RIF-PRD): Αν και δεν ανήκει στην οικογένεια των Λογικών διαλέκτων της RIF, είναι μια από τις σημαντικές διαλέκτους, η οποία συλλαμβάνει τις κύριες πτυχές διαφόρων Συστημάτων Παραγωγής Κανόνων (production rule systems). Έχουν γίνει σημαντικές προσπάθειες για την διασφάλιση όσο το δυνατόν περισσότερου

¹⁶Λογική Horn: Είναι μια τυπική γλώσσα που επιτρέπει την έκφραση κανόνων (και γεγονότων), και ερωτημάτων. Επιτρέπει αποτελεσματικό συμπερασμό μέσω μιας διαδικασίας που υπολογίζει απαντήσεις σε ερωτήματα από κανόνες (και γεγονότα). (163)

διαμοιρασμού με τις άλλες διαλέκτους και αυτός ο διαμοιρασμός αποτέλεσε τον κύριο λόγο για την ανάπτυξη της RIF Core Dialect

- The Core Dialect (RIF-Core): Είναι ένα υποσύνολο των RIF-BLD και RIF-PRD. Επιτρέπει την περιορισμένη ανταλλαγή Κανόνων μεταξύ των διαλέκτων Λογικών Κανόνων (logic rule dialects) και των Κανόνων Παραγωγής (production rules). Αντιστοιχεί στην Λογική Horn άνευ συναρτήσεων (συχνά αποκαλούμενη και Datalog) με έναν αριθμό επεκτάσεων για την υποστήριξη δυνατοτήτων, όπως αντικείμενα και πλαίσια, διεθνοποιημένοι αναγνωριστές πόρων (internationalized resource identifiers) για τους κατασκευαστές της γλώσσας και τους τύπους δεδομένων του σχήματος XML. (17)

1.6.3. SWRL

Η SWRL (Semantic Web Rule Language, Γλώσσα Κανόνων Σημασιολογικού Ιστού) είναι μια προτεινόμενη γλώσσα του Σημασιολογικού Ιστού, που συνδυάζει τις υπογλώσσες OWL DL και OWL Lite με τις υπογλώσσες λογικής Horn άνευ συναρτήσεων. (3) Αυτή επεκτείνει τα αξιώματα της OWL με τρόπο ώστε να περιλαμβάνουν horn-like κανόνες. Πιο συνοπτικά, θα μπορούσαμε να πούμε πως είναι μια γλώσσα Κανόνων του Σημασιολογικού Ιστού, που συνδυάζει την OWL και την RuleML¹⁷, επιτρέποντας στους κανόνες αυτούς να συνδυαστούν με τη βάση γνώσης της OWL. Οι προδιαγραφές της SWRL υποβλήθηκαν στο W3C τον Μάιο του 2004. (45)

Οι κανόνες της SWRL έχουν μια μορφή επαγωγής μεταξύ ενός σώματος (body) και μιας κεφαλής (head). Το επιδιωκόμενο νόημα ενός τέτοιου κανόνα μπορεί να θεωρηθεί ως: όταν οι συνθήκες που ορίζονται στο σώμα ενός κανόνα ισχύουν, τότε οι συνθήκες που ορίζονται στην κεφαλή πρέπει να ισχύουν και αυτές. (18) Το συντακτικό αυτών των Κανόνων στηρίζεται στην XML σύνταξη της RuleML και της OWL. (19) (3)

¹⁷ RuleML: Βλ. § 3.3

Η βασική πολυπλοκότητα της SWRL πηγάζει από το γεγονός ότι αυθαίρετες παραστάσεις της OWL (όπως για παράδειγμα οι περιορισμοί) μπορούν να εμφανιστούν στη κεφαλή ή στο σώμα ενός κανόνα, κάτι που ενισχύει την εκφραστική ικανότητα, αλλά σε βάρος της αδυναμίας λήψης αποφάσεων. Δηλαδή, δεν μπορεί να υπάρξει μηχανή συμπερασμού που να εξάγει τα ίδια συμπεράσματα με την σημασιολογία της SWRL. Πρόκληση είναι ο προσδιορισμός υπογλωσσών της SWRL, οι οποίες να εξισορροπούν την εκφραστική ισχύ με την υπολογιστική επιλυσιμότητα. (3)

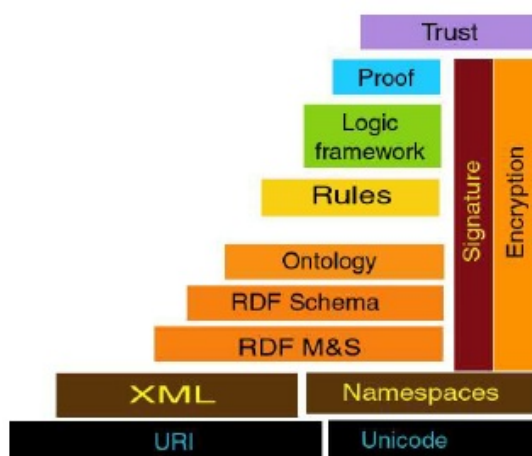
1.6.4. Προγράμματα Περιγραφικής Λογικής

Το Προγράμματα Περιγραφικής Λογικής (Description Logic Programs, DPL) είναι η τομή της OWL με τη λογική Horn άνευ συναρτήσεων και αποτελούν μια ακόμα πρόταση για την ενοποίηση των κανόνων και της OWL. (18) Η λογική Horn και οι Περιγραφικές Λογικές¹⁸ είναι ορθογώνιες, δηλαδή καμία από τις δύο δεν είναι υποσύνολο της άλλης. Οπότε με την τομή των δύο λογικών εννοούμε το τμήμα της μιας γλώσσας που μπορεί να μεταφραστεί στην άλλη με τέτοιο τρόπο ώστε να διατηρείται η σημασιολογία, και αντίστροφα ή πιο απλά το τμήμα της μιας γλώσσας που μπορεί να οριστεί από την άλλη γλώσσα. (3)

Η προσέγγιση των Προγραμμάτων Περιγραφικής Λογικής με αυτόν τον τρόπο έχει ως θετικό ότι υπάρχει ελευθερία χρήσης είτε της OWL ή των Κανόνων για τους σκοπούς της μοντελοποίησης, ακόμα και με τη μορφή Προγραμμάτων Συλλογιστικής με Περιγραφική Λογική ή Συστημάτων Επαγωγικών Κανόνων, με την επιπλέον δυνατότητα να μπορούμε να μοντελοποιήσουμε με το ένα πλαίσιο αλλά να χρησιμοποιήσουμε την μηχανή συλλογισμών (reasoning engine) του άλλου πλαισίου. Αυτά μας παρέχουν και ευελιξία στις κινήσεις μας και διαλειτουργικότητα στα εργαλεία που χρησιμοποιούμε. (3)

¹⁸ Περιγραφική Λογική: Βλ. § 3.1.3

1.7. Εναλλακτική Στοίβα Σημασιολογικού Ιστού



Εικόνα 10: «Πλήρης» γενικευμένη στοίβα
Σημασιολογικού Ιστού

Η δομή που αναπτύξαμε ως τώρα της στοίβας του Σημασιολογικού Ιστού έχει το χαρακτηριστικό της ενιαίας στοίβας. Δηλαδή, κάθε στοίβα περιέχει μια τεχνολογία και οποιαδήποτε επιπλέον τεχνολογία θα ανήκει σε άλλη στοίβα. Όλα αυτά, με ισχύουσα την «προς τα πάνω συμβατότητα», εξασφαλίζοντας την διατήρηση παλιών και νέων τεχνολογιών στο πύργο του Σημασιολογικού Ιστού όσο αυτός

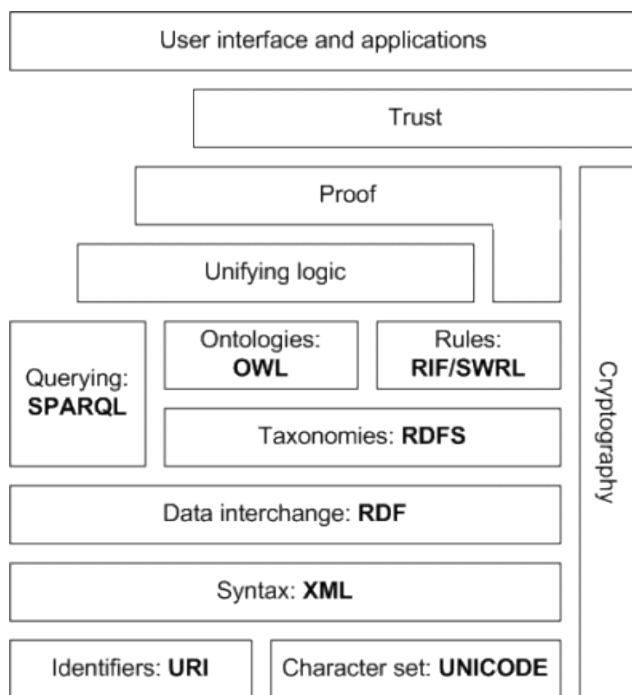
αναπτύσσεται και επεκτείνεται. Από την αρχική στοίβα μέχρι σήμερα έχουν προστεθεί αρκετές νέες τεχνολογίες για να συμβάλλουν στην υλοποίηση του οράματος του Σημασιολογικού Ιστού, κάποιες από τις οποίες αναφέραμε παραπάνω.

Οι κύριες διαφορές σε σχέση με την αρχική στοίβα είναι οι ακόλουθες:

- Το επίπεδο των οντολογιών συγκεκριμενοποιείται με δύο εναλλακτικές:
 - ο την ισχύουσα πρότυπη γλώσσα οντολογιών Ιστού (OWL)
 - ο μια γλώσσα βασισμένη σε κανόνες (rule-based language): η SWRL και η RIF.
- Τα Προγράμματα Περιγραφικής Λογικής.
- Η εμφάνιση μέσα στη στοίβα της SPARQL.

Με βάση αυτές τις διαφορές, μια πλήρης γενικευμένη στοίβα έχει την μορφή που βλέπουμε στην εικόνα 10 (χωρίς την SPARQL, η οποία δε μπορεί κατ' αυτό τον τρόπο να τοποθετηθεί μέσα στη στοίβα). Μια τέτοια στοίβα σαφώς είναι αισθητικά απλή και έχει μια απλουστευμένη διαλειτουργικότητα, όμως θεωρείται πως είναι μη ρεαλιστική και μη βιώσιμη.

Σύμφωνα με την πρώτη Αρχιτεκτονική του Σημασιολογικού Ιστού, οι κύριες γλώσσες που περιλάμβανε ο Σημασιολογικός Ιστός θα έπρεπε να αποτελούν μια ενιαία στοίβα και κάθε νέα εξέλιξη να τοποθετείται πάνω από το προηγούμενο γλωσσικό στρώμα. Τα παλαιότερα στρώματα της στοίβας πρέπει να είναι προς τα



Εικόνα 11: Εναλλακτική στοίβα Σημασιολογικού Ιστού

πάνω συμβατά με τις νέες εξελίξεις, εξασφαλίζοντας τη συμβατότητα των παλιών τεχνολογιών με τις νέες/μελλοντικές που θα εντάσσονται στο Σημασιολογικό Ιστό όσο αυτός ωριμάζει και επεκτείνεται.

Η αρχιτεκτονική της μιας ενιαίας στοίβας είναι αναμφίβολα αισθητικά απλή και διαλειτουργική. Όμως, πολλοί από τους εργαζόμενους στο τομέα αυτό, πιστεύουν ότι αυτή η αρχιτεκτονική είναι μη ρεαλιστική και μη βιώσιμη. Κάθε τεχνολογία δε θα

διατηρεί τα πλεονεκτήματά της για πάντα και δε μπορούμε να απαιτήσουμε ότι κάθε νέα εξέλιξη να είναι συμβατή με τις παλαιότερες.

Η εναλλακτική αρχιτεκτονική που προτάθηκε, αναγνωρίζει τις δυσκολίες (ακόμη και σε φιλοσοφικό επίπεδο) που εμφανίζονται στην αρχιτεκτονική ενιαίας-στοίβας (Single-Stack Architecture, SSA). Η ιδέα-κλειδί που κρύβεται πίσω από αυτή την αρχιτεκτονική είναι πως μια πιο ρεαλιστική αρχιτεκτονική πρέπει να επιτρέπει σε πολλαπλές τεχνολογικές στοίβες να υπάρχουν πλάι-πλάι. Το ιδανικό είναι οι διπλές στοίβες να είναι διαλειτουργικές σε ένα μεγάλο βαθμό και όταν αυτό δεν είναι εφικτό να είναι τουλάχιστο μερικώς διασυνδεδεμένες.

Η νέα αρχιτεκτονική του Σημασιολογικού Ιστού είναι περισσότερο ρεαλιστική. Κάθε τεχνολογία, συμπεριλαμβανομένου του σχεδιασμού γλωσσών, γίνεται τελικά απηρχαιωμένη, και καμία τεχνολογία δεν μπορεί να αντιμετωπίσει όλα τα προβλήματα. Στηριζόμενοι στην συνεχή εξέλιξη όλων των γλωσσών προγραμματισμού θα μπορούσαμε να πούμε πως είναι μια μάταιη ελπίδα ότι μια ενιαία «προς τα πάνω συμβατή» γλώσσα, που αναπτύχθηκε κατά τη βρεφική ηλικία του Σημασιολογικού Ιστού, θα αρκούσε για όλες τις μελλοντικές σημασιολογικές εργασίες που πρέπει να γίνουν στον Ιστό.

Η αρχιτεκτονική πολλών-στοιβών (Multi-Stack Architecture, MSA) είναι επίκαιρη σήμερα, καθώς οι περιορισμοί των σημερινών τυποποιημένων τεχνολογιών είναι ήδη αισθητοί στην Σημασιολογικό Ιστό και ειδικά στην υποπεριοχή των Κανόνων.

Η ανάγκη για κανόνες επεξεργασίας στο Σημασιολογικό Ιστό είχε προβλεφθεί πριν ακόμα την επινοήση του ονόματος «Σημασιολογικός Ιστός», στα πολύ αρχικά στάδια της ανάπτυξης του. Επίσης, η SHOE¹⁹ (Simple HTML Ontology Extension), ο συνδυασμός κανόνων και οντολογιών, είχε υλοποιηθεί και ήταν σε εφαρμογή πριν από την εμφάνιση της XML. Τώρα που η OWL έχει τυποποιήσει το βασικό επίπεδο της προδιαγραφής Οντολογιών και η RuleML παρέχει ένα πρότυπο στρώμα σειριοποίησης, οι Κανόνες στον Ιστό έχουν γίνει πλέον το επίκεντρο έντονης δραστηριότητας. Οι γλώσσες SWRL, SWSL-Rules, και WRL, είναι από αυτές που έχουν προταθεί πρόσφατα στο συγκεκριμένο τομέα.

Η SWRL είναι μια νέα τεχνολογία, η οποία επεκτείνει την OWL-DL²⁰ και επιτρέπει τη χρήση της Περιγραφικής Λογικής με ορισμένα είδη Κανόνων, αν και η Περιγραφική Λογική δεν είναι μια τεχνολογία που έρχεται στο μυαλό όταν κάποιος ακούει για «Κανόνες». Η χρήση των Κανόνων για την αναπαράσταση γνώσης και για τα ευφυή συστήματα πληροφοριών χρονολογείται πάνω από τριάντα χρόνια. Μέχρι τώρα είναι μια ώριμη τεχνολογία η οποία έχει δεκαετίες θεωρητικής ανάπτυξης, και πρακτικής και εμπορικής χρήσης. Η συσσωρευμένη εμπειρία στον τομέα αυτό υπερβαίνει την εμπειρία που αποκτήθηκε με τη χρήση των Περιγραφικών Λογικών και αυτό το πεδίο είναι πιθανώς το πιο ώριμο, όταν πρόκειται για τους Κανόνες.

Το πρόβλημα είναι ότι η ώριμη τεχνολογία για εφαρμογές στηριζόμενες στους Κανόνες (rule-based applications) βασίζεται στο λογικό προγραμματισμό²¹ και τη μη μονότονη συλλογιστική²² (nonmonotonic reasoning, LPNMR), τα οποία δεν είναι

¹⁹SHOE (Simple HTML Ontology Extension): Αποτελεί ένα μικρό σύνολο επεκτάσεων της HTML που επιτρέπει την επισήμανση των εγγράφων του Ιστού έτσι ώστε η γνώση που περιέχουν να γίνεται αντιληπτή από τις μηχανές. (145)

²⁰OWL-DL: Βλ. § 3.2.1

²¹Λογικός Προγραμματισμός: Στηρίζεται κυρίως πάνω στη Μαθηματική Λογική και αποτελεί ένα σημαντικό κλάδο της Τεχνητής Νοημοσύνης. (157)

²²Μη μονότονη συλλογιστική: Μας παρέχει τυπικές μεθόδους οι οποίες επιτρέπουν στα ευφυή συστήματα να λειτουργούν επαρκώς όταν έχουν ελλειπείς ή μεταβαλλόμενες πληροφορίες. (164)

πλήρως συμβατά με την κλασσική Λογική Πρώτης Τάξης²³ στο οποίο βασίζονται η OWL και η SWSL. Οι εν λόγω προτάσεις για διαδικτυακούς κανόνες (Web rules proposals), WRL και SWSL-Rules, βασίζονται σε μη μονοτονικούς συλλογισμούς. Συνεπώς, παρά το ότι η OWL είναι αδιαμφισβήτητα σημαντική και η SWRL θα βρει τα πεδία εφαρμογής της, η συντριπτική πλειοψηφία των εφαρμογών που στηρίζονται στους κανόνες δε μπορούν να γίνουν πρωταρχικά στην SWRL ή δεν υπάρχει τρόπος να γίνουν εύκολα και αποτελεσματικά. Αυτό το πρόβλημα δημιουργεί τη δεύτερη κύρια στοίβα του Σημασιολογικού Ιστού. (20)

Οι τεχνολογίες από το κάτω μέρος της στοίβας μέχρι την OWL είναι επί του παρόντος τυποποιημένες και αποδεκτές για το χτίσιμο εφαρμογών του Σημασιολογικού Ιστού. Δεν είναι ακόμα σαφές πώς η κορυφή της στοίβας πρόκειται να υλοποιηθεί, αλλά γνωρίζουμε πως όλα τα στρώματα της στοίβας πρέπει να υλοποιηθούν για την πλήρη επίτευξη του οράματος του Σημασιολογικού Ιστού. (12)

1.8. Στόχοι Σημασιολογικού Ιστού

Διακρίνουμε δύο κύριους στόχους στις υπάρχουσες εργασίες για το Σημασιολογικό Ιστό:

Ο Σημασιολογικός Ιστός ως Ιστός δεδομένων: Ως βασικό στόχο έχει την ενοποίηση των δομημένων και ημιδομημένων πηγών δεδομένων στο Διαδίκτυο. Ένας κυρίαρχος τρόπος για να επιτευχθεί αυτό, είναι να εκθέσουμε σύνολα δεδομένων στον Ιστό σε μορφή RDF και να χρησιμοποιούμε την γλώσσα RDFS για να εκφράσουμε την επιδιωκόμενη σημασιολογία αυτών των συνόλων δεδομένων. Έτσι, θα επιτρέψουμε την ενοποίηση και την εξειδικευμένη επαναχρησιμοποίησή τους. Για παράδειγμα, ας φανταστούμε μια εμπλουτισμένη πηγή πληροφοριών που θα συνδυάζει γεωγραφικά δεδομένα με ένα σύνολο αξιολογήσεων εστιατορίων από πελάτες.

²³ Λογική Πρώτης Τάξης: Βλ. § 3.1.2

Ο Σημασιολογικός Ιστός ως εμπλουτισμός του υπάρχοντος Ιστού: Ως βασικό στόχο έχει να βελτιώσει τον υπάρχοντα Παγκόσμιο Ιστό. Κάποια τυπικά παραδείγματα χρήσης αποτελούν: οι αναβαθμισμένες μηχανές αναζήτησης, η δυναμική εξατομίκευση των Ιστότοπων και ο σημασιολογικός εμπλουτισμός των υφιστάμενων σελίδων. Εδώ, οι πηγές των απαιτούμενων σημασιολογικών μεταδεδομένων θεωρείται ότι είναι αυτοματοποιημένες: εξαγωγή εννοιών, αναγνώριση επώνυμων οντοτήτων, αυτόματη ταξινόμηση κ.ά. ή και μη αυτοματοποιημένες (για παράδειγμα, παραγόμενη από ανθρώπους μέσω κοινωνικών μηχανισμών σε κοινότητες που παρέχουν σήμανση μεγάλης κλίμακας).

Και τα δύο παραπάνω βασίζονται στη χρήση σημασιολογικής σήμανσης, συνήθως με την μορφή μεταδεδομένων, που περιγράφονται από απλά σχήματα παρόμοια με οντολογίες. Μερικές σημαντικές διαφορές είναι οι διαφορετικοί στόχοι, οι διαφορετικές πηγές σημασιολογίας, οι διαφορετικές περιπτώσεις χρήσης και οι διαφορετικές τεχνολογίες.

Ο Σημασιολογικός Ιστός κατά καιρούς έχει δεχθεί διάφορες κριτικές και υπάρχουν διάφορες εσφαλμένες αντιλήψεις σχετικά με αυτόν. Κάποιες από αυτές είναι οι παρακάτω:

1. Ο Σημασιολογικός Ιστός προσπαθεί να επιβάλλει νόημα από την κορυφή: Αυτό αποτελεί μια εσφαλμένη αντίληψη σύμφωνα με την οποία ο Σημασιολογικός Ιστός επιβάλλει νόημα στους χρήστες μέσω προτύπων OWL και RDFS. Όμως, στην ουσία, το μοναδικό νόημα που επιβάλλεται από τις γλώσσες OWL και RDFS είναι των συνδετικών τους (connectives). Οι χρήστες έχουν την πλήρη ελευθερία επιλογής λεξιλογίου, αντιστοίχισης του δικού τους νοήματος στους όρους του λεξιλογίου και περιγραφής οποιουδήποτε πεδίου επιλέξουν. Οπότε, με βάση αυτή την οπτική, θα μπορούσαμε να χαρακτηρίσουμε τις OWL και RDFS ως εντελώς ουδέτερες.
2. Ο Σημασιολογικός Ιστός απαιτεί να συμφωνούν όλοι σε ένα μοναδικό και προκαθορισμένο νόημα των όρων που χρησιμοποιούνται: Το νόημα των όρων είναι κάτι το ρευστό, που έχει σχέση με τα συμφραζόμενα και δεν μπορεί να προκαθοριστεί για παγκόσμια χρήση. Ο σκοπός του Σημασιολογικού Ιστού είναι η ανάπτυξη πολλών οντολογιών και η δημιουργία αντιστοιχίσεων μεταξύ τους (εξαρτώμενες από τα συμφραζόμενα) και όχι η επιβολή μιας συγκεκριμένης οντολογίας.

3. Ο Σημασιολογικός Ιστός απαιτεί από τους χρήστες να κατανοούν τις πολύπλοκες λεπτομέρειες της τυποποιημένης αναπαράστασης της γνώσης: Όπως έχουμε ήδη παρατηρήσει, ορισμένες από τις θεμελιώδεις τεχνολογίες του Σημασιολογικού Ιστού βασίζονται σε περίπλοκες λεπτομέρειες των τυποποιημένων αναπαραστάσεων γνώσης. Η σχεδίαση καλών οντολογιών απαιτεί εξειδικευμένες γνώσεις. Όμως οι περισσότεροι χρήστες των ήδη υπαρχόντων και των μελλοντικών εφαρμογών του Σημασιολογικού Ιστού δε θα χρειαστεί να αντιμετωπίσουν ποτέ τις οντολογίες ή τις γλώσσες αναπαράστασής τους, καθώς οι μηχανές πλοήγησης ή εξατομίκευσης, που θα χρησιμοποιούν οι χρήστες, μπορούν να τροφοδοτούνται από υποκείμενες οντολογίες που έχουν εκφραστεί σε RDFS ή OWL.
4. Η μη αυτόματη σήμανση όλων των υπαρχόντων ιστοσελίδων, η οποία απαιτείται από το Σημασιολογικό Ιστό, είναι μια ανέφικτη εργασία: Η συντήρηση του περιεχομένου των Ιστοσελίδων είναι ήδη δύσκολη για τους περισσότερους κάτοχους Ιστοτόπων. Θα αποτελούσε ανεπιθύμητη εξέλιξη για το Σημασιολογικό Ιστό αν ήταν απαραίτητο να συντηρηθούν παράλληλες εκδόσεις των ίδιων πληροφοριών που είναι προσπελάσιμες από υπολογιστές, με την χρήση των RDF ή OWL. Οι εφαρμογές του Σημασιολογικού Ιστού βασίζονται στην αυτοματοποίηση μεγάλης κλίμακας για την εξαγωγή τέτοιας σημασιολογικής σήμανσης από τις ίδιες πηγές, η εξαγόμενη σημασιολογία της οποίας να μην θα είναι συχνά πολύ απλή αλλά θα επαρκεί για πολλές εφαρμογές.

Υπάρχουν αρκετά αντικείμενα της ερευνητικής ατζέντας, τα οποία αναπτύχθηκαν ελάχιστα, αλλά έχουν σημαντική συμβολή στη δυνατότητα επίτευξης του οράματος του Σημασιολογικού Ιστού:

- Αλληλεπίδραση μεταξύ των επεξεργάσιμων από υπολογιστές αναπαραστάσεων και της δυναμικής των κοινωνικών δικτύων των ανθρώπων-χρηστών.
- Μηχανισμοί για το χειρισμό της εμπιστοσύνης, της υπόληψης, της ακεραιότητας, και της προέλευσης με ημιαυτονομημένο τρόπο.
- Λειτουργίες συμπερασμού και ερωτημάτων που είναι αρκετά ανθεκτικές ώστε να λειτουργούν σε συνθήκες περιορισμένων πόρων (υπολογιστικού χρόνου, λανθάνοντος χρόνου δικτύου (network latency), μνήμης ή αποθηκευτικού χώρου) και οι οποίες μπορούν να λάβουν ευφυείς αποφάσεις για να αντισταθμίσουν τη χρήση των πόρων με την ποιότητα των αποτελεσμάτων. (3)

2. Αναπαράσταση γνώσης

Στη καθημερινότητά μας, η αναπαράσταση της γνώσης γίνεται με μια απλή περιγραφή σε φυσική γλώσσα. Παρατηρούμε ότι οι φυσικές γλώσσες είναι πραγματικά πολύ εκφραστικές. Η φυσική γλώσσα, αν και είναι το καλύτερο μέσο αναπαράστασης για τον άνθρωπο, είναι ακατάλληλη για την αναπαράσταση γνώσης σε υπολογιστικά συστήματα. Ο κυριότερος λόγος είναι πως η φυσική γλώσσα είναι πολυσήμαντη, δηλαδή κάθε φράση επιδέχεται πολλές διαφορετικές ερμηνείες που εξαρτώνται και από τα συμφραζόμενα, κάτι που θα μπορούσε να προκαλέσει δυσκολία για τη σκέψη. Οι φυσικές γλώσσες είναι, επίσης, μη συνθετικές, καθώς το νόημά τους εξαρτάται από το γλωσσικό περιβάλλον (τις προτάσεις πριν και μετά από την εξεταζόμενη). Το νόημα της πρότασης εξαρτάται και από την ίδια την πρόταση αλλά και από τα συμφραζόμενα της. Μέσω της σύνταξης και της σημασιολογίας της φυσικής γλώσσας γίνεται εφικτή η εμπειριστατωμένη περιγραφή του περιβάλλοντος. Παλιά έβλεπαν τη φυσική γλώσσα ως μια δηλωτική γλώσσα αναπαράστασης γνώσης και είχαν γίνει και απόπειρες για τον επακριβή προσδιορισμό της τυπικής σημασιολογίας της. Όμως τώρα είναι πιο εμφανές, πλέον, πως η φυσική γλώσσα είναι πιο πολύ ένα μέσο επικοινωνίας, παρά ένα μέσο αναπαράστασης. (4)

Επιπλέον, οι γλώσσες προγραμματισμού, όπως για παράδειγμα η C++ και η Java, είναι γλώσσες που αναπαριστούν άμεσα μόνο υπολογιστικές διαδικασίες και αποτελούν τη μεγαλύτερη κλάση τυπικών γλωσσών σε κοινή χρήση. Αυτό που λείπει από τις γλώσσες προγραμματισμού είναι ένας γενικός μηχανισμός για να συνδέονται τα γεγονότα. Σε μια δομή δεδομένων, κάθε ενημέρωση γίνεται με μια ειδική διαδικασία για το συγκεκριμένο πεδίο, τις λεπτομέρειες της οποίας συνάγει ο εκάστοτε προγραμματιστής μέσω της γνώσης που έχει για το αναφερόμενο πεδίο (η προσέγγιση αυτή αποκαλείται διαδικαστική ²⁴ (procedural)). Επίσης, οι δομές δεδομένων δεν έχουν την απαραίτητη εκφραστικότητα για να μπορούν να χειριστούν την μερική πληροφόρηση. Δηλαδή, τα προγράμματα μπορούν να αποθηκεύσουν μεμονωμένες τιμές στα πεδία τους και μερικά από αυτά υποστηρίζουν επιπλέον την

²⁴ Διαδικαστική προσέγγιση (procedural approach): αλγόριθμος

δυνατότητα να είναι η τιμή άγνωστη. Η διαδικαστική προσέγγιση μπορεί να αντιπαρατεθεί με την δηλωτική²⁵ (declarative) προσέγγιση.

Κατά τον Winston, η αναπαράσταση γνώσης είναι ένα σύνολο συντακτικών και σημασιολογικών παραδοχών που καθιστούν δυνατή την περιγραφή ενός κόσμου. Τα βασικά δομικά στοιχεία που έχει κάθε μέθοδος αναπαράστασης γνώσης είναι:

- Το συντακτικό (syntax): ο ορισμός των συμβόλων που χρησιμοποιεί και οι κανόνες για τους δυνατούς συνδυασμούς μεταξύ τους.
- Η σημασιολογία (semantic): ο καθορισμός των εννοιών που αποδίδονται στα σύμβολα και ο επιτρεπτός συνδυασμός συμβόλων που καθορίζεται μέσω του συντακτικού.

Τα κριτήρια με βάση τα οποία μπορεί να αξιολογηθεί μια μέθοδος αναπαράστασης γνώσης είναι συνοπτικά τα ακόλουθα:

- Επάρκεια αναπαράστασης (representation adequacy): Να μπορεί να αναπαραστήσει όλα τα είδη γνώσης και να έχει γνώση για τις αλλαγές που επέρχονται στο κόσμο που αναπαριστά.
- Επάρκεια συνεπαγωγής (inferential adequacy): Να μπορεί να συνεργαστεί με τους μηχανισμούς, που επεξεργάζονται τις υπάρχουσες δομές γνώσεις, για τη δημιουργία νέων δομών που αντιστοιχούν σε νέα γνώση.
- Αποδοτικότητα συνεπαγωγής (inferential efficiency): Να μπορεί να εισάγει επιπλέον πληροφορίες στις δομές γνώσεις, ώστε να κατευθύνει τους μηχανισμούς εξαγωγής συμπερασμάτων γρηγορότερα προς τη λύση.
- Αποδοτικότητα απόκτησης (acquisitional efficiency): Να επιτρέπεται η απόκτηση νέας γνώσης εύκολα και γρήγορα. (4)

2.1. Δημοφιλέστερες μέθοδοι Αναπαράστασης Γνώσης

Οι δημοφιλέστερες μέθοδοι αναπαράστασης γνώσης ανήκουν σε τέσσερις κατηγορίες: τη Λογική (την Προτασιακή Λογική, την Κατηγορηματική Λογική και τη

²⁵ Δηλωτική προσέγγιση (declarative approach): γνώση και συμπερασμός

Μη-Μονότονη Λογική), τις Δομημένες Αναπαραστάσεις Γνώσης (τα Σημασιολογικά Δίκτυα, τα Πλαίσια, τα Αντικείμενα, την Εννοιολογική Εξάρτηση, τα Σενάρια και τους Εννοιολογικούς Γράφους), τους Κανόνες και τη Διαδικαστική Αναπαράσταση. (4)

2.1.1. Λογική

Η Λογική²⁶ μάς παρέχει έναν τρόπο να τυποποιήσουμε τη διαδικασία της ανθρώπινης σκέψης και να την αποσαφηνίσουμε. Επίσης, μας προσφέρει μια σημαντική και εύχρηστη μεθοδολογία τόσο για την αναπαράσταση όσο και για την επίλυση προβλημάτων. Αποτελεί την πιο κλασική αλλά και ταυτόχρονα την πλέον τεκμηριωμένη και μαθηματικά τυποποιημένη μέθοδο αναπαράστασης.

Η Προτασιακή Λογική (Propositional Logic) αποτελεί την απλούστερη μορφή λογικής. Σε αυτήν, κάθε γεγονός του πραγματικού κόσμου αναπαρίσταται με μια λογική πρόταση, η οποία μπορεί να χαρακτηριστεί ως αληθής ή ψευδής.

Η Κατηγορηματική Λογική (Predicate Logic) επεκτείνει τη Προτασιακή Λογική, αντιμετωπίζοντας ταυτόχρονα το πρόβλημα της μη προσπελασιμότητας των στοιχείων των γεγονότων της.

Η Μη-Μονότονη Λογική (Non-Monotonic Logic): Οι κλασικές μορφές της λογικής αναμφίβολα προσφέρουν έναν ισχυρό μηχανισμό αναπαράστασης της γνώσης και εξαγωγής νέων συμπερασμάτων. Όμως είναι ακατάλληλες για κάποιες κατηγορίες εφαρμογών λόγω της μονοτονίας που εμφανίζουν. Ουσιαστικά, σε μια μονότονη λογική, υπάρχει η αρχική βάση γνώσης (ένα σύστημα αξιωμάτων S) και ένα σύνολο τύπων F που συνάγονται από αυτήν. Αποκτώντας νέα γνώση και προσθέτοντας ένα ή περισσότερα αξιώματα στο S , το σύνολο F αυξάνεται μονότονα με το πλεονέκτημα ότι δεν είναι απαραίτητη η καταγραφή των γεγονότων πάνω στα οποία βασίζεται η αλήθεια ενός νέου γεγονότος. Επίσης, η απόκτηση νέας γνώσης σε καμία περίπτωση δε θα μειώσει το πλήθος των συμπερασμάτων που μπορούν να εξαχθούν, σε αντίθεση με τη μη-μονότονη λογική, όπου η προσθήκη νέων αξιωμάτων είναι δυνατόν να μειώσει το σύνολο των δυνατών συμπερασμάτων, αφαιρώντας

²⁶ Περισσότερα για τη Λογική, τη Προτασιακή Λογική και τη Κατηγορηματική Λογική στο Κεφ. 3.

κάποια που αποδεικνύονται εσφαλμένα μετά την προσθήκη. Οι μη-μονότονες λογικές είναι ιδιαίτερα χρήσιμες για καταστάσεις οι οποίες εμφανίζονται συχνά στο πραγματικό κόσμο, όπως καταστάσεις για τις οποίες δεν έχουμε πλήρη γνώση ή η γνώση δημιουργείται κατά τη διάρκεια της εκτέλεσης ενεργειών (ενέργειες για τις οποίες δεν είμαστε βέβαιοι για την αναγκαιότητα ή την ορθότητά τους) ή ακόμα και μεταβάλλεται (λόγω μεταβολών στον κόσμο). Είναι χρήσιμη, επίσης, σε καταστάσεις στις οποίες το σύστημα χρησιμοποιεί υποθέσεις (assumptions). (4)

2.1.2. Δομημένες αναπαράστασεις γνώσης

Οι δομημένες μέθοδοι αναπαράστασης της γνώσης μας είναι απαραίτητες για τις αναπαράστασεις που δε μπορούν να γίνουν με τη κλασσική Λογική (όπως για παράδειγμα κλάσεις αντικειμένων), για τη μείωση του αριθμού των συμβόλων και των εκφράσεων που τα περιλαμβάνουν (μειώνοντας τον όγκο της γνώσης για ένα πρόβλημα) και για να προσεγγίσουμε την αναπαράσταση γνώσης περισσότερο διαισθητικά και λιγότερο αυστηρά.

Το Σημασιολογικό Δίκτυο (semantic network) αποτελείται από κόμβους (nodes), οποίοι υποδηλώνουν κλάσεις αντικειμένων (classes), αντικείμενα (objects), έννοιες (concepts), τιμές ιδιοτήτων (values), κ.ά., και από τους δεσμούς (links), οι οποίοι είτε αποτελούν τις σχέσεις μεταξύ των αντικειμένων ή τις ιδιότητες που συνδέουν αντικείμενα με τιμές. Η υλοποίηση ενός σημασιολογικού δικτύου στον υπολογιστή μπορεί να γίνει με δύο εναλλακτικές: με μια γλώσσα Προγραμματισμού που να επιτρέπει την αναπαράσταση συνόλων ή με ειδικές σχεδιαστικές γλώσσες (όπως για παράδειγμα η Λογική). Έχει ως πλεονέκτημα ότι τα αποτελέσματα μπορούν να εξαχθούν μέσω της ιεραρχικής δομής τους και καταλαμβάνει σχετικά μικρό χώρο αποθήκευσης, αλλά δεν υπάρχουν προκαθορισμένες δομικές σχέσεις για να στηριχτεί ο σχεδιασμός ενός Σημασιολογικού Δικτύου με κόμβους και τις σχέσεις μεταξύ τους, καθιστώντας την υλοποίησή του σχεδόν αυθαίρετη.

Τα Πλαίσια (frames) είναι μια αντικειμενοστρεφής (object oriented) προσέγγιση στην Αναπαράσταση Γνώσης και αποτελούν ένα προσφιλή τρόπο αναπαράστασης γνώσης. Τα Πλαίσια (ή σχήματα - schemata) ορίζονται σαν «δομές

δεδομένων για την αναπαράσταση στερεότυπων καταστάσεων» (Minsky). Μοιάζουν με τις εγγραφές (records) που υποστηρίζουν οι περισσότερες γλώσσες προγραμματισμού, με τις διαφορές πως δεν είναι αναγκαστικά όμοια μεταξύ τους, δεν περιέχουν ίδιου τύπου πληροφορίες και οργανώνονται σε ιεραρχικές δομές. Έχουν ως σκοπό την ομαδοποίηση πληροφοριών που σχετίζονται μεταξύ τους. Η βασική διαφορά που έχουν σε σχέση με τα Σημασιολογικά Δίκτυα είναι πως ένα Πλαίσιο περιέχει όλη την πληροφορία για την έννοια που αναπαριστά, ενώ στο Σημασιολογικό Δίκτυο η έννοια αναπαρίσταται μέσω ενός κόμβου και οι ιδιότητές της; Περιγράφονται σε άλλους κόμβους που συνδέονται με αυτόν.

Τα Αντικείμενα (objects) προήλθαν από την έρευνα για γλώσσες προσομοίωσης. Τα βασικά χαρακτηριστικά τους είναι πως έχουν κάποια κατάσταση η οποία αφορά κάποιες ιδιότητες (properties-fields), έχουν επίσης ένα σύνολο από μεθόδους (methods) οι οποίες ορίζουν τη συμπεριφορά του αντικειμένου, και αντιδρούν σε προκαθορισμένα μηνύματα (messages) ή γεγονότα (events) που λαμβάνουν από τον εξωτερικό κόσμο. Οργανώνονται σε κλάσεις οι οποίες περιλαμβάνουν το σύνολο των ιδιοτήτων των μεθόδων και των μηνυμάτων στα οποία αυτές αποκρίνονται. Παρουσιάζουν στοιχεία των Πλαισίων αλλά και κάποια επιπρόσθετα που παρέχουν μια ευελιξία η οποία οδηγεί όλο και περισσότερους χρήστες να χρησιμοποιούν τα Αντικείμενα αντί για τα Πλαίσια.

Η Εννοιολογική Εξάρτηση (conceptual dependency) είναι μια μέθοδος που έχει σαν κύριο στόχο την αναπαράσταση του νοήματος των προτάσεων φυσικής γλώσσας με σημασιολογικό δίκτυο έτσι ώστε να γίνει εύκολη η εξαγωγή συμπερασμάτων από προτάσεις και η αναπαράσταση να είναι ανεξάρτητη από τη γλώσσα διατύπωσης της πρότασης. Η μέθοδος αυτή βασίζεται στη δημιουργία ενός συνόλου από πρωταρχικές ή αρχέγονες σχέσεις (primitive relations), οι οποίες είναι σταθερές σχέσεις μεταξύ των αντικειμένων, υπό την προϋπόθεση βέβαια ότι κάθε μια από αυτές τις σχέσεις έχει καλά ορισμένη σημασιολογία. Ένα πλεονέκτημα των δικτύων εννοιολογικής εξάρτησης είναι ότι η εξαγωγή συμπερασμάτων είναι ευκολότερη αν οι έννοιες μιας φράσης αναπαριστώνται σε ένα αρχέγονο επίπεδο. Όμως, ένα βασικό μειονέκτημα της αναπαράστασης αυτής είναι ότι η γνώση δε μπορεί πάντα να αναλύεται σε πολλά μικρά κομμάτια αρχέγονης γνώσης.

Τα Σενάρια (scripts) αποτελούν μια στερεότυπη ακολουθία γεγονότων σε μια συγκεκριμένη δραστηριότητα. Το σενάριο έχει τα εξής βασικά μέρη: Συνθήκες

εισόδου (entry conditions, είναι οι συνθήκες που πρέπει να ικανοποιηθούν πριν εμφανιστούν τα γεγονότα που περιέχει το σενάριο), αποτελέσματα (results, είναι τα αποτελέσματα πρέπει να εμφανιστούν μετά τα γεγονότα του σεναρίου), σκηνικά (props, είναι τα αντικείμενα που εμφανίζονται στο σενάριο), ρόλοι (roles, είναι οντότητες που παίζουν κάποιο ρόλο στα γεγονότα του σεναρίου), παραπομπές (track, είναι κάποιες άλλες τυπικές καταστάσεις που μοιράζονται πολλά από τα γεγονότα του σεναρίου) και σκηνές (scenes, είναι η ακολουθία των). Κάθε μέρος από αυτά περιέχει κάποιες πληροφορίες με τις χαρακτηριστικές ιδιότητες των οντοτήτων ή των αντικειμένων και τις τιμές τους. Τα σενάρια δε μπορούν να αναπαραστήσουν όλα τα είδη γνώσης, αλλά λειτουργούν καλά για ειδικές εφαρμογές.

Οι Εννοιολογικοί γράφοι (conceptual graphs) αποτελούν μια από τις πλέον σύγχρονες και μαθηματικά τεκμηριωμένες μεθόδους αναπαράστασης γνώσης, η οποία έχει ρίζες στη μοντέρνα γλωσσολογία, στη ψυχολογία και στη φιλοσοφία και μας προσφέρει μεγάλη εκφραστικότητα, συνδυάζοντας τη λογική με τις δομημένες αναπαραστάσεις. Ένας εννοιολογικός γράφος είναι ένας πεπερασμένος γράφος που αποτελείται από διασυνδεδεμένους κόμβους εννοιών και σχέσεων, που εναλλάσσονται. Τα τόξα (arcs) στον γράφο, συνδέουν τις σχέσεις με τις έννοιες. Πιο συγκεκριμένα, ανήκουν στις σχέσεις ενώ είναι προσκολλημένα στις έννοιες (concepts), οι οποίες αποτελούν στιγμιότυπα των τύπων-εννοιών (concept-types) και αποτελούνται από έναν τύπο-έννοιας (μπορεί να θεωρηθεί ότι είναι μια κλάση ή μια κατηγορία οντοτήτων ή ενεργειών ή ιδιοτήτων ή αφηρημένων εννοιών, και άλλα) ο οποίος ίσως συνοδεύεται από μια αναφορά (referent) σε μια συγκεκριμένη οντότητα.

(4)

2.1.3. Κανόνες (if-then rules)

Οι Κανόνες αποτελούν τον πιο δημοφιλή και πρακτικό τρόπο αναπαράστασης. Είναι μια φυσιολογική προσέγγιση στην αναπαράσταση της γνώσης, καθώς η γνώση αναπαριστάται με ένα τρόπο που πλησιάζει την ανθρώπινη γνώση για πολλές από τις δραστηριότητες που απαιτούν νοημοσύνη. Επίσης, η εξαγωγή συμπερασμάτων γίνεται με έναν εύκολο τρόπο. Κάποια επιπλέον πλεονεκτήματα των κανόνων είναι τα

ακόλουθα: Κάθε κανόνας ορίζει ένα μικρό και «ανεξάρτητο» τμήμα της γνώσης για ένα πρόβλημα, η προσθήκη νέων κανόνων μπορεί να γίνει «ανεξάρτητα» από τους υπάρχοντες κανόνες και οι ήδη υπάρχοντες κανόνες σε ένα σύνολο κανόνων μπορούν να αλλάξουν «ανεξάρτητα» από άλλους κανόνες. Η «ανεξαρτησία» αυτή δεν είναι απόλυτη. Υπάρχει απόκλιση μεταξύ των επιθυμητών αυτών χαρακτηριστικών από την πραγματικότητα, καθώς οι αλλαγές στη βάση των κανόνων δεν είναι απλή διαδικασία.

Πίνακας 1: Μορφές κανόνων και η σημασιολογία τους

Μορφές Κανόνων	Εκφράζει	Επεξήγηση
ON συμβάν IF συνθήκες THEN ενέργειες Ενεργός κανόνας	Διαδικαστική γνώση	Όταν συμβεί το γεγονός αν οι συνθήκες αληθεύουν τότε εκτέλεσε τις ενέργειες
IF συνθήκες THEN ενέργειες Κανόνας παραγωγής	Διαδικαστική γνώση	Αν οι συνθήκες αληθεύουν τότε εκτέλεσε τις ενέργειες
IF συνθήκες THEN συμπέρασμα Συνεπαγωγικός κανόνας	Δηλωτική γνώση	Αν οι συνθήκες αληθεύουν τότε αληθεύει και το συμπέρασμα

Τα συστήματα κανόνων μπορούν να χωριστούν σε τρεις μεγάλες κατηγορίες:

- Τα συστήματα εξαγωγής συμπερασμάτων (deduction system): οι κανόνες εκφράζουν δηλωτική γνώση, αλλά δεν αναφέρει ρητά πότε και πως εφαρμόζεται.
- Τα συστήματα παραγωγής (production systems): οι κανόνες εκφράζουν διαδικαστική γνώση και οι εκτελούμενες ενέργειες επιφέρουν μη-αναστρέψιμα αποτελέσματα.
- Τα ενεργά συστήματα ((re-)active systems): οι κανόνες εκφράζουν με ακρίβεια το πότε ακριβώς ενεργοποιούνται (όταν συμβεί συγκεκριμένο συμβάν, τότε και μόνο τότε εξετάζεται αν ικανοποιείται η συνθήκη τους και εν συνεχεία εκκαλούνται οι ενέργειες τους). (4)

2.1.4. Διαδικαστική Αναπαράσταση

Η Διαδικαστική Αναπαράσταση (Procedural Attachment) είναι η μορφή αναπαράστασης γνώσης η οποία χρησιμοποιείται για να αναπαραστήσει οποιαδήποτε προβλήματα μπορούν να περιγραφούν σε μορφή αλγορίθμων (συγκεκριμένων προβλημάτων/υποπροβλημάτων), μέσω εξειδικευμένων διαδικασιών. Η κωδικοποίηση τέτοιων προγραμμάτων γίνεται σε γλώσσες προγραμματισμού όπως η C++, η Java, η Python και άλλες. Ένα διαδραστικό πρόγραμμα μπορεί να ενεργοποιηθεί είτε από άλλες διαδικασίες είτε από άλλα δεδομένα είτε λόγω ικανοποίησης κάποιας συνθήκης (demons – δαίμονες).

Στα πλεονεκτήματα συγκαταλέγεται η γρήγορη επεξεργασία της κωδικοποιημένης γνώσης. Όμως το πρόγραμμα δεν είναι τόσο διαφανές και υπάρχει μια επιπρόσθετη δυσκολία στην αναθεώρηση ή ακόμα και στη προσθήκη γνώσης. (4)

3. Η Λογική για το Σημασιολογικό Ιστό

3.1. Λογική

Η Λογική είναι μια γλώσσα για συλλογισμούς. Μια συλλογή Κανόνων που χρησιμοποιούμε όταν κάνουμε λογικούς συλλογισμούς. Οι ανθρώπινοι συλλογισμοί έχουν παρατηρηθεί και μελετηθεί επί αιώνες, τουλάχιστον από εποχή των Αρχαίων Ελλήνων, και τα πρότυπα που εμφανίζονται στη συλλογιστική (reasoning) έχουν αποσπαστεί, εξαχθεί και εκσυγχρονιστεί. (21) Μας προσφέρει μια σημαντική και εύχρηστη μεθοδολογία για την αναπαράσταση και επίλυση προβλημάτων. Μας επιτρέπει να συλλογιζόμαστε (reason) για την ορθότητα των προγραμμάτων, να αναπαριστούμε προβλήματα αλλά και να τα επιλύουμε.

Παρέχει έναν τρόπο για την αποσαφήνιση και την τυποποίηση της διαδικασίας της ανθρώπινης σκέψης. (22) Το αντικείμενο της είναι να εξετάσει τους ανθρώπινους συλλογισμούς και να διαμορφώσει κανόνες που θα εξασφαλίζουν ότι μια τέτοια συλλογιστική είναι σωστή. Η Λογική προσέγγιση περιλαμβάνει την έκφραση της ανθρώπινης γνώσης σε μια τυπική γλώσσα (formal language) η οποία συχνά αποκαλείται «γλώσσα αναπαράστασης γνώσης» και την εφαρμογή συστήματος κανόνων εκλογίκευσης σε αυτές τις τυπικές αναπαραστάσεις. (23) Η ανάγκη για μια φορμαλιστική αναπαράσταση της ανθρώπινης σκέψης προέρχεται από το γεγονός ότι η φυσική γλώσσα είναι, μέσα στα άλλα, βερμπαλιστική (verbose), ασαφής (vague), πολυσήμαντη (ambiguous) και περιέχει συμφραζόμενα (context sensitive). (22)

Έχει μεγάλη εκφραστική ισχύ και μας παρέχει μια γλώσσα υψηλού επιπέδου στην οποία μπορούμε να εκφράσουμε τη γνώση με «διαφανή» τρόπο. Η τυπική σημασιολογία της είναι κατανοητή και μας επιτρέπει να καθορίσουμε αν μια πρόταση προκύπτει σημασιολογικά από ένα σύνολο άλλων προτάσεων. Επιπρόσθετα, μέσω των συστημάτων αποδείξεων μπορούμε να εξάγουμε αυτόματα προτάσεις από ένα σύνολο προϋποθέσεων, μέσω της σύνταξής τους. Υπάρχουν και συστήματα αποδείξεων στα οποία η σημασιολογική λογική συνέπεια συμπίπτει με τη συντακτική προέλευση. Λόγω της ύπαρξης αυτών των συστημάτων απόδειξης είναι εφικτός ο εντοπισμός της απόδειξης που οδηγεί σε μια λογική συνέπεια. (3)

Ο τυπικός χαρακτήρας που έχει η Λογική κάνει εφικτή την πιθανότητα της υλοποίησης από υπολογιστή της διαδικασίας των ανθρώπινων συλλογισμών, χωρίς αυτό να σημαίνει υποχρεωτικά ότι μια τυπική γλώσσα μπορεί να μοντελοποιήσει κάθε πτυχή της ανθρώπινης συλλογιστικής, με ικανοποιητικό τρόπο. Όμως από αυτό είναι φανερό το γιατί η τυποποιημένη λογική (formal logic) είναι σημαντική για την Τεχνητή Νοημοσύνη, και για την αυτόματη συλλογιστική γενικότερα. (23)

Υπάρχουν διάφορα είδη λογικής, όπως Λογική των Προτάσεων (propositional logic), η Λογική των Αντικειμένων (predicate logic), η Λογική που περιλαμβάνει αβεβαιότητα, η Λογική που περιλαμβάνει ασάφεια, η Χρονική Λογική κ.α. Η Προτασιακή Λογική και η Κατηγορηματική Λογική είναι θεμελιώδους σημασίας για όλους τους τύπους της Λογικής. (21)

3.1.1. Προτασιακή Λογική

Η Προτασιακή Λογική (Propositional Logic) είναι το απλούστερο λογικό σύστημα, ένας φορμαλισμός κάποιων απλών μορφών συλλογιστικής. (24) Η μικρότερη μονάδα που συναντάμε στη Προτασιακή Λογική είναι η πρόταση. (25) Μια πρόταση (proposition) είναι μια δήλωση της γλώσσας που διατυπώνει κάτι για έναν εξωτερικό κόσμο. (23) Δεν ασχολούμαστε με το νόημα ή την ανάλυση μεμονωμένων προτάσεων αλλά με την αλήθεια ή το ψεύδος αυτών και υπάρχει ένα μεγάλο ενδιαφέρον για το εάν (και κατ' επέκταση «πώς») η αλήθεια ή το ψεύδος συγκεκριμένων προτάσεων προκύπτει από ένα σύνολο προτάσεων. (25) Οι προτάσεις αυτές είναι είτε αληθείς είτε ψευδείς αλλά όχι και τα δύο. (24) Συνεπώς, ερωτήσεις, εντολές, υποσχέσεις κ.α. δεν αποτελούν προτάσεις. (23)

Για κάθε τυπική γλώσσα, πρέπει να οριστούν η σύνταξή της (η οποία καθορίζει τις επιτρεπτές ακολουθίες συμβόλων) και η σημασιολογία της (η οποία καθορίζει τις μεταξύ τους σχέσεις). (24)

Η σύνταξη (syntax) της προϋποθέτει τον καθορισμό του αλφάβητου της γλώσσας, δηλαδή του συνόλου των συμβόλων με τα οποία μπορούν να κατασκευαστούν οι εκφράσεις της γλώσσας της λογικής των προτάσεων, οι οποίες είναι πεπερασμένες ακολουθίες των συμβόλων του αλφαβήτου.

Το αλφάβητο της προτασιακής λογικής αποτελείται από τα επόμενα σύμβολα:

- Ένα αριθμήσιμο σύνολο συμβόλων που αποκαλούνται σύμβολα προτάσεων ή προτασιακές μεταβλητές.
- Τέσσερα σύμβολα λογικών συνδέσμων:
 - ο $\neg$ (άρνηση),
 - ο $\wedge$ (σύζευξη),
 - ο $\vee$ (διάζευξη),
 - ο $\rightarrow$ (συνεπαγωγή),
- Την αριστερή και τη δεξιά παρένθεση, “(“ και “)” αντίστοιχα.

Με βάση αυτό το αλφάβητο ορίζεται ποιες προτάσεις είναι σωστές ή καλοφτιαγμένες εκφράσεις (ή προτασιακοί τύποι), μέσω ενός γενικευμένου επαγωγικού ορισμού. (26)

Η σημασιολογία (semantics) της προτασιακής λογικής αντιστοιχεί μία αληθοτιμή σ' έναν προτασιακό τύπο, στηριζόμενη σε μια ερμηνεία της γλώσσας²⁷. (24) Η αλήθεια και το ψεύδος αντιστοιχούν στις δύο αληθοτιμές T (true) και F (false) (αλλιώς αποκαλούμενες και ως «τιμές αληθείας») και η αληθοτιμή μιας σύνθετης πρότασης καθορίζεται από τις αληθοτιμές των επιμέρους προτάσεων που τις σχηματίζουν, σύμφωνα και με τους συνδέσμους που χρησιμοποιήθηκαν για να τις ενώσουν αυτές, με βάση τους αληθοπίνακες²⁸. (26)

Τα πλεονεκτήματα της Προτασιακής Λογικής συνοψίζονται στο ότι έχει μια απλή σύνταξη και στο ότι είναι καταληκτική (decidable), δηλαδή πάντα καταλήγει σε συμπέρασμα. Όμως έχει και κάποια μειονεκτήματα, όπως για παράδειγμα ότι υπάρχει έλλειψη της γενικότητας. Επίσης, υπονοεί ότι ο κόσμος αποτελείται μόνο από γεγονότα αληθή ή ψευδή και δεν υπάρχει δυνατότητα διαχωρισμού και προσπέρασης των οντοτήτων του κόσμου. (4)

²⁷ Μια ερμηνεία (interpretation) αντιστοιχεί τιμές αληθείας στις προτασιακές μεταβλητές και σε σύνθετους τύπους με χρήση ενός αληθοπίνακα. (24)

²⁸ Πίνακες στους οποίους προσδιορίζοντας τις αληθοτιμές των επιμέρους προτασιακών μεταβλητών, μπορούμε να αντιστοιχίσουμε αληθοτιμές στους προτασιακούς τύπους. (26)

3.1.2. Λογική Πρώτης Τάξης

Η Προτασιακή Λογική είναι μια δηλωτική γλώσσα, επειδή η σημασιολογία της βασίζεται σε μια σχέση αληθείας μεταξύ προτάσεων και δυνατών κόσμων. Μας παρέχει αρκετή εκφραστική ισχύ για την αντιμετώπιση και τη μερική πληροφόρηση (χρησιμοποιώντας την διάζευξη και την άρνηση). Επίσης, μας παρέχει την συνθετικότητα, μια επιθυμητή –για τις γλώσσες αναπαράστασης- ιδιότητα που διευκολύνει σε μεγάλο βαθμό τα συστήματα συλλογιστικής, σύμφωνα με την οποία το νόημα της εκάστοτε πρότασης να είναι συνάρτηση του νοήματος των επιμέρους μερών της. Όμως η Προτασιακή Λογική δεν παρέχει την απαιτούμενη εκφραστική ισχύ για την περιγραφή ενός περιβάλλοντος με πολλά αντικείμενα με ένα συνοπτικό τρόπο. (4) Συνεπώς η Προτασιακή Λογική δεν είναι αρκετά ισχυρή για να αντιπροσωπεύσει όλους τους τύπους των ισχυρισμών που χρησιμοποιούνται στις Επιστήμη των Υπολογιστών και στα Μαθηματικά ή να εκφράσει συγκεκριμένους τύπους σχέσεων μεταξύ προτάσεων, όπως για παράδειγμα την ισοδυναμία. (27)

Υπάρχουν είδη συλλογιστικών απαγωγών²⁹ που δε γίνεται να δικαιολογηθούν μόνο με βάση τη Προτασιακή Λογική. Είναι συλλογισμοί των οποίων η ορθότητα δε μπορεί να δοθεί μέσω αληθοπίνακα, καθώς δεν εξαρτάται από την τοποθέτηση των προτάσεων σε σχέση με τους λογικούς συνδέσμους. Εξαρτάται από τη δομή των επιμέρους προτάσεων και από κάποιες επιπλέον εκφράσεις που δεν ορίστηκαν στη Προτασιακή Λογική, όπως το «όλοι», το «για κάθε» και άλλα. (26)

Η Λογική Πρώτης Τάξης³⁰ επεκτείνει τη Προτασιακή Λογική (μπορεί να θεωρηθεί και γενίκευσή της) εισάγοντας επιπλέον έννοιες: τους όρους (terms), τα κατηγορήματα (predicates) και τους ποσοδείκτες (quantifiers). Είναι πιο πολύπλοκη από την Προτασιακή Λογική και μπορεί να θεωρηθεί μια γενίκευσή της. (28) Στη Λογική Πρώτης Τάξης μια δήλωση έχει μια συγκεκριμένη εσωτερική δομή η οποία αποτελείται από τους όρους και τα κατηγορήματα. Οι όροι δηλώνουν τα αντικείμενα και τα κατηγορήματα εκφράζουν τις ιδιότητες ή τις σχέσεις μεταξύ αυτών των

²⁹ Απαγωγική συλλογιστική (abductive reasoning): Είναι η εξαγωγή συμπερασμάτων κατά την οποία, έχοντας δεδομένα μια βάση γνώσης και μερικές παρατηρήσεις επιχειρείται η εύρεση υποθέσεων, οι οποίες μαζί με τη βάση γνώσης εξηγούν τις παρατηρήσεις. (4)

³⁰ Αλλιώς αποκαλούμενη και ως Κατηγορηματική Λογική, Predicate Logic. (28)

αντικειμένων. (23) Το σύνολο των συμβόλων της λογικής πρώτης τάξης (το αλφάβητο πρώτης τάξης), αποτελείται από:

- Τα λογικά σύμβολα:
 - ο Μεταβλητές: Ένα αριθμήσιμο σύνολο συμβόλων
 - ο Λογικούς Συνδέσμους: Τα σύμβολα:
 - $\neg$ (άρνηση),
 - $\wedge$ (σύζευξη),
 - $\vee$ (διάζευξη),
 - $\rightarrow$ (συνεπαγωγή),
 - $\leftrightarrow$ (ισοδυναμία)
 - ο Κόμμα, Παρενθέσεις: Τα σύμβολα “,” και “(,”)” αντίστοιχα
 - ο Ποσοδείκτες: $\exists$ (υπαρξιακός ποσοδείκτης) και $\forall$ (καθολικός ποσοδείκτης)
- Τα μη λογικά σύμβολα:
 - ο Σύμβολα Σταθερών: Ένα σύνολο σταθερών που η κάθε μια θα συμβολίζει ένα σταθερό αντικείμενο.
 - ο Σύμβολα Κατηγορημάτων: Κάθε σύμβολο κατηγορήματος έχει μια συγκεκριμένη τάξη και ανάλογα με αυτή θα προσδιορίζει είτε ιδιότητες των αντικειμένων (αν είναι τάξης 1) ή n-μελείς σχέσεις (αν είναι τάξης n).
 - ο Σύμβολα Συναρτήσεων: Τα Σύμβολα Συναρτήσεων n-θέσεων συμβολίζουν συναρτήσεις n μεταβλητών στο σύμπαν των αντικειμένων μας.

Η σύνταξη της είναι σχεδιασμένη για να διευκολύνει τις δηλώσεις σχετικά με τα αντικείμενα. (29) Ο ορισμός μιας γλώσσας γίνεται με βάση το σύνολο των μη λογικών συμβόλων από τα οποία αποτελείται, καθώς τα λογικά σύμβολα είναι σε όλες τις γλώσσες τα ίδια, και αυτό αποτελεί και το σημείο διαφοροποίησης τους. Μια οποιαδήποτε έκφραση της γλώσσας είναι μια πεπερασμένη ακολουθία συμβόλων της γλώσσας. Ορίζεται το πότε μια έκφραση είναι τύπος μιας συγκεκριμένης γλώσσας μέσω ενός επαγωγικού ορισμού. Όταν όλα τα σύμβολα της γλώσσας (για παράδειγμα σύνδεσμοι, ποσοδείκτες) έχουν την κανονική τους σημασία και αποδώσουμε μια πραγματικότητα στα βασικά σύμβολα της γλώσσας τότε η γλώσσα αυτή, μέσω της

ερμηνείας³¹ της, παίρνει νόημα. (26) Μοντέλο (model) ενός τύπου ονομάζεται μια ερμηνεία που καθιστά τον συγκεκριμένο τύπο αληθή. (30)

Το νόημα ενός τύπου ορίζεται ως προς έναν αφηρημένο κόσμο (“abstract world”) που ονομάζεται περιοχή (domain) ή σύμπαν αναφοράς (universe of discourse). Δοσμένης μιας ερμηνείας των όρων³² που τον αποτελούν, το νόημα αυτό είναι είτε αληθές είτε ψευδές. Όταν καθορίσουμε σε ποιες οντότητες αναφέρονται οι όροι³³ και ποιοι τύποι είναι αληθείς, είμαστε σε θέση να προσδιορίζουμε την τιμή αληθείας σύνθετων (composite) τύπων. (31) Η ύπαρξη των μεταβλητών στη Λογική Πρώτης Τάξης είναι το στοιχείο που επιτρέπει την αναπαράσταση της «γενικής» γνώσης. (4)

Για τη Λογική Πρώτης Τάξης υπάρχουν ακριβή και πλήρη συστήματα αποδείξεων, κάτι που δε διαθέτουν πιο εκφραστικές Λογικές (οι Λογικές υψηλότερης τάξης). Η Λογική Πρώτης τάξης, καθώς και οι ειδικές περιπτώσεις της, είναι μονοτονικές, υπό την έννοια ότι αν εξαχθεί κάποιο συμπέρασμα, τότε αυτό παραμένει έγκυρο ακόμα και αν προκύψει νέα γνώση. (3)

Στα θετικά στοιχεία της Λογικής Πρώτης Τάξης συγκαταλέγεται το ότι υπάρχει μια αντιστοιχία αυτής με τη φυσική γλώσσα. Επίσης, εμφανίζει μια ικανοποιητική έκφραση ποσοτικοποίησης των εννοιών με τους κατάλληλους ποσοδείκτες και έχει την ικανότητα να αντιληφθεί τη γενικότητα. Όμως, υπάρχει μια αδυναμία στην έκφραση ασάφειας, καθώς κάθε μια πρόταση μπορεί να είναι είτε αληθής ή ψευδής. Επίσης, το κάθε εξαγόμενο συμπέρασμα προστίθεται στη γνώση χωρίς να δίνεται η δυνατότητα αναθεώρησης του αν στη πορείαδειχθεί ότι είναι εσφαλμένο.

Η γλώσσα της Λογικής Πρώτης Τάξης είναι δομημένη γύρω από τα αντικείμενα και τις σχέσεις. Η κύρια διαφορά μεταξύ της Προτασιακής Λογικής και της Λογικής Πρώτης Τάξης έχει σχέση με την οντολογική δέσμευση (ontological commitment), δηλαδή τις παραδοχές σε σχέση με τη φύση της πραγματικότητας που

³¹ Ερμηνεία (Interpretation): Ορίζει την περιοχή της γλώσσας, αντιστοιχίζει όρους της γλώσσας σε οντότητες της περιοχής και αποδίδει σε κάθε (κλειστό) τύπο μια τιμή αληθείας (κλειστός τύπος: ο τύπος που δε περιέχει καμία μεταβλητή (26)). (30)

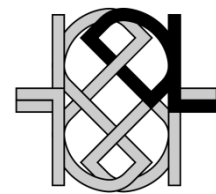
³² Ένας όρος στη γλώσσα, αναφέρεται σ' ένα στοιχείο (individual) του σύμπαντος αναφοράς. (31)

³³ Αυτό γίνεται με την ανάθεση όρων (term assignment), η οποία είναι μια απεικόνιση όρων σε “οντότητες” (entities) του σύμπαντος. (31)

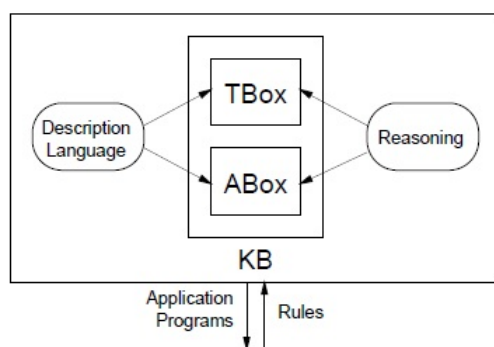
γίνεται από κάθε γλώσσα. Για παράδειγμα η Προτασιακή Λογική δέχεται ότι υπάρχουν γεγονότα που είτε ισχύουν ή δεν ισχύουν στον κόσμο (κάθε γεγονός είναι είτε αληθές ή ψευδές), ενώ η Λογική Πρώτης Τάξης δέχεται πως ο κόσμος αποτελείται από αντικείμενα με κάποιες σχέσεις μεταξύ τους, οι οποίες είτε ισχύουν ή δεν ισχύουν. Δηλαδή η Λογική Πρώτης Τάξης κάνει περισσότερες παραδοχές. (4)

3.1.3. Περιγραφικές Λογικές

Με τον όρο Περιγραφικές Λογικές (Description Logics - DL) αναφερόμαστε σε μια δομημένη μέθοδο αναπαράστασης γνώσης, με τη βοήθεια της οποίας μπορούμε να αναπαραστήσουμε έννοιες και συσχετίσεις ενός πεδίου εφαρμογής με φορμαλιστικό τρόπο που επιτρέπει εξαγωγή συμπερασμάτων μέσω συλλογιστικής. Οι Περιγραφικές Λογικές αποτελούν υποσύνολα της Λογικής Πρώτης Τάξης και είναι εφοδιασμένες με τυπικά ορισμένη σημασιολογία, η οποία είναι προσανατολισμένη στις διαδικασίες συμπερασμού (π.χ., έλεγχος συνέπειας, ικανοποιησιμότητας και στιγμιότυπων).



Εικόνα 12:
Λογότυπο των
Περιγραφικών
Λογικών (Enrico
Franconi).



Εικόνα 13: Βάση γνώσης σε Περιγραφικές
Λογικές

Τα βασικά στοιχεία που απαρτίζουν μια τέτοια γλώσσα είναι οι έννοιες, οι ρόλοι και τα χαρακτηριστικά (TBox), καθώς και τα στιγμιότυπά τους (ABox). Το TBox περιέχει τα αξιώματα που ορίζουν τις κλάσεις και τις συσχετίσεις αυτών σε μια οντολογία (Terminology box, σώμα ορολογίας). Το ABox περιέχει τους ισχυρισμούς σχετικά με τα άτομα του κόσμου και τις μεταξύ τους συσχετίσεις (Assumption box, σώμα ισχυρισμών). Μια βάση γνώσης σε Περιγραφικές Λογικές ορίζεται ως το ζεύγος $K=(T,A)$ (K: Βάση γνώσης), όπου T είναι ένα TBox και A ένα ABox. (32)

Σχεδόν όλες οι γενικεύσεις έχουν εξαιρέσεις ή ισχύουν μόνο μέχρι κάποιο βαθμό, κάτι που δεν επιτρέπει σε όλες τις πτυχές του πραγματικού κόσμου να μπορούν να αποτυπωθούν με τη λογική πρώτης τάξης εύκολα. Στα τυπικά συστήματα Λογικής Πρώτης Τάξης, η πρόβλεψη του χρόνου λύσης αρκετές φορές είναι αδύνατη, καθώς υπάρχουν σύνολα προτάσεων τα οποία αναγκάζουν το σύστημα να χρειάζεται εβδομάδες για να λύσει το πρόβλημα. Συχνά επαφίεται στα χέρια του χρήστη η επεξεργασία της περιγραφής, ώστε να παρακαμφθούν τέτοια τα σύνολα προτάσεων.

Η έρευνα για την εύρεση χρήσιμων υποσυνόλων της λογικής πρώτης τάξης για τα οποία ο συμπερασμός θα είναι υπολογιστικά βατός, οδήγησε στην ανάπτυξη των Περιγραφικών Λογικών. Οι Περιγραφικές Λογικές δίνουν έμφαση στη βατότητα του συμπερασμού. Για την επίλυση ενός στιγμιότυπου ενός προβλήματος, γίνεται πρώτα η περιγραφή και μετά ο έλεγχος για τον αν υπάγεται σε κάποια από τις δυνατές κατηγορίες λύσεων. Θέλουμε να εξασφαλίσουμε ότι ο έλεγχος της υπαγωγής μπορεί να επιλυθεί σε πολυωνυμικό χρόνο ως προς το μέγεθος των περιγραφών. Όμως αυτό μας αφήνει ουσιαστικά δύο επιλογές: τα δύσκολα προβλήματα είτε δε θα μπορούν καθόλου να διατυπωθούν είτε θα απαιτούν εκθετικά μεγάλες περιγραφές. (29)

Στις Περιγραφικές Λογικές, ακολουθείται ο εξής συμβολισμός:

- $\neg$: Άρνηση
- T : Το σύνολο των στοιχείων του συνόλου
- $\perp$: Το κενό σύνολο
- $\forall$: Για όλα τα στοιχεία (γενικός περιορισμός)
- $\exists$: Ύπαρξη ενός στοιχείου (υπαρξιακός περιορισμός)
- $\equiv$: Ταύτιση
- $\sqcap$: Τομή
- $\sqcup$: Ένωση
- $\leq$: Το-πολύ (περιορισμός πληθικότητας)
- $\geq$: Το-λιγότερο (περιορισμός πληθικότητας)

ανάλογα με το υποσύνολο το οποίο χρησιμοποιείται από τα παραπάνω σύμβολα, καθορίζεται και η εκφραστική ισχύς της περιγραφικής γλώσσας που αυτά ορίζουν. (32)

3.2. *Οντολογίες*

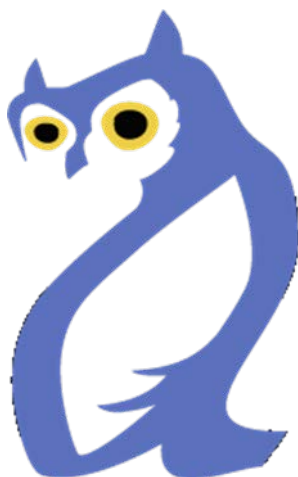
3.2.1. *OWL 1*

Από το ξεκίνημα του Σημασιολογικού Ιστού η ανάπτυξη γλωσσών για τη μοντελοποίηση οντολογιών ήταν ένας πολύ βασικός στόχος. Οι αρχικές προσπάθειες είχαν επικεντρωθεί στην RDF και στην RDFS αλλά αποδείχτηκε σύντομα πως και οι δύο είχαν μια αρκετά περιορισμένη εκφραστική δύναμη. (33)

Οι γλώσσες RDF και RDFS επιτρέπουν την αναπαράσταση ενός μέρους της οντολογικής γνώσης. Τις χαρακτηρίσαμε ως στοιχειώδεις γλώσσες οντολογιών. Τα κύρια θεμελιώδη στοιχεία μοντελοποίησης των RDF/RDFS αφορούν την οργάνωση των λεξιλογίων σε τυποποιημένες ιεραρχίες: σχέσεις υποκλάσης και υποϊδιότητας, περιορισμούς πεδίου ορισμού και συνόλου τιμών, καθώς και στιγμιότυπα κλάσεων. Όμως λείπουν αρκετές άλλες δυνατότητες, καθώς έχουμε:

- ❖ Τοπική εμβέλεια ιδιοτήτων: στην RDFS δε μπορούμε να δηλώσουμε περιορισμούς στο σύνολο τιμών, οι οποίοι θα ισχύουν μόνο για μερικές κλάσεις.
- ❖ Μη επικάλυψη κλάσεων: στην RDFS μπορούμε να δηλώσουμε μόνο σχέσεις υποκλάσεων και όχι σχέσεις κλάσεων.
- ❖ Περιορισμοί πληθικότητας: στην RDFS δε μπορούμε να επιβάλλουμε περιορισμούς στο πλήθος των διακριτών τιμών που μπορεί ή πρέπει να πάρει μια ιδιότητα.
- ❖ Ειδικά χαρακτηριστικά ιδιοτήτων: δε μπορούμε να δηλώσουμε τη μεταβατικότητα, τη μοναδικότητα ή την αντιστροφή μιας ιδιότητας. (3)

Συνεπώς, χρειαζόμαστε μια πιο πλούσια από την RDFS γλώσσα. Όμως θα πρέπει να επιλέξουμε μια που θα υποστηρίζεται από σχετικά αποδοτικά προγράμματα συλλογιστικής, ενώ θα είναι επαρκώς εκφραστική (για να μπορεί να εκφράζει μεγάλες κλάσεις οντολογιών και γνώσης), καθώς υπάρχει η πιθανότητα αν χρησιμοποιούμε μια πολύ εκφραστική γλώσσα να ξεπεράσουμε το όριο της μη υπολογισιμότητας. (3)



Εικόνα 14: Λογότυπο OWL

Μέσω του W3C σχηματίστηκε η ομάδα “Web Ontology Working Group” η οποία είχε ως κύριο σκοπό την ανάπτυξη μιας εκφραστικής γλώσσας που θα ήταν κατάλληλη για δημοσιεύσεις στο Σημασιολογικό Ιστό. Το αποτέλεσμα αυτής της προσπάθειας ήταν η OWL (Ontology Web Language), η οποία έγινε Σύσταση του W3C τον Φεβρουάριο του 2004. (33)

Οι Περιγραφικές Λογικές αποτελούν βάση της OWL. Απλουστευμένα θα μπορούσαμε να πούμε πως η OWL είναι η Γλώσσα Σήμανσης (Markup Language) που περιγράφει κανείς στον υπολογιστή (για να καταλαβαίνει) ό,τι περιγράφει σε περιγραφική λογική. Η OWL βασίζεται στις γλώσσες RDF και RDFS και χρησιμοποιεί τη σύνταξη της RDF που βασίζεται στην XML. Για να μπορέσει κάποιος να γράψει οντολογίες που να καταλαβαίνει ο υπολογιστής, αρκεί να ξέρει περιγραφική λογική και από κει και πέρα με την χρήση ενός Ontology Editor αναπαράγεται αυτόματα ο κώδικας σε OWL.

Μια οντολογία ορίζει τους όρους που χρησιμοποιούνται για την περιγραφή και την αναπαράσταση ενός πεδίου γνώσης. Αποτελείται από χρησιμοποιήσιμους από τη μηχανή ορισμούς των βασικών εννοιών στο πεδίο και των μεταξύ τους σχέσεων. Συγκεκριμένα μια οντολογία παρέχει περιγραφές για:

- Κλάσεις (γενικές έννοιες) στο πεδίο αναφοράς.
- Ιδιότητες (ή χαρακτηριστικά) που περιγράφουν τις κλάσεις και συχνά τις συσχετίζουν μεταξύ τους.
- Στιγμιότυπα των κλάσεων αυτών.

Οι γλώσσες οντολογιών επιτρέπουν στους χρήστες να γράφουν ρητές, τυπικές επίνοιες (conceptualizations) για μοντέλα πεδίων. Οι κύριες απαιτήσεις είναι: καλά ορισμένη σύνταξη (αναγκαία συνθήκη για την επεξεργασία πληροφοριών από υπολογιστές), τυπική σημασιολογία (η σημασιολογία να μην αναφέρεται σε υποκειμενικές αντιλήψεις, ούτε επιδέχεται διαφορετικές ερμηνείες από διαφορετικά άτομα ή υπολογιστές), επαρκής εκφραστική ισχύ (θέλουμε εκφραστικότητα εκεί που τη χρειαζόμαστε, ο οποία εξαρτάται από το υποσύνολο των συμβόλων που ορίζει την γλώσσα μας), ευκολία στην έκφραση (να θεωρείται ως πιο εύκολη μέθοδος αναπαράστασης γνώσης συγκριτικά με άλλες).

Η σημασιολογία αποτελεί προϋπόθεση για την υποστήριξη συλλογισμών. Για την οντολογική γνώση μπορούν να γίνουν συλλογισμοί σχετικά με τη συμμετοχή σε κλάσεις, την ισοδυναμία κλάσεων, την συνέπεια και την ταξινόμηση. Η υποστήριξη συλλογισμών είναι σημαντική επειδή επιτρέπει:

- ❖ Τον έλεγχο της συνέπειας της οντολογίας και της γνώσης,
- ❖ Τον έλεγχο ανεπιθύμητων σχέσεων ανάμεσα σε κλάσεις,
- ❖ Την αυτόματη ταξινόμηση των στιγμιότυπων σε κλάσεις

Με την αυτόματη υποστήριξη συλλογισμών μπορούμε να ελέγξουμε πολύ περισσότερες περιπτώσεις από αυτές που θα μπορούσαν να ελεγχθούν με μη αυτόματο τρόπο. Οι περιγραφικές λογικές αποτελούν ένα υποσύνολο της κατηγορηματικής λογικής για το οποίο είναι δυνατό να υπάρξει αποδοτική υποστήριξη συλλογισμών.

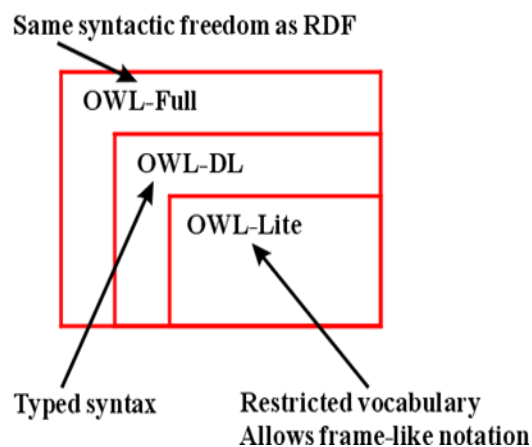
Η OWL δεν αποτελεί «άμεση» επέκταση της RDFS καθώς κάτι τέτοιο θα λειτουργούσε ενάντια στην εκφραστική ισχύ και την αποδοτική συλλογιστική. Το πλήρες σύνολο των απαιτήσεων για μια γλώσσα οντολογιών μοιάζει ανέφικτο: αποδοτική υποστήριξη συλλογισμών και ευκολία στην έκφραση για μια γλώσσα τόσο ισχυρή όσο είναι ο συνδυασμός της RDFS με μια πλήρη λογική. Οι απαιτήσεις αυτές οδήγησαν την Ομάδα Εργασίας Οντολογιών Ιστού του οργανισμού W3C να ορίσει την OWL (OWL 1.0) ως τρεις διαφορετικές υπογλώσσες³⁴, κάθε μια από τις οποίες έχει προσαρμοστεί να ανταποκρίνεται σε διαφορετικές πτυχές του πλήρους συνόλου των απαιτήσεων.

OWL Full: Είναι η πλήρης γλώσσα, η οποία χρησιμοποιεί όλα τα γλωσσικά θεμελιώδη στοιχεία της OWL και επιτρέπει το συνδυασμό τους με οποιονδήποτε αυθαίρετο τρόπο, χρησιμοποιώντας τις γλώσσες RDF και RDFS, με μοναδικό περιορισμό το παραγόμενο έγγραφο να είναι έγκυρο σε RDF. Συνεπώς, όπως και στην RDF, υπάρχει η δυνατότητα αλλαγής του νοήματος των προκαθορισμένων θεμελιωδών στοιχείων (της RDF ή της OWL), με την εφαρμογή των γλωσσικών θεμελιωδών στοιχείων μεταξύ τους. Στα θετικά της στοιχεία ανήκει το ότι είναι πλήρως συμβατή προς τα επάνω (up-ward compatible) με την RDF, τόσο συντακτικά όσο και σημασιολογικά. Δηλαδή, κάθε έγκυρο έγγραφο RDF είναι ταυτόχρονα και

³⁴ Οι υπογλώσσες είναι προφίλ της γλώσσας. Αποτελούν περικομμένες εκδόσεις της γλώσσας οι οποίες ανταλλάσσουν κάποια από την εκφραστική ισχύ για την ανάκτηση της αποδοτικότητας των συλλογισμών. (33)

έγκυρο έγγραφο OWL Full, και κάθε έγκυρο συμπέρασμα σε RDF/RDFS είναι επίσης έγκυρο συμπέρασμα σε OWL. Αλλά η γλώσσα έχει γίνει τόσο ισχυρή ώστε δημιουργεί πλέον διλήμματα στις αποφάσεις, εκμηδενίζοντας κάθε ελπίδα πλήρους (ή αποδοτικής) υποστήριξης συλλογισμών.

OWL DL: Μια υπογλώσσα της OWL Full, που περιορίζει τον τρόπο χρήσης των «κατασκευαστών» (constructors) των OWL και RDF, με σκοπό να αξιοποιηθεί υπολογιστική επιλυσιμότητα της περιγραφικής λογικής. Ουσιαστικά, δεν επιτρέπει την εφαρμογή ενός «κατασκευαστή» της OWL σε έναν άλλον, διασφαλίζοντας έτσι ότι η γλώσσα αντιστοιχεί σε μια καλά ορισμένη Περιγραφική Λογική. Επιτρέπεται η αποδοτική υποστήριξη συλλογισμών, αλλά χάνεται η πλήρης συμβατότητα με την RDF. Ένα έγγραφο RDF θα χρειαστεί γενικά να επεκταθεί με ορισμένους τρόπους και να περιοριστεί με άλλους για να θεωρηθεί έγκυρο έγγραφο OWL DL. Κάθε έγκυρο έγγραφο OWL DL είναι και έγκυρο έγγραφο RDF.



Εικόνα 15: Υπογλώσσες της OWL 1

OWL Lite: Σε σχέση με την OWL DL θα πρέπει να ικανοποιεί ακόμα περισσότερους περιορισμούς. Αποκλείει τις απαριθμητές κλάσεις (enumerated classes), τις προτάσεις μη επικάλυψης (disjointness statements) και την αυθαίρετη πληθικότητα (arbitrary cardinality). Αυτό είναι ένας πρόσθετος περιορισμός που περιορίζει την OWL DL σε ένα υποσύνολο των «κατασκευαστών» της γλώσσας. Η γλώσσα είναι ευκολότερη στην κατανόηση (για τους χρήστες) αλλά και στην υλοποίηση (για τους κατασκευαστές εργαλείων). Άλλα έχει περιορισμένη εκφραστικότητα.

Οι προγραμματιστές οντολογιών που υιοθετούν την OWL θα πρέπει να αναλογιστούν ποια υπογλώσσα ταιριάζει καλύτερα στις ανάγκες τους:

- OWL Lite – OWL DL: η OWL Lite είναι πιο εκφραστική από την OWL DL.
- OWL DL – OWL Full: Η OWL Full προσφέρει τις λειτουργίες μετα-μοντελοποίησης της RDFS αλλά οι πλήρεις υλοποιήσεις σε OWL Full δεν

είναι εφικτές, καθώς η υποστήριξη συλλογισμών είναι λιγότερο προβλέψιμη, συγκριτικά με την OWL DL.

Υπάρχουν αυστηρά πλαίσια συμβατότητας προς τα επάνω ανάμεσα σε αυτές τις τρεις υπογλώσσες:

- ✓ Κάθε έγκυρη οντολογία OWL Lite είναι και έγκυρη οντολογία OWL DL
- ✓ Κάθε έγκυρη οντολογία OWL DL είναι και έγκυρη οντολογία OWL Full
- ✓ Κάθε έγκυρο συμπέρασμα OWL Lite είναι και έγκυρο συμπέρασμα OWL DL
- ✓ Κάθε έγκυρο συμπέρασμα OWL DL είναι και έγκυρο συμπέρασμα OWL Full

Από τις υπογλώσσες που αναφέραμε πιο πάνω, η μόνη συμβατή προς τα κάτω για την OWL, είναι η OWL Full. Στη συγκεκριμένη, κάθε εργαλείο επεξεργασίας OWL θα παρέχει σωστές ερμηνείες για οποιοδήποτε έγγραφο RDFS με τίμημα την υπολογιστική δυσεπιλυσιμότητα.

Η Περιγραφική Λογική είναι τμήμα της Λογικής Πρώτης Τάξης. Αφού η OWL βασίζεται στη Περιγραφική Λογική, κληρονομεί πολλές από τις ιδιότητες της συγκεκριμένης Λογικής Πρώτης Τάξης. Ειδικότερα, κληρονομεί την υπόθεση του ανοιχτού κόσμου. Σύμφωνα με την υπόθεση του ανοιχτού κόσμου (open-world assumption), δε μπορούμε να συνάγουμε το ψεύδος από την απουσία της αλήθειας (μια πρόταση δε μπορεί να θεωρηθεί αληθής λόγω απουσίας της απόδειξής της). Με την αντίθετη υπόθεση (υπόθεση του κλειστού κόσμου – closed-world assumption) μπορούμε να συνάγουμε το ψεύδος εξ' αιτίας της αδυναμίας να συνάγουμε την αλήθεια (μια πρόταση είναι αληθής όταν δε μπορεί να αποδειχθεί η άρνησή της). Υπάρχει, επίσης, η υπόθεση του μοναδικού ονόματος (unique-name assumption), σύμφωνα με την οποία δύο μεμονωμένα στοιχεία με διαφορετικά ονόματα, είναι στη πραγματικότητα δύο διαφορετικά μεμονωμένα στοιχεία.

Συστήματα όπως οι βάσεις δεδομένων και τα συστήματα λογικού προγραμματισμού έχουν την τάση να υποστηρίζουν τους κλειστούς κόσμους και τα μοναδικά ονόματα, ενώ συστήματα αναπαράστασης γνώσης και απόδειξης θεωρημάτων υποστηρίζουν τους ανοικτούς κόσμους και τα μη μοναδικά ονόματα. Η OWL χρησιμοποιεί την υπόθεση του ανοιχτού κόσμου, ενώ δε χρησιμοποιεί την υπόθεση του μοναδικού ονόματος. Όμως οι οντολογίες χρειάζονται μερικές φορές το πρώτο ζευγάρι και άλλες φορές το δεύτερο.

Αυτά οδήγησαν στη δημιουργία της OWL DLP, μιας υπογλώσσας της OWL DL, η οποία δεν ανήκει στα επίσημα είδη OWL του οργανισμού W3C. Η συγκεκριμένη γλώσσα είναι αρκετά αδύναμη ώστε να μη παρατηρούνται διαφορές ανάμεσα σε αυτές τις δύο επιλογές. Επιτρέπει τη χρήση κατασκευαστών της OWL, όπως η ισοδυναμία κλάσεων και ιδιοτήτων (η ισότητα), η ανισότητα μεταξύ μεμονωμένων στοιχείων (οι αντίστροφες, μεταβατικές, συμμετρικές και συναρτησιακές ιδιότητες) και η τομή κλάσεων, αλλά όχι την τομή και τους αυθαίρετους περιορισμούς πληθικότητας. Αυτά μας παρέχουν χρήσιμη εκφραστικότητα σε πολλές πρακτικές περιπτώσεις και εγγυώνται τη σωστή ανταλλαγή ανάμεσα στα προγράμματα συλλογιστικής της OWL, ανεξάρτητα από τις υποθέσεις του κλειστού κόσμου και του μοναδικού ονόματος. Το θετικό είναι ότι οι άνθρωποι ή οι εφαρμογές που επιθυμούν να κάνουν διαφορετικές επιλογές σχετικά με τις υποθέσεις αυτές, μπορούν χωρίς πρόβλημα να ανταλλάσουν οντολογίες σε OWL DLP. Αλλά αν «ξεφύγουν» από τα πλαίσια της OWL DLP θα παρατηρήσουν ενδεχομένως διαφωνία στη σημασιολογία, δηλαδή μπορεί να εξάγονται διαφορετικά συμπεράσματα από τις ίδιες προτάσεις. (3)

3.2.2. Από την OWL 1 στην OWL 2

Η τυποποίηση της OWL έδωσε το έναυσμα για την ανάπτυξη και την υιοθέτηση ενός πλήθους εκλογικευτών (reasoners)³⁵ και οντολογικών επεξεργαστών (ontology editors). Οντολογίες της OWL έχουν χρησιμοποιηθεί σε πολυποίκιλους τομείς, αποκαλύπτοντας όμως ταυτόχρονα και κάποιες ελλείψεις στον αρχικό σχεδιασμό της. Αυτό οδήγησε στην ιδέα για τη δημιουργία μιας επανέκδοσης της OWL, την OWL 1.1.

Τα προβλήματα που είχαν παρουσιαστεί στο σχεδιασμό της OWL 1 δεν ήταν σημαντικά, αλλά ήταν αρκετά για να οδηγήσουν σε μια αναθεώρηση της OWL 1. Τα προβλήματα αυτά περιελάμβαναν περιορισμούς εκφραστικότητας και ζητήματα με τη σύνταξη. Η OWL 1.1 θα είχε ως σκοπό την διευθέτηση ορισμένων από αυτών των

³⁵ Εκλογικευτής (reasoner) είναι ένας αλγόριθμος που αναλαμβάνει την επίλυση αντιφάσεων και τη εξαγωγή νέων ισχυρισμών πάνω σε μια βάση γνώσης.

προβλημάτων, μέσω της εκμετάλλευσης πρόσφατων εξελίξεων στο τομέα της έρευνας των Περιγραφικών Λογικών για την μερική αντιμετώπιση των περιορισμών στην εκφραστικότητα της γλώσσας. Το σύνολο των νέων χαρακτηριστικών που συμφωνήθηκε πως θα έχει αυτή η γλώσσα ήταν η τομή των όσων ήθελαν οι χρήστες με το τι πίστευαν οι θεωρητικοί πως ήταν πιθανό και τι πίστευαν οι δημιουργοί των εφαρμογών πως ήταν πρακτικό. Όσο ο σχεδιασμός της OWL 1.1 προχωρούσε, έγινε εκκαθάριση της γλώσσας και των προδιαγραφών της με σκοπό να παραχθεί μια πιο εύρωστη πλατφόρμα για πιθανή μελλοντική ανάπτυξη.

Αρχικά η ανάπτυξη της OWL 1.1 γινότανε από μια ανεπίσημη ομάδα από χρήστες της γλώσσας και προγραμματιστές. Όταν οι προδιαγραφές της γλώσσας έγιναν σταθερές και κυκλοφόρησαν οι πρώτες υλοποιήσεις, η πρόταση της OWL 1.1 υποβλήθηκε στο W3C με την πρόθεση αυτή να αποτελέσει την αφορμή για τη δημιουργία μιας νέας ομάδας “W3C Working Group”. Τελικά, η ομάδα εργασίας δημιουργήθηκε επίσημα το Σεπτέμβριο του 2007. Η νέα γλώσσα κατάφερε να εξελιχθεί πολύ γρήγορα, κάτι που οδήγησε την Ομάδα Εργασίας να αποφασίσει τελικά τον Απρίλιο του 2008 να την ονομάσει OWL 2, υποδηλώνοντας με αυτό τον τρόπο ένα σημαντικό βήμα στην εξέλιξη της γλώσσας. (33)

3.2.3. OWL 2

Από τις 27 Οκτωβρίου του 2009, η OWL 2 αποτελεί το πρότυπο για την περιγραφή οντολογιών στο Σημασιολογικό Ιστό. (13) Η OWL 2 προσθέτει νέες λειτουργίες σε σχέση με την OWL 1. Μερικά από τα νέα χαρακτηριστικά δεν έχουν κάποια ουσιαστική χρησιμότητα, ενώ κάποια άλλα προσφέρουν επιπλέον εκφραστικότητα, μεταξύ των οποίων: κλειδιά, αλυσίδες ιδιοτήτων, πλουσιότεροι τύποι δεδομένων και πεδία ορισμού, συγκεκριμένοι περιορισμοί πληθικότητας, ασύμμετρες, ανακλαστικές και ξένες ιδιότητες και ενισχυμένες ιδιότητες σχολιασμού. (34)

Οι οντολογίες και τα στοιχεία τους προσδιορίζονται χρησιμοποιώντας Internationalized Resource Identifiers³⁶ (IRIs). Έτσι, η OWL 2 επεκτείνει την OWL 1, η οποία χρησιμοποιεί Uniform Resource Identifiers (URIs). Κάθε URI πρέπει να είναι απόλυτο και όχι σχετικό. Δύο URIs είναι δομικά ισοδύναμα αν και μόνο αν οι αλφαριθμητικές αναπαραστάσεις τους είναι ταυτόσημες. (13)

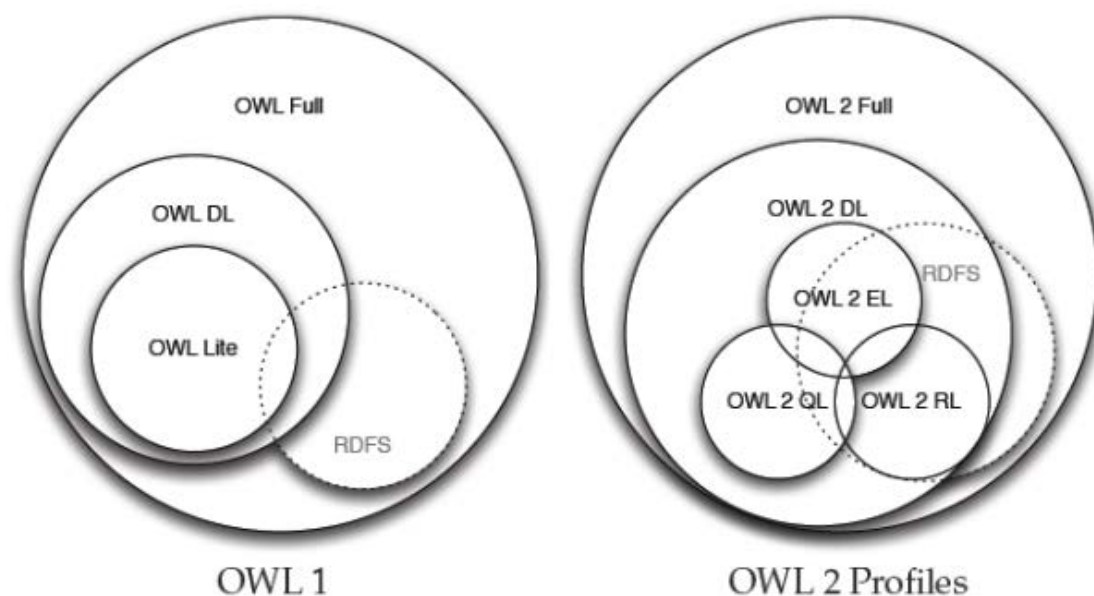
Για τη διευθέτηση των προβλημάτων που αναφέρθηκαν σε προηγούμενο κεφάλαιο, η OWL 2 μας παρέχει τρία νέα προφίλ (υπογλώσσες): την OWL 2 EL, την OWL 2 QL, και την OWL 2 RL. Το κάθε ένα από αυτά προσδιορίζεται εφαρμόζοντας περιορισμούς στη Συναρτησιακού Τύπου Σύνταξη (Functional-Style Syntax) της OWL 2, μας παρέχει διαφορετική εκφραστική ισχύ και στοχεύει σε διαφορετικές εφαρμογές. Οποιαδήποτε οντολογία γράφεται σε κάποιο από αυτά τα προφίλ είναι μια έγκυρη OWL 2 οντολογία. (33)

OWL 2 EL: Υποσύνολο της OWL 2 για επιτάχυνση της συλλογιστικής. Είναι ιδιαίτερα χρήσιμη για εφαρμογές που έχουν οντολογίες με μεγάλο αριθμό κλάσεων ή και ιδιοτήτων. Περιλαμβάνει την εκφραστική δύναμη που χρησιμοποιείται από πολλές παρόμοιες οντολογίες. Η οντολογική συνέπεια, τα αξιώματα περιγραφής των κλάσεων και ο έλεγχος των στιγμιότυπων μπορούν να αποφασιστούν σε πολυωνυμικό χρόνο.

OWL 2 QL: Υποσύνολο της OWL 2 για πρόσβαση σε δεδομένα αποθηκευμένα σε Βάσεις Δεδομένων. Είναι χρήσιμη σε εφαρμογές που χρησιμοποιούν πολύ μεγάλο όγκο στιγμιότυπων δεδομένων (instance data) και όποτε η απάντηση των ερωτημάτων είναι η πιο σημαντική συλλογιστική διαδικασία. Πιο συγκεκριμένα, αυτό το προφίλ αποτελεί την τομή της RDFS και της OWL 2 DL³⁷. Έχει σχεδιαστεί με τέτοιο τρόπο, ώστε η ορθή και πλήρης διαδικασία της απάντησης ενός ερώτημα (query) ανήκει στο LOGSPACE (ή με μεγαλύτερη ακρίβεια στο AC⁰) σε σχέση με το μέγεθος των δεδομένων (ισχυρισμών).

³⁶Internationalized Resource Identifiers (IRIs): Αποτελεί μια ακολουθία χαρακτήρων από το Universal Character Set, ένα μεγάλο σύνολο χαρακτήρων που καλύπτει πολλά αλφάβητα και έχει ενιαία κωδικοποίηση (Unicode/ISO10646). (165)

³⁷OWL 2 DL: έχει κατασκευαστεί αντίστοιχα με την OWL 1 DL.



Εικόνα 16: OWL 1 και OWL 2 profiles

OWL 2 RL: Υποσύνολο της OWL 2 που επιτρέπει συλλογιστική μέσω της χρήσης κανόνων σε OWL (forward-chaining rules). Απευθύνεται σε εφαρμογές που απαιτούν κλιμακωτή συλλογιστική (scalable reasoning) χωρίς να θυσιάζουν την εκφραστική δύναμη. (13) Έχει σχεδιαστεί για εφαρμογές της OWL 2, οι οποίες είναι σε θέση να ανταλλάξουν την πλήρη εκφραστικότητα της γλώσσας για την αποδοτικότητα, καθώς και για τις RDF(S) εφαρμογές οι οποίες χρειάζονται κάποια επιπρόσθετη εκφραστικότητα από την OWL 2. (Περισσότερες πληροφορίες για την πολυπλοκότητα των υπογλωσσών βρίσκονται στο Παράρτημα 6.1).

Εκτός από όσα προσδιορίστηκαν παραπάνω, υπάρχουν και άλλα πιθανά προφίλ της OWL 2. Ένα παράδειγμα που μπορούμε να αναφέρουμε είναι πως υπάρχει μια ολόκληρη οικογένεια προφίλ η οποία επεκτείνει την OWL QL. Επίσης, η OWL Lite μπορεί να θεωρηθεί προφίλ της OWL 2, επειδή κάθε οντολογία OWL Lite είναι και οντολογία OWL 2. Ένα τελευταίο παράδειγμα είναι η OWL 1 DL, η οποία για αντίστοιχο λόγο με την OWL Lite μπορεί να θεωρηθεί προφίλ της OWL 2. (13)

Οι βελτιώσεις που έγιναν στις προδιαγραφές της OWL 2, διευκόλυναν στην υλοποίηση εργαλείων που υποστηρίζουν την OWL 2 και την επιτακτική ανάγκη για την επιπρόσθετη λειτουργικότητα στη πράξη, που παρέχεται από την OWL 2. (33) Επίσης, με την OWL 2 έγινε εισαγωγή ενός επιπλέον συντακτικού. Το Manchester OWL είναι ένα φιλικό προς τον χρήστη συντακτικό για OWL 2 περιγραφές, αλλά μπορεί να χρησιμοποιηθεί και για να γραφούν ολόκληρες OWL 2 οντολογίες. Η αρχική έκδοση του Manchester OWL συντακτικού δημιουργήθηκε από την OWL 1

και αναβαθμίστηκε για την OWL 2. Το Manchester συντακτικό χρησιμοποιείται στο Protégé 4 και στο TopBraid Composer (εργαλεία πλήρως συμβατά με τα Πρότυπα του W3C που επιτρέπουν τη συγγραφή/επεξεργασία οντολογιών και τη δημιουργία σημασιολογικών εφαρμογών), κυρίως για εισαγωγή και προβολή περιγραφών συσχετισμένες με κατηγορίες. Μερικά εργαλεία (πχ το Protégé 4) επεκτείνουν τη σύνταξη για να επιτρέψουν ακόμα περισσότερη σύμπτυξη στην αναπαράσταση (πχ για κατανόηση) ή για αντικατάσταση των IRIs από «τιμών-ετικετών» (label values). Όμως, δεν είναι δυνατόν να εκφραστούν όλες οι OWL και OWL 2 οντολογίες με αυτό. (13)

Η OWL 2 έχει υιοθετηθεί από διάφορους τομείς εφαρμογών, όπως η βιοπληροφορική, κλινικές επιστήμες, τη μηχανική και το δίκαιο. Αυτό οφείλεται, μέσα στα άλλα, στην ύπαρξη των επεξεργαστών (editors) και των εκλογικευτών (reasoners) για την OWL 2. Εκτιμάται ότι όσο βελτιώνονται τα εργαλεία της OWL 2, όλο και περισσότεροι χρήστες θα την χρησιμοποιήσουν. Επιπλέον αναμένεται πως η προτυποποίηση των προφίλ και η διαθεσιμότητα των μηχανών εκλογίκευσης, που επιτρέπουν ανιχνεύσιμες μορφές συλλογιστικής, θα διευκολύνουν τη χρήση της OWL από χρήστες που στην αρχή αποθαρρύνονταν από θέματα κλιμάκωσης και αποτελεσματικότητας. (33)

3.2.4. Πιθανή μελλοντική ανάπτυξη της OWL

Ο στόχος της ανάπτυξης της OWL 2 ήταν εκτός από την επίλυση των έντονων προβλημάτων που είχαν σχέση με την εκφραστικότητα, η παροχή μιας εύρωστης πλατφόρμας με σκοπό την δυνατότητα της περαιτέρω μελλοντικής ανάπτυξης. Παρακάτω παραθέτουμε κάποια τμήματα στα οποία υπάρχει δυνατότητα μελλοντικής ανάπτυξης της OWL 2:

Συντακτική επεκτασιμότητα (Syntactic extensibility): Έχουν υποβληθεί διάφορες προτάσεις για προσθήκη «συντακτικής ζάχαρης». Πιο συγκεκριμένα, θα μπορούσε να θεωρηθεί χρήσιμη, για παράδειγμα, η επέκταση της OWL με σύστημα “macro”, το οποίο θα επιτρέπει στους χρήστες τον προσδιορισμό των δικών τους συντακτικών συντομεύσεων.

Γλώσσες ερωτημάτων (Query languages): Οι εκφραστικές γλώσσες ερωτημάτων είναι πολύ σημαντικές για τις πρακτικές εφαρμογές της OWL. Έχουν προταθεί συνδυαστικά ερωτήματα (conjunctive queries), τα οποία έχουν επεκταθεί σε αλγεβρικές γλώσσες (algebraic languages). Τέτοια παραδείγματα αποτελούν η EQL-Lite και η nRQL. Η πρόκληση που έχει να αντιμετωπίσει η κοινότητα που αναπτύσσει την OWL έχει να κάνει με τον ακριβή προσδιορισμό των σχέσεων τόσο μεταξύ αυτών των προτάσεων μεταξύ τους, όσο με την SPARQL, την γλώσσα ερωτημάτων της RDF.

Αλληλεπίδραση με Κανόνες (Integration with rules): Μια από τις ομάδες εργασίας της RIF (Rules Interchange Format, Πρότυπο Ανταλλαγής Κανόνων) του W3C προσπαθεί να δημιουργήσει ένα πρότυπο ανταλλαγής για τους κανόνες στον Ιστό. Πολλές προσεγγίσεις για την ενσωμάτωση κανόνων με γλώσσες βασισμένες στη Περιγραφική Λογική έχουν προταθεί τα τελευταία χρόνια. Συνεπώς, η ενοποιημένη ανάπτυξη της RIF με την OWL είναι ένα πολύ σημαντικό έργο και για τις δύο Κοινότητες. Ένα σύνολο κανόνων που έχουν ιδιαίτερη σχέση είναι οι «DL-safe» κανόνες, οι οποίοι αποκαλούνται έτσι επειδή παρέχουν χρήσιμη εκφραστική δύναμη, ενώ παράλληλα επιτρέπουν πρακτικά εφικτές υλοποιήσεις.

Μη μονοτονικές επεκτάσεις (Nonmonotonic extensions): Οι μη μονοτονικές επεκτάσεις της OWL έχουν ζητηθεί από τους χρήστες, καθώς είναι πολύ σημαντικές για την επίλυση των πιο δύσκολων προβλημάτων σχεδιασμού, όπως για παράδειγμα εξαιρέσεις ή περιορισμούς τύπων παρόμοιους με αυτών των βάσεων δεδομένων. (33)

3.3. RuleML

3.3.1. Κανόνες

Οι Κανόνες έχουν χρησιμοποιηθεί στην θεωρητική επιστήμη των υπολογιστών, στη τεχνολογία των διερμηνευτών (compiler technology), τις βάσεις δεδομένων, τον λογικό προγραμματισμό και την τεχνητή νοημοσύνη. (35)

Χρησιμοποιούνται, επίσης, για πολλούς διασυνδεδεμένους σκοπούς, συλλαμβάνοντας τις παρατυπίες σε τομείς εφαρμογών όπως οι ακόλουθοι: (36)

- Μηχανική: Κανόνες Διάγνωσης (Diagnosis Rules)
 - Εμπόριο: Επιχειρηματικοί Κανόνες (Business rules)
 - Δίκαιο: Νομική Συλλογιστική (Legal reasoning)
 - Διαδίκτυο: Έλεγχος ταυτότητας για πρόσβαση (Access authentication)
- (35)

Γενικά θα μπορούσαμε να πούμε πως οι κανόνες αποτελούν μονάδες γνώσης, οι οποίες περιλαμβάνουν κάποιου είδους λογική. Μπορούν, για παράδειγμα, να καθορίζουν:

- Στατικούς ή δυναμικούς περιορισμούς ακεραιότητας (για παράδειγμα, για τον περιορισμό του χώρου καταστάσεων ή των ιστορικών των εκτελέσεων ενός συστήματος).
- Παραγωγές (για παράδειγμα, για τον ορισμό παραγόμενων εννοιών).
- Αντιδράσεις (για παράδειγμα, για τον καθορισμό της αντιδραστικής συμπεριφοράς ενός συστήματος ως απόκριση σε γεγονότα). (37)

Στο πλαίσιο του Σημασιολογικού Ιστού οι κανόνες μπορούν να κατασκευαστούν με βάση την F-Λογική³⁸ για συλλογιστική πάνω στην RDF. Μπορούν να χρησιμοποιηθούν για την ενίσχυση του περιεχομένου των Ιστοσελίδων και των XML εγγράφων με διάφορους τρόπους, όπως για παράδειγμα με τους κανόνες παραγωγής (derivation rules), οι οποίοι επιτρέπουν τη δυναμική συμπερίληψη των εξαγόμενων γεγονότων, ενώ οι κανόνες αντίδρασης επιτρέπουν την προδιαγραφή της συμπεριφοράς σε απόκριση σε συμβάντα του περιηγητή. (35)

³⁸ F-λογική (F-logic): Frame-logic: είναι μια μορφή αναπαράστασης γνώσης και μια γλώσσα οντολογιών που συνδυάζει τα θετικά στοιχεία της εννοιολογικής μοντελοποίησης με τις αντικειμενοστρεφείς γλώσσες βασισμένες σε πλαίσια, έχοντας μια δηλωτική, συμπαγή και απλή σύνταξη. (169)

3.3.2. RuleML

Η RuleML (Rule Markup Language, Γλώσσα Σήμανσης Κανόνων) χρησιμοποιείται για τον διαμοιρασμό βάσεων κανόνων (rule bases) στηριζόμενους στην XML και την δημοσίευση τους στον Ιστό. (38) Είναι μια σημαντική προσπάθεια τυποποίησης για τη σήμανση των κανόνων στον Ιστό. Δεν είναι μια γλώσσα μόνο σήμανσης, αλλά μια οικογένεια γλωσσών που αντιστοιχούν σε διαφορετικά είδη κανόνων. (3)



Εικόνα 17: Λογότυπο RuleML

Η Πρωτοβουλία RuleML, η οποία άρχισε τον Αύγουστο του 2000 κατά τη διάρκεια Διεθνούς Συνεδρίου χωρών του δακτυλίου του Ειρηνικού σχετικά με την Τεχνητή Νοημοσύνη (Pacific Rim International Conference on Artificial Intelligence, PRICAI 2000), αναπτύσσει μία ανοικτή, ουδέτερη από επιχειρήσεις γλώσσα κανόνων, η οποία είναι βασισμένη σε XML/RDF. Αυτό θα επιτρέψει την ανταλλαγή των κανόνων μεταξύ των διάφορων συστημάτων συμπεριλαμβανομένων των κατανεμημένων τμημάτων λογισμικού στο Διαδίκτυο, των ετερογενών client-server συστημάτων που βρίσκονται μέσα στις μεγάλες εταιρίες κ.λ.π. Η γλώσσα RuleML προσφέρει τη σύνταξη της XML για την αναπαράσταση γνώσης κανόνων, η οποία είναι διαλειτουργική μεταξύ σημαντικών εμπορικών και μη εμπορικών συστημάτων κανόνων. (35)

Το Διεθνές αυτό Συνέδριο συγκέντρωσε ομάδες ειδικών από διάφορες χώρες, συμπεριλαμβανομένων ηγετών από τον ακαδημαϊκό χώρο, καθώς και από τη βιομηχανία στις γλώσσες αναπαράστασης γνώσης και σήμανσης. Ανάμεσα στους συμμετέχοντες από τον χώρο της βιομηχανίας, υπήρχαν και προμηθευτές μηχανών κανόνων, προμηθευτές τεχνολογίας Ιστού, προμηθευτές εργαλείων XML/RDF καθώς και χρήστες τεχνολογίας όπως οι οικονομικές εταιρίες, επιχειρήσεις τηλεπικοινωνιών και μερικές από τις σημαντικότερες δικτυακές πύλες και ASPs³⁹. Η Πρωτοβουλία RuleML συνεργάζεται με πολυάριθμες συναφείς προσπάθειες και μέσω αυτής της συνεργασίας δίνεται η δυνατότητα στη RuleML να διαμοιράσει τους μηχανισμούς και

³⁹ ASP (application service provider): Επιχείρηση που παρέχει στους πελάτες, μέσω ενός δικτύου, υπηρεσίες βασισμένες στους υπολογιστές, (158)

να παράξει μια γλώσσα κανόνων στα ήδη υπάρχοντα καθώς και στα αναδυόμενα βιομηχανικά πρότυπα. (35)

Η Πρωτοβουλία RuleML έχει συζητήσει με το W3C για την πιθανότητα δημιουργίας μιας ομάδας εργασίας η οποία θα αφοσιωθεί στους κανόνες του Ιστού (αξιώματα) ή σε ένα συνδυασμό προσπαθειών οντολογιών του Ιστού όπως εκφράζεται με την σχέση που δηλώνει πως μια οντολογία είναι μια ταξινόμια με αξιώματα. Κάτι τέτοιο ίσως θα δημιουργούσε μια ομοιόμορφη γλώσσας οντολογιών με μια ταξινόμια περιγραφικής λογικής και κανόνες Horn λογικής. (35)

Η ανταλλαγή βάσεων κανόνων RuleML σε μια ευρεία κλίμακα, θα απαιτήσει μια ταξινόμια των σχέσεων που καθορίζονται στη βάση κανόνων, όπου μια σχέση με τα επιχειρήματά της γίνεται μια κλάση με τις υποδοχές της (slots). Αυτό θα δίνει τη δυνατότητα στους συμμετέχοντες, σε μια ανταλλαγή βάσεων κανόνων να ευθυγραμμίζουν τις αντίστοιχες ιεραρχίες σχέσεων με σκοπό να ανιχνεύουν ασυμβατότητες πριν από τη συγχώνευση και την πυροδότηση των ορισμών κανόνων τους. Οι μεγάλης κλίμακας ταξονομίες DAML+OIL θα απαιτήσουν από ένα σύστημα κανόνων να αντλεί ή να χρησιμοποιεί ορισμένες υπονοούμενες πληροφορίες, οι οποίες δε συλλαμβάνονται από την ταξινόμια μόνο. Οι απαραίτητοι κανόνες θα μπορούσαν να σημειωθούν σύμφωνα με την κατάλληλη εκφραστική κατηγορία RuleML. Οι DAML+OIL ταξονομίες και τα αξιώματα RuleML πρέπει να εκφραστούν με τους συμβατούς τρόπους ή ιδανικότερα, σε μια ενοποιημένη γλώσσα. Για να το επιτύχουν αυτό, οι σημάνσεις RuleML και DAML+OIL θα μπορούσαν να συναναπτυχθούν στην ομάδα οντολογιών Ιστού προς μια κοινή έκδοση, κάτι που δεν είναι εύκολο να γίνει γιατί υπάρχουν κάποια σημαντικά ζητήματα πάνω σε αυτή τη συνεργασία που πρέπει πρώτα να επιλυθούν. (35)

Η RuleML ξεκίνησε με βάση προϋπάρχουσες γλώσσες σήμανσης κανόνων (rule markup languages). Ως μερικά παραδείγματα μπορούμε να αναφέρουμε τις γλώσσες RFML⁴⁰, URML⁴¹ και AORML⁴² (35) Η πρώτη έκδοση ήταν η 0.7 στις

⁴⁰ RFML (Relational-Functional Markup Language): Είναι μια διαδικτυακή μορφή εξόδου για τις σχεσιακές-λειτουργικές βάσεις γνώσεων που υλοποιήθηκαν ως τμήμα του συστήματος Relfun. (35)

⁴¹ URML (Universal Rule Markup Language): Η URML προσπαθεί να ενοποιήσει τον αντικειμενοστρεφή, βασισμένο στους κανόνες, προγραμματισμό (Object Oriented Rule-based programming) με την XML, και να παράξει μια βάση για την υλοποίηση των διαδικτυακών αντικειμένων και το χειρισμό τους σε κανόνες. (35)

25/1/2001 και στις 9/12/2011 βγήκε η τελευταία έκδοση της RuleML, η 1.0. (39) Η έκδοση 1.0 της RuleML αποκαλείται «Στήλη της Ροζέτας» (Rosetta Stone)⁴³. Αυτό, επειδή στη συγκεκριμένη έκδοση έχουμε δύο διαφορετικές γλώσσες σχήματος (schema languages), την XSD⁴⁴ και την Relax NG⁴⁵, οι οποίες λειτουργούν ανεξάρτητα, για να τυποποιήσουν τη σύνταξη στο μέτρο του δυνατού στο πλαίσιο της κάθε γλώσσας.

Στη κατασκευή πραγματικών διαδικτυακών οντολογιών φαίνεται σοφό να αρχίσει κανείς με τις λιγότερο εκφραστικές κλάσεις και από την πλευρά της ταξινόμιας, αλλά και από την πλευρά των αξιωμάτων, καθώς μια κοινότητα δημιουργών δεν είναι σε θέση να μπορεί να προσδιορίσει τις απαιτήσεις και των μελλοντικών κοινοτήτων χρηστών. Το οντολογικό περιεχόμενο πρέπει να συσκευαστεί σε μια όσο το δυνατόν πιο ελαφριά γλώσσα οντολογίας, ώστε τελικά να γίνει διαθέσιμο στο μέγιστο πιθανό αριθμό χρηστών. Η Πρωτοβουλία RuleML δοκίμασε να φτιάξει μια τέτοια μεθοδολογία κατασκευάζοντας ένα από κάτω προς τα πάνω σύστημα υπογλωσσών για τριπλέτες RDF που περιγράφουν λογική Horn με εξισώσεις συν URI άτομα και συσχετίσεις. Αυτό θα μπορούσε να συμπληρωθεί από μια από κάτω προς τα πάνω ταξονομική ανακατασκευή της γλώσσας και να συγκεντρωθούν μέσω κοινών εργασιών σε μια οντολογική επιπεδοποίηση. (35)

Οι κυριότερες διακρίσεις στην οικογένεια της RuleML αποτελούν οι (derivation rules), (queries) και (integrity constraints) όπως επίσης (production rules) (reaction rules). Αποτελούν το πιο ανεπτυγμένο παρακλάδι ομάδων γλωσσών Κανόνων που από μόνες τους έχουν μια διαδικτυακή Datalog⁴⁶ γλώσσα στον πυρήνα τους. Η Hornlog RuleML προσθέτει συναρτησιακές εκφράσεις σαν όρους. Στη

⁴² AORML (Agent-Oriented Rule Markup Language): Καθορίζει μια γλώσσα σήμανσης για τους προσανατολισμένους σε πράκτορες επιχειρησιακούς κανόνες στο πλαίσιο των προτύπων AOR (Agent Object Relationship). (35)

⁴³ Στήλη της Ροζέτας: Αρχαία Αιγυπτιακή στήλη που χρονολογείται στον 2^ο αιώνα π.Χ. Επάνω της έχει εγχάρακτη μια επιγραφή σε δύο γλώσσες (αιγυπτιακή και ελληνική) και τρία συστήματα γραφής (ιερογλυφικά, δημόδη αιγυπτιακή και ελληνική). (159)

⁴⁴ XSD: Χρησιμοποιεί μια προσέγγιση που βασίζεται στα δένδρα. (39)

⁴⁵ Relax NG: Χρησιμοποιεί μια προσέγγιση που βασίζεται στα πλέγματα (lattices). Αποτελεί ένα νέο συστατικό της RuleML. (39)

⁴⁶ Datalog: Είναι μια γλώσσα ερωτήσεων και κανόνων για λογικές βάσεις δεδομένων (deductive databases), η οποία συντακτικά είναι υποσύνολο της Prolog. (166)

Hornlog με ισότητα, τέτοιες μη διερμηνευμένες (τύπου κατασκευαστών) συναρτήσεις συμπληρώνονται από διερμηνευμένες (ορισμένες από εξισώσεις) συναρτήσεις. Αυτό το παρακλάδι Κανόνων Παραγωγής επεκτείνεται προς τα πάνω, στη κατεύθυνση της Λογικής Πρώτης Τάξης, έχει υποπαρακλάδια για την άρνηση ως αποτυχία⁴⁷ (negation-as-failure), την ισχυρή άρνηση⁴⁸ (strong-negation) ή τις συνδυασμένες γλώσσες και τις γλώσσες με πρόσθετα τύπου «μέλη». (40)

3.3.3. Πρότυπα σήμανσης και πρωτοβουλίες σχετικά με τη RuleML

Οι Κανόνες μπορούν να δηλωθούν ή με φυσική γλώσσα ή με κάποιο τυπικό συμβολισμό ή με ένα συνδυασμό και των δύο. Η Πρωτοβουλία RuleML κινείται προς τη κατεύθυνση των γλωσσών σήμανσης που στηρίζονται στην XML, κάτι το οποίο επιτρέπει διαδικτυακή αποθήκευση των κανόνων, ανταλλαγή, ανάκληση και εκτέλεση/εφαρμογές.

Τα πρότυπα σήμανσης και οι πρωτοβουλίες που σχετίζονται με την RuleML περιλαμβάνουν: (36)

- Mathematical Markup Language (MathML):
- DARPA Agent Markup Language (DAML):
- Predictive Model Markup Language (PMMLs)
- Attribute Grammars in XML (AG-markup):
- Extensible Stylesheet Language Transformations (XSLT). (36)

⁴⁷ Άρνηση ως αποτυχία: η άρνηση ως αποτυχία είναι ένας μη μονοτονικός κανόνας του λογικού προγραμματισμού ο οποίος χρησιμοποιείται για να συμπεράνει το “όχι p” από αποτυχία να συμπεράνει το “p”. (154)

⁴⁸ Ισχυρή άρνηση: η άρνηση ισχύει ελλείψει πληροφοριών. Για την ισχυρή άρνηση θα πρέπει να έχουμε πληροφορία ότι δεν ισχύει κάτι. (155)

3.3.3.1. *Mathematical Markup Language (MathML)*



Εικόνα 18: Λογότυπο MathML

Mathematical Markup Language (MathML): Είναι μια χαμηλού επιπέδου προδιαγραφή που χρησιμοποιείται για την περιγραφή των μαθηματικών. Προσφέρει μια αναγκαία βάση για την ένταξη Μαθηματικών

εκφράσεων σε Ιστοσελίδες και είναι ιδιαίτερα χρήσιμη και σε περιπτώσεις που έχουμε ροή εργασίας δημοσιεύσεων για την επιστήμη και την τεχνολογία και γενικά οπουδήποτε γίνεται χειρισμός των μαθηματικά μέσω λογισμικού. (41)

Έχει ως στόχο να διευκολύνει τη χρήση και την επαναχρησιμοποίηση του μαθηματικού και του επιστημονικού παρεχομένου στον Ιστό και σε άλλες εφαρμογές, όπως για παράδειγμα τα αλγεβρικά συστήματα υπολογιστών (computer algebra systems) και τη σύνθεση φωνής. Μπορεί να χρησιμοποιηθεί για να κωδικοποιήσει τόσο το μαθηματικό περιεχόμενο όσο και την παρουσίαση των μαθηματικών συμβολισμών (mathematical notation). Η σήμανση του περιεχομένου μέσω της MathML είναι καταλληλότερη για το προσδιορισμό συναρτήσεων και όχι για σχέσεις ή για γενικούς κανόνες. (36)

Η MathML 1.0 εκδόθηκε ως μια Σύσταση του W3C τον Απρίλιο του 1998 (42) και η MathML 3.0 εκδόθηκε ως μια Σύσταση του W3C στις 21 Οκτωβρίου 2010. (41) Αποτελείται από ένα πλήθος XML ετικετών που μπορούν να χρησιμοποιηθούν για τη σήμανση μιας εξίσωσης με τους όρους που παρουσιάζονται και επίσης με τη σημασιολογία τους. Η MathML έχει ως βαθύτερο στόχο να εντοπίσει κάποιο κομμάτι του νοήματος πίσω από τις εξισώσεις και δεν μένει στην οπτική τους παρουσίαση. (42)

3.3.3.2. *DARPA Agent Markup Language (DAML)*

DARPA Agent Markup Language (DAML): Το όνομά DARPA προήλθε από ένα πρόγραμμα χρηματοδότησης στο Defense Advanced Research Projects Agency των Ηνωμένων Πολιτειών. (43) Το πρόγραμμα DAML, ξεκίνησε επίσημα το 2000 στη Βοστώνη και ως στόχο είχε την ανάπτυξη μιας γλώσσας και εργαλείων που θα

διευκόλυναν το όραμα του Σημασιολογικού Ιστού. (44) Ένα κύριο αποτέλεσμα του



Εικόνα 19: Λογότυπο DAML

DAML προγράμματος ήταν η γλώσσα σήμανσης πρακτόρων (agent markup language) DAML.

(43) Η γλώσσα DAML έχει αναπτυχθεί ως μια επέκταση της XML και της RDF. (44) Παρέχει μια βασική υποδομή, η οποία επιτρέπει στις μηχανές να κάνουν τα ίδια είδη των απλών

συμπερασμάτων που κάνουν και οι άνθρωποι. (45)

Η τελευταία έκδοση της γλώσσας, η οποία αποκαλείται DAML+OIL, παρέχει ένα πλούσιο σύνολο κατασκευαστών για τη δημιουργία οντολογιών και για τη σήμανση πληροφοριών με τρόπο που να είναι αναγνώσιμες και κατανοήσιμες από τις μηχανές. (44) Το 2002 έγινε η υποβολή της DAML+OIL γλώσσας στο W3C και αυτό αποτέλεσε την αφετηρία για τη δημιουργία μιας γλώσσας οντολογιών από την ομάδα εργασίας οντολογιών του W3C, την OWL. (43) Η DAML+OIL έχει τις ρίζες της σε τρεις κύριες γλώσσες: την DAML, την OIL⁴⁹ (Ontology Inference Layer) και την SHOE⁵⁰ (Simple HTML Ontology Extensions). (43)

3.3.3.3. Predictive Model Markup Language (PMML)



Εικόνα 20: Λογότυπο PMML

Predictive Model Markup Language

(PMML): Είναι μια γλώσσα η οποία παρέχει στις εφαρμογές τρόπους για να καθορίσουν στατιστικά μοντέλα (statistical models) καθώς και μοντέλα εξόρυξης δεδομένων (data mining models), και επιτρέπει επίσης τον διαμοιρασμό των μοντέλων μεταξύ συμβατών με την γλώσσα PMML

εφαρμογών. (46) Με αυτή τη γλώσσα, που στηρίζεται στην XML, μπορεί κάποιος να προσδιορίσει και να διαμοιράσει διάφορα μοντέλα για τα αποτελέσματα της εξόρυξης

⁴⁹ OIL: Μπορεί να θεωρηθεί ως μια Οντολογική υποδομή για τον Σημασιολογικό Ιστό. Κατασκευάστηκε με σημείο εκκίνησης την RDFS και ενσωματώνει στοιχεία από τα συστήματα που βασίζονται στα Πλαίσια (Frame) και τις Περιγραφικές Λογικές. (146)

⁵⁰ SHOE: Αποτελεί μια μικρή επέκταση της HTML η οποία επιτρέπει στους κατόχους των Ιστοσελίδων να «σημάνουν» το περιεχόμενο των έγγραφων που έχουν στον Ιστό με γνώση αναγνώσιμη από τις μηχανές. (145)

δεδομένων (data-mining results), συμπεριλαμβανομένων των κανόνων συσχέτισης⁵¹ (association rules). (36)

Παλαιότερα, αυτό ήταν ιδιαίτερα δύσκολο, αλλά με την PMML η ανταλλαγή των μοντέλων μεταξύ συμβατών εφαρμογών είναι τώρα απλή, καθώς η PMML παρέχει στις εφαρμογές μια ανεξάρτητη κατασκευαστή (vendor) μέθοδο για τον προσδιορισμό μοντέλων. Συνεπώς, τα ζητήματα ιδιοκτησίας (proprietary issues) και οι ασυμβατότητες δε θα αποτελούν πλέον εμπόδιο για την ανταλλαγή μοντέλων μεταξύ εφαρμογών. (46) Η PMML χρησιμοποιεί την XML και έχει δομή που περιγράφεται μέσω της XML Schema. Κάθε έγγραφο PMML είναι ένα XML έγγραφο το οποίο μπορεί να περιέχει ένα ή περισσότερα μοντέλα εξόρυξης δεδομένων. (47) Επιτρέπει στους χρήστες την δημιουργία μοντέλων, στο πλαίσιο της εφαρμογής ενός κατασκευαστή και την χρήση εφαρμογών άλλων κατασκευαστών για την απεικόνιση, την ανάλυση, την αξιολόγηση ή ακόμα και την χρησιμοποίηση με άλλους τρόπους των μοντέλων. (46)

Η πρώτη έκδοση της PMML ήταν η 0.7 τον Ιούλιο του 1997 και φτάνουμε στο σήμερα, στην έκδοση 4.1 που εκδόθηκε στις 31 Δεκεμβρίου 2011. Ανάμεσα σε αυτές τις δύο βρίσκονται ακόμα 9 εκδόσεις, οι οποίες προσθέτουν όλο και περισσότερα νέα χαρακτηριστικά σε σχέση με την εκάστοτε προηγούμενη έκδοση. (48)

3.3.3.4. *Attribute Grammars in XML (AG-markup)*

Attribute Grammars in XML (AG-markup): Για τους σημασιολογικούς κανόνες των Attribute Grammars, υπάρχουν πολλές δυνατές σημάνσεις XML (XML markups). (36) Μια attribute grammar είναι μια γραμματική χωρίς συμφραζόμενα (context-free grammar) που έχει επεκταθεί με χαρακτηριστικά (attributes), σημασιολογικούς κανόνες (semantic rules) και συνθήκες (conditions) προκειμένου να παρέχει πληροφορίες με συμφραζόμενα (context-sensitive information), προσθέτοντας χαρακτηριστικά σε κάποια από τα μη τερματικά σύμβολα

⁵¹ Κανόνες συσχέτισης (association rules): Κανόνες που συμβάλλουν στην ανακάλυψη των συσχετίσεων που υπάρχουν μεταξύ των γνωρισμάτων του συνόλου δεδομένων. (4)

(nonterminals⁵²) της. Κάθε διακριτό σύμβολο της γραμματικής συσχετίζεται με ένα πεπερασμένο, ίσως κενό, σύνολο γνωρισμάτων (49)

Οι Attribute Grammars και η XML είναι και οι δύο επεκτάσεις γραμματικών χωρίς συμφραζόμενα (context-free grammar), μέσω των ετικετών (tags), DTDs και των φύλλων στυλ (stylesheets). Χρησιμοποιούν το ίδιο εργαλείο για έλεγχο συντακτικών λαθών και για τη δημιουργία των στοιχείων των δένδρων τους. (50)

3.3.3.5. Extensible Stylesheet Language Transformations (XSLT)

Extensible Stylesheet Language Transformations (XSLT): Τα έγγραφα XML δεν περιέχουν πληροφορίες μορφοποίησης, κάνοντας ευμετάβλητο το τελικό οπτικό αποτέλεσμα, καθώς ένα δεδομένο έγγραφο XML μπορεί να παρουσιαστεί με διάφορους τρόπους, όταν εφαρμοστούν σε αυτό διαφορετικά φύλλα στυλ. (3)



Εικόνα 21: Εικόνα XSLT

Η XSLT είναι μια γλώσσα για μετατροπή εγγράφων XML σε άλλα XML έγγραφα και έχει σχεδιαστεί για χρήση ως τμήμα του XSL⁵³. (51) Ορίζει κανόνες με τους οποίους ένα έγγραφο XML μετασχηματίζεται σε ένα άλλο έγγραφο XML, σε HTML ή σε απλό κείμενο. Τα έγγραφα XSLT είναι έγγραφα XML, άρα η XSLT είναι μια εφαρμογή XML. (3) Ένας μετασχηματισμός στην XSLT

γλώσσα εκφράζεται στη μορφή φύλλων στυλ και αποτελεί ένα συντακτικά καλά δομημένο XML έγγραφο⁵⁴. Σε αντίθεση με την XSLT, η XSL περιλαμβάνει ένα XML λεξιλόγιο για τον προσδιορισμό της μορφοποίησης. Η XSL προσδιορίζει τη μορφοποίηση ενός XML εγγράφου χρησιμοποιώντας την XSLT για να περιγράψει πως το έγγραφο αυτό μετατρέπεται σε ένα άλλο XML έγγραφο που χρησιμοποιεί το λεξιλόγιο μορφοποίησης. (51)

⁵² Μη τερματικά σύμβολα (nonterminals): Τα σύμβολα τα οποία μπορούν να αντικατασταθούν μέσω ενός γραμματικού κανόνα μιας γλώσσας.

⁵³ Η XSL (EXtensible Stylesheet Language) περιλαμβάνει μια γλώσσα μετασχηματισμού (την XSLT) και μια γλώσσα μορφοποίησης. (3)

⁵⁴ Ένα έγγραφο XML είναι έγκυρο αν είναι σωστά δομημένο, χρησιμοποιεί δομικές πληροφορίες και τις εφαρμόζει. (3)

Η γλώσσα XSLT μπορεί να χρησιμοποιηθεί ανεξάρτητα από τη γλώσσα μορφοποίησης. (3) Όμως, δεν υπάρχει ο στόχος να χρησιμοποιηθεί για ως μια γενικής χρήσης γλώσσα μετασχηματισμού XML εγγράφων. Έχει κατά κύριο λόγο σχεδιαστεί για το είδος των μετασχηματισμών που την χρειάζονται ως τμήμα της XSL. (51) Αποτελεί ένα πολύτιμο εργαλείο για όσες εφαρμογές βασίζονται στην XML, καθώς παρέχει τη δυνατότητα μετατροπής δεδομένων και μεταδεδομένων από μια αναπαράσταση XML σε μια άλλη. (3)

Η πρώτη έκδοση της XSLT αποτελούσε μια Σύσταση του W3C, που έγινε στις 16 Νοεμβρίου το 1999. Έκτοτε βρήκε και η δεύτερη έκδοση, πάλι ως Σύσταση του W3C στις 23 Ιανουαρίου το 2007. Η δεύτερη έκδοση χρησιμοποιεί τις αναβαθμισμένες τεχνολογίες, όπως το Xpath 2.0⁵⁵, και έρχεται για να προσθέσει πολλές νέες λειτουργίες και να διορθώσει κάποιες δυσλειτουργίες της πρώτης έκδοσης. (52)

3.3.4. Το Σχέδιο της RuleML

Όπως περιγράψαμε, στόχος της RuleML είναι να επιτραπεί η επαναχρησιμοποίηση και η ανταλλαγή κανόνων σε ένα υψηλό επίπεδο. Συνεπώς, αντί για τη δημιουργία μιας επιπλέον γλώσσας κανόνων, η RuleML προσφέρει μία οικογένεια υπογλωσσών μαζί με ένα μοντέλο διαμοιραζόμενων δεδομένων, που μπορεί να επωφεληθεί από τις υπάρχουσες και τις μελλοντικές γλώσσες. (35)

Έχοντας ως δεδομένο το γλωσσικό πλούτο και τις πολύπλοκες δυναμικές των επιχειρησιακών πεδίων, μπορούμε να διακρίνουμε ξεκάθαρα ότι οποιοσδήποτε μαθηματικός υπολογισμός των κανόνων, όπως για παράδειγμα στις προτάσεις Horn κλασικής λογικής, θα έπρεπε να θεωρείται ως μια περιορισμένη περιγραφική θεωρία, η οποία αποτυπώνει μόνο ένα μέρος του συνολικού χώρου των κανόνων. Αντιθέτως, αυτό που χρειαζόμαστε είναι μία πολύπλευρη προσέγγιση για τον ετερογενή εννοιολογικό χώρο των κανόνων. Κατά συνέπεια, στη RuleML ορίζεται ως μία οικογένεια γλωσσών που συλλαμβάνει τα κυριότερα είδη κανόνων. Ενώ αυτές οι

⁵⁵ XML Path Language (XPath): Είναι μια γλώσσα ερωτημάτων για έγγραφα XML. Η δεύτερη έκδοση αυτής εκδόθηκε στις 14 Δεκεμβρίου το 2010 και αποτελεί Σύσταση του W3C. (144)

γλώσσες συνοδεύονται από μια προτεινόμενη τυπική σημασιολογία, κάποιες από τις βάσεις κανόνων τους ενδεχομένως να έχουν και μια διαφορετική αποδεκτή σημασιολογία. Αυτό εξυπηρετεί ποικίλες τυποποιήσεις οι οποίες βασίζονται στη μη τυποποιημένη λογική (non-standard logics), υποστηρίζοντας τη χρονική⁵⁶ (temporal), την ασαφή⁵⁷ (fuzzy), την αναιρέσιμη⁵⁸ (defeasible), καθώς και άλλες μορφές συλλογιστικής. (37)

Κάθε μια από τις υπογλώσσες της οικογένειας έχει ένα XML Schema για τον καθορισμό του και διευθυνσιοδοτείται στον Ιστό μέσω ενός URI, το οποίο επιτρέπει την κληρονομικότητα μεταξύ των σχημάτων των υπογλωσσών και την ακριβή αναφορά στην απαιτούμενη εκφραστικότητα. Η δομή της οικογένειας των υπογλωσσών παρέχει μια ιεραρχία ως προς την εκφραστικότητα για τις υπογλώσσες, καθώς τα URIs είναι υποκείμενα του σημασιολογικού χαρακτηρισμού (θεωρητικό μοντέλο). (40)

```
rules: rule
  reaction rules: react
  transformation rules: trans
    derivation rules: imp
      facts: fact
      queries: query
        integrity constraints: ic
```

Εικόνα 22: Υπογλώσσες της οικογένειας της RuleML

Η ιεραρχία των γενικών κανόνων της RuleML διακλαδώνεται σε δύο βασικές κατηγορίες, στους αντιδραστικούς κανόνες και στους κανόνες μετασχηματισμού. Στο επόμενο επίπεδο, οι κανόνες μετασχηματισμού εξειδικεύονται στην υποκατηγορία των κανόνων εξαγωγής συμπερασμάτων. Οι κανόνες εξαγωγής συμπερασμάτων έχουν και τις υποκατηγορίες των

γεγονότων και των ερωτημάτων. Τελικά, τα ερωτήματα εξειδικεύονται σε περιορισμούς ακεραιότητας. Περισσότερες υποδιαιρέσεις είναι υπό μελέτη, ιδίως στους αντιδραστικούς κανόνες. Μια γραφική όψη των κανόνων της RuleML φαίνεται

⁵⁶ Χρονική Λογική (temporal logic): περιγράφει κάθε σύστημα κανόνων και συμβολισμού για την αναπαράσταση και συλλογιστική προτάσεων που χρησιμοποιούν το χρόνο. (167)

⁵⁷ Ασαφής Λογική (fuzzy logic): Μια πρόταση δεν είναι απαραίτητα απλά αληθής ή ψευδής. Μπορεί, για παράδειγμα, να είναι αληθής με κάποιο βαθμό αληθείας.

⁵⁸ Αναιρέσιμη Λογική (defeasible logic): Είναι μια απλή αλλά αποδοτική προσέγγιση στη μη-μονότονη λογική. (4)

στην εικόνα 22. Παρακάτω έχουμε μια πιο ολοκληρωμένη παρουσίαση των κανόνων αυτών: (53)

- Αντιδραστικοί κανόνες (reaction rules): γεγονός-συνθήκη-ενέργεια
- Κανόνες μετασχηματισμού (transformation rules): λειτουργικοί-εξισωτικοί κανόνες
- Κανόνες εξαγωγής συμπερασμάτων (derivation rules): επαγωγικοί-συμπερασματικοί κανόνες
- Κανόνες που εξειδικεύονται σε γεγονότα (facts): κανόνες εξαγωγής συμπερασμάτων χωρίς προϋποθέσεις
- Ερωτήματα (queries): κανόνες εξαγωγής συμπερασμάτων χωρίς συμπεράσματα.
- Περιορισμοί ακεραιότητας (integrity constraints): κανόνες διατήρησης-συνέπειας (53)

4. Εφαρμογές

Οι τεχνολογίες του Σημασιολογικού Ιστού μπορούν να εφαρμοστούν για διάφορες χρήσεις. Κάποιες από αυτές είναι η δημοσίευση πληροφοριών, η ενοποίηση δεδομένων, η διαχείριση γνώσης, η τηλεεκπαίδευση, οι υπηρεσίες του Ιστού, η δεικτοδότηση πολυμεσικών συλλογών, οι προμήθειες μέσω δικτύου και η διαλειτουργικότητα των συσκευών. (3)

Παρακάτω θα δούμε κάποια συγκεκριμένα και ευρείας χρήσης παραδείγματα.

4.1. DBpedia



Εικόνα 23: Λογότυπο DBpedia

Η DBpedia αποτελεί μια κοινοτική προσπάθεια απόσπασης και επαναχρησιμοποίησης δομημένων πληροφοριών που προέρχονται από τη Wikipedia και να γίνουν αυτά τα δεδομένα διαθέσιμα στο Web. (54)

Η DBpedia προέρχεται από τη Wikipedia και διανέμεται υπό τους ίδιους όρους αδειοδότησης με την ίδια την Wikipedia. Η Wikipedia, πλέον, έχει διπλή αδειοδότηση (dual-licensing), οπότε το ίδιο πρέπει να ακολουθήσει και η άδεια της DBpedia αρχίζοντας από την έκδοση 3.4. Τα δεδομένα που περιλαμβάνει η έκδοση 3.4 της DBpedia (καθώς και οι επόμενες εκδόσεις) έχουν αδειοδοτηθεί σύμφωνα με τους όρους της άδειας Creative Commons Attribution-ShareAlike 3.0⁵⁹ και την GNU Free Documentation License⁶⁰,

⁵⁹ Creative Commons Attribution-ShareAlike 3.0: Επιτρέπεται ο διαμοιρασμός, η μεταποίηση και η εμπορευματοποίηση των δεδομένων, αρκεί να γίνει κάποια συνεισφορά κατά τον τρόπο που το ορίζει ο συγγραφέας ή ο αδειοδότης και η προκύπτουσα εργασία να διανεμηθεί κάτω από την ίδια ή μια παρόμοια με αυτή την άδεια. (140)

⁶⁰ GNU Free Documentation License: Κάθε αντίγραφο του υλικού που βασίζεται σε αυτή την άδεια, έστω και τροποποιημένο, συνεχίζει να φέρει την ίδια άδεια. Τα αντίγραφα αυτά μπορούν να

ενώ τα δεδομένα που περιλαμβάνουν οι εκδόσεις ως και την 3.3 της DBpedia υπόκεινται μόνο στους όρους της άδειας GNU Free Documentation License. (55)

Η Wikipedia έχει εξελιχθεί σε μία από τις κεντρικές πηγές της γνώσης της ανθρωπότητας, η οποία συντηρείται από χιλιάδες συντελεστές. Η DBpedia αξιοποιεί αυτή τη γιγάντια πηγή γνώσης, εξάγοντας δομημένες πληροφορίες από τη Wikipedia και καθιστώντας τις πληροφορίες αυτές προσβάσιμες στο Διαδίκτυο σύμφωνα με τους όρους της άδειας «Creative Commons Attribution-ShareAlike 3.0 License» και την «GNU Free Documentation License». (54) Μέσω της συγκεκριμένη αδειοδότησης, επιτρέπεται στους Πράκτορες του Σημασιολογικού Ιστού να παρέχουν λογικά συμπεράσματα και προηγμένη αναζήτηση στα σύνολα δεδομένων που προέρχονται από τη Wikipedia (τα δεδομένα αυτά δημοσιεύονται σε RDF) διευκολύνοντας τη διασύνδεση, την επαναχρησιμοποίηση και την επέκταση σε άλλες πηγές. (56)

Η DBpedia επιτρέπει στον χρήστη να ρωτήσει εξελιγμένα ερωτήματα στη Wikipedia. Για παράδειγμα "δώσε μου όλες τις πόλεις στην Ελλάδα με περισσότερους από 10.000 κατοίκους" ή "δώσε μου όλους τους Έλληνες συνθέτες από τον 19ο αιώνα", και να συνδεθούν άλλα σύνολα δεδομένων στο διαδίκτυο με τα δεδομένα της Wikipedia. Στόχος είναι να δώσει νέους τρόπους αξιοποίησης του πλούτου των πληροφοριών της Wikipedia, καταστρώνοντας ευκολότερη τη χρήση αυτής, με νέους και ενδιαφέροντες τρόπους. Συνολικά, οι πιθανές χρήσεις της βάσης γνώσης της DBpedia είναι διαδεδομένες και ποικίλουν από τη διαχείριση της γνώσης σε επιχειρήσεις, μέσω αναζητήσεων στον Παγκόσμιο Ιστό, σε επαναστατικές αναζητήσεις στη Wikipedia. Επίσης, θα μπορούσε να εμπνεύσει νέους μηχανισμούς για την πλοήγηση, τη διασύνδεση και τη βελτίωση της ίδιας της εγκυκλοπαίδειας.

Η DBpedia είναι ο πυρήνας των Συνδεδεμένων Δεδομένων (Linked Data⁶¹) και σαν αποτέλεσμα όλο και περισσότεροι "εκδότες" προσπαθούν να την έχουν ως βάση αναφοράς για τα δικά τους δεδομένα. (57) Η βάση γνώσης της DBpedia περιγράφει σήμερα πάνω από 3.640.000 πράγματα, εκ των οποίων 1.830.000 ταξινομούνται σε μια σταθερή Οντολογία, συμπεριλαμβανομένων 416.000 ατόμων,

πωληθούν, αλλά εάν παραχθούν σε ποσότητα επιβάλλεται να γίνουν διαθέσιμα σε μορφή που επιδέχεται περαιτέρω επεξεργασία. (142)

⁶¹ Διασυνδεδεμένα Δεδομένα: Είναι μια μέθοδος δημοσίευσης δεδομένων στο Διαδίκτυο και διασύνδεσης μεταξύ διαφορετικών πηγών. (141)

526.000 τοποθεσιών, 106.000 μουσικών άλμπουμ, 60.000 ταινιών, 17.500 ηλεκτρονικών παιχνιδιών και 169.000 οργανώσεις. Το σύνολο δεδομένων της DBpedia διαθέτει ετικέτες και περιλήψεις για αυτά τα 3.640.000 πράγματα σε έως και 97 διαφορετικές γλώσσες, 2.724.000 συνδέσμους προς εικόνες και 6.300.000 συνδέσμους με εξωτερικές ιστοσελίδες, 6.200.000 εξωτερικές συνδέσμους σε άλλες βάσεις δεδομένων RDF, 740.000 κατηγορίες της Wikipedia, και 2.900.000 κατηγορίες της Yago⁶². Η βάση γνώσης της DBpedia αποτελείται συνολικά από πάνω από 1,2 δισεκατομμύρια κομμάτια πληροφοριών (τριπλότυπο RDF), από τα οποία τα 385 εκατομμύρια προήλθαν από την αγγλική έκδοση της Wikipedia και 665 εκατ. προήλθαν από εκδόσεις της Wikipedia σε άλλες γλώσσες. (54)

4.1.1. DBpedia Live

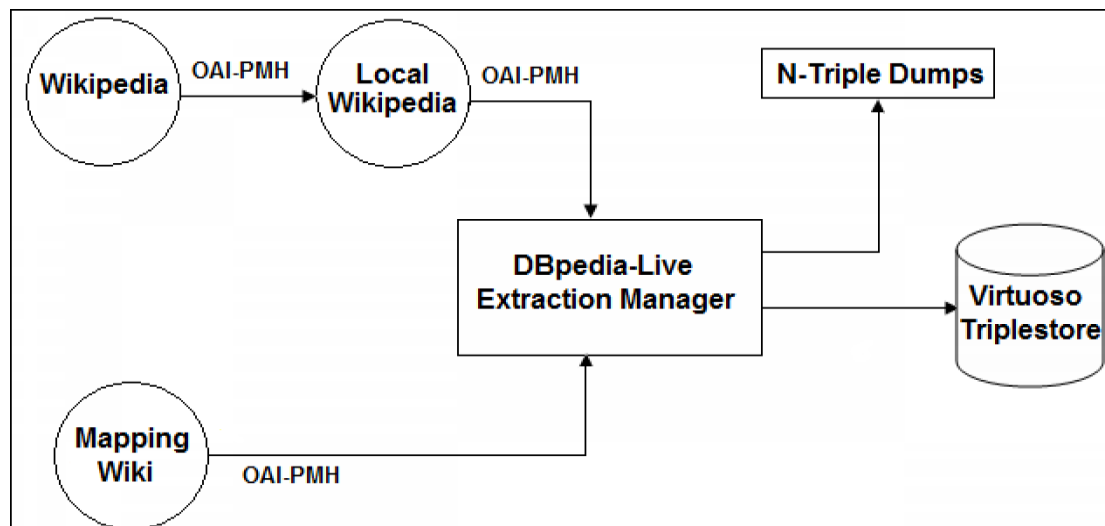
Η DBpedia θεωρείται ο καθρέφτης του Σημασιολογικού Ιστού της Wikipedia. Με το χρόνο, τα άρθρα της Wikipedia αναθεωρούνται, το οποίο καθιστά τα δεδομένα στη DBpedia ξεπερασμένα. Για αυτό το σκοπό, υπάρχει η DBpedia-Live. Ο κύριος στόχος της DBpedia-Live είναι να κρατήσει τη DBpedia πάντα σε συγχρονισμό με την Wikipedia. Σύμφωνα με τα στατιστικά, 1.4 άρθρα της Wikipedia τροποποιούνται ανά δευτερόλεπτο που έχει ως αποτέλεσμα 84 άρθρα ανά λεπτό. Η DBpedia-Live μπορεί κατά μέσο όρο να επεξεργαστεί περίπου 105 σελίδες ανά λεπτό. Συνεπώς, η ταχύτητα της επεξεργασίας της DBpedia-Live μπορεί να ανταπεξέλθει την ταχύτητα της ροής των δεδομένων. Έχει προβλεφθεί πως σε περίπτωση διακοπής, η DBpedia-Live θα ξεκινήσει από το τελευταίο σημείο στο οποίο σταμάτησε.

Ο πυρήνας της DBpedia αποτελείται από μια διαδικασία εξαγωγής «Infobox». Τα Infoboxes αποτελούν πρότυπα (templates) που περιέχουν κάποιες δεδομένες πληροφορίες και περιέχονται σε πολλά άρθρα της Wikipedia (εμφανίζονται συνήθως στην πάνω δεξιά γωνία των άρθρων). Εκτός από την εξόρυξη Infobox, το framework έχει σήμερα 19 εξαγωγείς που επεξεργάζονται τους ακόλουθους τύπους περιεχομένου της Wikipedia, κάποιοι από τους οποίους είναι οι ακόλουθοι: ετικέτες, διαγλωσσικοί

⁶² YAGO: μια επεκτάσιμη και ελαφριά οντολογία με υψηλή κάλυψη και ποιότητα. Περιέχει στη παρούσα φάση πάνω από ένα εκατομμύριο οντότητες και πέντε εκατομμύρια γεγονότα. (139)

σύνδεσμοι, εικόνες, ανακατευθύνσεις, εξωτερικοί σύνδεσμοι, γεω-συντεταγμένες, ID σελίδας, ID διασκευής, κατηγορίες άρθρου, αντιστοιχίες (mappings), infobox.

Η αρχιτεκτονική του συστήματος της DBpedia-Live φαίνεται στο ακόλουθο σχήμα. Τα κύρια συστατικά του συστήματος της DBpedia-Live είναι τα εξής:



Εικόνα 24: Τα κύρια συστατικά του συστήματος της DBpedia-Live

- Τοπική Wikipedia (Local Wikipedia): Έχει γίνει μια τοπική εγκατάσταση της Wikipedia με τρόπο που να είναι σε συνεχή συγχρονισμό με την online Wikipedia. Το Πρωτόκολλο OAI-PMH⁶³ (Open Archives Initiative Protocol for Metadata Harvesting) επιτρέπει σε μια εφαρμογή να λαμβάνει μια συνεχή ροή ενημερώσεων από ένα wiki. Επίσης, χρησιμοποιείται για να τροφοδοτήσει ενημερώσεις στο DBpedia-Live Extraction Manager.
- Αντιστοίχιση Wiki (Mapping Wiki): Αναφερόμαστε στη δημιουργία αντιστοιχίσεων μεταξύ Προτύπων της Wikipedia και οντολογιών της DBpedia. Οι αντιστοιχίσεις της DBpedia (DBpedia mappings) μπορούν να βρεθούν στο wiki της DBpedia⁶⁴. Θα μπορούσαμε χρησιμοποιώντας το OAI-PMH να πάρουμε τη ροή των ενημερώσεων σε αντιστοιχίσεις DBpedia. Όμως μια οποιαδήποτε αλλαγή σε αντιστοίχιση

⁶³ Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH: Είναι ένα διαλειτουργικό πλαίσιο, ανεξάρτητο εφαρμογής, που στηρίζεται στη συλλογή μεταδεδομένων. Τα δεδομένα αυτά προέρχονται είτε από διαχειριστικά συστήματα παροχής δεδομένων που το υποστηρίζουν ως μέσο έκθεσης δεδομένων ή από παρόχους υπηρεσιών που χρησιμοποιούν τα συλλεχθέντα μεταδεδομένα μέσω αυτού ως ένα τρόπο για να κατασκευάσουν υπηρεσίες προστιθέμενης αξίας. (137)

⁶⁴ <http://mappings.dbpedia.org>

θα επηρεάσει αρκετές σελίδες της Wikipedia, οι οποίες θα πρέπει να επανεπεξεργαστούν.

- DBpedia-Live Extraction Manager: Αυτό το στοιχείο είναι το ουσιαστικό πλαίσιο εξαγωγής (extraction framework) της DBpedia-Live. Όταν πρέπει να υποβληθεί σε επεξεργασία μια σελίδα, το πλαίσιο εφαρμόζει τους εξαγωγείς σε αυτήν. Όταν ολοκληρωθεί η επεξεργασία της σελίδας, οι νέες τριπλέτες που εξάγονται (γράφονται επίσης ως αρχείο N-Triples⁶⁵) εισάγονται μέσα σε ένα σύστημα υποστήριξης του triplestore (Virtuoso), αντικαθιστώντας τις παλιές τριπλέτες. Οποιαδήποτε άλλη εφαρμογή ή καθρέφτης της DBpedia-Live (DBpedia-Live mirror), που πρέπει να είναι σε συνεχή συγχρονισμό με τη DBpedia-Live, μπορεί να κατεβάσει αυτά τα αρχεία και να τα ενσωματώσει στο δικό της triplestore. (58)

4.1.2. Ελληνική DBpedia



Εικόνα 25: Λογότυπο Ελληνικής DBpedia

Στις 12 Απριλίου 2011, το Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης ανακοίνωσε την Ελληνική DBpedia. (59) Η Ελληνική DBpedia δραστηριοποιείται στο πλαίσιο της διεθνούς κοινότητας της DBpedia (<http://dbpedia.org>). Η ελληνική έκδοση της εγκυκλοπαίδειας Wikipedia (η Βικιπαίδεια) περιλαμβάνει αυτή τη στιγμή 66.821 άρθρα και 169.162 σελίδες συνολικά (αυτές οι πληροφορίες ισχύουν ως τις 21:32 την 26 Νοεμβρίου 2011) (60).

Η ελληνική DBpedia έχει τη συμπαράσταση του Αριστοτέλειου Πανεπιστημίου Θεσσαλονίκης και προέρχεται μέσα από το Πρόγραμμα Μεταπτυχιακών Σπουδών στην Επιστήμη του Διαδικτύου του Τμήματος Μαθηματικών του Α.Π.Θ., που εδράζει στο Δήμο Βέροιας. Η δημιουργία της πρώτης έκδοσης της ελληνικής DBpedia έγινε από μεταπτυχιακούς φοιτητές του ΠΜΣ (Σ.

⁶⁵ N-triples: κείμενο σε μορφή απλού κειμένου που χρησιμοποιείται για την κωδικοποίηση ενός RDF γράφου. Κάθε τριπλέτα RDF ορίζεται σε μια μόνο σειρά του παραγόμενου αρχείου. (138)

Αλεξίου, Α. Καραλή, Γ. Καρβουνά, Δ. Κοντοκώστα και Π. Ξανθοπούλου) στο πλαίσιο ομαδικής εργασίας τους. Η επιστημονική ομάδα του Α.Π.Θ. για την Ελληνική DBpedia αποτελείται από τον μεταπτυχιακό φοιτητή κ. Δ. Κοντοκώστα, τον Επιστημονικό Συνεργάτη και Διδάσκοντα Δρ. Χ. Μπράτσα και τους Καθηγητές Ι. Αντωνίου και Γ. Μητακίδη. (61)

Με δεδομένο πως η DBpedia βρίσκεται στο κέντρο των Ανοιχτών Διασυνδεδεμένων Δεδομένων, η ελληνική DBpedia μπορεί να αποτελέσει τον πυρήνα της δημιουργίας των Ελληνικών Ανοιχτών Διασυνδεδεμένων Δεδομένων⁶⁶. Έως τώρα, η σύνδεση, η δημοσιοποίηση και η σημασιολογική επεξεργασία δεδομένων ήταν εφικτή μόνο για διευθύνσεις και επισημάνσεις σε λατινικούς χαρακτήρες. Η ερευνητική ομάδα του Α.Π.Θ. (Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης) ανέπτυξε αλγόριθμο και λογισμικό συμβατό με κάθε φυσική γλώσσα, που κατ' επέκταση αποτελεί μια υλοποιήσιμη πρόταση προς την Κοινοπραξία του Παγκοσμίου Ιστού (World Wide Web Consortium), για την ενσωμάτωση στο Web 3.0 γλωσσών με μη λατινικούς χαρακτήρες. (62)

Συνεπώς, σύμφωνα με όλα τα παραπάνω, η βάση γνώσης της DBpedia έχει πολλά πλεονεκτήματα σε σχέση με υπάρχουσες βάσεις γνώσης:

- καλύπτει πολλά πεδία,
- αντιπροσωπεύει πραγματική κοινοτική συμφωνία,
- εξελίσσεται αυτόματα όταν η Wikipedia αλλάζει,
- είναι πραγματικά πολύγλωσση. (54)

4.2. *Linking Open Data*

Το έργο Linking Open Data (Σύνδεση Ανοιχτών Δεδομένων) αποτελεί ένα κοινοτικό project, που γίνεται υπό την ηγεσία του W3C. (63) Ιδρύθηκε τον Ιανουάριο του 2007 και υποστηρίχθηκε από το Semantic Web Education and Outreach Group του W3C. Ο σκοπός αυτού του έργου ήταν και είναι να συνυφάνει τον Ιστό των

⁶⁶ Ανοιχτά Διασυνδεδεμένα Δεδομένα: Βλ. § 4.2

Δεδομένων αναγνωρίζοντας τα υπάρχοντα σύνολα δεδομένων, τα οποία διατίθενται με ανοιχτές άδειες (open licenses), μετατρέποντάς τα σε RDF σύμφωνα με τους κανόνες των Συνδεδεμένων Δεδομένων⁶⁷ και δημοσιεύοντάς τα στον Ιστό. (64) Η κίνηση “Open Data” (Open Data Movement) στοχεύει στο να καταστήσει ελεύθερη σε όλους την διάθεση των δεδομένων.



Εικόνα 26: Λογότυπο Linking Open Data

Το RDF παρέχει ένα ευέλικτο μοντέλο δεδομένων για την ενσωμάτωση πληροφοριών από διαφορετικές πηγές, με δυνατότητες σύνδεσης που

δεν παρέχονται από οποιοδήποτε άλλο μοντέλο δεδομένων. Οι RDF συνδέσεις επιτρέπουν την πλοήγηση από μια πηγή δεδομένων σε συσχετισμένα στοιχεία δεδομένων μέσα σε άλλες πηγές, χρησιμοποιώντας ένα πρόγραμμα περιήγησης του Σημασιολογικού Ιστού. Επίσης, μπορούν να ακολουθούνται από τους ανιχνευτές (crawlers) των μηχανών αναζήτησης του Σημασιολογικού Ιστού, το οποίο μπορεί να παρέχει εξελιγμένες δυνατότητες αναζήτησης και εκτέλεσης ερωτημάτων πάνω σε ανιχνεύσιμα δεδομένα. Δεδομένου ότι τα αποτελέσματα των αναζητήσεων είναι δομημένα δεδομένα και όχι μόνο σύνδεσμοι σε σελίδες HTML, μπορούν να χρησιμοποιηθούν και σε άλλες εφαρμογές.

Υπάρχουν ήδη διάφορα ενδιαφέροντα ανοιχτά σύνολα δεδομένων (open data sets) και αρκετές προσπάθειες για τη δημοσίευση δεδομένων. Κάποια παραδείγματα είναι το έργο DBpedia.org (DBpedia.org project), η Οντολογία Geonames (Geonames Ontology), ο διακομιστής D2R εκδίδοντας την βιβλιογραφία DBLP (DBLP bibliography) και ο διακομιστής μουσικής dbtune (dbtune music server). Τα σύνολα δεδομένων αυτών δημοσιεύονται κάτω από τις άδειες Creative Commons⁶⁸ ή

⁶⁷ 1. Χρήση των URIs για των προσδιορισμό πραγμάτων. 2. Χρήση των HTTP URIs ώστε να μπορούν οι άνθρωποι να αναζητήσουν αυτά τα ονόματα. 3. Να παρέχονται χρήσιμες πληροφορίες, όταν κάποιος αναζητάει ένα URI, χρησιμοποιώντας πρότυπα όπως η RDF και η SPARQL. 4. Να υπάρχουν σύνδεσμοι προς άλλα URIs, έτσι ώστε να διευκολύνεται η εύρεση νέων πραγμάτων. (147)

⁶⁸ Creative Commons: Με τη χρήση αυτής της άδειας, παραμένετε δικαιούχος πνευματικής ιδιοκτησίας, αλλά παρέχετε το δικαίωμα στον καθένα να αντιγράψει και διανέμει το έργο σας υπό την προϋπόθεση ότι θα σας αναφέρει ως δημιουργό/δικαιούχο και θα πληροί τους όρους χρήσης που αναφέρετε στην άδεια αυτή. (148)

TALIS⁶⁹. Έχουν γίνει κάποιες πρώτες προσπάθειες για να διασυνδέσουν αυτές τις πηγές δεδομένων. Για παράδειγμα, οι RDF περιγραφές της DBpedia για τις πόλεις περιλαμβάνουν συνδέσεις «owl:sameAs» σε δεδομένα Geonames σχετικά με την πόλη. (65)

Αρχικά, οι συμμετέχοντες του project ήταν κυρίως ερευνητές και προγραμματιστές σε πανεπιστημιακά ερευνητικά εργαστήρια και μικρές εταιρίες. Ο ανοιχτός χαρακτήρας του έργου έπαιξε μεγάλο ρόλο στην ταχεία ανάπτυξή του, καθώς ο καθένας μπορεί να συμμετέχει δημοσιεύοντας ένα σύνολο δεδομένων σύμφωνα με τους κανόνες των Συνδεδεμένων Δεδομένων και διασυνδέοντάς το με άλλα υπάρχοντα σύνολα δεδομένων. Όσο το project αποκτούσε όλο και μεγαλύτερες διαστάσεις, εμπλέκονταν σε αυτό όλο και μεγαλύτεροι οργανισμοί, όπως το BBC, το Thomson Reuters και η Βιβλιοθήκη του Κογκρέσου. (64)

Ένα τεράστιο διασυνδεδεμένο σύνολο δεδομένων θα ήταν επωφελές για διάφορους τομείς της ανάπτυξης του Σημασιολογικού Ιστού, συμπεριλαμβανομένων των Προγραμμάτων Περιήγησης του Σημασιολογικού Ιστού και άλλων διεπαφών χρήστη, των Ανιχνευτών του Σημασιολογικού Ιστού (Semantic Web crawlers), των αποθετηρίων RDF και των μηχανών συλλογιστικής (reasoning engines). Το να υπάρχει μια ποικιλία από online χρήσιμα δεδομένα, θα ενθαρρύνει τους ανθρώπους να συνδεθούν σε αυτό και θα μπορούσε να βοηθήσει επίσης στη συνύφανση στο σύνολο του Σημασιολογικού Ιστού. (65)

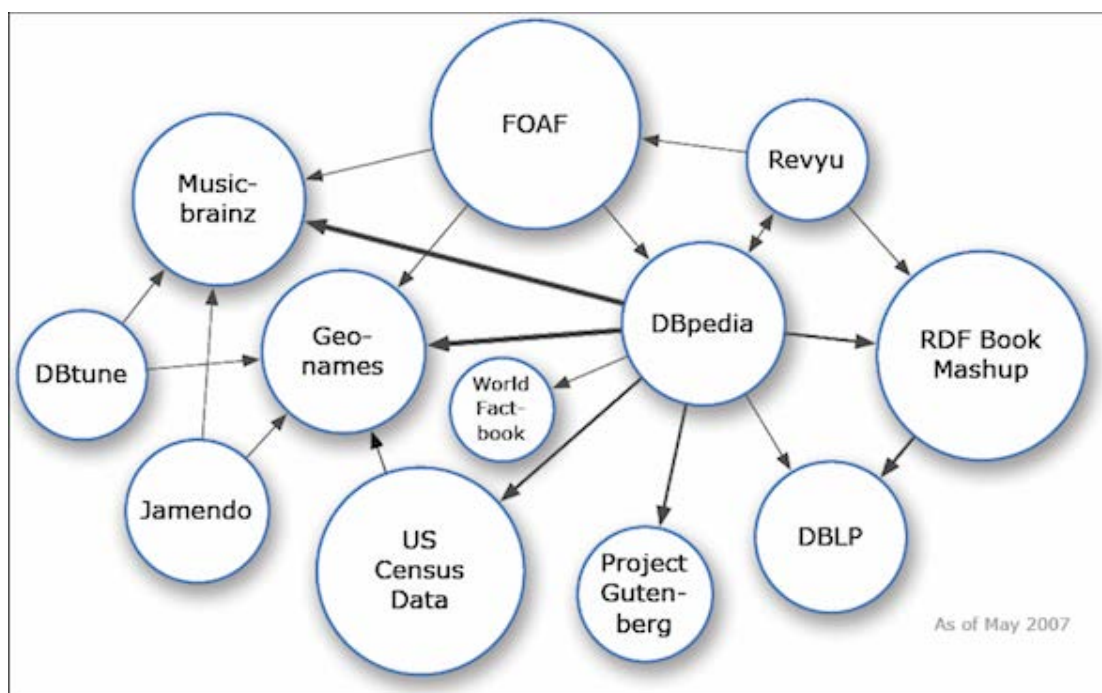
Αυτά τα σύνολα δεδομένων διασυνδέονται με RDF συνδέσμους στη μορφή τριπλετών, συνδέοντας ένα υποκείμενο URI από ένα σύνολο δεδομένων με ένα αντικείμενο URI από ένα άλλο σύνολο δεδομένων. Κάνοντας χρήση αυτών των συνδέσμων μπορεί ο καθένας να πλοηγηθεί. Για παράδειγμα:

- από έναν επιστήμονα υπολογιστών στη DBpedia, στις δημοσιεύσεις του στη βάση δεδομένων DBLP,
- από ένα βιβλίο της DBpedia, σε αξιολογήσεις και προσφορές αγοράς για το βιβλίο αυτό που παρέχονται από το RDF Book Mashup,

⁶⁹ TCL (Talis Community License): Αυτή η μορφή άδειας προορίζεται για την προστασία των δεδομένων, σύμφωνα με την ευρωπαϊκή έννοια για το δίκαιο μιας βάσης δεδομένων, σε αντίθεση με τα πνευματικά δικαιώματα που παρέχονται σε έργα δημιουργίας από την πλειονότητα άλλων ισοδύναμων αδειών. (149)

- από ένα συγκρότημα στη DBpedia, στη λίστα τραγουδιών τους που παρέχεται από το Musicbrainz ή το DBtune. (63)

Το έργο αυτό μπορεί εύκολα να ενσωματωθεί σε άλλα ευρέως διαδεδομένα συστήματα. Τα δεδομένα είναι άμεσα προσβάσιμα από προγράμματα περιήγησης του Σημασιολογικού Ιστού (όπως Tabulator, Disco, OpenLink RDF Browser, και OpenLink Data Explorer). Επιπλέον, οι Ανιχνευτές του Σημασιολογικού Ιστού



Εικόνα 27: LOD Cloud, Μάιος 2007

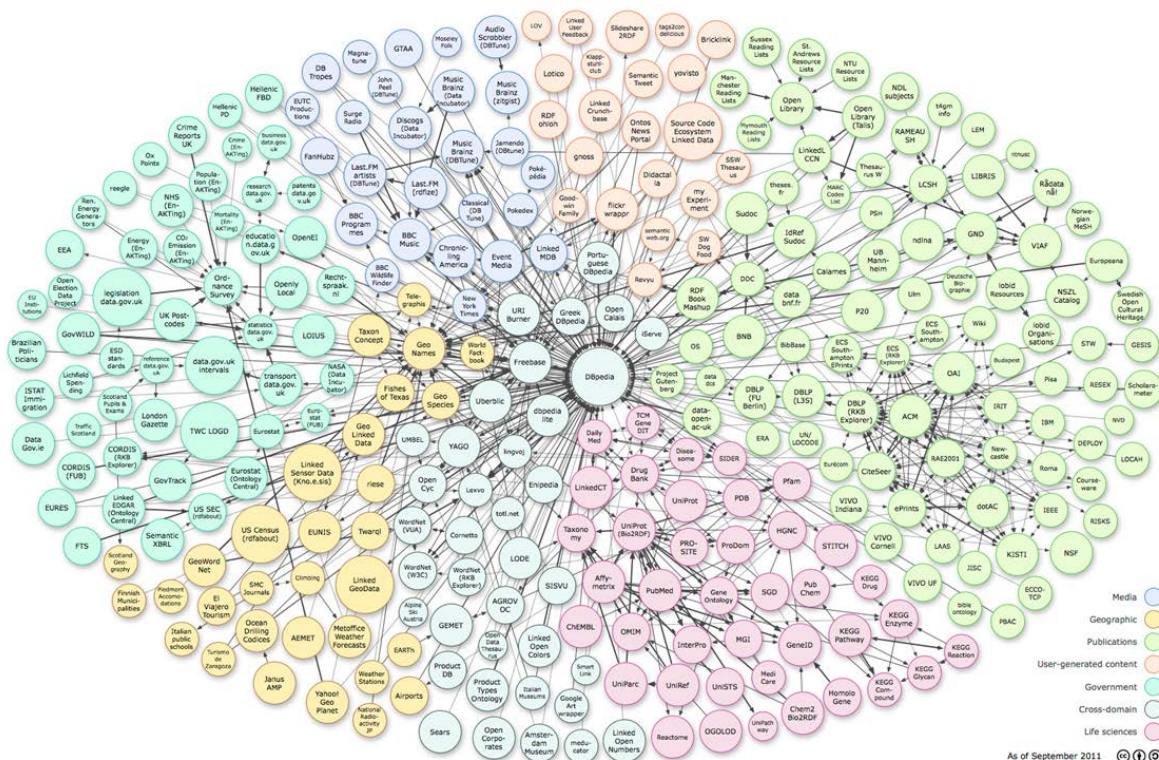
(Semantic Web crawlers), όπως το SWSE⁷⁰ (Semantic Web Search Engine) και το Swoogle⁷¹ παρέχουν μια ολοκληρωμένη άποψη σχετικά με τα δεδομένα και εκλεπτυσμένες διεπαφές αναζήτησης. (65)

Μια διαγραμματική μορφή των μεμονωμένων συνόλων δεδομένων, καθώς και των σχέσεων που έχουν μεταξύ τους, αποτελούν τα σύννεφα LOD (LOD cloud diagram). (66) Το περιεχόμενο του σύννεφου παρατηρούμε πως έχει ποικίλη φύση. Το μέγεθος κάθε κύκλου αντιστοιχεί στο πλήθος των τριπλετών που υπάρχουν σε κάθε σύνολο δεδομένων, τα οποία είτε παρέχονται από το δημιουργό είτε υπολογίζονται με προσεγγιστικές διαδικασίες. Τα βέλη στα παραπάνω σχήματα υποδεικνύουν ότι υπάρχουν τουλάχιστον 50 σύνδεσμοι μεταξύ δύο συνδεμένων

⁷⁰ <http://swse.deri.org/>

⁷¹ <http://swoogle.umbc.edu/>

συνόλων δεδομένων και όσο πιο «παχιά» είναι τα βέλη, τόσο περισσότερους συνδέσμους έχουμε. Τα αμφίδρομα βέλη μεταξύ δύο συνόλων δεδομένων υποδεικνύουν ότι αυτά τα δύο αντικατοπτρίζονται. (67)



Εικόνα 28: LOD Cloud, Σεπτέμβριος 2011

Ένα από τα πρώτα διαγράμματα δημιουργήθηκε την 1 Μαΐου το 2007 και αποτελούνταν από 12 σύνολα δεδομένων. (67) Ως τις 19 Σεπτεμβρίου του 2011 έχουν δημοσιοποιηθεί και διασυνδεθεί από το έργο 295 σύνολα δεδομένων. Αυτά αποτελούνται από 31 δισεκατομμύρια τριπλέτες RDF, οι οποίες είναι διασυνδεδεμένες με περίπου 504 εκατομμύρια RDF συνδέσμους. (66)

Μερικά σύνολα δεδομένων είναι ως συνδετικοί κόμβοι στον Ιστό των Δεδομένων. Ως ένα εμφανές παράδειγμα μπορούμε να αναφέρουμε το σύνολο δεδομένων της DBpedia, το οποίο αποτελείται από RDF τριάδες που έχουν εξαχθεί από τα «infoboxes» (τα οποία βρίσκονται συνήθως στη δεξιά πλευρά των άρθρων της Wikipedia), ενώ το Geonames ⁷² παρέχει RDF περιγραφές εκατομμυρίων γεωγραφικών τοποθεσιών παγκοσμίως. Συνεπώς, καθώς αυτά τα δύο σύνολα δεδομένων παρέχουν URIs και RDF περιγραφές για πολλές συνηθισμένες οντότητες

⁷² <http://www.geonames.org/ontology/>

ή έννοιες, συχνά αναφέρονται και σε άλλα πιο εξειδικευμένα σύνολα δεδομένων και κατ' αυτό τον τρόπο έχουν εξελιχθεί σε κόμβους στους οποίους συνδέονται όλο και περισσότερα σύνολα δεδομένων. (64)

Μετά την διεθνοποίηση της DBpedia, το πρώτο Ελληνικό “νεφος” (cloud) Διασυνδεδεμένων Δεδομένων (Linked Open Data - LOD) έγινε μέρος του παγκόσμιου LOD Cloud στις 20/9/2011. (68) Η Ελληνική DBpedia αποτέλεσε το πρώτο βήμα προς τη διεθνοποίηση των Συνδεδεμένων Δεδομένων (Linked Data) και την πρώτη επιτυχημένη απόπειρα να εξυπηρετήσει Συνδεδεμένα Δεδομένα με «dereferencable» IRI's, κάτι που θα μπορούσε να αποτελέσει έναν γενικό οδηγό για LOD (Linking Open Data) εκδόσεις σε μη Λατινογενείς γλώσσες. (69)

Με την προσθήκη του Ελληνικού Νέφους Διασυνδεδεμένων Δεδομένων, η πληροφοριακή αξία των Παγκόσμιων Διασυνδεδεμένων Δεδομένων (LOD Cloud) εμπλουτίστηκε με τα Ελληνικά Δημόσια, Ανοικτά Δεδομένα του OpenGov, με τα δεδομένα της Ελληνικής Αστυνομίας, της Πυροσβεστικής, και σύντομα θα προστεθούν τα δεδομένα της Διαύγειας⁷³. Ο σκοπός είναι η χρησιμοποίηση αυτών των Διασυνδεδεμένων Δεδομένων σε συνδυασμό και με άλλα, ώστε να δημιουργηθούν κάποιες πρότυπες εφαρμογές, που θα έχουν σαν απώτερο σκοπό την προσφορά υπηρεσιών που θα βελτιώσουν ποιοτικά την διοίκηση και την διαχείριση γνώσης στις δημόσιες υπηρεσίες. (68)

4.3. FOAF

Μια διάσημη εφαρμογή του Σημασιολογικού Ιστού είναι η «Φίλος ενός Φίλου» (Friend of a Friend, Foaf), η οποία δημιουργεί σελίδες «αναγνώσιμες» από τους υπολογιστές, στις οποίες περιγράφονται άτομα, οι σχέσεις τους με άλλα άτομα, καθώς και πράγματα που ασχολούνται ή δημιουργούν. (70) Είναι ένα περιγραφικό λεξιλόγιο που εκφράζεται χρησιμοποιώντας την RDF και την OWL. (71) Θα

⁷³Διαύγεια: Μέσω της ιστοσελίδας <http://diavgeia.gov.gr> παρέχεται πρόσβαση σε όλους στους νόμους και τις αποφάσεις που εκδίδουν τα κυβερνητικά όργανα, οι φορείς του δημόσιου τομέα και οι Ανεξάρτητες Αρχές. (170)



Εικόνα 29: Λογότυπο FOAF

μπορούσαμε πιο συνοπτικά να το ορίσουμε ως ένα τρόπο περιγραφής ατόμων και σχέσεων σε υπολογιστές. (72) Η βασική ιδέα πίσω από αυτό είναι η σύνδεση πραγμάτων. (73)

Το project ξεκίνησε νωρίς το 2000 από τους Dan Brickley και Libby Miller ως ένα πειραματικό project συνδεδεμένων πληροφοριών. (70) Η έμπνευση σχετικά με το FOAF προήλθε μέσα από την σελίδα του Dan

Brickley το 1998, η οποία ήταν μια απλή HTML σελίδα που περιείχε μια RDF περιγραφή του ίδιου, ενώ το RDF ήταν μια νέα και δυσνόητη τεχνολογία. Η ιδέα αυτή αποτέλεσε αρχή για την δημιουργία του FOAF περίπου ένα χρόνο αργότερα, όταν εμφανίστηκαν κάποια πραγματικά RDF εργαλεία για την επεξεργασία τέτοιων περιγραφών. (74)

Το όνομα FOAF επιλέχτηκε για να αντικατοπτρίζει τον προβληματισμό σχετικά με τα δίκτυα κοινωνικής δικτύωσης και το διαδίκτυο, τους αστικούς μύθους, την εμπιστοσύνη και τις συνδέσεις. Η χρήση του συγκεκριμένου ονόματος (FOAF) είναι για να συμπληρώσει και όχι για να αντικαταστήσει αυτές τις προηγούμενες χρήσεις. Ουσιαστικά, τα FOAF έγγραφα περιγράφουν τα χαρακτηριστικά και τις σχέσεις ανάμεσα σε φίλους των φίλων, και τους φίλους τους, και τις ιστορίες που λένε. (75) Μπορεί να θεωρηθεί η πρώτη κοινωνική εφαρμογή του Σημασιολογικού Ιστού, υπό την έννοια ότι συνδυάζει την τεχνολογία του RDF με ανησυχίες πάνω στο «Social Web».

Μέσω του FOAF, το κάθε άτομο μπορεί να περιγράψει τον εαυτό του. Και οι ομάδες ανθρώπων να περιγράψουν και κοινωνικά δίκτυα, χωρίς υπάρχει ανάγκη μιας κεντρικής βάσης δεδομένων. Κάθε προφίλ έχει ένα μοναδικό αναγνωριστικό (όπως για παράδειγμα η ηλεκτρονική διεύθυνση ενός ατόμου), το οποίο χρησιμοποιείται όταν προσδιορίζονται αυτές οι σχέσεις του ατόμου με τα άλλα άτομα και τα «αντικείμενα» γύρω του. (71)

Το FOAF συλλέγει μια ποικιλία δεδομένων, κάποια από τα οποία περιγράφουν άτομα, κάποια άλλα ομάδες και κάποια έγγραφα. Η κάθε εφαρμογή μπορεί να επιλέξει ποια από τα στοιχεία που ορίζονται μέσα από το FOAF. Οι κυρίαρχοι όροι του FOAF μπορούν να χωριστούν σε τρεις κατηγορίες:

- Πυρήνας (Core): Περιγράφουν χαρακτηριστικά των ανθρώπων και των κοινωνικών ομάδων, τα οποία είναι ανεξάρτητα από την πάροδο του χρόνου και την τεχνολογία.
- Κοινωνικός Ιστός (Social Web): Είναι ένα πλήθος χαρακτηριστικών που χρησιμοποιείται κατά την περιγραφή Διαδικτυακών Λογαριασμών (Internet accounts), Βιβλίων Διευθύνσεων (addressbooks) και άλλων δραστηριοτήτων που στηρίζονται στο Ιστό.
- Βοηθητικά Προγράμματα Συνδεδεμένων Δεδομένων (Linked Data utilities): Η βασική ιδέα πίσω από αυτό είναι πως υπάρχει ένα σύνολο στοιχείων που είναι χρήσιμα στην δικτυακή κοινότητα, δίνοντας παράλληλα έμφαση στη κεντρική ιδέα του FOAF, η οποία είναι η σύνδεση δικτύων πληροφοριών με δίκτυα ατόμων. (75)

Υπάρχουν χιλιάδες άνθρωποι, οι οποίοι περιγράφουν τον εαυτό τους και τις ζωές τους στην προσωπική σελίδα τους με περιεχόμενο που είναι κατανοητό μόνο από άλλους ανθρώπους. Το FOAF δίνει την ικανότητα και στις μηχανές να μπορούν να αντιληφθούν αυτό το περιεχόμενο σε βαθμό που να αντιλαμβάνονται τις σχέσεις που συνδέουν τους ανθρώπους ή τα μέρη και τα πράγματα που περιγράφουν στον Ιστό. (73)

Μέσα στη περιγραφή FOAF εντάσσονται το όνομα, η ηλεκτρονική διεύθυνση (email), και οι φίλους ενός συγκεκριμένου ατόμου. Το παραγόμενο αρχείο είναι σε XML και RDF, κάτι που επιτρέπει σε εφαρμογές να επεξεργαστούν αυτές τις περιγραφές, ανακαλύπτοντας πληροφορίες για τα άτομα αυτά και τις ομάδες στις οποίες ανήκουν. (76) Αυτό γίνεται με τον καθορισμό των σχέσεων μεταξύ των ατόμων. Οι υπολογιστές μπορούν να χρησιμοποιήσουν αυτά τα προφίλ FOAF να βρουν, για παράδειγμα, μια λίστα με όλους τους ανθρώπους που ζουν σε μια συγκεκριμένη περιοχή ή μια λίστα με όλους κοινούς γνωστούς δύο ατόμων. (71)

Το FOAF έχει σχεδιαστεί, όντας ένα RDF λεξιλόγιο, να μην απαιτεί ένα συγκεντρωτικό μοντέλο επεκτάσεων. Η χρήση της RDF συμβάλλει σε αυτό, καθώς υπήρχε μια ανάγκη για μια «ελαφριά», συλλογιστική προσέγγιση, ώστε να μπορεί να είναι διαλειτουργικό ακόμα και σε ανεξάρτητες επεκτάσεις των αρχείων FOAF. (74) Προσφέρει την προοπτική υποστήριξης σε πολλές καινοτόμες και ενδιαφέρουσες εξελίξεις στο χώρο των εφαρμογών διαδικτυακών κοινοτήτων. (76) Αλλά, δεν υιοθετήθηκε ευρέως στο Διαδίκτυο, παρά το απλό πρότυπό του και τις σχετικά απλές

περιπτώσεις χρήσης του. (71) Όμως, κάποιες ευρέως χρησιμοποιούμενες εφαρμογές ενδιαφέρθηκαν για τα FOAF δεδομένα και αυτό οδήγησε στη δημιουργία κάποιων projects, όπως:

- Υποστήριξη FOAF από RDF ευρετήρια (όπως το Swoogle), αρχειοθέτηση και αναζήτηση των δημόσιων δεδομένων. (77)
- Εξαγωγή των προφίλ των μελών από Ιστότοπους όπως τα Live Journal και DeadJournal blogging sites, και η κοινότητα My Opera (εξαγωγή προφίλ μελών και ομάδων).
- Η υποστήριξη της FOAF για τον χρήστη (client-side support) είναι διαθέσιμη στο πρόγραμμα περιήγησης Safari, στο πρόγραμμα περιήγησης Firefox (πρόσθετο Semantic Radar), και στο πρόγραμμα περιήγησης Google Chrome (πρόσθετο RDF Detective).
- Δημιουργία μονάδων (modules) και πρόσθετων (plugins) που θα υποστηρίζουν τα FOAF προφίλ ή την FOAF+SSL εξουσιοδότηση για γλώσσες προγραμματισμού, καθώς και για συστήματα διαχείρισης περιεχομένου. (71)

4.3.1. FOAF-a-matic

Γενικά, υπάρχουν αρκετά εργαλεία που σχετίζονται με το FOAF.

- Εργαλεία δημιουργίας περιγραφών FOAF, όπως τα: FOAF-a-matic, Foaf Creator, foafme, scifoaf, Elmo, vIdentity, Export Address Book to FOAF, Morla, FOAFDrive.
- Εργαλεία που συγκεντρώνουν και παρουσιάζουν (για παράδειγμα, μέσω διεπαφής χρήστη) περιγραφές FOAF: FoafNaut, Elmo, myfoafexplorer, FOAF-inator, FOAFMap.
- Εργαλεία που χρησιμοποιούν το FOAF για πιστοποίηση σε εφαρμογές του διαδικτύου: SharedID. (78)

Στη παρούσα φάση, ένας απλός τρόπος δημιουργίας αρχείων FOAF είναι το FOAF-a-matic, συμπληρώνοντας κάποιες απλές φόρμες. (72) Το FOAF-a-matic είναι μια απλή JavaScript εφαρμογή του Leigh Dodd, που επιτρέπει σε κάποιο άτομο την

γρήγορη και απλή δημιουργία περιγραφών FOAF του εαυτού του. Αυτό γίνεται χρησιμοποιώντας μια φόρμα και συμπληρώνοντας στοιχεία όπως: ονοματεπώνυμο, ηλεκτρονική διεύθυνση, πληροφορίες επικοινωνίας, πληροφορίες εκπαίδευσης, γνωστά άτομα (με ονοματεπώνυμο και ηλεκτρονική διεύθυνση). Ο σκοπός για τον οποίο το κάθε άτομο προσθέτει στη περιγραφή του άτομα που γνωρίζει, είναι για να μπορεί κάποιος επεξεργαστής FOAF, όταν διαβάσει και καταλογογραφίσει την περιγραφή, να το συσχετίσει σαν ένα δίκτυο ατόμων. (76)

Αυτή τη περίοδο επεκτείνεται η παραπάνω εφαρμογή στην έκδοση FOAF-a-Matic Mark 2, η οποία έχει κάποια σημαντικά πλεονεκτήματα ως προς τους χρήστες σε σχέση με την πρώτη της έκδοση. Είναι μια επανέκδοση του αρχικού Javascript με κύριες διαφορές πως υπάρχει η δυνατότητα εγκατάστασης αυτού ως πρόγραμμα σε κάθε υπολογιστή και πως η δημιουργία FOAF προφίλ γίνεται μέσω γραφικού περιβάλλοντος το οποίο έχει τα πεδία που καλείται να συμπληρώσει ο χρήστης και στο τέλος εξάγει το RDF/FOAF αρχείο. (79)

4.4. SIOC



Εικόνα 30: Λογότυπο SIOC

Τα παραδοσιακά μέσα πληροφόρησης (βιβλιοθήκες, εκδόσεις) έχουν αντικατασταθεί από τους Ιστοτόπους Κοινοτήτων (online communities) όπως τα blogs, wikis, fora κτλ. Αυτά τα νέα μέσα πληροφόρησης παρέχουν βελτιωμένη και διαδραστική επαφή στο πλαίσιο μιας κοινότητας και πολλές φορές αποτελούν πολύτιμες πηγές αναζήτησης. Οι τεχνολογίες του Σημασιολογικού Ιστού θα μπορούσαν να χρησιμοποιηθούν για τον εμπλουτισμό αυτής της πηγής πληροφοριών, καθιστώντας αυτές αναγνώσιμες όχι μόνο από τους ανθρώπους αλλά και από τις μηχανές.

Οι Ιστότοποι Κοινοτήτων πολλές φορές παρέχουν στοχευόμενες πληροφορίες για κάποιο θέμα που μας ενδιαφέρει και παρέχουν ένα μέρος για στήριξη, συνεργασία και συνεισφορά μεταξύ των μελών τους. Όμως, αυτοί οι Ιστότοποι δεν είναι με κάποιον τρόπο ενωμένοι μεταξύ τους, ακόμα και αν πραγματεύονται παρόμοιες πληροφορίες. Γενικά, τα Online Community Sites χαρακτηρίζονται ως «νησιά χωρίς γέφυρες να τα ενώνουν». Όταν υπάρχει κάποια μερική πληροφορία σε ένα forum, πρόσθετα κομμάτια αυτής θα υπάρχουν σίγουρα σε κάποια άλλη Ιστοσελίδα, κάτι που δε θα εμφανιστεί αυτόματα αλλά θα πρέπει να το εξορύξει ο ίδιος ο χρήστης με κάποιον τρόπο. (80)

Το έργο Semantically-Interlinked Online Communities (SIOC, προφέρεται "shock") είναι μια τεχνολογία του Σημασιολογικού Ιστού η οποία ξεκίνησε το 2004 από τους John Breslin και Uldis Bojars στο DERI και παρέχει μεθόδους για διασύνδεση των μεθόδων συζήτησης, μέσω ενός λεξιλογίου όρων και σχέσεων το οποίο μοντελοποιεί χώρους δεδομένων του Ιστού. (81) Μερικά παραδείγματα τέτοιων χώρων δεδομένων είναι: τα φόρουμ συζητήσεων, τα blogs, οι λίστες ηλεκτρονικού ταχυδρομείου, κοινόχρηστοι σελιδοδείκτες και οι γκαλερί εικόνων.

Οι Κοινοτικές Ιστοσελίδες ενεργούν ως «τόπος συνάντησης» για τους ομοϊδεάτες ανθρώπους και τους πράκτορες λογισμικού, όπου μπορούν να ανταλλάξουν τις ιδέες και τις συνδέσεις. Μια κοινοτική περιοχή από τη σκοπιά SIOC είναι λιγότερο μια επίσημη και περισσότερο μια χαλαρή σύνδεση μεταξύ των ανθρώπων με το ίδιο ενδιαφέρον. Κάθε Κοινότητα έχει τις δικές της απαιτήσεις. Διάφορα εργαλεία που χρησιμοποιούν οι on-line κοινότητες σήμερα είναι:

- Λίστες Ταχυδρομείου (Mailing Lists): Το E-mail αποτελεί το πιο διαδεδομένο μέσο επικοινωνίας στο Internet. Οι Mailing Lists αποτελούν ένα εργαλείο επικοινωνίας για τις online Κοινότητες, το οποίο μπορεί να στηθεί γρήγορα. Ουσιαστικά, είναι η συλλογή ηλεκτρονικών διευθύνσεων που χρησιμοποιείται για να σταλεί κάποιο υλικό σε πολλαπλούς παραλήπτες. Η δομή τους βασίζεται κυρίως σε κείμενο.
- Usenet: Είναι το πιο διάσημο σύστημα newsgroup⁷⁴ στο οποίο οποιοσδήποτε μπορεί να εγκαταστήσει το απαιτούμενο λογισμικό, μπορεί

⁷⁴ Newsgroups: μια συλλογή από κατανεμημένες ανά διακομιστή (server) ομάδες συζήτησης, η οποία αποτελεί ένα από τα παλιότερα αρχέτυπα δημιουργίας Κοινοτήτων που υπάρχουν. (80)

να συμμετάσχει. Το κάθε μήνυμα έχει το ένα δικό του μοναδικό αναγνωριστικό (id), παρόμοιο με αυτό του e-mail.

- Πίνακας Ανακοινώσεων (Bulletin Board): Αποτελεί έναν από τους πιο γρήγορους και γνωστούς τρόπους καθιέρωσης μιας on-line κοινότητας για άτομα με κοινά ενδιαφέροντα. Συνήθως περιέχουν ένα σύνολο από fora χωρισμένα σε κατηγορίες. Δύο από τα πιο γνωστές πλατφόρμες είναι οι vBulletin και phpBB.
- Chat: Είναι ο πιο σύγχρονος τρόπος επικοινωνίας. Κάποια χαρακτηριστικά παραδείγματα είναι το IRC (Internet Relay Chat), το IM (Internet Message) και τα Web chats. Ο πρωταρχικός σκοπός είναι η αποστολή και η λήψη μηνυμάτων κειμένου σε αληθινό χρόνο απόκρισης στο πλαίσιο Κοινοτήτων, αλλά σε πολλά από αυτά είναι δυνατή και η αποστολή αρχείων.
- Ιστολόγιο (Weblogs): Είναι Ιστοσελίδες που ενημερώνονται συνήθως αποκλειστικά από τους δημιουργούς τους και περιέχουν καταχωρήσεις με χρονολογική σειρά. Μπορούν να δημιουργηθούν και από άτομα χωρίς ιδιαίτερη γνώση προγραμματισμού, χρησιμοποιώντας μια απλή διεπαφή (web-based interface).
- Wiki: Ένα wiki είναι μια συλλογικά επεξεργάσιμη Ιστοσελίδα, που επιτρέπει στα άτομα μιας Κοινότητας να ανοίγουν, να διαβάζουν και να γράφουν στις σελίδες της Ιστοσελίδας. Ένα χαρακτηριστικό παράδειγμα είναι η Wikipedia, η οποία είναι μια πολύγλωσση και ανοιχτού περιεχομένου online εγκυκλοπαίδεια. (80)

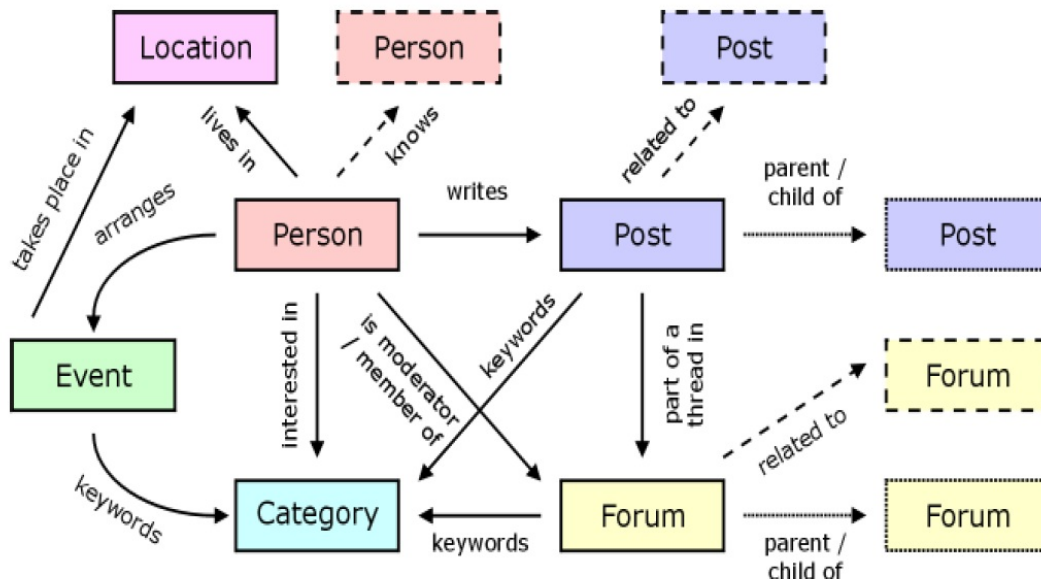
Οι οντολογίες και οι τεχνολογίες του Σημασιολογικού Ιστού μπορούν να εφαρμοστούν σε αυτή τη περίπτωση παράγοντας συνδέσεις μεταξύ των διάφορων εφαρμογών και των τύπων πληροφοριών που παρέχουν. Αυτό θα έχει σαν αποτέλεσμα οι διαθέσιμες πληροφορίες στο διαδίκτυο να γίνονται πιο ολοκληρωμένες, βελτιώνοντας την παροχή πιο σύνθετων υπηρεσιών και πληροφοριών.

Το έργο Semantically-Interlinked Online Communities αποτελείται από την οντολογία SIOC (ένα ανοιχτό πρότυπο αρχείου για διάβασμα από μηχανή για την έκφραση των πληροφοριών που περιέχονται τόσο με άμεσο όσο και με έμμεσο τρόπο στις μεθόδους συζήτησης στο Διαδίκτυο), από τους παραγωγούς μεταδεδομένων

SIOC για ένα πλήθος διάσημων πλατφορμών για Ιστολόγια και συστημάτων διαχείρισης περιεχομένου ⁷⁵ και από αποθηκευτικά συστήματα και συστήματα περιήγησης-αναζήτησης για την αξιοποίηση αυτών των SIOC δεδομένων. (81)

Η οντολογία SIOC συνδυάζει την ορολογία από τα ήδη υπάρχοντα λεξιλόγια με τους νέους όρους που απαιτούνται για να περιγράψουν τις σχέσεις μεταξύ των εννοιών στη σφαίρα των διάφορων online communities. (80) Το λεξιλόγιο SIOC βασίζεται στην RDF και προσδιορίζεται χρησιμοποιώντας την RDFS. Στα έγγραφα SIOC μπορούν να χρησιμοποιηθούν και άλλες υπάρχουσες οντολογίες για τον εμπλουτισμό των πληροφοριών που περιγράφονται, όπως για παράδειγμα να γίνει χρήση του FOAF λεξιλογίου για προσθήκη πληροφοριών σχετικά με το δημιουργό μιας δημοσίευσης. (81)

Υπάρχουν διαφορετικών ειδών ρόλοι στις online Κοινότητες: Υπάρχουν αυτοί που δημοσιεύουν τις πληροφορίες (πάροχοι πληροφοριών), αυτοί που τις χρησιμοποιούν (καταναλωτές πληροφοριών) και τέλος αυτοί που παρέχουν την αναγκαία υποδομή που χρειάζεται για την δημοσίευση και ανταλλαγή αυτών των πληροφοριών (πάροχοι υποδομής). Το δύσκολο κομμάτι είναι να πειστεί ο κόσμος να



Εικόνα 31: Διασύνδεση Κοινωνικών Ιστοσελίδων.

⁷⁵ Σύστημα Διαχείρισης Περιεχομένου (Content Management System): Είναι ένα σύστημα που παρέχει μια συλλογή χειροκίνητων ή και αυτόματων διαδικασιών που χρησιμοποιούνται για τη διαχείριση της ροής εργασίας σε ένα περιβάλλον συνεργασίας. (143)

υιοθετήσει και να χρησιμοποιήσει αυτές τις τεχνολογίες. Για να υπάρξει πρόσβαση σε πληροφορίες σε RDF στο Σημασιολογικό Ιστό, θα πρέπει οι πάροχοι υποδομής (webmasters, mailing list administrators) να «ανοίξουν» τις βάσεις τους για τις τεχνολογίες πρόσβασης του Σημασιολογικού Ιστού.

Η οντολογία SIOC μπορεί να διαχειριστεί και να προσαρμοστεί σε όλα τα εργαλεία που αναφέρθηκαν έως τώρα. Τα περισσότερα από αυτά τα εργαλεία αποθηκεύουν τα δεδομένα τους σε μια ημιδομημένη μορφή και οι πιο προηγμένες Ιστοσελίδες όπως τα bulletin boards, τα weblogs, και τα wikis χρησιμοποιούν σχεσιακές βάσεις δεδομένων⁷⁶ για να αποθηκεύουν τα δεδομένα τους. Σε περιπτώσεις που τα δεδομένα δεν είναι αρχειοθετημένα, όπως για παράδειγμα στα κανάλια του IRC, τα bots⁷⁷ μπορούν να δώσουν πρόσβαση σε όσα λέγονται μέσα σε μια συνομιλία. Δε χρειάζεται κάποια κεντρική αρχή που να μπορεί να παρέχει όλες τις υπηρεσίες που χρειαζόμαστε, αλλά να αρχίσουμε από την κοινοτική περιοχή μας και ανακαλύψουμε τις σχετικές περιοχές μέσω των συνδέσεων που προστίθενται χειροκίνητα. Ο στόχος του SIOC είναι να καταγράψει όλες τις σχετικές με τα community sites πληροφορίες, δημιουργώντας ένα μεγάλο εύρος πληροφοριών και ταυτόχρονα μια απλή οντολογία για να πλοηγούνται εύκολα οι χρήστες. Στην Εικόνα 31 παρατηρούμε τις σχέσεις μεταξύ των κύριων κλάσεων που χρειάζονται για τη διασύνδεση Κοινοτικών Ιστοσελίδων. (80)

Το όραμα του SIOC είναι η δημιουργία μιας οντολογίας η οποία να είναι σε θέση να περιγράψει πλήρως το περιεχόμενο και τη δομή των περισσότερων από τις online Κοινοτικές Ιστοσελίδες και επιπλέον τη δημιουργία νέων συνδέσεων μεταξύ καναλιών συζητήσεων (discussion channels) και δημοσιεύσεων (posts), με σκοπό να δημιουργηθούν νέοι και ποιο ενδιαφέροντες τρόποι αναζήτησης, με χρήση των παραπάνω εργαλείων. Επίσης, έχει ως σκοπό να αντιμετωπίσει το πρόβλημα στο Σημασιολογικό Ιστό σύμφωνα με το οποίο δεν υπάρχουν εφαρμογές χωρίς δεδομένα και δεν υπάρχουν δεδομένα χωρίς εφαρμογές, καθιστώντας εύκολη τη δημιουργία SIOC δεδομένων. (82)

⁷⁶Σχεσιακή βάση δεδομένων: μια συλλογή δεδομένων σε συσχετισμένους πίνακες με δυνατότητα μηχανισμού ανάγνωσης, εγγραφής, τροποποίησης και πολλών άλλων πολυπλοκότερων λειτουργιών.

⁷⁷Τα bots στο IRC αποτελούν κομμάτια κώδικα (scripts) που εμφανίζονται σαν άτομα μέσα στα κανάλια αλλά μπορούν να εκτελούν εντολές είτε αυτόματα ή χειροκίνητα (παρέχοντας πρόσβαση στις λειτουργίες τους σε συγκεκριμένους χρήστες), σύμφωνα με το πώς είναι προγραμματισμένα.

Έχει ήδη προταθεί η μελλοντική δυνατότητα για μια peer-to-peer⁷⁸ έκδοση του SIOC, όπου το προφίλ του χρήστη και οι δημοσιεύσεις θα διατηρούνται στην δική τους αρχική μηχανή σε αντίθεση με μια κεντρική Ιστοσελίδα. Συνεπώς, η οντολογία SIOC αποτελεί ένα βήμα προς στη σωστή κατεύθυνση, καθώς η προτεινόμενη δομή αποτελείται από πολλές διασυνδεδεμένες κοινότητες που μπορούν εν δυνάμει να αναλυθούν περαιτέρω σε κοινότητες των διασυνδεδεμένων μεμονωμένων ατόμων. (80)

4.4.1. Υλοποιήσεις και Εφαρμογές

Υπάρχουν διάφορες υλοποιήσεις και εφαρμογές που δημιουργούν ή/και χρησιμοποιούν SIOC δεδομένα.

Για τη δημιουργία SIOC δεδομένων μπορούμε να δημιουργήσουμε μέσω php ή java κάποιους εξαγωγείς SIOC ή να χρησιμοποιήσουμε τους ήδη έτοιμους που υπάρχουν (όπως για παράδειγμα για το WordPress: WordPress SIOC Exporter, για το Drupal: Drupal SIOC Exporter, για την phpBB 2.x: phpBB 2.x SIOC Exporter, για τις Mailing Lists: Mailing List Archives και Mailing List Exporter, για το IRC: IRC2RDF, για το Twitter: Twitter2RDF, για τα Media Wiki: SIOC-MediaWiki exporter, κ.α.).

Για τη χρήση των SIOC δεδομένων, τα εργαλεία που υπάρχουν χωρίζονται στις εξής υποκατηγορίες: Αυτά που χρησιμοποιούν την SIOC για νέα δεδομένα (Fishtank, IkeWiki, Talis Engage, κ.α.), αυτά που εκτελούν ανίχνευση SIOC δεδομένων (SIOC Crawler, SIOC Browsers, SIOC Explorer κ.α.), αυτά που εντοπίζουν δεδομένα SIOC μέσω ερωτημάτων (query) ή χρησιμοποιώντας SPARQL endpoints⁷⁹ (ODS demo server, MyOpenLink.net, #B4mad.Net SPARQL endpoint, κ.α.) και αυτά που επαναχρησιμοποιούν τα ήδη υπάρχοντα SIOC δεδομένα (IKHarvester κ.α.).

⁷⁸ Peer-to-peer: Είναι μια αρχιτεκτονική διασύνδεσης υπολογιστών ή δικτύων, στην οποία οι πόροι μοιράζονται ισοδύναμα και οι κόμβοι έχουν ίσα δικαιώματα μεταξύ τους.

⁷⁹ SPARQL endpoint: Ένα SPARQL endpoint επιτρέπει στους χρήστες να κάνουν ερωτήματα πάνω σε μια βάση γνώσης με χρήση της γλώσσας SPARQL. (152)

Τέλος, υπάρχουν εφαρμογές που εντοπίζουν και αρχειοθετούν τα SIOC δεδομένα, όπως το Semantic Radar (πρόσθετο του περιηγητή Firefox) και το PingTheSemanticWeb.com. (81) Και τα δύο διαθέτουν τη δυνατότητα να ειδοποιούν τους χρήστες για τον εντοπισμό μεταδεδομένων για επιπλέον πηγές από αυτές που έχουν ήδη εντοπιστεί, κάτι που γνωρίζουν όσοι δημιουργούν ή ανεβάζουν μεταδεδομένα στην Ιστοσελίδα τους (83) ή γνωρίζουν για την ύπαρξή τους στο πλαίσιο μιας Κοινότητας (84) .

4.5. NextBio



Εικόνα 32: Λογότυπο Nextbio

Το NextBio είναι μια ιδιωτική εταιρεία λογισμικού που παρέχει μια πλατφόρμα για τις φαρμακευτικές εταιρείες και τους ερευνητές των επιστημών της υγείας, με σκοπό να αναζητούν, να ανακαλύπτουν και να μοιράζονται τη γνώση μεταξύ δημόσιων δεδομένων, αλλά και όσα υπόκεινται σε αποκλειστική εκμετάλλευση. Ιδρύθηκε το 2004 από τους Saeid Akhtari, Ilya Kupersmidt και Mostafa Ronaghi και εδρεύει στο Cupertino, στην Καλιφόρνια των Η.Π.Α. και έλαβε το 2008 το βραβείο Frost & Sullivan (North American Life Sciences Customer Value Enhancement).

Ουσιαστικά, αποτελεί μια βάση δεδομένων για την ενοποίηση υψηλής απόδοσης πειραματικών δεδομένων βιοεπιστημών, τα οποία επισημαίνονται και συνδέονται μέσω οντολογιών της βιοϊατρικής. Περιλαμβάνει στοιχεία από διαφορετικούς οργανισμούς, πλατφόρμες, τύπους δεδομένων και ερευνητικά πεδία, που είναι ενσωματωμένα και συσχετίζονται μέσα σε ένα ενιαίο περιβάλλον αναζήτησης και επεκτείνεται σταθερά για να υποστηρίζει και άλλους τύπους βιολογικών δεδομένων.

Παρέχει ένα ενοποιημένο περιβάλλον εργασίας για τους ερευνητές για να διατυπώσουν και να δοκιμάσουν νέες υποθέσεις σε τεράστιες συλλογές πειραματικών δεδομένων. Μέσω μιας ενοποιημένης διεπαφής, οι ερευνητές μπορούν να διατυπώνουν και να εξετάζουν νέες υποθέσεις σε τεράστιες συλλογές πειραματικών

δεδομένων. Είναι προβάσιμο μέσω ενός search engine interface στην ηλεκτρονική διεύθυνση <http://www.nextbio.com/b/nextbio.nb>. (85)

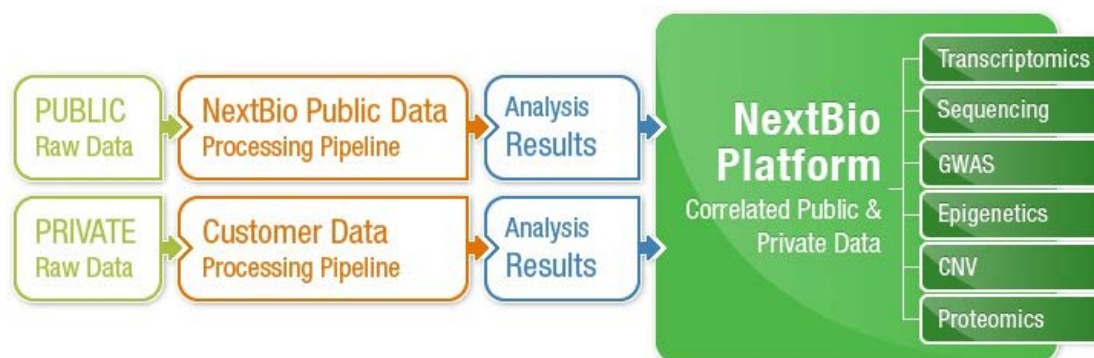
Το NextBio υπάρχει σε τρεις εκδόσεις, για να ανταποκρίνεται καλύτερα στις επιμέρους ανάγκες των επιστημόνων και των ερευνητικών ομάδων. Κάθε μια από αυτές έχει όλο και περισσότερες δυνατότητες:

- NextBio Basic: Είναι δωρεάν και διαθέσιμη μέσω διαδικτυακής εφαρμογής.
- NextBio Professional: Είναι σχεδιασμένο για τους μεμονωμένους ερευνητές ή τις ομάδες ερευνητών για να μπορέσουν να αξιοποιήσουν με το πιο αποτελεσματικό τρόπο τόσο τις δημοσίως διαθέσιμα δεδομένα όσο και τα εσωτερικά παραγόμενα πειραματικά δεδομένα για νέες επιστημονικές ανακαλύψεις.
- NextBio Enterprise: Είναι μια πλήρως ενοποιημένη έκδοση Enterprise, η οποία έχει σχεδιαστεί για τις φαρμακευτικές και βιοτεχνολογικές εταιρείες, τους εμπορικούς και μη κερδοσκοπικούς οργανισμούς και τα ακαδημαϊκά ιδρύματα.

Η Πλατφόρμα NextBio είναι ένα σημασιολογικό πλαίσιο που βασίζεται στις οντολογίες. Συνδέει σε μεγάλο βαθμό ετερογενή στοιχεία και πληροφορίες σε μορφή κειμένου. Το σημασιολογικό πλαίσιο βασίζεται σε οντολογίες γονιδίων, ιστών, ασθενειών και χημικών ενώσεων. (86)

Η Enterprise έκδοση (έκδοση για επιχειρήσεις) επιτρέπει στα εσωτερικά ιδιόκτητα δεδομένα (proprietary data) να ενσωματωθούν στη βάση δεδομένων του Nextbio των δημοσίως διαθέσιμων δεδομένων. Σύμφωνα με την εταιρεία, η Enterprise version της πλατφόρμας NextBio χρησιμοποιείται στην έρευνα των επιστημών της υγείας και την ανάπτυξη και την ανάπτυξη φαρμάκων από ερευνητές και οι νοσοκομειακούς γιατρούς σε: Merck Pharmaceutical, Johnson & Johnson Pharmaceutical Research & Development, L.L.C., Celgene, Genzyme, Eli Lilly and Company, και Regeneron Pharmaceuticals. (85) Επίσης, το University of Southern California (USC), ένα από τα κυρίαρχα ερευνητικά ιδρύματα της Αμερικής, χρησιμοποιεί την έκδοση NextBio Enterprise για ερευνητική χρήση από τους φοιτητές και από το προσωπικό. Το NextBio προσφέρει ήδη πρόσβαση στα ακαδημαϊκά ιδρύματα: στη μεγάλη συλλογή των συνεχώς επιμελημένων

γονιδιωματικών πληροφοριών και των ενισχυμένων δεδομένων και στις λειτουργίες αναζήτησης βιβλιογραφίας. (87)



Εικόνα 33: Πλατφόρμα NextBio

Η τεχνολογία που αναπτύχθηκε για το NextBio επιτρέπει στους ερευνητές στις βασικές επιστήμες, στην ανακάλυψη φαρμάκων και στις κλινικές έρευνες να αποκτήσουν μια συσχετισμένη άποψη από τις μεγάλες ποσότητες των επεξεργασμένων δημόσιων δεδομένων, σε συνδυασμό με τις δικές τους βάσεις δεδομένων. Χρησιμοποιώντας NextBio, οι ερευνητές μπορούν να εξερευνήσουν αυτές τις διαφορετικές μορφές των δεδομένων σε ένα ενιαίο περιβάλλον με πρωτοφανή ταχύτητα και εύρος, κάνοντας μόνο μερικά κλικ του ποντικιού. (88) Οι ερευνητές μπορούν, επίσης, να συνεισφέρουν τα ευρήματά τους για να ενσωματωθούν στη βάση. (9)

4.6. Cys

Η Cycorp αποτελεί τον κυρίαρχο πάροχο σημασιολογικών τεχνολογιών, που φέρνουν ένα νέο επίπεδο ευφυΐας και συλλογιστικής της κοινής λογικής σε μια ποικιλία εφαρμογών λογισμικού. Έχει έδρα το Όστιν του Τέξας και γραφεία στη Λιουμπλιάνα της Σλοβενίας.

Το Cys λογισμικό είναι μια μακροχρόνια προσπάθεια για την ανάπτυξη πραγματικής τεχνητής νοημοσύνης. Συνδυάζει μια απaráμιλλη οντολογία κοινής λογικής (common sense ontology), μια βάση γνώσης με μια ισχυρή συλλογιστική

μηχανή, και διεπαφές στη φυσική γλώσσα με σκοπό να καταστήσει εφικτή την ανάπτυξη καινοτόμων εφαρμογών.

Ξεκίνησε σαν ερευνητικό έργο πριν το 1984 στο Atari research center, με ιδρυτή και επικεφαλής της προσπάθειας του τον Douglas Lenat. Το 1994, ιδρύθηκε η Cycorp για την περαιτέρω ανάπτυξη, εμπορευματοποίηση και για την εφαρμογή της τεχνολογίας του Cyc. Η Cycorp προσφέρει μια άνευ κόστους άδεια στη σουίτα ανάπτυξης λογισμικού (development toolkit) των σημασιολογικών τεχνολογιών της στην κοινότητα των ερευνητών με σκοπό να προωθήσει την ανάπτυξη της σημασιολογικής συλλογιστικής (semantic reasoning) από την ερευνητική κοινότητα. Επίσης, ο πυρήνας της οντολογίας Cyc είναι δημοσίως προσβάσιμος. (89)

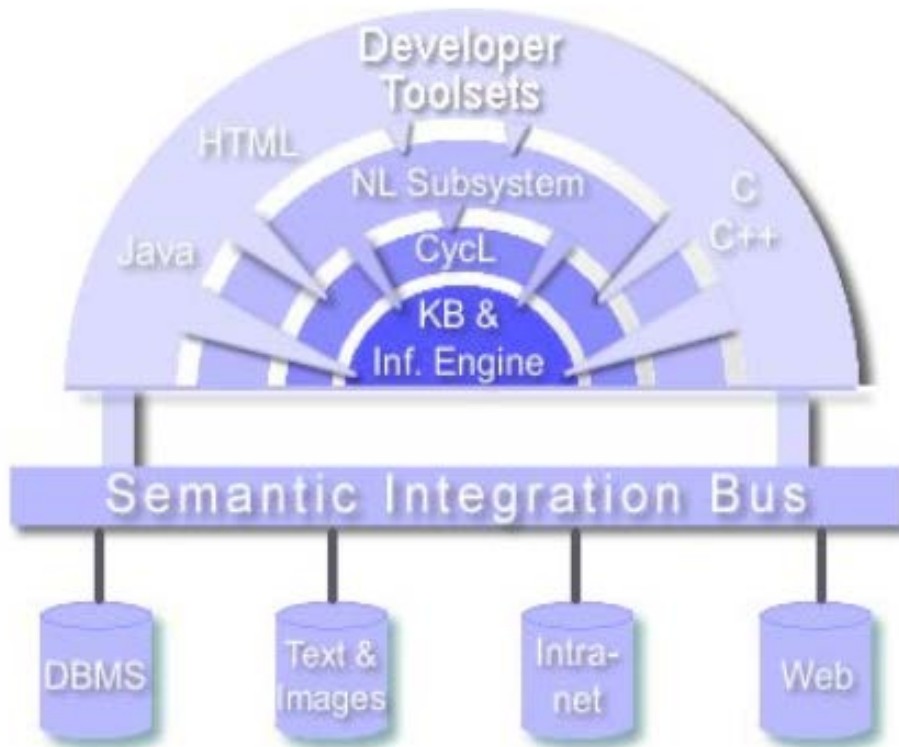
4.6.1. O Cyc Knowledge Server

Ο Cyc Knowledge Server είναι μια πολύ μεγάλη “multi-contextual” βάση γνώσης και μηχανή εξαγωγής συμπερασμάτων. Αναπτύχθηκε από την CyCorp με στόχο να λύσει το πρόβλημα του «εύθραυστου» λογισμικού, κατασκευάζοντας μια θεμελίωση από βασική γνώση «κοινής λογικής» (δηλαδή, ένα σημασιολογικό υπόστρωμα από όρους, κανόνες και σχέσεις), που θα οδηγήσει στη δημιουργία μιας ποικιλίας προϊόντων και υπηρεσιών, που να στηρίζονται σε μεγάλο βαθμό στην γνώση που περιέχουν. Το Cyc έχει σκοπό την παραγωγή ενός «βαθέος» στρώματος κατανόησης, το οποίο θα μπορεί να χρησιμοποιηθεί και από τα άλλα προγράμματα για την αύξηση της ευελιξίας τους. (90)

Η τεχνολογία του Cyc περιλαμβάνει τα παρακάτω βασικά συστατικά:

- Η βάση γνώσης της Cyc - the Cyc knowledge base: είναι μια τυποποιημένη αναπαράσταση μιας τεράστιας ποσότητας στοιχειώδους ανθρώπινης γνώσης (fundamental human knowledge): γεγονότα, εμπειρικούς και ευρεστικούς κανόνες, για συλλογισμούς γύρω από αντικείμενα και γεγονότα της καθημερινής ζωής. Η βάση γνώσης αποτελείται από όρους (οι οποίοι αποτελούν το λεξιλόγιο του CycL) και ισχυρισμούς (assertions), που συσχετίζουν τους όρους και περιλαμβάνουν απλούς θεμελιώδεις ισχυρισμούς, καθώς επίσης και κανόνες (rules). Έως

σήμερα, περιέχει περίπου πεντακόσιες χιλιάδες όρους, συμπεριλαμβανομένων περίπου δεκαπέντε χιλιάδων ειδών σχέσεων, και



Εικόνα 34: O Cyc knowledge server

πέντε εκατομμύρια δηλώσεων που αφορούν αυτούς τους όρους. (91)

- Η μηχανή εξαγωγής συμπερασμάτων της Cyc - the Cyc Inference Engine: Εκτελεί γενική λογική συνεπαγωγή⁸⁰ (logical deduction), modus ponens⁸¹, modus tollens⁸², καθολική και υπαρξιακή ποσοτικοποίηση (universal and existential quantification) με πολύ γνωστούς μηχανισμούς εξαγωγής συμπερασμάτων από την τεχνητή νοημοσύνη, όπως την κληρονομικότητα (inheritance), την αυτόματη ταξινόμηση (automatic classification) κ.ά. σαν ειδικές περιπτώσεις.

⁸⁰ Λογική συνεπαγωγή (logical deduction): Αποτελεί μία βασική ιδιότητα των γλωσσικών εκφράσεων και προκύπτει από τα λεγόμενα που δηλώνονται ρητά.

⁸¹ Modus ponens: Εάν μια δήλωση ή πρόταση επάγει μια δεύτερη, και η πρώτη δήλωση είναι αληθής, τότε και η δεύτερη είναι αληθής. (161)

⁸² Modus tollens: Εάν μια δήλωση ή πρόταση επάγει μια δεύτερη και η πρώτη δεν είναι αληθής, τότε και η δεύτερη δε μπορεί να είναι αληθής. (162)

- Η γλώσσα αναπαράστασης CycL - the CycL Representation Language: Η CycL είναι μια μεγάλη και ευέλικτη γλώσσα αναπαράστασης γνώσης. Είναι μια τυπική γλώσσα, το συντακτικό της οποίας προέρχεται από τον Πρωτοβάθμιο Κατηγορηματικό Λογισμό⁸³ (first-order predicate calculus) (FOPC) και την Lisp⁸⁴. Αποτελεί μια προσαύξηση του Πρωτοβάθμιου Κατηγορηματικού Λογισμού με επεκτάσεις, έτσι ώστε να μπορεί να χειριστεί την ισότητα, τους προκαθορισμένους συλλογισμούς, το skolemization⁸⁵ και τις δυνατότητες της Λογικής Δεύτερης Τάξης. Η CycL χρησιμοποιεί και ένα είδος περιορισμού (circumscription), περιλαμβάνει την υπόθεση μοναδικού ονόματος (unique name assumptions) και επίσης κάνει χρήση της υπόθεσης του κλειστού κόσμου (closed world assumption) όποτε αυτό είναι απαραίτητο. (92)
- Το υποσύστημα της επεξεργασίας της φυσικής γλώσσας - the Natural Language Processing Subsystem: Η επεξεργασία φυσικής γλώσσας είναι μια από τις πιο μελετημένες προκλήσεις της ανάπτυξης λογισμικού, που θα παρέχει την δυνατότητα στα συστήματα φυσικής γλώσσας να διαβάσουν και να βγάλουν νόημα από απλά αγγλικά κείμενα. Έχουν γίνει προσπάθειες για την παραγωγή συστημάτων φυσικής γλώσσας (Natural Language – NL systems) από πολλές ομάδες, αλλά μέχρι τώρα καμία δεν το έχει πετύχει σε σημαντικό βαθμό πέρα από περιορισμένους προκατασκευασμένους τομείς.
- Cyc Semantic Integration Bus: Η πληροφορία που υπάρχει σε ένα υπολογιστικό σύστημα, αποτελείται από δεδομένα που είναι δομημένα (Βάσεις Δεδομένων), ημιδομημένα (Λογιστικά φύλα, Ιστοσελίδες), και αδόμητα (Αρχεία κειμένου) και μπορεί να αποθηκευτεί σε πολλές μορφές. Η Cyc μπορεί να μετατρέψει ένα τμήμα αυτής της πληροφορίας σε χρήσιμη γνώση και το υπόλοιπο μέρος της πληροφορίας μπορεί να σημειωθεί για ευκολότερη πρόσβαση από τους ανθρώπους. Αντιμετωπίζει

⁸³ Πρωτοβάθμιος Κατηγορηματικός Λογισμός (FOPC, First-Order Predicate Calculus): Μια εναλλακτική/παλαιότερη ονομασία της Λογικής Πρώτης Τάξης.

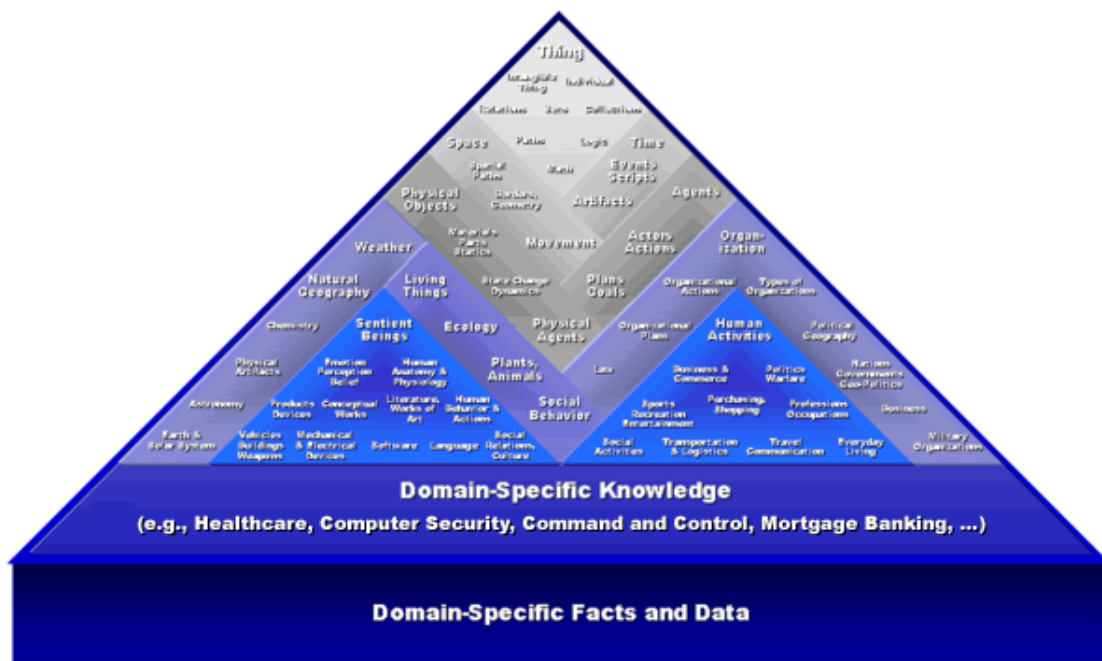
⁸⁴ Lisp: Οικογένεια γλωσσών προγραμματισμού υπολογιστών υψηλού επιπέδου. (168)

⁸⁵ Skolemization: Είναι μια μέθοδος για την αφαίρεση υπαρξιακών περιορισμών από δηλώσεις της τυπικής λογικής. (156)

κάθε εγγραφή της βάσης δεδομένων σαν να είναι ένας υπονοούμενος ισχυρισμός στη βάση γνώσης, ο οποίος είναι μετά διαθέσιμος κατά τη διάρκεια της εξαγωγής συμπερασμάτων. Κατά έναν παρόμοιο τρόπο, τα πεδία κειμένου μπορούν να διαβαστούν με τη χρήση του επεξεργαστή φυσικής γλώσσας, για να διαπιστωθεί εάν περιέχουν κάποιους χρήσιμους υπονοούμενους ισχυρισμούς, οι οποίοι ενδεχομένως να περιγράφουν τι πραγματεύεται το κείμενο. Η Cyc κάνει χρήση αυτής της πληροφορίας για τον εντοπισμό και την αναφορά πηγών πληροφοριών, που ο χρήστης μπορεί να χρησιμοποιήσει για να απαντήσει σε συγκεκριμένα ερωτήματα.

- Εργαλεία διεπαφής Cyc - Cyc Developer Toolsets: Το σύστημα της Cyc, περιλαμβάνει επίσης μια πληθώρα εργαλείων διεπαφής, που επιτρέπουν στο χρήστη την περιήγηση, την επεξεργασία και την επέκταση της λειτουργίας της βάσης γνώσης, τη διενέργεια ερωτημάτων στην μηχανή διεπαφών και την αλληλεπίδραση με τα “natural language” και “database integration” modules. Το πιο δημοφιλές εργαλείο είναι ο HTML browser, που επιτρέπει στο χρήστη να δει την βάση γνώσης σε μια hypertext διάσταση. Οι HTML σελίδες, που περιγράφουν ορολογία της Cyc παράγονται στη στιγμή από το σύστημα. (91)

Το ακρωνύμιο Cyc προέρχεται από την αγγλική λέξη "encyclopedia" (εγκυκλοπαίδεια). Το Cyc project έχει δημιουργηθεί για να αναπαριστά σε χρήσιμη



Εικόνα 35: Στιγμιότυπο της βάσης γνώσης της Cyc

για τον υπολογιστή μορφή τη γνώση που απαιτείται, για να γίνει κατανοητή σε αυτόν μια εγκυκλοπαίδεια. Όχι τη γνώση στην εγκυκλοπαίδεια αυτήν καθ' αυτήν, αλλά τα πράγματα που ο δημιουργός της εγκυκλοπαίδειας διαισθάνεται ότι είναι τόσο ευρέως αντιληπτά, που τα έχει θεωρήσει τετριμμένα και παραλείπει να τα αναφέρει. Αυτή η γνώση μπορεί να ληφθεί ως δεδομένη σύμφωνα με αυτό που αποκαλούμε «κοινή λογική». (93)

Η οντολογία Cyc φαίνεται να είναι «καθολική», υπό την έννοια ότι οτιδήποτε μπορεί να φανταστεί κάποιος (όσο γενικό ή ειδικό ή πεζό ή ακόμα και απόκρυφο μπορεί να είναι) είναι δυνατό να συνδεθεί σωστά με οντολογία στα κατάλληλα σημεία. (94) Η βάση γνώσεων έχει δημιουργηθεί πάνω σε ένα πυρήνα 1.000.000 δηλώσεων (ή αλλιώς κανόνων) που σχεδιάστηκαν για να καλύψουν ένα μεγάλο τμήμα αυτού που θεωρείται η «κοινή γνώση» (consensus knowledge) γνώση για τον κόσμο. (95) Η βάση γνώσης της Cyc είναι τόσο μεγάλη, που φαίνεται δύσκολο να μπορέσει να πλοηγηθεί κάποιος σε αυτήν με την πρώτη ματιά. (96)

Η ανάπτυξη του Cyc αποτέλεσε ένα μακροχρόνιο και υψηλού κινδύνου στοίχημα, το οποίο πλέον έχει αρχίσει να αποδίδει καρπούς. Το Cyc έχει αρκετές εφαρμογές που είναι ήδη διαθέσιμες ή υπό ανάπτυξη, όπως: επεξεργασία της φυσικής γλώσσας (natural-language processing), ενοποίηση των ετερογενών βάσεων δεδομένων (integration of heterogeneous databases), ενισχυμένη ανάκτηση γνώσης από πληροφορίες επικεφαλίδων, διαχείριση εγκυκλοπαιδικής γνώσης, διανεμημένα τεχνητή ευφυΐα και διαδικτυακή ανάκτηση πληροφοριών. Υπάρχουν επίσης και άλλες πιθανές εφαρμογές, όπως «έξυπνες» διεπαφές, ευφυής προσομοίωση χαρακτήρα για παιχνίδια (intelligent character simulation for games), ενισχυμένη εικονική πραγματικότητα, βελτιωμένη αυτόματη μετάφραση, βελτιωμένη αναγνώριση ομιλίας, προηγμένη προσομοίωση χρήστη και σημασιολογική εξόρυξη δεδομένων (semantic data mining). (97)

4.6.2. OpenCyc

Η OpenCyc αποτελεί μια έκδοση ανοιχτού κώδικα της τεχνολογίας Cyc (μπορεί να αντιγραφεί, να διανεμηθεί και να χρησιμοποιηθεί για εμπορικούς και μη

σκοπούς, σύμφωνα με τους όρους της άδειας της (93)), η οποία θεωρείται η μεγαλύτερη και πληρέστερη βάση γενικής γνώσης και μηχανή συλλογισμού της κοινής λογικής. Περιέχει το σύνολο όλων των μη ιδιόκτητων (non-proprietary) όρων της Cys, όπως επίσης και εκατομμύρια ισχυρισμών, τα οποία προσφέρονται άνευ κόστους από την Cycorp. (98) Η OpenCyc είναι γραμμένη στη γλώσσα αναπαράστασης CycL, όπως και η Cys. Μπορεί να θεωρηθεί ως βάση πολλών «ευφυών» εφαρμογών, όπως για παράδειγμα:

- η ταχεία επιπεδοποιημένη ανάπτυξη μιας οντολογίας,
- η ιεράρχηση της προτεραιότητας, η δρομολόγηση, η περίληψη, η επισημείωση των μηνυμάτων του ηλεκτρονικού ταχυδρομείου (email),
- τα έμπειρα συστήματα,
- τα παιχνίδια. (99)

Η συνεργασία της Cys με την XML παρέχει αμφίδρομα οφέλη: Η κοινότητα της XML παρέχει μια ισχυρή σύνταξη για την προσθήκη δομής στο διαδίκτυο και η Cys με τη σειρά της ενισχύει την XML παρέχοντας μια ισχυρή και καθολική σημασιολογία για τη μοντελοποίηση των αντικειμένων που περιγράφονται μέσω της XML. Ο συνδυασμός τους παρέχει τη δυνατότητα για τη δημιουργία νέων και ισχυρών εφαρμογών που θα προκύψουν με βάση την κατανόηση της γνώσης από τις μηχανές, κάτι που πρωτότερα ήταν αδύνατο. (93)

Η πρώτη έκδοση της OpenCyc κυκλοφόρησε την άνοιξη του 2002. Περιελάμβανε μόλις 6.000 έννοιες και 60.000 γεγονότα. (100) Σήμερα, κυκλοφορεί η έκδοση 2.0 της OpenCyc η οποία εκδόθηκε στις 27/7/2009 και είναι διαθέσιμη για λήψη από το SourceForge⁸⁶. Η έκδοση αυτή περιλαμβάνει:

- Ολόκληρη την Cys οντολογία: Περιέχει εκατοντάδες χιλιάδες λέξεις, μαζί με εκατομμύρια ισχυρισμών που συσχετίζουν τις έννοιες μεταξύ τους, σχηματίζοντας μια οντολογία της οποίας το πεδίο είναι όλη η ανθρώπινη κοινή πραγματικότητα.
- Πάνω από εκατό χιλιάδες ευρύτερους ισχυρισμούς, για να μπορεί να συλλάβει επιπρόσθετες έννοιες μεταξύ των εννοιών.
- Αγγλικές συμβολοσειρές (strings) οι οποίες αντιστοιχούν σε κάθε όρο έννοιας για να βοηθήσουν τόσο στην αναζήτηση όσο και στην προβολή.

⁸⁶ <http://sourceforge.net/projects/opencyc/files/>

- Η μηχανή εξαγωγής συμπερασμάτων της Cyc και ο πλοηγητής της βάσης γνώσης της Cyc στηρίζονται πλέον στη Java, για βελτιωμένη απόδοση και αυξημένη φορητότητα της πλατφόρμας
- Έγγραφα και υλικό εκπαίδευσης που μπορεί να ασχολείται ο καθένας με τον δικό του προσωπικό ρυθμό, για να βοηθήσει τους χρήστες μα κατανοήσουν από το βασικό επίπεδο ως και ένα ενδιάμεσο επίπεδο τα ζητήματα της αναπαράστασης γνώσης και των εφαρμογών που αναπτύσσονται χρησιμοποιώντας την Cyc.
- Μια προδιαγραφή της CycL.
- Μια προδιαγραφή της Cyc API⁸⁷ για την ανάπτυξη εφαρμογών.
- Σύνδεσμοι μεταξύ των εννοιών της Cyc και τα σύνολα των συνωνύμων όρων (synsets) του WordNet⁸⁸.
- Σύνδεσμοι μεταξύ εννοιών της Cyc (συμπεριλαμβανομένων των κατηγορημάτων) και της οντολογίας FOAF.
- Συνδέσμους μεταξύ εννοιών της Cyc και άρθρων της Wikipedia. (99)

Υπάρχουν σχέδια, ώστε να μπορέσει να κατανοήσει και να παράγει στη φυσική γλώσσα. Τον Απρίλιο του 2002 είχε δημιουργηθεί η δυνατότητα να παράγει σε φυσική γλώσσα, αλλά δεν είχε κατασκευαστεί η συντακτική ανάλυση. Σε αυτή την προσπάθεια αναμένεται να βοηθήσει και η Κοινότητα του OpenCyc. (93)

Η OpenCyc συνδυάζεται, με εξαιρετικά γόνιμο τρόπο, με την DBpedia. Για την Cyc το δεδομένο είναι πως η γνώση ανά τομέα αλλάζει γρηγορότερα, οπότε η προσέγγιση που ακολουθεί είναι να ασχολείται πρώτα με πιο αφηρημένες έννοιες και οντότητες που εκπροσωπούνται και στη συνέχεια άρχισε να περιλαμβάνει περισσότερες γνώσεις γύρω από τους τομείς. Όμως, για τη DBpedia η γνώση ανά τομέα είναι αυτό που χρειάζονται συνήθως οι άνθρωποι για να επιλύσουν πραγματικά προβλήματα μέσα στους τομείς και συνεπώς περιλαμβάνει πρώτιστα γνωστικά πεδία. (101)

⁸⁷ Cyc API: Cyc Application Programmers Interface

⁸⁸ WordNet: Αποτελεί ένα σημασιολογικό λεξικό

4.7. Μεταδεδομένα σε Ψηφιακές Βιβλιοθήκες

«Ένας ανεπίσημος ορισμός μιας ψηφιακής βιβλιοθήκης είναι μια διοικούμενη συλλογή πληροφοριών, με σχετικές υπηρεσίες, όπου οι πληροφορίες αποθηκεύονται σε ψηφιακές μορφές και είναι προσβάσιμες μέσω ενός δικτύου. Ένα βασικό μέρος αυτού του ορισμού είναι ότι οι πληροφορίες είναι διαχειριζόμενες. Μια ροή δεδομένων που στέλνεται στη γη από ένα δορυφόρο δεν είναι βιβλιοθήκη. Τα ίδια δεδομένα, όταν οργανώνονται συστηματικά, γίνονται μια συλλογή ψηφιακής βιβλιοθήκης.» William Arms (102)

Οι ψηφιακές βιβλιοθήκες είναι οι βιβλιοθήκες οι οποίες έχουν αποθηκευμένο όλο το υλικό τους σε ψηφιακή μορφή. (103) Αρχικά είχε χρησιμοποιηθεί ο όρος «ηλεκτρονική βιβλιοθήκη», ο οποίο γρήγορα αντικαταστάθηκε από τον όρο «ψηφιακή βιβλιοθήκη». Οι ψηφιακές βιβλιοθήκες, ως τώρα, έχουν πολλούς τομείς εφαρμογής (όπως για παράδειγμα η διαχείριση γνώσης και περιεχομένου, η σχολική επικοινωνία, εκπαίδευση και έρευνα, η αρχειοθέτηση και συντήρηση, και η ηλεκτρονική διακυβέρνηση (104)), ενώ παράλληλα παρατηρείται μια συνεχής διερεύνηση του περιβάλλοντος δραστηριοποίησής τους. (105)

Μια απλή αναζήτηση ενός εγγράφου δεν είναι πάντα αποτελεσματική. Υπάρχει μεγάλη πιθανότητα ένα τεράστιο κομμάτι των αποτελεσμάτων να είναι έγγραφα που να μην έχουν κάποια στενή σχέση με το επιθυμητό αποτέλεσμα. Τα μεταδεδομένα σε αυτή την περίπτωση, περιγράφοντας με έναν αξιόπιστο τρόπο τα στοιχεία της βιβλιοθήκης, έρχονται να βοηθήσουν τον εκάστοτε ενδιαφερόμενο να βρει τις πληροφορίες που επιθυμεί, απαλλαγμένες από όλα τα περιττά στοιχεία.

Τα μεταδεδομένα μπορούν να προσφέρουν πολλά στην ανάπτυξη των ψηφιακών βιβλιοθηκών, καθώς μπορούν να περιγράψουν το περιεχόμενο και τις ιδιότητες ενός οποιουδήποτε αντικειμένου μιας ψηφιακής βιβλιοθήκης. Αποτελούν το κλειδί στην εύρεση των πηγών και τη χρήση οποιουδήποτε στοιχείου. Αυτό δεν αποτελεί κάτι που είναι μη οικείο στους βιβλιοθηκάρους, από την οπτική ότι και οι ίδιοι χρησιμοποιούν αυτές τις περιγραφές, για να δημιουργήσουν εγγραφές σε καταλόγους που περιγράφουν τα έγγραφα της βιβλιοθήκης.

Υπάρχουν κάποια επίσημα πρότυπα βιβλιοθηκών για μεταδεδομένα, τα AACR⁸⁹, αλλά τέτοιες εγγραφές είναι χρονοβόρες στη δημιουργία και πρέπει να γίνουν από ειδικά εκπαιδευμένο προσωπικό. Η καταχώρηση χειροκίνητα από τους ανθρώπους να μεν είναι καλύτερη, αλλά αποτελεί μια αρκετά επίπονη διαδικασία, ειδικά εάν αναλογιστούμε το ήδη μεγάλο και γρήγορα επεκτάσιμο περιβάλλον πληροφοριών. Συνεπώς, προτείνονται ως λύσεις απλούστερα σχέδια για τα μεταδεδομένα. (106)

Υπάρχουν τρεις κατηγορίες μεταδεδομένων που χρησιμοποιούνται συχνά για να περιγράψουν τα έγγραφα σε μια ψηφιακή βιβλιοθήκη:

1. Περιγραφικά: Είναι οι πληροφορίες που περιγράφουν το περιεχόμενο του αρχείου, όπως για παράδειγμα οι εγγραφές καταλόγου σε MARC⁹⁰ και τα βοηθήματα ανεύρεσης. Η χρήση τους είναι για βιβλιογραφικούς σκοπούς καθώς και για αναζήτηση-ανάκτηση.
2. Δομικά: Είναι οι πληροφορίες που συνδέουν ένα αρχείο με κάποια άλλα αρχεία, με σκοπό να αποτελούν λογικές μονάδες. Για παράδειγμα οι πληροφορίες που συνδέουν την κάθε εικόνα που υπάρχει σε ένα βιβλίο με τις υπόλοιπες εικόνες του βιβλίου.
3. Διοικητικά: Είναι οι πληροφορίες που χρησιμοποιούνται για τη διαχείριση του αρχείου ή τον έλεγχο της πρόσβασης σε αυτό. Περιλαμβάνουν για κάθε ψηφιακό αρχείο τις πληροφορίες σάρωσης, τον τύπο στον οποίο είναι αποθηκευμένα, τα δικαιώματα πνευματικής ιδιοκτησίας και τις πληροφορίες για τη μακροπρόθεσμη συντήρησή τους. (107)

⁸⁹ AACR: American Association for Cancer Research

⁹⁰ MARC (MACHINE Readable Cataloguing record, εγγραφή καταλόγου αναγνώσιμη από υπολογιστή): Οι εγγραφές MARC είναι πρότυπα που χρησιμοποιούνται για την αναπαράσταση και την επικοινωνία όχι μόνο βιβλιογραφικών αλλά και άλλων πληροφοριών, σε μορφή αναγνώσιμη από τους υπολογιστές. (151)

4.7.1. Πρωτόκολλο Μεταδεδομένων Dublin Core

Τα τελευταία χρόνια γίνεται προσπάθεια για την εύρεση μηχανισμών οι οποίοι να είναι σε θέση να περιγράψουν τα τεχνικά, τα δομικά και τα διοικητικά χαρακτηριστικά των ψηφιακών αντικειμένων. Αν και οι εργασίες σε αυτό τον τομέα είναι ακόμα σε αρχικό στάδιο, έχουν προκύψει διάφορα σχέδια. Αυτό που ξεχωρίζει πιο πολύ, είναι το Dublin Core (πήρε το όνομά του από το Δουβλίνο του Οχάιο), το οποίο αποτελεί μια προσπάθεια καθορισμού των βασικών στοιχείων που απαιτούνται για την περιγραφή των πληροφοριών.

Το Dublin Core καθόρισε ένα σύνολο δεκαπέντε στοιχείων μεταδεδομένων, τα οποία σχετίζονται κυρίως με το περιεχόμενο και την πνευματικά ιδιοκτησία και είναι πολύ πιο απλά από εκείνα που χρησιμοποιούνται στις καταχωρήσεις. Έχουν σχεδιαστεί έτσι ώστε να είναι αρκετά απλά και αρκετά περιγραφικά, ώστε ταυτόχρονα να χρησιμοποιούνται από συντάκτες και παράλληλα να είναι χρήσιμα στην ανακάλυψη των πληροφοριών. (106) Τα στοιχεία αυτά, υπάρχουν στο Παράρτημα 6.2 και παρέχονται από τον δημιουργό της πηγής.

Τα δεκαπέντε αυτά στοιχεία είναι τμήμα ενός μεγαλύτερου συνόλου λεξιλογίου μεταδεδομένων και τεχνικών προδιαγραφών, που συντηρούνται από την Πρωτοβουλία Μεταδεδομένων του Dublin Core (Πρωτοβουλία Μεταδεδομένων του Dublin Core, DCMI). (108)

Έχει δημιουργηθεί ήδη ένα σώμα ανάπτυξης προγραμμάτων εφαρμογής σε πάνω από 10 χώρες, για το Dublin Core. Επιπλέον, έχει ήδη καθοριστεί μια εννοιολογική προσέγγιση, στην οποία τα στοιχεία ενσωματώνονται σε ένα έγγραφο HTML χρησιμοποιώντας την ετικέτα "META". Όμως, δεν υπάρχει ακόμα ένα κοινό πρότυπο μεταδεδομένων, κάτι που προκαλεί ένα επιπρόσθετο εμπόδιο στην πρόσβαση και τη χρήση πληροφοριών σε μια ψηφιακή βιβλιοθήκη ή σε ένα συντονισμένο ψηφιακό σχέδιο βιβλιοθηκών. (106)

4.8. Σημασιολογικό Δίκτυο Αισθητήρων

Ο αισθητήρας είναι μια συσκευή η οποία μετράει κάποιο μετρήσιμο χαρακτηριστικό και το αποτέλεσμα αυτής της μέτρησης το μετατρέπει σε σήμα το οποίο είναι αναγνώσιμο από έναν παρατηρητή ή από ένα όργανο. Αισθητήρες χρησιμοποιούμε στην καθημερινότητά μας, όπως για παράδειγμα στις οθόνες αφής. (109) Πολλές εφαρμογές και υπηρεσίες πληροφορικής και τηλεπικοινωνιών βασίζονται στην αξιοποίηση πληροφοριών από αισθητήρες, όπως για παράδειγμα αισθητήρες θερμοκρασίας, κίνησης, κάμερες, μικρόφωνα και δίκτυα αισθητήρων και ο αριθμός αυτών των εφαρμογών αυξάνεται συνεχώς με ολοένα ταχύτερο ρυθμό. Όμως, παρά την εκθετικά αυξανόμενη χρήση αυτών, δεν έχει δημιουργηθεί ακόμα κάποιος εύκολος τρόπος για την ανάπτυξη εφαρμογών που να αξιοποιούν όχι μόνο τα δεδομένα αλλά και το πολυμεσικό περιεχόμενο που προέρχεται από πολλαπλούς αισθητήρες. (110)

Το Σημασιολογικό Δίκτυο Αισθητήρων (Semantic Sensor Web, SSW) αποτελεί μια προσέγγιση για τη συμπλήρωση των δεδομένων των αισθητήρων με χωρικά, χρονικά και θερμικά σημασιολογικά μεταδεδομένα. Η τεχνική αυτή βασίζεται στις τρέχουσες προσπάθειες τυποποίησης στο πλαίσιο των Open Geospatial Consortium (OGC) και Sensor Web Enablement (SWE) και τις επεκτείνει με τεχνολογίες Σημασιολογικού Ιστού με σκοπό να παρέχουν βελτιωμένες περιγραφές και πρόσβαση στα δεδομένα των αισθητήρων.

Σήμερα, αισθητήρες σε ολόκληρη την υδρόγειο συγκεντρώνουν τεράστιους όγκους δεδομένων. Με τέτοιους όγκους δεδομένων, καταλήγουμε να έχουμε υπερβολικά πολλά δεδομένα και όχι αρκετή γνώση, κάτι που εντείνεται με την ταχεία ανάπτυξη και εξάπλωση της τεχνολογίας των αισθητήρων. Η συμπλήρωση των δεδομένων αυτών με σημασιολογικά δεδομένα θα οδηγούσε σε άμβλυνση του παραπάνω προβλήματος. Κατ' αυτό τον τρόπο, θα αυξανόταν η διαλειτουργικότητα μεταξύ ετερογενών δικτύων αισθητήρων και θα παρείχαν συναφείς πληροφορίες οι οποίες είναι απαραίτητες για την επίγνωση καταστάσεων (Situation awareness)⁹¹. Οι

⁹¹ Η επίγνωση της κατάστασης, είναι η αντίληψη των περιβαλλοντικών στοιχείων ως προς το χρόνο και / ή τον χώρο, η κατανόηση της σημασίας τους, και η προβολή της κατάστασής τους μετά από αλλαγή κάποιας μεταβλητής, όπως για παράδειγμα ο χρόνος. (150)

τεχνικές του Σημασιολογικού Ιστού μπορούν να συμβάλουν σημαντικά στο πρόβλημα της ενσωμάτωσης και της ανακάλυψης δεδομένων, καθώς βοηθούν στη χαρτογράφηση μεταξύ των διαφορετικών σχημάτων μεταδεδομένων με ένα δομημένο τρόπο. (111)

4.8.1. Sensor Wiki

Αναπτύσσεται, μια επέκταση σε πραγματικό χρόνο της ιδέας του Semantic Sensor Web, με την ονομασία Sensor Wiki, με σκοπό να μπορεί να γίνει σε πραγματικό χρόνο η περιήγηση του φυσικού κόσμου μέσω μιας μυριάδας αισθητήρων. Με την ύπαρξη του "Sensing Cloud", δηλαδή ένα σύνολο από αισθητήρες διαφόρων τύπων που έχουν ενεργοποιημένη IP διεύθυνση και είναι τοποθετημένοι σε όλη τη Γη, είναι εφικτή η ενσωμάτωση στο διαδίκτυο πληροφοριών από πολλαπλούς αισθητήρες, κάτι που είναι κατ' επέκταση χρήσιμο σε διάφορα ερωτήματα που γίνονται μέσω ανθρώπων ή και μέσω μηχανών, στο πλαίσιο του Sensor Wiki.

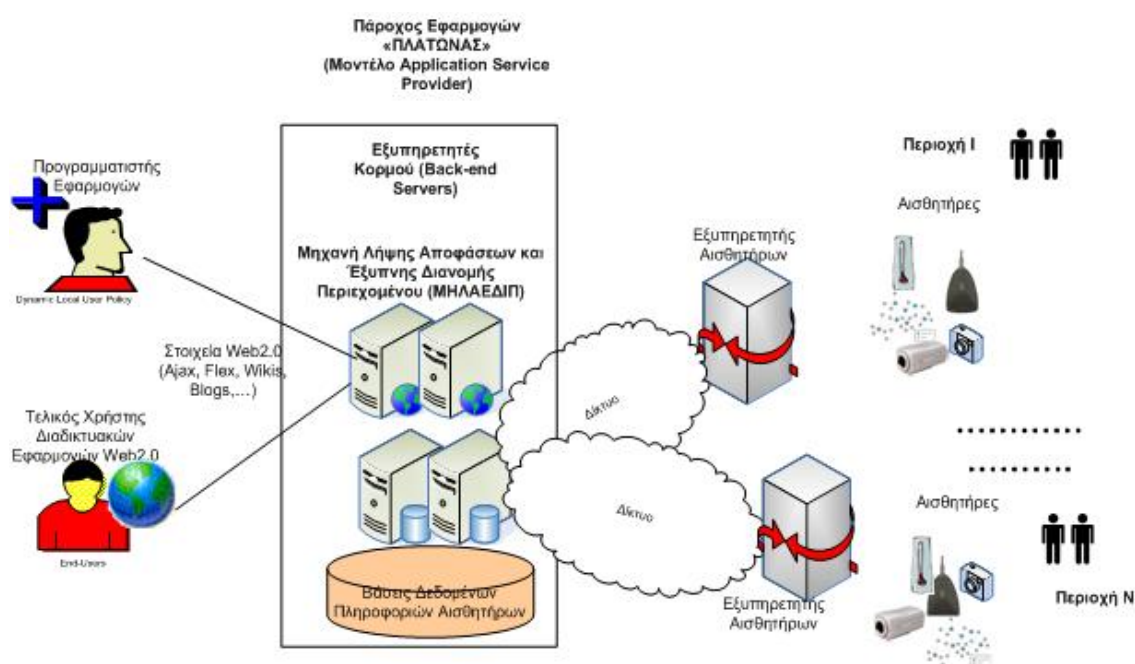
Σε ένα Wiki αισθητήρων ένας ή περισσότεροι αισθητήρες προσφέρουν πληροφορίες σε πραγματικό χρόνο σε σελίδες με κατάλληλα θέματα και δομή, που να επιτρέπει στους χρήστες του Sensor Wiki να αναζητούν διαδραστικά πληροφορίες για αντικείμενα, γεγονότα ή σημεία ενδιαφέροντος. (111)

4.8.2. Έργο ΠΛΑΤΩΝΑΣ

Το έργο ΠΛΑΤΩΝΑΣ, το οποίο ξεκίνησε το Μάρτιο του 2011, έχει ως στόχο την ανάπτυξη μιας πλατφόρμας λογισμικού, η οποία θα διευκολύνει την ανάπτυξη εφαρμογών που χειρίζονται και παρουσιάζουν το περιεχόμενο που προέρχεται από αισθητήρες και θα είναι βασισμένη σε ανοιχτά πρότυπα, όπως για παράδειγμα η γλώσσα SensorML. Επίσης, θα δημιουργήσει ένα σύνολο εργαλείων που θα υποστηρίζει ολόκληρο τον κύκλο ανάπτυξης σύνθετων εφαρμογών πολλαπλών

ετερογενών αισθητήρων. Το έργο εκτιμάται πως θα ολοκληρωθεί σε 24 μήνες από την ημερομηνία έναρξής του.

Η πλατφόρμα ΠΛΑΤΩΝΑΣ θα υποστηρίζει τη συλλογή, την επεξεργασία και την ανάλυση των δεδομένων πολλαπλών αισθητήρων. Όμως, θα βοηθάει και την ανάπτυξη «έξυπνων» εφαρμογών, συμπεριλαμβανομένων εφαρμογών επιχειρησιακής ευφυΐας (business intelligence). Μέσω των εργαλείων που θα αναπτυχθούν θα είναι εφικτή και η ενεργοποίηση συνόλων κανόνων (business rules) για την επεξεργασία των δεδομένων στην βάση δημιουργίας και διαχείρισης «πολιτικών» (policies). (110)



Εικόνα 36: Αρχιτεκτονική προτεινόμενης πλατφόρμας "ΠΛΑΤΩΝΑΣ"

Κατάλληλη έκδοση του εξυπηρετητή αισθητήρων θα δοθεί στη μορφή πρωτότυπου ελεύθερου λογισμικού ανοιχτού κώδικα (ΕΛΛΑΚ) (112), καταστρώντας την αξιοποιήσιμη και από άλλες εταιρίες του κλάδου της πληροφορικής στην Ελλάδα, οι οποίες σχεδιάζουν, αναπτύσσουν και ολοκληρώνουν εφαρμογές αισθητήρων. Πέραν αυτού, θα ωφεληθούν στο σύνολο οι χρήστες των εφαρμογών, καθώς θα μειωθεί το κόστος ανάπτυξης και εν τέλει το κόστος παροχής της λύσης.

Η ανάπτυξη εφαρμογών με βάση την πλατφόρμα «ΠΛΑΤΩΝΑΣ» και τα σχετικά εργαλεία, στοχεύουν σε διαδικτυακές εφαρμογές, με ιδιαίτερη έμφαση στην ανάπτυξη εφαρμογών συμβατών με τις τεχνολογίες του Web2.0. (113) Τα υποσυστήματα σύμφωνα με την αρχιτεκτονική της προτεινόμενης πλατφόρμας του έργου, τα οποία διακρίνονται και στην εικόνα 36 είναι τα ακόλουθα:

- Εξυπηρετητής αισθητήρων (Sensor Edge Server): παρέχει τις δυνατότητες συλλογής, φιλτραρίσματος, επεξεργασίας, κατηγοριοποίησης και αποθήκευσης δεδομένων από πολλαπλούς αισθητήρες.
- Μηχανή Λήψης Αποφάσεων και Έξυπνης Διανομής Περιεχομένου (ΜΗΛΑΕΔΠ): καθορίζει, σύμφωνα με τους επιχειρησιακούς κανόνες και «πολιτικές» (policies), τον τρόπο και τη μορφή που το περιεχόμενο αισθητήρων θα κατευθύνεται στις διάφορες εφαρμογές.
- Εργαλεία ανάπτυξης εφαρμογών: μέσω αυτών γίνεται η δημιουργία και η διαχείριση επιχειρησιακών κανόνων και πολιτικών, για τη διανομή και για την απεικόνιση περιεχομένου. Αυτά θα ολοκληρωθούν σε ένα ενιαίο περιβάλλον. (114)

4.9. Μεταδεδομένα σε Video

Τα αντικείμενα βίντεο είναι πολύ δύσκολο να αναζητηθούν στο διαδίκτυο, καθώς το μεγαλύτερο μέρος της τεχνολογίας που χρησιμοποιείται για την άντληση της σημασίας σε αυτά εξαρτάται από το κείμενο που είναι διαθέσιμο στη σελίδα στην οποία είναι αναρτημένα. Υπάρχει, δηλαδή, η έλλειψη μιας απλής ταξινόμησης για τους χρήστες. (115) Τα μεταδεδομένα είναι ιδιαίτερα χρήσιμα στα video, εκεί όπου οι πληροφορίες σχετικά με το περιεχόμενο του κάθε video (όπως για παράδειγμα οι περιγραφές των σκηνών) δεν είναι τόσο κατανοητές από έναν υπολογιστή, αλλά εμείς θέλουμε να έχουμε αποδοτικές αναζητήσεις μέσα στο πλήθος των video. (116)

Έχουν εμφανιστεί κάποια εργαλεία βίντεο που επιτρέπουν σε συγκεκριμένες πληροφορίες να εντάσσονται από μόνες τους στα αρχεία. Εναλλακτικά, μπορεί ο ίδιος ο χρήστης να εισαγάγει τέτοια μεταδεδομένα πριν την δημιουργία του αρχείου. Όμως το πιο σημαντικό είναι η κατασκευής ενός βασικού πυρήνα συνόλου στοιχείων για την ενδυνάμωση της διαλειτουργικότητας των μεταδεδομένων.

Πιθανότατα ένα λεξιλόγιο βασισμένο στις πληροφορίες που μπορούν να προέλθουν από το ίδιο το αντικείμενο, ενώ παρέχει επίσης ένα πλαίσιο για τους χρήστες για να συμπεριλαμβάνουν πρόσθετα στοιχεία σχετικά με το αντικείμενο

πολυμέσων, εφόσον το επιθυμούν, θα ήταν πολύ χρήσιμο. Ο κυρίαρχος στόχος ενός τέτοιου λεξιλογίου θα ήταν να καταστεί δυνατή η συμβατότητα μεταδεδομένων μεταξύ των υφιστάμενων και νέων τύπων υπηρεσιών και εφαρμογών. Δηλαδή, να δημιουργηθεί ένα σύνολο στοιχείων που να μπορεί να χρησιμοποιηθεί από πολλές διαφορετικές παραλλαγές των εφαρμογών και υπηρεσιών που βασίζονται σε μεταδεδομένα.

Η κυρίαρχη ιδέα για την υλοποίηση της παραπάνω πρότασης, είναι να δημιουργηθεί ένα πρακτικό και εύκολο στη χρήση σύνολο στοιχείων που θα μπορεί και να υλοποιείται γρήγορα από τους προγραμματιστές λογισμικού και τους αρχειονόμους, αλλά και να είναι διαισθητικό σε όλες τις κατηγορίες των καλλιτεχνών. Όταν ένα σύστημα δεδομένων είναι λογικό, απλό και κατανοητό στα απλά άτομα, τα ενθαρρύνει να γίνουν πιο ενεργά και να προσθέσουν περισσότερες πληροφορίες, στη συγκεκριμένη περίπτωση σχετικά με τα έργα τους.

Για εναρκτήριο σύνολο μπορεί να χρησιμοποιηθεί ένα DC-συμβατό⁹² υποσύνολο το οποίο θα χρησιμοποιείται ήδη σε αναρίθμητες άλλες εφαρμογές και θα εγγυάται τη συμβατότητας με όλα τα Πρότυπα Μεταδεδομένων της W3C, με σκοπό την μετέπειτα κατασκευή ενός μεγαλύτερου πλαισίου που θα είναι σε θέση να καλύπτει όλα τα αντικείμενα πολυμέσων. Όσες προσπάθειες γίνουν, πρέπει να είναι συμβατές με τα λεξιλόγια: Dublin Core Metadata Element Set, SMIL (Synchronized Multimedia Integration Language), Media RSS, και το Internet Archive's metadata format. Πιθανότατα σε αυτή τη λίστα να υπάρχει η δυνατότητα προσθήκης και ορισμένων ακόμα λεξιλογίων.

Κάποιος μπορεί να θέλει να συγκρίνει τις πληροφορίες ενός συγκεκριμένου αρχείου, το οποίο βρίσκεται σε δύο ή και περισσότερους διαφορετικούς τύπους, οπότε καλό είναι το μοντέλο που θα σχεδιάσουμε να είναι συμβατό με τα περισσότερα μοντέλα μεταδεδομένων ήχου και στατικής εικόνας, καθώς είναι πιθανό να δημιουργηθούν και να συσχετιστούν τέτοιοι τύποι αρχείων με τα αντικείμενα βίντεο που θέλουμε να περιγράψουμε. Για αυτούς τους λόγους είναι σημαντικό να υπάρχει, όποτε αυτό είναι εφικτό, κάποια ένα-προς-ένα σχέση των στοιχείων μεταξύ των σημασιολογιών διαφορετικών λεξιλογίων. (115)

⁹² DC: Dublin Core

4.9.1. Αρχαιοθετημένα λεξιλόγια για μεταδεδομένα σε Video

Κάποια υπάρχοντα αρχειοθετημένα λεξιλόγια για μεταδεδομένα σε Video (Video Metadata Archive Vocabularies) είναι τα ακόλουθα:

DCMI (Dublin Core Metadata Element Set): Είναι ένα πρότυπο, το οποίο αποτελείται από ένα σύνολο 15 ιδιοτήτων για την περιγραφή πληροφοριών διατηρηματικών πόρων. (108) (115) Ως πληροφορία ενός πόρου ορίζεται να είναι «ο,τιδήποτε έχει ταυτότητα». Δεν υπάρχουν κάποιοι θεμελιώδεις περιορισμοί στους τύπους των πόρων στους οποίους τα μεταδεδομένα του Dublin Core μπορούν να ανατεθούν.

Media RSS: Είναι ένα λεξιλόγιο για χρήση με τις ετικέτες (tags) του RSS 2.0s⁹³ ώστε να μπορούν να περιλαμβάνονται περισσότερο λεπτομερή μεταδεδομένα στα αντικείμενα πολυμέσων.

SMIL (Synchronized Multimedia Integration Language): Αποτελεί Σύσταση του W3C για μια βασισμένη στην XML γλώσσα για την έκφραση των ιδιοτήτων μεγάλων διαδικτυακών πολυμεσικών παρουσιάσεων. Εκτός από τις μεθόδους για τη συσχέτιση υπερσυνδέσεων με ένα πολυμεσικό αντικείμενο, περιλαμβάνει επίσης προσωρινή συμπεριφορά, καθώς και τη περιγραφή της παρουσίασης της διάταξης στην οθόνη.

Internet Archive: Πρόκειται για μια φιλόδοξη προσπάθεια να αρχειοθετηθεί κάθε είδους κειμένου, μαζί με τη βιβλιογραφία του, ακόμα κι αν δεν ήταν ποτέ γραπτό να υπάρχουν στο Διαδίκτυο. Παρόμοια, το Moving Images Archive δίνει στους χρήστες τη δυνατότητα να μεταφορτώνουν τις δικές τους τηλεοπτικές δημιουργίες και να συσχετίζουν ορισμένα βασικά μεταδεδομένα, ώστε να παρέχεται βελτιστοποιημένη αναζήτηση.

OurMedia: Παρέχει χώρο φιλοξενίας για τα ενημερωτικά μέσα, που δημιουργούνται από τους χρήστες. Σημειώνεται ότι η μορφή των μεταδεδομένων του OurMedia συμμορφώνεται απόλυτα με τη μορφή των μεταδεδομένων του Internet Archive.

⁹³RSS 2.0 (Really Simple Syndication): βασίζεται στην XML. Δημιουργήθηκε για την ανταλλαγή περιεχομένου, όπως τίτλοι ειδήσεων, σύνδεσμοι, περιγραφές κ.ά. (171)

SOMA (Shared Online Media Archive): Πρόκειται για ένα μη οριστικό πρότυπο μεταδεδομένων, που αποσκοπεί στην ανταλλαγή μεταδεδομένων πολυμεσικών αρχείων μεταξύ χρηστών. Το συγκεκριμένο πρότυπο βασίζεται κατά κύριο λόγο στο Dublin Core 1.1, καθώς και στο EBU Tech 3273⁹⁴.

The Showcase: Αποτελεί μια online πλατφόρμα αρχειοθέτησης πολυμεσικών παραγωγών, η οποία ξεκίνησε να λειτουργεί το 2003 και υλοποιεί το πρότυπο των μεταδεδομένων SOMA.

MIDI Time Code (MTC), βασισμένο στο SMPTE Time Code: Αποτελεί μια χαμηλού επιπέδου (low-level) δυαδική κωδικοποίηση, η οποία χρησιμοποιείται από πολλά πακέτα λογισμικού και συσκευές αναπαραγωγής MIDI. Η χρήση του είναι ευρέως διαδεδομένη είτε για σύνθεση μουσικών έργων για ταινίες και ζωντανά θεατρικά δρώμενα είτε για απλή οικιακή ηχογράφηση. Το SMPTE αποτελεί αναφορά στο MPEG7⁹⁵. Μέσω αυτής της προδιαγραφής δίνεται η δυνατότητα να κάνει κάποιος «broadcast engineering», ακριβή έλεγχο στο χρόνο, έναρξη/λήξη οργάνου/φωνής και σημεία έναρξης.

Creative Commons: Παρέχει ένα σύνολο εναλλακτικών αδειών πνευματικής ιδιοκτησίας και ένα σχήμα μεταδεδομένων (metadata format) για την έκφραση των όρων των αδειών αυτών. (115) Οι άδειες της Creative Commons αποτελούν ένα βασικό σύνολο από έξι άδειες δικαιωμάτων πνευματικής ιδιοκτησίας, οι οποίες προσφέρουν έναν εύκολο τρόπο διαχείρισης των όρων των πνευματικών δικαιωμάτων που αποδίδονται αυτόματα σε όλα τα δημιουργικά έργα που βρίσκονται υπό πνευματική ιδιοκτησία. Οι άδειες αυτές μπορούν να εφαρμοστούν σε διάφορα είδη έργων, όπως εκπαιδευτικούς πόρους, μουσική, φωτογραφία, βάσεις δεδομένων, πληροφορίες του δημόσιου και του ιδιωτικού τομέα, και άλλα. (117)

TV-Anytime: Είναι ένα πρότυπο σχεδιασμένο για την υποστήριξη της καταγραφής ήχου ή/και βίντεο, καθώς και video on demand⁹⁶. Βασίζεται στο MPEG-

⁹⁴EBU Tech 3273: <http://tech.ebu.ch/docs/tech/tech3321.pdf>

⁹⁵MPEG7: Είναι ένα πρότυπο για την περιγραφή του περιεχομένου των πολυμεσικών δεδομένων που προσπαθεί να υποστηρίξει ένα όσο το δυνατόν ευρύτερο φάσμα εφαρμογών. Υποστηρίζει σε κάποιο βαθμό την κατανόηση από τους υπολογιστές αυτών των περιγραφών και έχει ονομαστεί επίσημα «Multimedia Content Description Interface». (172)

⁹⁶ Οι υπηρεσίες οι οποίες επιτρέπουν την αναπαραγωγή video οποτεδήποτε το θελήσει ο χρήστης, όπως για παράδειγμα το youtube.

7 και παρέχει ένα ευρύ μοντέλο δεδομένων που στηρίζεται στην XML για την περιγραφή του οπτικοακουστικού περιεχομένου και τη σχέση μεταξύ των στοιχείων του περιεχομένου. Η προδιαγραφή των μεταδεδομένων περιλαμβάνει ένα είδος σχήματος ταξινόμησης (classification scheme). (115)

4.9.2. Προτεινόμενα βασικά στοιχεία

Οι νέοι τύποι μεταδεδομένων βίντεο είναι σημαντικό να είναι συμβατοί με τους παλαιότερους, που έχουν ήδη εκτεταμένη χρήση στο πλαίσιο των υπαρχόντων αρχείων πολυμέσων (media archives). Κάποια από αυτά έχουν σχεδιαστεί για να είναι συμβατά με το Dublin Core Metadata Element Set (όπως για παράδειγμα το Internet Archive's format), συνεπώς η συμβατότητα επέρχεται εύκολα. Κάποια άλλα, όμως χρησιμοποιούν διαφορετικά ονόματα, αλλά τα μοντέλα είναι παρόμοια, και συνεπώς είναι εύκολο να γίνει μια ένα προς ένα αντιστοίχιση μεταξύ των διαφόρων στοιχείων.

Παρακάτω παρατίθεται μια λίστα με διάφορα στοιχεία που θα μπορούσαν να περιέχονται σε ένα αρχείο μεταδεδομένων βίντεο. Θεωρείται δεδομένο πως κάθε καταχώριση μεταδεδομένων θα περιέχει τουλάχιστον το URL του αρχείου της ταινίας (αναγνωριστικό), την ημερομηνία υποβολής του URL (ημερομηνία), καθώς και το όνομα του προσώπου που υπέβαλε τα μεταδεδομένα (αποστολέα ή δημιουργός). Η παρακάτω λίστα αναπαριστά ένα «βασικό» υποσύνολο για πιο απλές υλοποιήσεις. Προτεινόμενα βασικά στοιχεία:

- Αναγνωριστικό (DC): Η διεύθυνση URL του αρχείου της ταινίας.
- Τίτλος (DC): Ο τίτλος της ταινίας.
- Δημιουργός (DC): Ο δημιουργός της ταινίας.
- Ημερομηνία (DC): Η ημερομηνία που η ταινία έγινε διαθέσιμη μέσω ενός URL.
- Αντικείμενο (DC): Μια λίστα με λέξεις-κλειδιά που περιγράφουν το περιεχόμενο της ταινίας.
- Format (DC): format του αρχείου βίντεο: Quicktime, Windows Media, κλπ.

- Δικαιώματα (DC): Μια διεύθυνση URL που θα κατευθύνει προς μια γραπτή περιγραφή των όρων αδειοδότησης (licensing terms) της ταινίας.
- Περιγραφή – Το σημείο που περιέχεται η «περίληψη».
- Σχόλιο
- Γεωγραφικό πλάτος
- Γεωγραφικό μήκος
- Αυτός που συνεισφέρει (contributor (DC)): Ένα άτομο ή ένας οργανισμός που έχει συμβάλει στο περιεχόμενο του πόρου.
- Δικαιούχος (DC): Ένα άτομο ή ένας οργανισμός που κατέχει ή διαχειρίζεται τα δικαιώματα επί του πόρου. μέγεθος αρχείου.
- Όνομα αρχείου
- Διάρκεια
- Hash: Μια τιμή της συνάρτησης κατακερματισμού SHA-1⁹⁷ για το αρχείο.
- Απόσπασμα Video (clip) / έλεγχος χρόνου (time control): Μόνο ένα τμήμα της αναπαραγωγής.

Κάποια επιπλέον προαιρετικά στοιχεία που θα μπορούσαν να συμπεριληφθούν είναι: κωδικοποιητής, πρόσωπο, εικόνα (για αυτό το πρόσωπο), σκηνοθέτης, παραγωγός, αναλογίες πλευρών (aspect ratio), ρυθμός καρέ (frame rate), αριθμός καναλιών ήχου (number of audio channels), ρυθμός δειγματοληψίας (sample rate). (115)

4.10. Inspire

Η INSPIRE (Infrastructure for Spatial Information in the European Community, Υποδομή για τη Γεωχωρική Πληροφορία στην Ευρωπαϊκή Ένωση)

⁹⁷ Secure Hash Algorithm 1 (SHA-1): Μια συνάρτηση κατακερματισμού είναι μια συνάρτηση η οποία θα παράξει ίδιο αποτέλεσμα για διαφορετικές εισόδους με πάρα πολύ μικρή πιθανότητα. Είναι μια συνάρτηση κατακερματισμού η οποία παράγει μια τιμή μεγέθους 160 Bit. Το αποτέλεσμα της συνάρτησης κατακερματισμού είναι πολύ ευαίσθητο σε μικρές αλλαγές της εισόδου.

αποτελεί μια Οδηγία της Ευρωπαϊκής Ένωσης, που θεσπίστηκε το Μάιο του 2007. Η σταδιακή εφαρμογή της Οδηγίας δεσμεύει όλα τα κράτη-μέλη της Ευρωπαϊκής



Εικόνα 37: Λογότυπο Inspire

Ένωσης, με στόχο την πλήρη εφαρμογή της έως τα τέλη του 2019. Με τον όρο «Υποδομή Γεωχωρικών Πληροφοριών» εννοείται ένα θεσμικό και τεχνικό πλαίσιο συντονισμού των διαδικασιών συλλογής, χρήσης και διάθεσης γεωχωρικών δεδομένων. (109)

Η συγκεκριμένη Οδηγία δημιουργεί το νομικό πλαίσιο για την ίδρυση και λειτουργία της Υποδομής της γεωχωρικής πληροφορίας στην Ευρώπη με σκοπό τη διαμόρφωση, εφαρμογή,

διαχείριση και εκτίμηση των πολιτικών της Ευρωπαϊκής Ένωσης σε όλα τα επίπεδα, όπως και για την παροχή πληροφοριών προς το ευρύτερο κοινό. Για την υλοποίηση της Οδηγίας, η Ευρωπαϊκή Επιτροπή είναι ο Συντονιστής σε Ευρωπαϊκό επίπεδο και τα κράτη-μέλη είναι αρμόδια για την υλοποίηση της Οδηγίας σε εθνικό επίπεδο. Η Οδηγία INSPIRE περιλαμβάνει γεωγραφικές και περιβαλλοντικές πληροφορίες, συνολικά σε 34 θεματικές ενότητες. Οι ενότητες αυτές ομαδοποιούνται σε τρία Παραρτήματα του κειμένου της Οδηγίας, που αντιστοιχούν περίπου και στην προτεραιότητα ένταξής τους στο πληροφοριακό σύστημα. Ο λεπτομερής ορισμός και το πεδίο εφαρμογής που καλύπτει κάθε θεματική ενότητα περιλαμβάνεται στο τεύχος «Definition of Annex Themes and Scope». Η Οδηγία έχει σχεδιαστεί ώστε να βελτιστοποιήσει τις δυνατότητες αξιοποίησης των χωρικών δεδομένων μέσω της τεκμηρίωσής τους, της λειτουργίας υπηρεσιών που αποσκοπούν στη διευκόλυνση της πρόσβασης σ' αυτά, στην αύξηση της διαλειτουργικότητάς τους και στην αντιμετώπιση των δυσκολιών στις οποίες προσκρούει η χρήση τους. (118)

Η INSPIRE εφαρμόζεται στις πληροφορίες που συνδέονται με ένα γεωγραφικό πλαίσιο, όπως περιβαλλοντικές παρατηρήσεις, στατιστικές κ.λπ., οι οποίες τηρούνται σε ηλεκτρονική μορφή από τις δημόσιες αρχές ή για δικό τους λογαριασμό. Οι πληροφορίες αυτές αφορούν περιοχές επάνω στις οποίες ένα κράτος μέλος έχει ή ασκεί δικαιοδοσία και καλύπτουν θέματα όπως τα διοικητικά σύνορα, οι παρατηρήσεις της ποιότητας του αέρα, των υδάτων ή των εδαφών, η βιοποικιλότητα, η χρήση της γης, τα δίκτυα μεταφοράς, η υδρογραφία, το υψόμετρο, η γεωλογία, η

κατανομή του πληθυσμού ή των ειδών, οι κατοικίες, οι βιομηχανικοί τόποι ή ακόμη οι ζώνες φυσικών κινδύνων.

Οι πληροφορίες αυτές πρέπει να συνοδεύονται από πλήρη μεταδεδομένα (δηλαδή πληροφορίες, οι οποίες περιγράφουν σύνολα χωρικών δεδομένων και υπηρεσίες χωρικών δεδομένων και καθιστούν δυνατή την εξεύρεση, την απογραφή και τη χρήση τους). Τα μεταδεδομένα αυτά αφορούν στους όρους που ισχύουν για την πρόσβαση και τη χρήση των χωρικών πληροφοριών, την ποιότητα και την ισχύ των πληροφοριών αυτών, στους όρους πρόσβασης, καθώς και στις δημόσιες αρχές που είναι αρμόδιες για τις πληροφορίες αυτές. (119)

Τα κράτη-μέλη πρέπει να θέτουν στη διάθεση ιδιωτών και δημόσιων οργανισμών χρηστών διαδικτυακές υπηρεσίες, παρέχοντάς τους κυρίως τη δυνατότητα να αναζητούν, να συμβουλευούνται και να κατεβάζουν χωρικές πληροφορίες από την εκάστοτε υπηρεσία. (120) Η πρόσβαση στις υπηρεσίες αυτές γίνεται μέσω της διαδικτυακής πύλης γεωγραφικών πληροφοριών INSPIRE, την οποία διαχειρίζεται σε κοινοτικό επίπεδο η Επιτροπή. (121) Κάθε κράτος έχει το δικαίωμα να περιορίζει την πρόσβαση του κοινού σε κάποιες χωρικές πληροφορίες, όταν για παράδειγμα η πρόσβαση αυτή ενδέχεται να επηρεάσει αρνητικά τις διεθνείς σχέσεις, τη δημόσια ασφάλεια, την εθνική άμυνα ή τον εμπιστευτικό χαρακτήρα των εργασιών των δημόσιων αρχών ή ορισμένων εμπορικών ή βιομηχανικών πληροφοριών, την τήρηση των δικαιωμάτων πνευματικής ιδιοκτησίας και τον εμπιστευτικό χαρακτήρα των προσωπικών δεδομένων ή ακόμα την προστασία του περιβάλλοντος. (119)

Η INSPIRE βασίζεται σε ένα σύνολο κοινών αρχών. Πιο συγκεκριμένα:

- Τα δεδομένα πρέπει να συλλέγονται μόνο μία φορά και να φυλάσσονται εκεί όπου θα μπορούν να διατηρηθούν πιο αποτελεσματικά.
- Πρέπει να είναι δυνατός ο συνδυασμός ασυσχέτιστων χωρικών πληροφοριών από διαφορετικές πηγές σε όλη την Ευρώπη και να το μοίρασμα αυτών σε πολλούς χρήστες και εφαρμογές.
- Θα πρέπει να είναι δυνατό οι πληροφορίες που συλλέγονται σε ένα συγκεκριμένο επίπεδο ή μια κλίμακα να μπορούν να μοιράζονται με όλα τα επίπεδα ή τις κλίμακες, Σε γενικό ή ακόμα και σε λεπτομερές πλαίσιο, ανάλογα με τις εκάστοτε ανάγκες.

- Οι γεωγραφικές πληροφορίες που χρειάζονται σε όλα τα επίπεδα της καλής διακυβέρνησης, πρέπει να είναι άμεσα και διαφανώς διαθέσιμες.
- Πρέπει να είναι εύκολος ο εντοπισμός των διαθέσιμων γεωγραφικών πληροφοριών, όπως ο τρόπος χρήσης τους για την κάλυψη μιας συγκεκριμένης ανάγκης και οι προϋποθέσεις κάτω από τις οποίες μπορούν να αποκτηθούν και να χρησιμοποιηθούν. (120)

4.10.1. Inspire Geoportal

Το INSPIRE geoportal⁹⁸ παρέχει τα μέσα για την αναζήτηση συνόλων χωρικών δεδομένων και υπηρεσίες χωρικών δεδομένων, και υπόκειται σε περιορισμούς πρόσβασης σχετικά με τα σύνολα χωρικών δεδομένων από κράτη μέλη της ΕΕ σύμφωνα με την οδηγία INSPIRE. Μια πρώτη έκδοση αυτού δόθηκε στη δημοσιότητα πριν το Νοέμβριο του 2011 που αποτελούσε και τη λήξη της προθεσμίας. Στις επόμενες εκδόσεις αναμένεται να ενσωματωθούν πρόσθετες λειτουργίες.

Το INSPIRE Geoportal αποτελείται από τρεις εφαρμογές:

- Discovery / Viewer
- Validator
- Metadata Editor (121)

4.10.1.1. Discovery / Viewer

Είναι μια εφαρμογή που επιτρέπει την αναζήτηση, την ανακάλυψη και την πρόσβαση σε γεωγραφικές πληροφορίες, οι οποίες παρέχονται από τους εμπορικούς και μη εμπορικούς Ευρωπαϊκούς κυβερνητικούς οργανισμούς. (121)

⁹⁸ <http://inspire-geoportal.ec.europa.eu>

4.10.1.2. Validator

Το INSPIRE Metadata Validator έχει ως σκοπό να ελέγξει τη συμβατότητας των μεταδεδομένων με βάση τον Κανονισμό Μεταδεδομένων INSPIRE. Ο Validator δέχεται τα μεταδεδομένα τα οποία ακολουθούν τις τεχνικές προδιαγραφές για μεταδεδομένα οι οποίες είναι κωδικοποιημένες στο σχήμα EN ISO 19139 και δημιουργεί μια αναφορά σχετικά με τη συμβατότητά τους. Ελέγχει, επίσης, αν το έγγραφο είναι καλοσχεδιασμένο ή αν περιέχει σφάλματα για τα οποία πρέπει να ενημερώσει τον χρήστη. (121)

4.10.1.3. Metadata Editor

Ο EUOSME (European Open Source Metadata Editor) είναι μια διαδικτυακή εφαρμογή που έχει γραφτεί σε Java και έχει ως σκοπό τη δημιουργία μεταδεδομένων συμβατών με τον Κανονισμό του INSPIRE και τον Κανονισμό Μεταδεδομένων του INSPIRE σε κάθε μια από τις 22 επίσημες Ευρωπαϊκές γλώσσες (Βουλγαρική, Τσεχική, Δανική, Ολλανδική, Αγγλική, Εσθονική, Φινλανδική, Γαλλική, Γερμανική, Ελληνική, Ουγγρική, Ιταλική, Λεττονική, Λιθουανική, Μαλτεζική, Πολωνική, Πορτογαλική, Ρουμανική, Σλοβακική, Σλοβενική, Ισπανική, Σουηδική) εκτός από τα Ιρλανδικά (η νομοθεσία που αφορά τα μεταδεδομένα δεν έχει μεταφραστεί σε αυτή τη γλώσσα) και έχει αναπτυχθεί ως τμήμα του EuroGEOSS project. Πιο συγκεκριμένα, αυτή η εφαρμογή δίνει τη δυνατότητα να περιγραφεί ένα σύνολο χωρικών δεδομένων, μια σειρά συνόλων δεδομένων χωρικών δεδομένων ή μια υπηρεσία χωρικών δεδομένων συμβατή με τα standards ISO 19115:2003 (αναθεώρηση 2003/Cor.1:2006) και ISO 19119:2005. Η πρώτη έκδοση αυτού ήταν η 1.0.0 και εκδόθηκε στις 2011-03-25, ενώ η τωρινή έκδοση είναι η 1.0.3. (121) (122)

4.10.2. GIMED: Ελληνικός Metadata Editor

Το GIMED είναι Ελληνικό Ελεύθερο Λογισμικό, που υλοποιεί έναν Metadata Editor για γεωχωρικά δεδομένα, υπό την άδεια GPL-3. (123) Υλοποιήθηκε από τον

Άγγελο Τζώτσο, υποψήφιο Διδάκτορα της σχολής Αγρονόμων Τοπογράφων Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου. Είναι υλοποιημένο σε C και C# και αποτελεί ένα πλήρως προσαρμόσιμο και επεκτάσιμο πρόγραμμα με Ελληνικό μενού, το οποίο είναι απλό στη χρήση και φιλικό προς το χρήστη. Επίσης, υποστηρίζεται τόσο από Windows όσο και από Linux. Το γραφικό του περιβάλλον είναι συμβατό με το INSPIRE Geportal και υποστηρίζει το προφίλ μεταδεδομένων των INSPIRE και ISO19139. (124) Το project αυτό καταχωρήθηκε στο SourceForge.net⁹⁹ στις 3 Νοεμβρίου το 2010. (123)

Το πρόγραμμα υποστηρίζει δυνατότητες όπως την αυτόματη ανάγνωση γεωγραφικής θέσης διανυσματικών και εικονιστικών δεδομένων από ΕΓΣΑ87 σε WGS84 και την δημιουργία αρχείων μεταδεδομένων XML συμβατά με καταλόγους CSW (GeoNetwork, MDWeb κλπ). Επίσης, επιτρέπει την επεξεργασία μεταδεδομένων και την επικαιροποίηση τους, απαραίτητα εργαλεία για τους Δημόσιους Οργανισμούς εν όψη δημιουργίας της ΕΥΓΕΠ¹⁰⁰ και της εφαρμογής του INSPIRE.

Το πρόγραμμα στην παρούσα φάση έχει δημοσιοποιηθεί στην έκδοση 1.3.0. Στο πλαίσιο των μελλοντικών επεκτάσεων υπάρχουν προγραμματισμένα για την έκδοση 1.5 να υλοποιηθούν η υποστήριξη μαζικών εργασιών δημιουργίας και η υποστήριξη μαζικών εργασιών επεξεργασίας, και για την έκδοση 2.0 η υποστήριξη περισσότερων συστημάτων αναφοράς και η υποστήριξη Σειρών Δεδομένων (Dataset Series). Επιπλέον, μια πιθανή μελλοντική υλοποίηση αποτελεί η ενσωμάτωση Ελληνικού Προφίλ Μεταδεδομένων και η αυτόματη εισαγωγή σε καταλόγους OGC-CSW ISO (Geonetwork, MDWeb, Deegree κλπ). (124)

⁹⁹ <http://sourceforge.net/>

¹⁰⁰ Εθνική Υποδομή Γεωχωρικών Πληροφοριών (ΕΥΓΕΠ): επιτρέπει την άμεση πρόσβαση στο σύνολο της ψηφιακά διαθέσιμης γεωπληροφορίας ολόκληρης της μέσω του διαδικτύου. (160)

5. Κατασκευή οντολογίας με τη βοήθεια βάσης δεδομένων

5.1. *Protégé*

Το Protégé αποτελεί ένα ελεύθερο και ανοιχτού κώδικα πρόγραμμα που επεξεργάζεται Οντολογίες. (125) Έχει αναπτυχθεί από το Stanford Center for Biomedical Informatics Research και το Stanford University School of Medicine. (126) Βασίζεται στη Java και υποστηρίζεται από μια όλο και αυξανόμενη Κοινότητα με ένα σύνολο εργαλείων που συμβάλουν στην κατασκευή μοντελοποίησης οντολογιών υψηλού επιπέδου και εφαρμογών που βασίζονται στη γνώση με τη χρήση οντολογιών. (127) Τα αρχεία που εξάγει μπορούν να έχουν διάφορους τύπους, όπως RDF(S), OWL, και XML Schema. (125)

Τη παρούσα στιγμή υπάρχουν σε παράλληλη ανάπτυξη δύο εκδόσεις του Protégé. Η έκδοση 3.4.7 η οποία ανήκει στην οικογένεια εκδόσεων 3.X και χρησιμοποιεί την OWL 1 και συνεπώς μας παρέχει όλα τα χαρακτηριστικά της. Επιτρέπει την εισαγωγή βάσεων δεδομένων αλλά δεν επιτρέπει τον ορισμό συγκεκριμένων τιμών πληθικότητας, Η άλλη έκδοση είναι η 4.1 που ανήκει στην οικογένεια εκδόσεων 4.X. Χρησιμοποιεί την OWL 2 και επιτρέπει τον ορισμό συγκεκριμένων τιμών πληθικότητας, αλλά δεν επιτρέπει την εισαγωγή βάσεων δεδομένων.

5.2. *Υλοποίηση οντολογίας*

Η οντολογία που θέλαμε να υλοποιήσουμε έχει σχέση με τη δομή και το αντικείμενο της Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών. Για την υλοποίηση αρχικά χρειαστήκαμε μια βάση δεδομένων που να μας παρέχει ένα επαρκές πλήθος από δεδομένα σχετικά με τη δομή της σχολής και του Προγράμματος

Σπουδών αυτής. Για αυτό το λόγο, χρησιμοποιήσαμε τη βάση δεδομένων του Ιστότοπου semfe.gr, την ανάπτυξη του λογισμικού του οποίου επιμελήθηκε ο Κωνσταντίνος Φίλιος.

Έγινε μια εξαγωγή των επιμέρους πινάκων που περιείχαν τα στοιχεία που χρειαζόμασταν (σχήμα βάσης και δεδομένα) και στήσαμε αυτή μια βάση τοπικά με τη βοήθεια του xampp¹⁰¹ και του phpmyadmin¹⁰². Μετά αφαιρέθηκαν από αυτά τα προσωπικά δεδομένα των χρηστών. Προσθέσαμε, όμως, κάποια επιπλέον τυχαία δεδομένα στους χρήστες (στον πίνακα `user_subjects`) ώστε να έχει στοιχεία και να μας επιτρέπει να ελέγξουμε την ορθότητα της προκύπτουσας οντολογίας. Στη συνέχεια αλλάξαμε τον τύπο μηχανής (engine) σε InnoDB, από MyISAM που ήταν εξ' αρχής, με σκοπό την δημιουργία κάποιων ξένων κλειδιών (foreign keys) που υποδηλώνουν τις σχέσεις μεταξύ των στοιχείων της βάσης. Οι σχέσεις που δημιουργήσαμε ήταν 1:n (από ένα στοιχείο προς πολλά). Τα κλειδιά που δημιουργήσαμε είναι τα ακόλουθα, μαζί με τις αντίστοιχες σχέσεις που ορίζουν:

- `uni_fields.field_id` → `uni_subject_field_field_ptr`
`uni_subject_fields.subject_field_field_ptr` REFERENCE
- `uni_subjects.subject_id` → `uni_subject_field_subject_ptr`
`uni_subject_fields.subject_field_subject_ptr` REFERENCE
- `uni_branches.branch_id` → `uni_subjects.subject_branch_ptr`
`uni_subjects.subject_branch_ptr` REFERENCE
- `uni_departments.department_id` → `uni_teachers.teacher.department`
`uni_teachers.teacher_department` REFERENCE
- `uni_subjects.subject_id` → `user_subject.user_subject_subject_ptr`
`user_subject.user_subject_subject_ptr` REFERENCE

Ο σκοπός πίσω από τη δημιουργία των παραπάνω κλειδιών, είναι αυτά να μπορέσουν να εισαχθούν στο Protégé ως ιδιότητες αντικειμένων (object properties). Το σχήμα της βάσης που δημιουργήσαμε μαζί με κάποια περιγραφικά στοιχεία παρατίθεται στο Παράρτημα 6.3.

Τη προκύπτουσα βάση την εισάγαμε στο Protégé 3.4.7, δημιουργώντας μια νέα οντολογία OWL DL σε OWL/RDF και εισάγαμε με τη βοήθεια του πρόσθετου

¹⁰¹<http://www.apachefriends.org/en/xampp.html>

¹⁰²http://www.phpmyadmin.net/home_page/index.php

Στην οντολογία που δημιουργούμε, ορίζουμε σύμφωνα με τη βάση που έχουμε και τα αντικείμενα του κόσμου μας που παράγονται από αυτήν:

- Τις κατευθύνσεις, τα εξάμηνα, τα υποχρεωτικά, τα επιλογής, τα προαιρετικά, τα μεταβατικά μαθήματα που υπάρχουν σε όλη τη διάρκεια των σπουδών.
- Τις ροές με βάση την κατεύθυνση, τα υποχρεωτικά/επιλογής μαθήματα που έχει η κάθε ροή και τα μαθήματα της κατηγορίας T.
- Τις βαθμίδες των καθηγητών και τους τομείς του ιδρύματος.
- Τους χρήστες με όσα μαθήματα έχουν δηλώσει και ποια από αυτά έχουν ή δεν έχουν προσβάσιμο βαθμό.

Κατασκευάστηκαν 3 υποκλάσεις που ορίζουν τον Κορμό, την κατεύθυνση του Μαθηματικού και τη κατεύθυνση Φυσικού Εφαρμογών.

Κατασκευάσαμε 9 υποκλάσεις που αντιστοιχούν στις 5 ροές του Φυσικού Εφαρμογών (Υπολογιστική & Θεωρητική Φυσική, Πυρηνική Φυσική & Στοιχειώδη Σωματίδια, Οπτοηλεκτρονική & LASER,

```

classDiagram
    uni_branches --|> uni_subjects
    uni_subjects --|> user_subjects
    uni_subjects -- uni_subject_fields
    uni_subjects -- uni_teachers
    uni_subjects -- uni_departments
    uni_subject_fields -- uni_fields

```

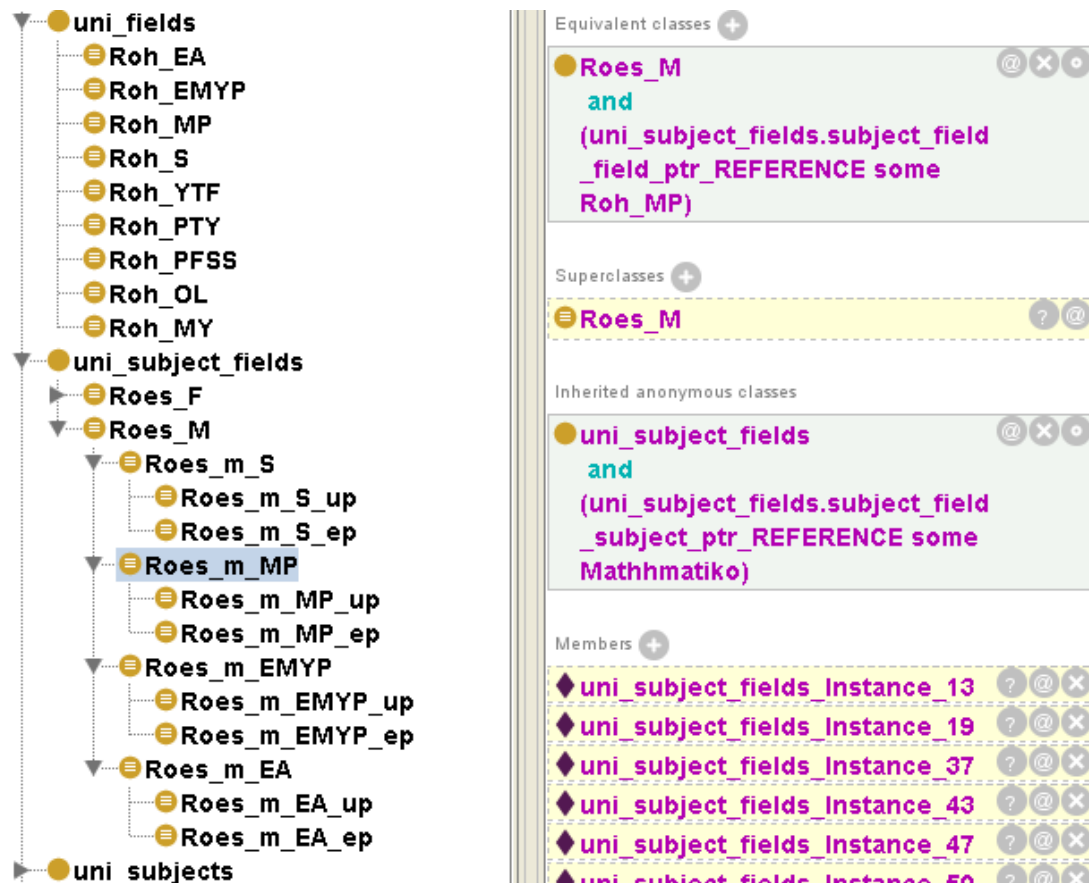
The diagram illustrates the following relationships:

- uni_branches** and **uni_subjects**: Generalization (dashed line with an open triangle pointing to uni_subjects).
- uni_subjects** and **user_subjects**: Generalization (dashed line with an open triangle pointing to user_subjects).
- uni_subjects** and **uni_subject_fields**: Association (dashed line with open arrowheads at both ends).
- uni_subjects** and **uni_teachers**: Association (dashed line with an open arrowhead pointing to uni_teachers).
- uni_subjects** and **uni_departments**: Association (dashed line with an open arrowhead pointing to uni_departments).
- uni_subject_fields** and **uni_fields**: Association (dashed line with an open arrowhead pointing to uni_fields).

Εικόνα 38: Σχέσεις μεταξύ των κύριων κλάσεων

Στη `uni_subject_fields`: Κατασκευάσαμε τις 9 στο σύνολο ροές της σχολής, χωρισμένες πρώτα ανά κατεύθυνση. Συνεπώς δημιουργήσαμε αρχικά δύο υποκλάσεις χρησιμοποιώντας για τον ορισμό τους την ιδιότητα αντικειμένου `uni_subject_fields.subject field field.ptr REFERENCE`, σε συνδυασμό με τις

υποκλάσεις που δημιουργήσαμε στο uni_fields. Στις υποκλάσεις αυτών εξειδικεύσαμε τον ορισμό των κλάσεων έτσι ώστε να έχουμε μια ροή ανά κλάση. Προχωρήσαμε σε περαιτέρω εξειδίκευση χωρίζοντας τις κλάσεις αυτές σε υποκλάσεις που ορίζουν τα υποχρεωτικά και τα προαιρετικά μαθήματα της εκάστοτε ροής. (Στιγμιότυπο protégé [1])



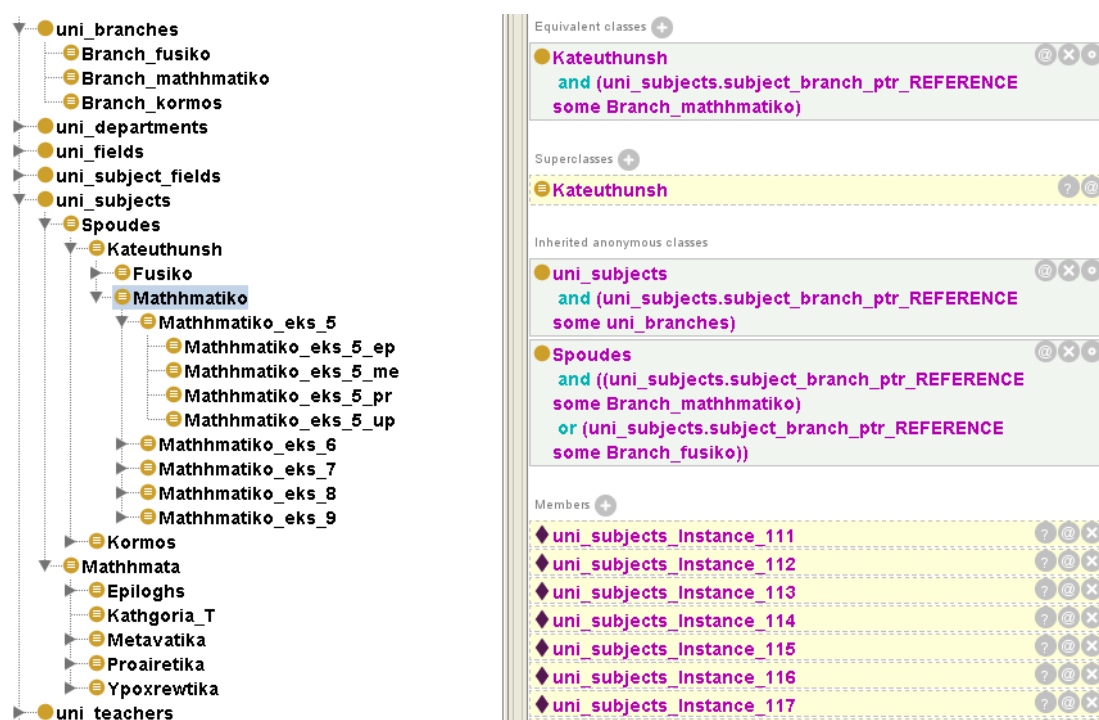
Εικόνα 39: Στιγμιότυπο protégé [1]

Στο uni_subjects: Κατασκευάσαμε δύο κύριες υποκλάσεις. Στη μια χωρίσαμε τα μαθήματα με βάση τα θεμελιώδη στοιχεία τους (υποχρεωτικά, επιλογής, προαιρετικά, μεταβατικά, κατηγορία T, υποχρεωτικά/προαιρετικά μαθήματα ροών) και στην άλλη ορίσαμε τα έτη, τις κατευθύνσεις, τα εξάμηνα και στο τέλος την υποχρεωτικότητα των μαθημάτων (υποχρεωτικά, επιλογής, προαιρετικά, μεταβατικά).

Στη πρώτη κατηγορία, χωρίσαμε το σύνολο των μαθημάτων σε υποχρεωτικά, επιλογής, προαιρετικά, μεταβατικά, κατηγορία T, μαθήματα ροών, κατασκευάζοντας τις υποκλάσεις Υποχρεωτικά, Epiloghs, Proairetika, Metavatika, Kathgoria_T. Με βάση συγκεκριμένους κωδικούς μαθημάτων (subject_id) κατασκευάσαμε και κλάση

με μαθήματα της κατηγορίας T της κατεύθυνσης του Μαθηματικού Εφαρμογών, καθώς δεν υπάρχει κάποιος ορισμός για αυτή τη κατηγορία των μαθημάτων στην βάση δεδομένων.

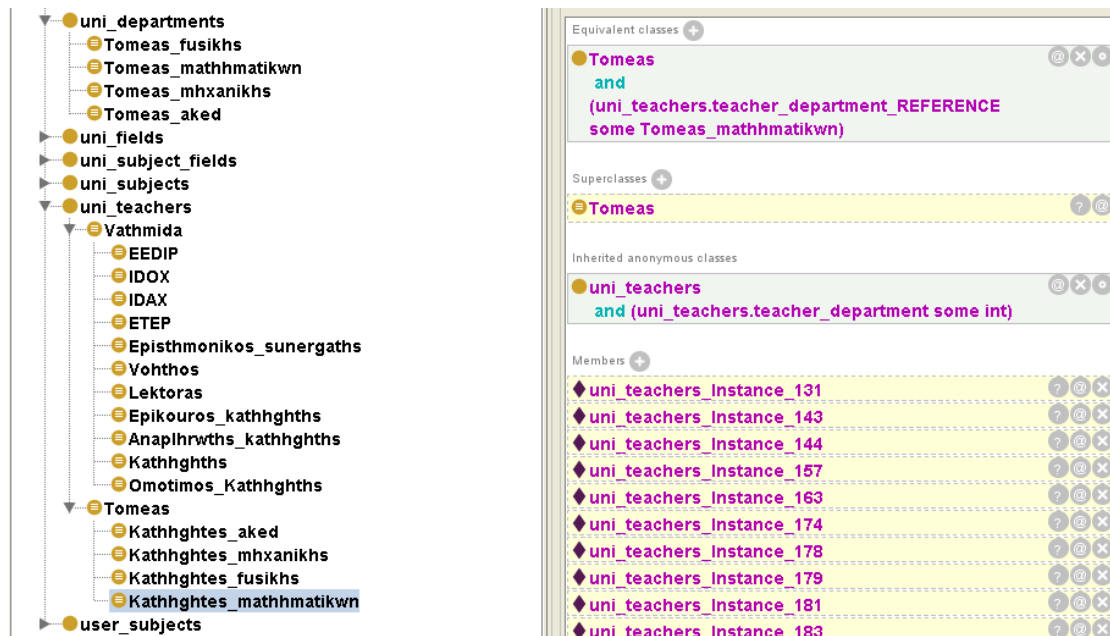
Στη δεύτερη κύρια υποκλάση κατασκευάσαμε κλάσεις ανά κύρια κατεύθυνση (με βάση τις υποκλάσεις που δημιουργήσαμε στο uni_branches και την ιδιότητα αντικειμένου «uni_subjects.subject_branch_ptr_REFERENCE») και μετά στις υποκλάσεις προσθέσαμε επιπλέον περιορισμούς για την κατασκευή εξαμήνων (subject_semester). Στη συνέχεια ορίσαμε και τα υποχρεωτικά, επιλογής, προαιρετικά και μεταβατικά μαθήματα ανά εξάμηνο, ως την τομή των μαθημάτων του εξαμήνου με την αντίστοιχη από τις τέσσερις κύριες κλάσεις των μαθημάτων (Ypoxrewtika, Epiloghs, Proairetika, Metavatika). Στιγμιότυπο protégé [2])



Εικόνα 40: Στιγμιότυπο protégé [2]

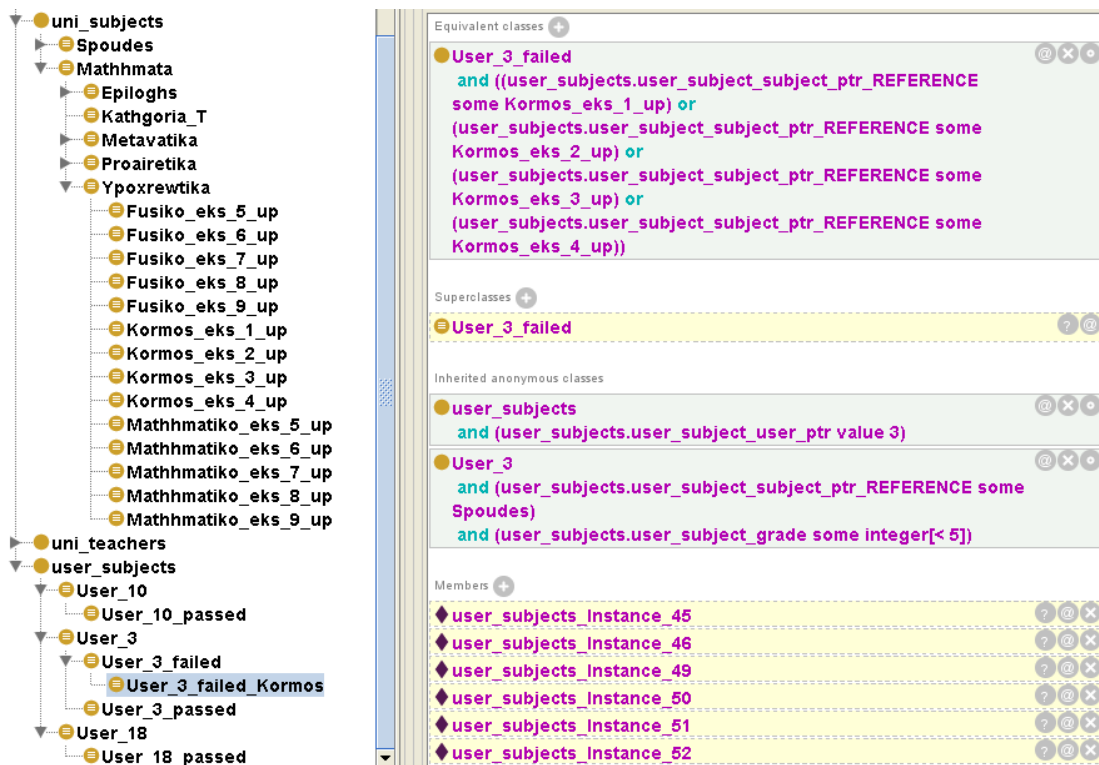
Στο uni_teachers: Κάναμε αρχικά δύο κύριες υποκλάσεις (Vathmida, Tomeas) για να χωριστούν οι καθηγητές αρχικά ανά Βαθμίδα και μετά ανά Τομέα (για όσους αυτά τα στοιχεία είναι γνωστά, θέτοντας τους περιορισμούς «uni_teachers.teacher_title some int» και «uni_teachers.teacher_department some int» αντίστοιχα στο σύνολο των καθηγητών). Ανά βαθμίδα ορίσαμε κλάσεις για τις βαθμίδες Ομότιμος Καθηγητής, Καθηγητής, Αναπληρωτής Καθηγητής, Επίκουρος Καθηγητής, Λέκτορας, Βοηθός, Επιστημονικός Συνεργάτης, Ε.Τ.Ε.Π., ΙΔΑΧ, ΙΔΟΧ

και Ε.Ε.ΔΙ.Π. με βάση το uni_teachers.teacher_title. Ανά τομέα ορίσαμε τέσσερις υποκλάσεις που διαχωρίζουν τους καθηγητές σύμφωνα με τον τομέα στον οποίο ανήκουν, με τη βοήθεια της ιδιότητας αντικειμένου uni_teachers.teacher_department_REFERENCE. (Στιγμιότυπο protégé [3])



Εικόνα 41: Στιγμιότυπο protégé [3]

Στο user_subjects: Δημιουργήσαμε μια υποκλάση που δέχεται περιορισμό κατάλληλο ώστε να μας δίνει τα μαθήματα ενός μόνο συγκεκριμένου ατόμου-id και από αυτά δημιουργήσαμε μια επιπλέον υποκλάση στην οποία ανήκουν μόνο τα προσβάσιμα μαθήματα (user_subject_grade some integer[>= 5]). Μπορούμε να ορίσουμε επίσης, για παράδειγμα, τα μαθήματα του Κορμού στα οποία ο φοιτητής έχει προσβάσιμους βαθμούς, ως υποκλάση των περασμένων μαθημάτων και με χρήση της ιδιότητας αντικειμένου «user_subject.user_subject_subject_ptr_REFERENCE some X», όπου X όποια υποκλάση του uni_subjects που θέλουμε. Αυτό που έχουμε υλοποιήσει είναι να μας δίνει για τρεις επιλεγμένους χρήστες τα μαθήματα στο σύνολο των σπουδών που έχουν πάρει προσβάσιμο βαθμό και για ένα συγκεκριμένο χρήστη από τους 3 τα μαθήματα στα οποία δεν είχε προσβάσιμο βαθμό. Μια υποκλάση αυτού που δημιουργήθηκε, είναι τα μαθήματα με μη προσβάσιμο βαθμό που είναι υποχρεωτικά στον Κορμό των σπουδών. (Στιγμιότυπο protégé [4])



Εικόνα 42: Στιγμιότυπο protégé [4]

Κάποια σχεδιαγράμματα που δημιουργήθηκαν από την προκύπτουσα οντολογία βρίσκονται στο Παράρτημα 6.3.

5.3. Συμπεράσματα

Αν και ο κυρίαρχος σκοπός της οντολογίας είναι η δημιουργία λεξιλογίου, στην παρούσα εργασία η οντολογία κλήθηκε να δώσει σχεδόν τα ίδια αποτελέσματα που θα μπορούσαμε να πάρουμε από μια βάση δεδομένων. Μια οντολογία έχει ως στόχο την περιγραφή των συσχετίσεων μεταξύ των κλάσεων και των αντικειμένων του κόσμου, ενώ μια βάση δεδομένων έχει ως στόχο την αποθήκευση δεδομένων σε αυτήν. Μπορεί να θεωρηθεί σαν μια πιο “εξελιγμένη βάση δεδομένων” εννοώντας πως αφήνει πίσω της την υπόθεση του κλειστού κόσμου που διέπει ένα σύστημα βάσης δεδομένων και τις συνέπειες του γεγονότος πως μια βάση δεδομένων είναι εν γένει μη πλήρης (για παράδειγμα σε μια βάση δεδομένων έχουμε δύο πίνακες που ο

ένας περιέχει τους ανθρώπους και ο άλλος τους καθηγητές – και προφανώς όλοι οι καθηγητές είναι άνθρωποι – μπορεί κάποιος καθηγητής να μην αναφέρεται στον πίνακα των ανθρώπων). Μέσω των συσχετίσεων και των αξιωμάτων που περιγράφονται σε μια οντολογία, “αποκαλύπτονται” μέσω της εκλογίκευσης συσχετίσεις οι οποίες προηγούμενα δεν ήταν δηλωμένες αλλά μπορούν να εννοηθούν μέσω αυτών (που στο παράδειγμα θα είχε ως αποτέλεσμα όλοι οι καθηγητές που δεν ήταν στον πίνακα των ανθρώπων, αν είχε δηλωθεί πως όλοι οι καθηγητές είναι άνθρωποι, τότε θα ανήκαν μετά τη συλλογιστική και στον πίνακα των ανθρώπων).

Τα υπάρχοντα συστήματα βάσεων δεδομένων θα μπορούσαν να χαρακτηριστούν ως επίπεδα μιας και δε μπορούν να ενσωματώσουν τη σημασιολογία στα δεδομένα τους και δεν έχουν γνώση για να αντιληφθούν κάτι πιο πέρα από όλα όσα δηλώνονται με ρητό τρόπο σε αυτά. Αντίθετα, μια οντολογία είναι σε θέση να αναγνωρίσει αυτόματα μη ορισμένες εξ’ αρχής σχέσεις μεταξύ των αντικειμένων του κόσμου της. Οπότε το να διαλέξουμε μεταξύ οντολογίας και βάσης δεδομένων είναι μια επιλογή μεταξύ ταχύτητας - μεγάλων δυνατοτήτων¹⁰³ (βάση δεδομένων) και πληρότητας - απλότητας (οντολογία), ελπίζοντας φυσικά πως η «ταχύτητα» είναι κάτι που θα βελτιωθεί κατά την περαιτέρω ανάπτυξη των οντολογιών και των εργαλείων τους.

¹⁰³ Σε μια βάση δεδομένων μπορούμε με ερωτήματα τα οποία πρέπει να τα γράψουμε σε SQL να πάρουμε το οποιοδήποτε επιθυμητό αποτέλεσμα. Όμως, κάποιες φορές τα ερωτήματα αυτά είναι ιδιαίτερα πολύπλοκα για να γραφούν, οπότε και συνήθως καταφεύγουμε σε πολλά μικρότερα ερωτήματα.

6. Παράρτημα

6.1. Πολυπλοκότητα των προφίλ της OWL 2

Στον Πίνακα που ακολουθεί στην επόμενη σελίδα, παρατηρούμε την Πολυπλοκότητα των προφίλ της OWL 2. Η ερμηνεία των Πολυπλοκοτήτων είναι η ακόλουθη:

- Not Applicable: Μη εφαρμόσιμο
- Undecidable: Μη αποφάνσιμο: Δε μπορεί να ολοκληρωθεί η συμπερασματολογία.
- Decidable, but complexity open (NP-Hard): Αποφάνσιμο, αλλά δεν έχουμε αποφασίσει τι πολυπλοκότητα έχει – Το NP-Hard στη παρένθεση δηλώνει το κατώτερο όριο της πολυπλοκότητας του προβλήματος (NP: Non deterministic Polynomial time, η κλάση των προβλημάτων που λύνονται (το πολύ) σε πολυωνυμικό χρόνο από ένα μη ντετερμινιστικό αλγόριθμο).
- Decidability open: Ανοιχτή αποφανσιμότητα - δε ξέρουμε αν έχει ή δεν έχει λύση.
- X-complete: Είναι τα X-Hard τα οποία ανήκουν στη κατηγορία X (όπου X μια κλάση πολυπλοκότητας). Είναι τα πιο δύσκολα προβλήματα μιας κλάσης πολυπλοκότητας X.
 - ο PTIME-complete: PTIME-πλήρες - Τα προβλήματα τα οποία λύνονται από ντετερμινιστικό αλγόριθμο (το πολύ) σε πολυωνυμικό χρόνο.
 - ο NP-complete: NP-πλήρες - Τα προβλήματα που λύνονται από μη ντετερμινιστικό αλγόριθμο (το πολύ) σε πολυωνυμικό χρόνο.
 - ο PSPACE-complete: PSACE-πλήρες – Τα προβλήματα που λύνονται με ντετερμινιστικό αλγόριθμο και χρησιμοποιούν (το πολύ) πολυωνυμικό χώρο (μνήμη).
 - ο NLogSpace-complete: Το LOGSPACE είναι τα προβλήματα που λύνονται με ντετερμινιστικό αλγόριθμο και χρησιμοποιούν (το

πολύ) λογαριθμικό χώρο (μνήμη). Το NLOGSPACE είναι το μη ντετερμινιστικό ανάλογο αυτής της κλάσης.

- ο In AC^0 : Είναι μια γνήσια υποκλάση της LOGSPACE και δεν είναι ορισμένη μέσω μηχανών Turing¹⁰⁴ αλλά μέσω κυκλωμάτων. Πιο συγκεκριμένα, οι AC^0 είναι η κλάση των προβλημάτων τα οποία ορίζονται χρησιμοποιώντας μια οικογένεια κυκλωμάτων βάθους $O(1)$ και πολωνυμικού μεγέθους τα οποία αποτελούνται από πύλες AND και OR με απεριόριστες εισόδους, και τα οποία προβλήματα μπορούν να παραχθούν από μια ντετερμινιστική μηχανή Turing σε λογαριθμικό χρόνο. Διαισθητικά, η AC^0 μας επιτρέπει να χρησιμοποιήσουμε πολωνυμικά πολλούς επεξεργαστές, αλλά ο χρόνος εκτέλεσης πρέπει να είναι σταθερός. Ένα παράδειγμα ενός προβλήματος που ανήκει στην AC^0 είναι η αποτίμηση των πρωτοβάθμιων ερωτημάτων σε βάσεις δεδομένων (ή ο έλεγχος μοντέλων πρωτοβάθμιων προτάσεων σε πεπερασμένα μοντέλα), όπου η βάση (πρωτοβάθμιο μοντέλο) θεωρείται ως η είσοδος και το ερώτημα (πρωτοβάθμια πρόταση) θεωρείται ότι είναι σταθερό.
- ο NEXPTIME-complete: NEXPTIME-πλήρη – Τα προβλήματα που λύνονται με μη ντετερμινιστικό αλγόριθμο (το πολύ) σε εκθετικό χρόνο.
- ο 2NEXPTIME-complete: NEXPTIME-πλήρη – Τα προβλήματα που λύνονται με μη ντετερμινιστικό αλγόριθμο (το πολύ) σε διπλά εκθετικό χρόνο.

Τα αποτελέσματα που υπάρχουν στον παρακάτω πίνακα αναφέρονται στην πολυπλοκότητα της χειρότερης περίπτωσης αυτών των προβλημάτων συλλογιστικής.
(13)

¹⁰⁴ Μηχανή Turing: Είναι ένα υπολογιστικό μοντέλο που μπορεί να προσομοιώσει οποιονδήποτε υπολογιστή. Τα προβλήματα που δεν μπορούν να επιλυθούν από μια μηχανή Turing δεν μπορούν να επιλυθούν (είναι μη επιλύσιμα). (153)

Language	Reasoning Problems	Taxonomic Complexity	Data Complexity	Query Complexity	Combined Complexity
OWL 2 RDF-Based Semantics	Ontology Consistency, Class Expression Satisfiability, Class Expression Subsumption, Instance Checking, Conjunctive Query Answering	Undecidable	Undecidable	Undecidable	Undecidable
OWL 2 Direct Semantics	Ontology Consistency, Class Expression Satisfiability, Class Expression Subsumption, Instance Checking	2NEXPTIME-complete (NEXPTIME if property hierarchies are bounded)	Decidable, but complexity open (NP-Hard)	Not Applicable	2NEXPTIME-complete (NEXPTIME if property hierarchies are bounded)
	Conjunctive Query Answering	Decidability open	Decidability open	Decidability open	Decidability open
OWL 2 EL	Ontology Consistency, Class Expression Satisfiability, Class Expression Subsumption, Instance Checking	PTIME-complete	PTIME-complete	Not Applicable	PTIME-complete
	Conjunctive Query Answering	PTIME-complete	PTIME-complete	NP-complete	PSPACE-complete
OWL 2 QL	Ontology Consistency, Class Expression Satisfiability, Class Expression Subsumption, Instance Checking	NLogSpace-complete	In AC^0	Not Applicable	NLogSpace-complete
	Conjunctive Query Answering	NLogSpace-complete	In AC^0	NP-complete	NP-complete
OWL 2 RL	Ontology Consistency, Class Expression Satisfiability, Class Expression Subsumption, Instance Checking	PTIME-complete	PTIME-complete	Not Applicable	PTIME-complete
	Conjunctive Query Answering	PTIME-complete	PTIME-complete	NP-complete	NP-complete
OWL 1 DL	Ontology Consistency, Class Expression Satisfiability, Class Expression Subsumption, Instance Checking	NEXPTIME-complete	Decidable, but complexity open (NP-Hard)	Not Applicable	NEXPTIME-complete
	Conjunctive Query Answering	Decidability open	Decidability open	Decidability open	Decidability open

Εικόνα 43: Πολυπλοκότητα των προφίλ της OWL 2

6.2. Στοιχεία μεταδεδομένων Dublin Core

1. Τίτλος	Ένα όνομα που δίνεται στην πηγή
2. Δημιουργός	Πρόσωπο ή οργανισμός υπεύθυνος του περιεχομένου της πηγής
3. Εκδότης	Πρόσωπο ή οργανισμός που κάνει την πηγή διαθέσιμη στη βιβλιοθήκη
4. Θέμα	Χρησιμοποιούνται συνήθως λέξεις-κλειδιά
5. Περιγραφή	Περίληψη, πίνακας περιεχομένων, κτλ
6. Πηγή	Αριθμός ή αλφαριθμητικό σύμφωνο με κάποιο επίσημο σύστημα προσδιορισμού πηγών
7. Γλώσσα	Η γλώσσα του περιεχομένου της πηγής κωδικοποιημένη
8. Συνάφεια	Αναφορά σε μια συναφή πηγή με τη χρήση ενός προσδιοριστικού αριθμού ή αλφαριθμητικού
9. Θεματική κάλυψη	Η έκταση ή το φάσμα του περιεχομένου της πηγής (π.χ. γεωγραφικές συντεταγμένες, χρονική περίοδος κτλ)
10. Ημερομηνία	Συνήθως ημερομηνία δημιουργίας ή έκδοσης της πηγής
11. Τύπος	Η φύση του περιεχομένου της πηγής
12. Μορφή	Μπορεί να περιλαμβάνει υλικό ή λογισμικό που απαιτείται για την εμφάνιση της πηγής, μέγεθος, διάρκεια κτλ
13. Προσδιοριστικό	Αριθμός ή αλφαριθμητικό σύμφωνο με κάποιο επίσημο σύστημα προσδιορισμού πηγών, όπως π.χ. ο ISBN για βιβλία
14. Χορηγός	Πρόσωπο ή οργανισμός υπεύθυνος για την ύλη του περιεχομένου της πηγής
15. Δικαιώματα	Πληροφορίες σχετικά με τα δικαιώματα χρήσης της πηγής

Εικόνα 44: Στοιχεία μεταδεδομένων Dublin Core

6.3. Τροποποιημένο σχήμα βάσης Εργαλείου *semfe.gr*

Αυτό είναι μια περιγραφή της τροποποιημένης βάσης που χρησιμοποιήθηκε στη παρούσα πτυχιακή. Αφαιρέθηκαν κάποιοι πίνακες από την αρχική βάση, αλλά δεν έχουν αλλάξει τα πεδία των πινάκων.

1. uni_branches: (ορισμός κορμού/κατευθύνσεων και ελάχιστο πλήθος μαθημάτων για πτυχίο)

branch_id - τιμές 1, 2, 3

branch_caption - Οι επιλογές Κορμός/Μαθηματικό/Φυσικό

branch_short - Συντομογραφία "επιλογής"

branch_first_semester - 1ο εξάμηνο "επιλογής"

branch_last_semester - τελευταίο εξάμηνο "επιλογής"

branch_subjects - ελάχιστο πλήθος μαθημάτων για πτυχίο

2. uni_departments: (τομείς)

department_id – id τομέα

department_caption – όνομα τομέα

department_web – ηλεκτρονική διεύθυνση τομέα

department_chief_teacher_ptr – id υπεύθυνου καθηγητή τομέα

department_description – περιγραφή τομέα

department_tel – τηλέφωνα τομέα

department_fax – fax τομέα

department_email – email τομέα

3. uni_fields: (ροές)

field_id - id ροής

field_caption - όνομα ροής

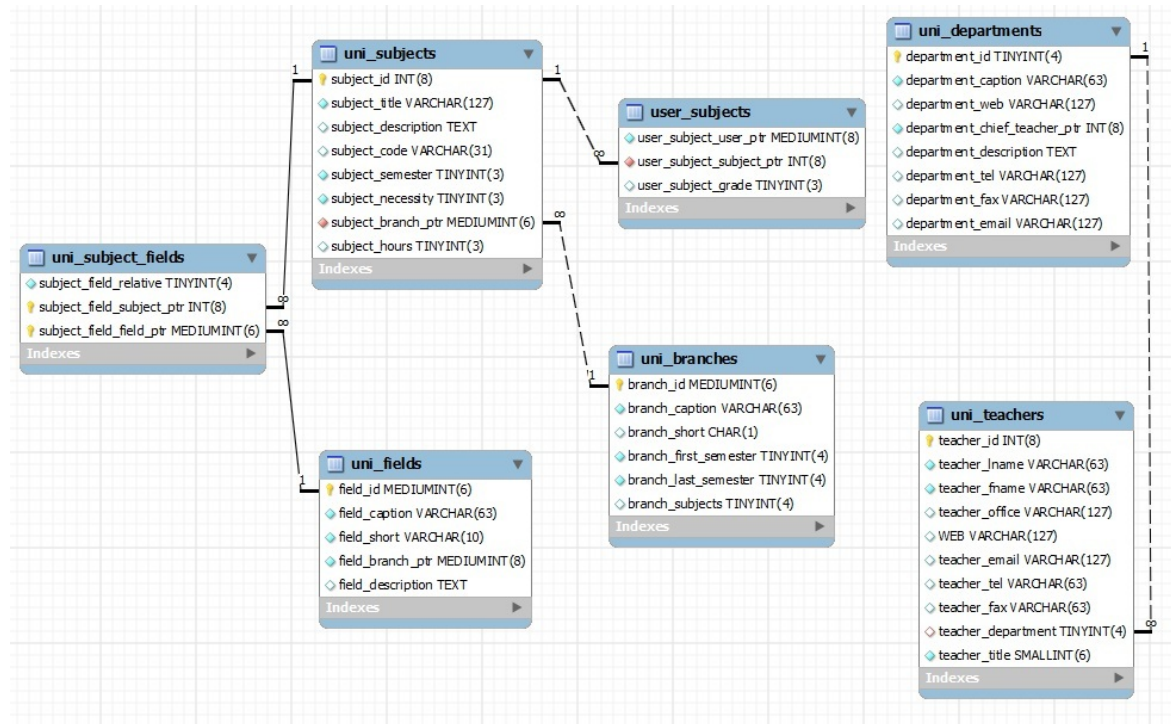
field_short - συντομογραφία ροής

field_branch_ptr - κατεύθυνση (τιμές 2 και 3)

field_description - περιγραφή

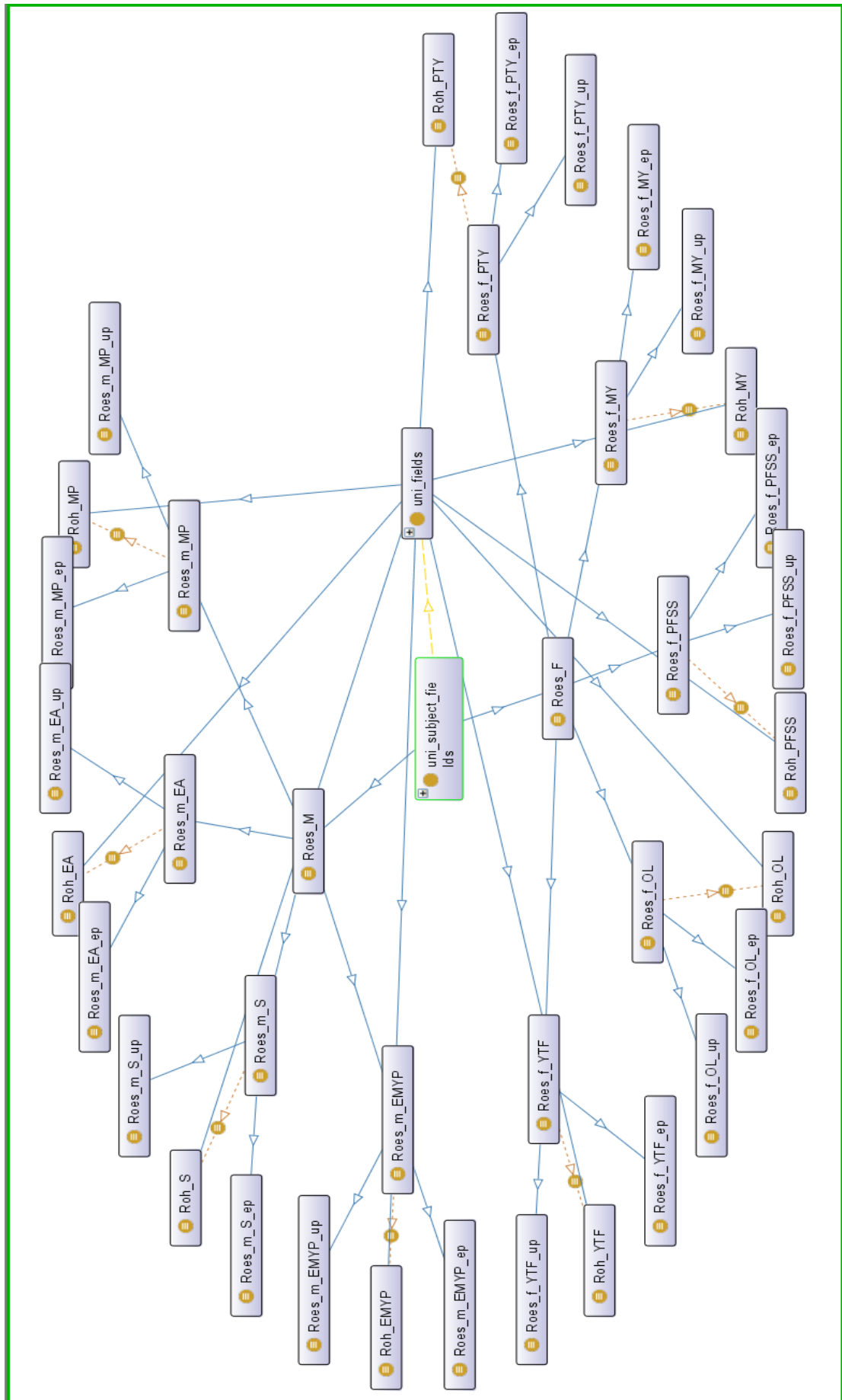
4. uni_subjects: (πληροφορίες οδηγού σπουδών για τα μαθήματα)
subject_id - id μαθήματος
subject_title - τίτλος μαθήματος
subject_description - περιγραφή μαθήματος
subject_code - κωδικός μαθήματος (κατά τον οδηγό σπουδών)
subject_semester - εξάμηνο μαθήματος (1-9)
subject_necessity - 1 Υποχρεωτικό (Κορμός+Κατευθύνσεις), 2 Κατ' Επιλογήν (Κορμός+Κατευθύνσεις), 3 Προαιρετικά, 4 Μεταβατικά
subject_branch_ptr - 1 Κορμός, 2 Φυσικό, 3 Μαθηματικό
subject_hours - ώρες διδασκαλίας
5. uni_subject_fields: (μαθήματα ροών)
subject_field_relative - 0 υποχρεωτικό, 1 κατ' επιλογήν
subject_field_subject_ptr - id μαθήματος
subject_field_field_ptr - ροή (1-9)
6. uni_teachers: (καθηγητές ΣΕΜΦΕ)
teacher_id – id καθηγητή
teacher_lname – επίθετο
teacher_fname – όνομα
teacher_office - γραφείο
WEB - ιστοσελίδα
teacher_email - email
teacher_tel - τηλέφωνο
teacher_fax - fax
teacher_department – τομέας (τιμές 1-4)
teacher_title – βαθμίδα (τιμές 1-11)
7. user_subjects: (δηλωμένα μαθήματα φοιτητών)
user_subject_user_ptr - user id
user_subject_subject_ptr - κωδικός μαθήματος
user_subject_grade - βαθμός μαθήματος (ακόμα και <5)

Το ακόλουθο σχεδιάγραμμα του σχήματος της βάσης δημιουργήθηκε με τη βοήθεια του προγράμματος MySQL Workbench 5.2.

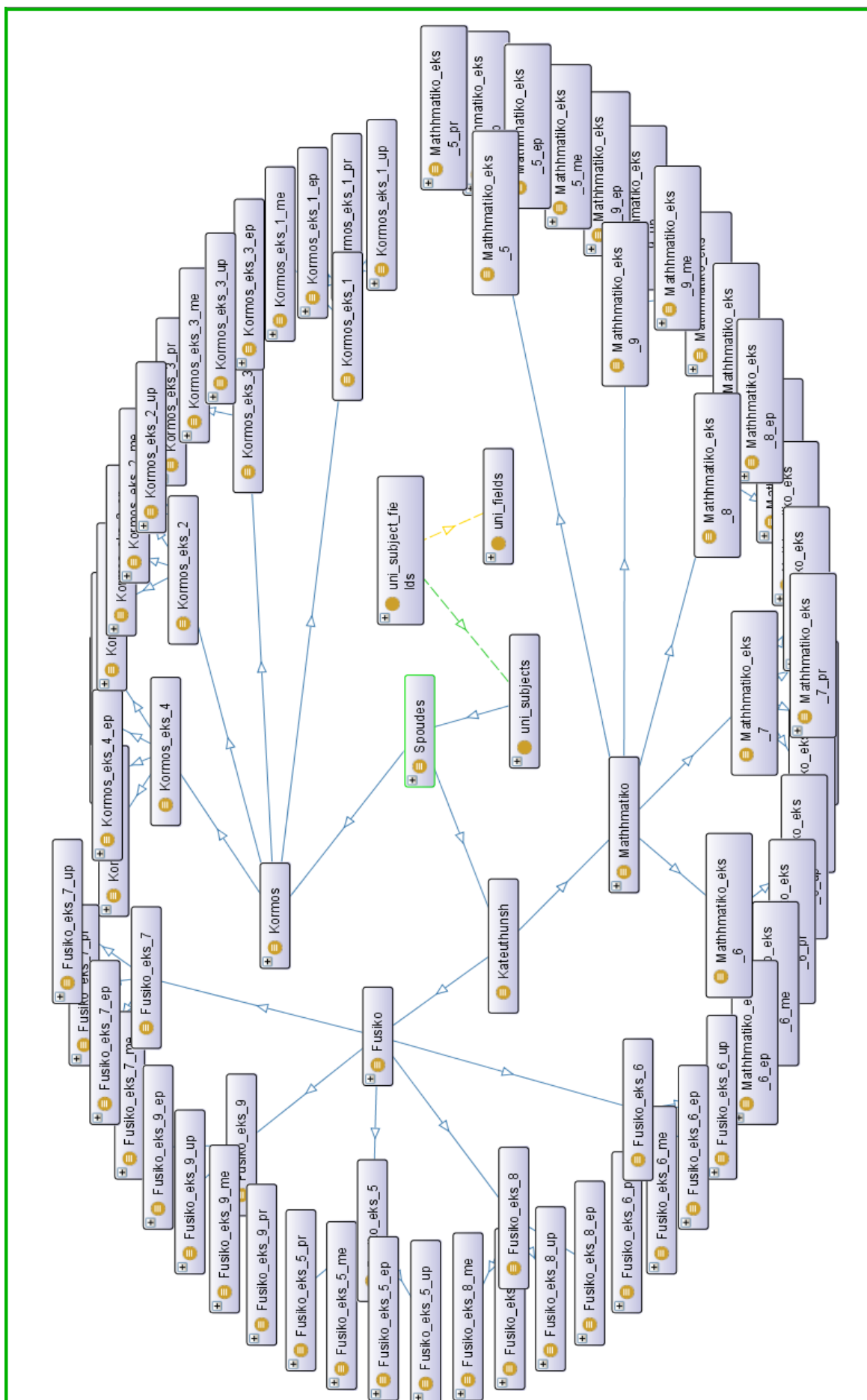


Εικόνα 45: Σχήμα βάσης με foreign keys

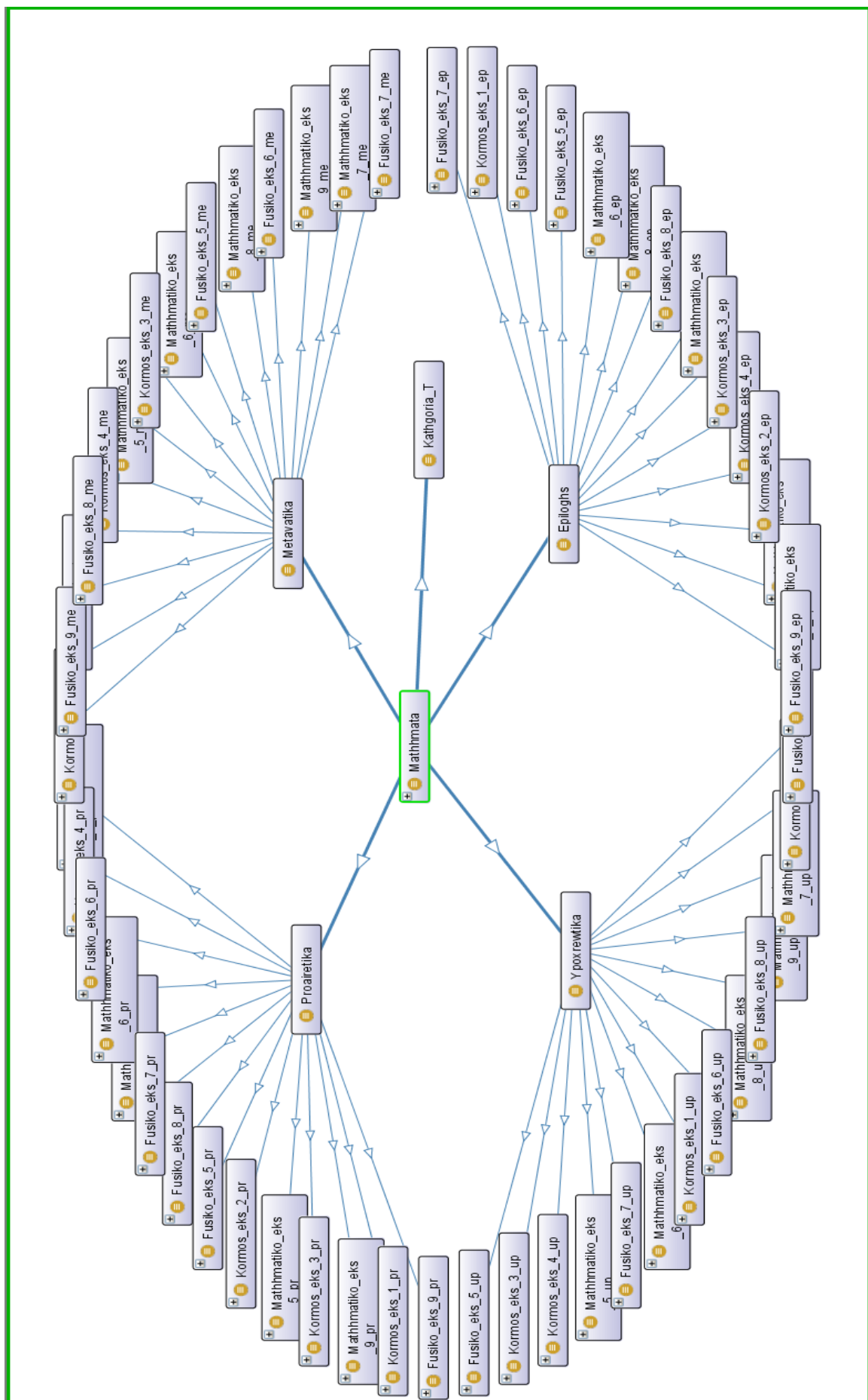
Τα ακόλουθα σχεδιαγράμματα δημιουργήθηκαν με τη βοήθεια του OntoGraph που βρίσκεται μέσα στο Protégé 4.1.



Εικόνα 46: Σχέσεις κλάσεων/υποκλάσεων των ροών



Εικόνα 48: Σχέσεις κατευθύνσεων/ροών με κατηγορίες μαθημάτων ανά εξάμηνο



Εικόνα 49: Σχέσεις κατηγοριών μαθημάτων και των υποκλάσεων τους

Βιβλιογραφία

1. Fischetti M, Berneres-Lee T. Weaving the Web New York: HarperCollins Publishers; 2000.
2. Wikipedia. [Online]. [cited 2011 12 22. Available from: http://en.wikipedia.org/wiki/World_Wide_Web.
3. Antoniou G, van Harmelen, F. A SemanticWeb Primer United States of America: Massachusetts Institute of Technology; 2003.
4. Βλαχάβας Ι, Κεφαλάς Π, Βασιλειάδης Ν, Κόκκορας Φ, Σακελλαρίου Η. Τεχνητή Νοημοσύνη - Γ' Έκδοση: Εκδόσεις Πανεπιστημίου Μακεδονίας; 2011.
5. Spivack N. lifeboat.com/. [Online]. [cited 2012 1 7 [[Αναφορά στο Web 3.0]]. Available from: <http://lifeboat.com/ex/web.3.0>.
6. Appelquist D, Brickley D, Carvahlo M, Ianella R, Passant A, Perey C, et al. W3C. [Online].; 2010 [cited 2011 12 24. Available from: <http://www.w3.org/2005/Incubator/socialweb/XGR-socialweb-20101206/>.
7. Cormode G, Krishnamurthy B. First Monday. [Online].; 2008 [cited 2011 12 24 [[Σχέση Web 1.0 με Web 2.0]]. Available from: <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2125/1972>.
8. Shannon V. A 'more revolutionary' Web. 2006 May 23.
9. Wikipedia. [Online]. [cited 2012 1 7 [[Αναφορά στο Σημασιολογικό Ιστό στις Εφαρμογές αυτού και στο Web 3.0]]. Available from: http://en.wikipedia.org/wiki/Semantic_Web.
10. Βικιπαίδεια. [Online]. [cited 2011 12 22 [[Ορισμός Οντολογίας]]. Available from: [http://el.wikipedia.org/wiki/Οντολογία_\(Πληροφορική\)](http://el.wikipedia.org/wiki/Οντολογία_(Πληροφορική)).
11. Kunjithapatham A, Sheshagiri M, Finin T, Joshi A, Peng Y. Personal Agents on the Semantic Web. Baltimore: University of Maryland, Department of Computer Science and Electrical Engineering.
12. Wikipedia. [Online]. [cited 2012 1 2 [[Αναφορά στη Διεπαφή Χρήστη]]. Available from: http://en.wikipedia.org/wiki/Semantic_Web_Stack.
13. W3C. [Online]. [cited 2011 12 24 [[Πληροφορίες για την OWL 2]]. Available from: <http://www.w3.org/TR/owl-profiles/>.

14. Wikipedia. [Online]. [cited 2012 1 17 [[Πληροφορίες για την RIF]]. Available from: http://en.wikipedia.org/wiki/Rule_Interchange_Format.
15. W3C. [Online]. [cited 2012 1 17 [[Πληροφορίες για την RIF]]. Available from: <http://www.w3.org/2005/rules/wiki/Primer>.
16. W3C. [Online]. [cited 2012 1 17 [[Πληροφορίες για τη RIF]]. Available from: <http://www.w3.org/TR/rif-ucr/>.
17. W3C. [Online]. [cited 2012 1 17. Available from: <http://www.w3.org/TR/rif-overview/>.
18. Wikipedia. [Online]. [cited 2012 1 17 [[Πληροφορίες για την SWRL]]. Available from: http://en.wikipedia.org/wiki/Semantic_Web_Rule_Language.
19. Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosz B, Dean M. W3C. [Online].; 2004 [cited 2011 12 24 [[Πληροφορίες για την SWRL]]. Available from: <http://www.w3.org/Submission/SWRL/>.
20. Kifer M, de Bruijn J, Boley H, Fensel D. A Realistic Architecture for the Semantic Web. ; 2005.
21. www.cs.odu.edu. [Online]. [cited 2012 1 12 [[Πληροφορίες για τη Λογική]]. Available from: <http://www.cs.odu.edu/~toida/nerzic/content/logic/>.
22. lpis.csd.auth.gr. [Online]. [cited 2012 1 13 [[Πληροφορίες για τη Λογική]]. Available from: http://lpis.csd.auth.gr/prolog/Logic_Programming/node3.html.
23. Kuper J. Overview of proposition and predicate logic. ; 2005.
24. lpis.csd.auth.gr. [Online]. [cited 2012 1 13 [[Πληροφορίες για τη Προτασιακή Λογική]]. Available from: http://lpis.csd.auth.gr/prolog/Logic_Programming/node4.html.
25. cs.odu.edu. [Online]. [cited 2012 1 13 [[Πληροφορίες για τη Προτασιακή Λογική]]. Available from: http://www.cs.odu.edu/~toida/nerzic/content/logic/prop_logic/intr_to_prop_logic.html.
26. Κολέτσος Γ. Εισαγωγή στη Μαθηματική Λογική: Εκκρεμές; 2009.
27. cs.odu.edu. [Online]. [cited 2012 1 13 [[Πληροφορίες για τη Προτασιακή Λογική]]. Available from: http://www.cs.odu.edu/~toida/nerzic/content/logic/pred_logic/intr_to_pred_logic.html.

[c.html](#).

28. [lpis.csd.auth.gr](#). [Online]. [cited 2012 1 13 [[Πληροφορίες για τη Λογική Πρώτης Τάξης]]. Available from: http://lpis.csd.auth.gr/prolog/Logic_Programming/node5.html.
29. Russell SJ, Norvig P. Τεχνητή νοημοσύνη: μια σύγχρονη προσέγγιση: Κλειδάριθμος; 2004.
30. [lpis.csd.auth.gr](#). [Online].; 2012 [[Αναφορά στις Ερμηνείες και τα Μοντέλα της Κατηγορηματικής Λογικής]]. Available from: http://lpis.csd.auth.gr/prolog/Logic_Programming/node7.html.
31. [lpis.csd.auth.gr](#). [Online]. [cited 2012 1 15 [[Πληροφορίες για τη Σημασιολογία στη Κατηγορηματική Λογική]]. Available from: http://lpis.csd.auth.gr/prolog/Logic_Programming/node6.html.
32. Nutt W, Baader F. Basic Description Logics. ; 2003.
33. Grau BC, Horrocks I, Motik B, Parsia B, Patel-Schneider P, Sattler U. OWL 2: The next step for OWL. Web Semantics: Science, Services and Agents on the World Wide Web. 2008.
34. w3.org. [Online]. [cited 2012 2 3 [[Πληροφορίες για την OWL 2]]. Available from: <http://www.w3.org/TR/owl2-new-features/>.
35. Boley H, Tabet S, Wanger G. cin.ufpe.br. [Online]. [cited 2012 2 16 [[Πληροφορίες για την RuleML]]. Available from: <http://www.cin.ufpe.br/~compint/aulas-IAS/artigos/BoleyTabetWagnerRuleML.html>.
36. RuleML. [Online]. Available from: <http://ruleml.org/>.
37. Wagner G, Antoniou G, Tabet S, Boley H. The Abstract Syntax of RuleML - Towards a General Web Rule Language Framework. National Research Council Canada, Institute for Information Technology.
38. Boley H, Athan T. ruleml.org. [Online]. [cited 2012 1 10 [[Πληροφορίες για τη RuleML]]. Available from: <http://ruleml.org/papers/Primer/RuleMLPrimer2011-12-11.html>.
39. ruleml.org. [Online]. [cited 2011 12 25. Available from: <http://ruleml.org>.
40. Boley H. The RuleML Family of Web Rule Languages..

41. w3.org. [Online]. [cited 2012 1 10 [[Πληροφορίες για τη MathML]]. Available from: <http://www.w3.org/Math/>.
42. w3.org. [Online]. [cited 2012 1 10 [[Πληροφορίες για την MathML]]. Available from: <http://www.w3.org/Math/whatIsMathML.html>.
43. Wikipedia. [Online]. [cited 2012 1 11 [[Πληροφορίες για τη DAML]]. Available from: http://en.wikipedia.org/wiki/DARPA_Agent_Markup_Language.
44. About DAML. [Online]. [cited 2012 1 11 [[Πληροφορίες για τη DAML]]. Available from: <http://www.daml.org/about.html>.
45. daml.org. [Online]. [cited 2012 1 11 [[Πληροφορίες για τη DAML]]. Available from: <http://www.daml.org/2002/04/why.html>.
46. www.dmg.org. [Online]. [cited 2012 1 15 [[Πληροφορίες για την PMML]]. Available from: <http://www.dmg.org/pmml-v4-1.html>.
47. dmg.org. [Online]. [cited 2012 1 15 [[Πληροφορίες για την PMML]]. Available from: <http://www.dmg.org/v4-1/GeneralStructure.html>.
48. Wikipedia. [Online]. [cited 2012 1 15 [[Πληροφορίες για την PMML]]. Available from: http://en.wikipedia.org/wiki/Predictive_Model_Markup_Language.
49. divms.uiowa.edu. [Online]. [cited 2012 1 17 [[Πληροφορίες για την Attribute Grammar]]. Available from: <http://www.divms.uiowa.edu/~hzhang/c123/Lecture3.pdf>.
50. Boley H. dfki.uni-kl.de. [Online]. [cited 2012 2 4 [[Πληροφορίες για το AG-markup]]. Available from: <http://www.dfki.uni-kl.de/~boley/xmlag/attgramm/>.
51. W3C. [Online]. [cited 2011 12 28 [[Πληροφορίες για την XSLT]]. Available from: <http://www.w3.org/TR/xslt>.
52. W3C. [Online]. [cited 2012 1 10 [[Αναφορά στην XSLT 2.0]]. Available from: <http://www.w3.org/TR/xslt20/>.
53. Boley H, Grosz B, Sintek M, Tabet S, Wagner G. ruleml.org. [Online]. [cited 2012 2 2 [[Πληροφορίες για την RuleML]]. Available from: <http://ruleml.org/indesign.html#RuleML-hierarchy>.
54. DBpedia. [Online]. [cited 2011 12 25 [[Πληροφορίες σχετικά με την

- DBpedia]]. Available from: http://wiki.dbpedia.org/About?show_comments=1.
55. DBpedia. [Online]. [cited 2011 12 25 [[Άδεια χρήσης της DBpedia]]. Available from: <http://wiki.dbpedia.org/Imprint>.
 56. DBpedia. [Online]. [cited 2011 12 20 [[Πληροφορίες για την DBpedia]]. Available from: <http://wiki.dbpedia.org/Datasets>.
 57. DBpedia el.Greece. [Online]. [cited 2012 1 2 [[Αναφορά στη DBpedia]]. Available from: <http://el.dbpedia.org/>.
 58. DBpedia. [Online]. [cited 2011 12 25 [[Στοιχεία πάνω στη DBpedia Live]]. Available from: <http://wiki.dbpedia.org/DBpediaLive>.
 59. DBpedia el.Greece. [Online]. [cited 2012 1 3 [[Αναφορά στην Ελληνική έκδοση της DBpedia]]. Available from: <http://el.dbpedia.org/blog/2011/04/14/ανακοίνωση-τύπου-της-ελληνικής-dbpedia>.
 60. Βικιπαίδεια. [Online]. [cited 2011 12 25 [[Στατιστικά στοιχεία Βικιπαίδειας]]. Available from: <http://el.wikipedia.org/wiki/Βικιπαίδεια:Σχετικά>.
 61. Master in Web Science. [Online]. [cited 2012 1 3 [[Πληροφορίες για τη Διεθνοποίηση της DBpedia]]. Available from: <http://webscience.auth.gr/?q=el/content/η-ελληνική-dbpedia-ως-πρωτο-βημα-προς-τη-διεθνοποιηση-της-dbpedia>.
 62. Βικιπαίδεια:DBpedia. [Online].; Βικιπαίδεια [cited 2012 1 3. Available from: <http://el.wikipedia.org/wiki/Βικιπαίδεια:DBpedia>.
 63. Bizer C, Heath T, Ayers D, Raimond Y. Interlinking Open Data on the Web..
 64. Bizer C, Heath T, Berners-Lee T. tomheath.com. [Online]. [cited 2012 2 4 [[Πληροφορίες για τα Linking Open Data]]. Available from: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.
 65. W3C. [Online]. [cited 2012 1 9 [[Πληροφορίες για το project Linking Open Data]]. Available from: <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.
 66. Wikipedia. [Online]. [cited 2012 1 9 [[Αναφορά στο LOD Cloud 2011]]. Available from: http://en.wikipedia.org/wiki/Linked_data.

67. richard.cyganiak.de. [Online]. [cited 2012 1 11 [[Πληροφορίες για το Linking Open Data Cloud]]. Available from: <http://richard.cyganiak.de/2007/10/loa/>.
68. DBpedia el.Greece. [Online]. [cited 2012 1 9 [[Πληροφορίες για το Ελληνικό Νέφος Διασυνδεδεμένων Δεδομένων]]. Available from: <http://el.dbpedia.org/node/32>.
69. DBpedia el.Greece. [Online]. [cited 2012 1 9 [[Αναφορά στην Ελληνική DBpedia]]. Available from: <http://wiki.el.dbpedia.org/en>.
70. foaf project. [Online]. [cited 2012 1 4 [[Αναφορά στο FOAF]]. Available from: <http://www.foaf-project.org/about>.
71. Wikipedia. [Online]. [cited 2012 1 4 [[Αναφορά στο FOAF]]. Available from: [http://en.wikipedia.org/wiki/FOAF_\(software\)](http://en.wikipedia.org/wiki/FOAF_(software)).
72. wiki.foaf-project.org. [Online]. [cited 2012 1 5 [[Αναφορά στο FOAF]]. Available from: <http://wiki.foaf-project.org/w/FAQ>.
73. foaf project. [Online]. [cited 2012 1 5 [[Αναφορά στο FOAF]]. Available from: <http://www.foaf-project.org/original-intro>.
74. wiki.foaf-project.org. [Online]. [cited 2012 1 4 [[Ιστορία FOAF]]. Available from: <http://wiki.foaf-project.org/w/FoafHistory>.
75. http://xmlns.com. [Online]. [cited 2012 1 4 [[Χρήση ονόματος FOAF]]. Available from: <http://xmlns.com/foaf/spec/>.
76. ldodds.com. [Online]. [cited 2012 1 4 [[Αναφορά στο FOAF-a-Matic]]. Available from: <http://www.ldodds.com/foaf/foaf-a-matic>.
77. foaf project. [Online]. [cited 2012 1 5 [[Projects FOAF]]. Available from: <http://www.foaf-project.org/projects>.
78. wiki.foaf-project.org. [Online]. [cited 2012 1 5 [[Εργαλεία FOAF]]. Available from: <http://wiki.foaf-project.org/w/Tools>.
79. [Online]. [cited 2012 1 4 [[Αναφορά στο FOAF-a-Matic Mark 2]]. Available from: <http://www.ldodds.com/wordtin/Wiki.jsp?page=FOAFaMaticMark2>.
80. Harth A, Breslin JG, O'Murchu I, Decker S. Linking Semantically-Enabled Online Community Sites. Galway, Ireland: National University of Ireland, Digital Enterprise Research Institute.
81. Wikipedia. [Online]. [cited 2012 1 5 [[Αναφορά στο SIOC]]. Available from:

- http://en.wikipedia.org/wiki/Semantically-Interlinked_Online_Communities.
82. sioc-project.org/faq. [Online]. [cited 2012 1 9 [[Πληροφορίες για το SIOC project]]. Available from: <http://sioc-project.org/faq>.
 83. pingthesemanticweb.com. [Online]. [cited 2012 1 9 [[Αναφορά στο pingthesemanticweb.com]]. Available from: <http://www.pingthesemanticweb.com/>.
 84. sioc-project.org. [Online]. [cited 2012 1 9 [[Αναφορά στο Semantic Radar]]. Available from: <http://sioc-project.org/firefox>.
 85. Wikipedia. [Online]. [cited 2012 1 5 [[Πληροφορίες για τη Nextbio]]. Available from: <http://en.wikipedia.org/wiki/Nextbio>.
 86. Nextbio. [Online]. [cited 2012 1 24 [[Πληροφορίες για το Nextbio]]. Available from: <http://www.nextbio.com/b/corp/pressReleases.nb#pr1>.
 87. Nextbio. [Online]. [cited 2012 1 24 [[Πληροφορίες για το Nextbio]]. Available from: <http://www.nextbio.com/b/corp/pressReleases.nb>.
 88. Nextbio. [Online]. [cited 2012 1 24 [[Πληροφορίες για το Nextbio]]. Available from: <http://www.nextbio.com/b/corp/pressReleases.nb>.
 89. cyc.com. [Online]. [cited 2012 2 13 [[Πληροφορίες για την Cyc]]. Available from: <http://cyc.com/cyc/company/about>.
 90. cyc.com. [Online]. [cited 2012 2 19 [[Πληροφορίες για την Cyc]]. Available from: <http://cyc.com/cyc/technology/whatisuncyc>.
 91. cyc.com. [Online]. [cited 2012 2 19 [[]]. Available from: http://cyc.com/cyc/technology/whatisuncyc_dir/whatsincyc.
 92. cyc.com. [Online]. [cited 2012 2 20 [[Πληροφορίες για την Cyc]]. Available from: http://cyc.com/cyc/technology/technology/whatisuncyc_dir/howdoesuncycreason.
 93. opencyc.org. [Online]. [cited 2012 2 5 [[Πληροφορίες για την Cyc και την OpenCyc]]. Available from: http://www.opencyc.org/faq/opencyc_faq.
 94. Stubkjær E. Integrating Ontologies: Assessing the use of the Cyc Ontology for Cadastral Applications. Fibigerstræde: Aalborg University, Dept. of Development and Planning.
 95. Cover Pages. [Online]. [cited 2012 2 19 [[Πληροφορίες για την Cyc]].

- Available from: <http://xml.coverpages.org/ni2001-03-06-a.html>.
96. cyc.com. [Online]. [cited 2012 2 19 [[Πληροφορίες για την Cyc]]. Available from: http://cyc.com/cyc/technology/whatis_cyc_dir/whatdoes_cyc_know.
 97. Grant P. Computer Science Foundations of Artificial Intelligence..
 98. <http://cyc.com>. [Online]. [cited 2012 2 13 [[Πληροφορίες για την OpenCyc]]. Available from: <http://cyc.com/cyc/opencyc/overview>.
 99. opencyc.org. [Online]. [cited 2012 2 13 [[Πληροφορίες για την OpenCyc]]. Available from: <http://www.opencyc.org/>.
 100. Wikipedia. [Online]. [cited 2012 2 5 [[Πληροφορίες για την OpenCyc και την Cyclopedia]]. Available from: <http://en.wikipedia.org/wiki/Cyc>.
 101. DBpedia. [Online]. [cited 2012 2 5 [[Πληροφορίες για την Cyclopedia]]. Available from: <http://wiki.dbpedia.org/OpenCyc>.
 102. Ψηφιακές Βιβλιοθήκες Ανοιχτού Κώδικα. [Online]. [cited 2012 1 29 [[Αναφορά στις Ψηφιακές Βιβλιοθήκες]]. Available from: <http://www.it.uom.gr/project/dl/intro.php?define=yes>.
 103. Wikipedia. [Online]. [cited 2012 1 29 [[Αναφορά στις Ψηφιακές Βιβλιοθήκες]]. Available from: http://en.wikipedia.org/wiki/Digital_library.
 104. Ψηφιακές Βιβλιοθήκες Ανοιχτού Κώδικα. [Online]. [cited 2012 1 29 [[Αναφορά στους τομείς λειτουργίας των Ψηφιακών Βιβλιοθηκών]]. Available from: <http://www.it.uom.gr/project/dl/intro.php?range=yes>.
 105. Ψηφιακές Βιβλιοθήκες Ανοιχτού Κώδικα. [Online]. [cited 2012 1 29 [[Αναφορά στις Ψηφιακές Βιβλιοθήκες]]. Available from: <http://www.it.uom.gr/project/dl/intro.php>.
 106. Ψηφιακές Βιβλιοθήκες Ανοιχτού Κώδικα. [Online]. [cited 2012 1 29 [[Πληροφορίες για τις Ψηφιακές Βιβλιοθήκες]]. Available from: http://www.it.uom.gr/project/dl/standards.php?dublin_core=yes.
 107. Βικιπαίδεια. [Online]. [cited 2012 1 29 [[Πληροφορίες για τις Ψηφιακές Βιβλιοθήκες]]. Available from: <http://el.wikipedia.org/wiki/Μεταδεδομένα>.
 108. Dublin Core Metadata Initiative. [Online]. [cited 2012 2 1 [[Πληροφορίες για το Dublin Core]]. Available from: <http://dublincore.org/documents/dces/>.
 109. Wikipedia. [Online]. [cited 2012 1 25 [[Αναφορά στους Αισθητήρες]].

- Available from: <http://en.wikipedia.org/wiki/Sensor>.
110. telesto.gr. [Online]. [cited 2012 1 16 [[Πληροφορίες για το Semantic Sensor Web και το έργο Πλάτωνας]]. Available from: http://telesto.gr/plato/?page_id=54.
111. Wikipedia. [Online]. [cited 2012 1 16 [[Πληροφορίες για το Semantic Sensor Web]]. Available from: http://en.wikipedia.org/wiki/Semantic_Sensor_Web.
112. Έργο ΠΛΑΤΩΝΑΣ. [Online]. [cited 2012 1 30 [[Πληροφορίες για το Έργο ΠΛΑΤΩΝΑΣ]]. Available from: http://telesto.gr/plato/?page_id=18.
113. Έργο ΠΛΑΤΩΝΑΣ. [Online]. [cited 2012 1 30 [[Πληροφορίες για το Έργο ΠΛΑΤΩΝΑΣ]]. Available from: http://telesto.gr/plato/?page_id=11.
114. Έργο ΠΛΑΤΩΝΑΣ. [Online]. [cited 2012 1 30 [[Πληροφορίες για το Έργο ΠΛΑΤΩΝΑΣ]]. Available from: http://telesto.gr/plato/?page_id=13.
115. Rein L, Winningham T. microformats.org. [Online]. [cited 2012 2 1 [[Πληροφορίες για τα Μεταδεδομένα σε Video]]. Available from: <http://microformats.org/wiki/video-metadata-model>.
116. Wikipedia. [Online]. [cited 2012 1 25 [[Πληροφορίες για τα Μεταδεδομένα σε Video]]. Available from: <http://en.wikipedia.org/wiki/Metadata>.
117. Creative Commons. [Online]. [cited 2012 2 5 [[Πληροφορίες για την Creative Commons]]. Available from: <http://wiki.creativecommons.org/FAQ>.
118. ΕΥΓΕΠ. [Online]. [cited 2012 2 25 [[Πληροφορίες για το Inspire]]. Available from: http://www.inspire.okxe.gr/index.php?option=com_content&view=article&id=54&Itemid=68.
119. Europa. [Online]. [cited 2012 2 25 [[Πληροφορίες για το Inspire]]. Available from: http://europa.eu/legislation_summaries/environment/general_provisions/128195_el.htm.
120. Inspire. [Online]. [cited 2012 2 25 [[Πληροφορίες για το Inspire]]. Available from: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/48>.
121. European Commission, INSPIRE Geoportal. [Online]. [cited 2012 2 25 [[Πληροφορίες για το Inspire Geoportal]]. Available from: [150](http://inspire-</p></div><div data-bbox=)

geoportal.ec.europa.eu/.

122. eurogeoss.eu. [Online]. [cited 2012 2 4 [[Πληροφορίες για το INSPIRE]]. Available from: http://www.eurogeoss.eu/Documents/EuroGEOSS_D_2_2_3.pdf.
123. sourceforge.net. [Online]. [cited 2012 2 4 [[Πληροφορίες για το GIMED]]. Available from: <http://gimed.sourceforge.net/>.
124. Τζώτσος Α, Βάκκας Θ. conf.ellak.gr. [Online]. [cited 2012 2 4 [[Πληροφορίες για το GIMED]]. Available from: <http://conf.ellak.gr/geodatacamp2010/wp-content/uploads/2010/11/geodatacamp10-inspire3.pdf>.
125. protégé. [Online]. [cited 2012 3 1 [[Πληροφορίες για το protégé]]. Available from: <http://protege.stanford.edu/>.
126. protégé. [Online]. [cited 2012 3 1 [[Πληροφορίες για το protégé]]. Available from: <http://protege.stanford.edu/aboutus/aboutus.html>.
127. protégé. [Online]. [cited 2012 3 1 [[Πληροφορίες για το protégé]]. Available from: <http://protege.stanford.edu/overview/>.
128. Wikipedia. [Online]. [cited 2011 12 22 [[Αναφορά στο Πρωτόκολλο TCP/IP]]. Available from: <http://en.wikipedia.org/wiki/TCP/IP>.
129. Wikipedia. [Online]. [cited 2011 12 22 [[Αναφορά στον Tim Berners-Lee]]. Available from: http://en.wikipedia.org/wiki/Tim_Berners-Lee.
130. Wikipedia. [Online]. [cited 2011 12 22 [[Αναφορά στο DECnet]]. Available from: <http://en.wikipedia.org/wiki/DECnet>.
131. Wikipedia. [Online]. [cited 2011 12 22 [[Αναφορά στην HTML]]. Available from: <http://en.wikipedia.org/wiki/HTML>.
132. Wikipedia. [Online]. [cited 2011 12 22 [[Αναφορά στο Πρωτόκολλο HTTP]]. Available from: http://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol.
133. Wikipedia. [Online]. [cited 2011 12 22 [[Αναφορά στο URL]]. Available from: http://en.wikipedia.org/wiki/Uniform_resource_locator.
134. Wikipedia. [Online]. [cited 2011 12 22 [[Αναφορά στο URI]]. Available from: http://en.wikipedia.org/wiki/Uniform_resource_identifier.
135. Wikipedia. [Online]. [cited 2011 12 22 [[Αναφορά στην SGML]]. Available from: http://en.wikipedia.org/wiki/Standard_Generalized_Markup_Language.

136. Wikipedia. [Online]. [cited 2011 12 22 [[Αναφορά στον Μιχάλη Δερτούζο]]. Available from: http://en.wikipedia.org/wiki/Michael_Dertouzos.
137. openarchives.org. [Online].; 2008 [cited 2011 12 25 [[Αναφορά στο Open Archives Initiative Protocol for Metadata Harvesting]]. Available from: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
138. W3C. [Online]. [cited 2011 12 25 [[Αναφορά στο N-Triples]]. Available from: <http://www.w3.org/2001/sw/RDFCore/ntriples/>.
139. Suchanek FM, Kasneci G, Weikum G. YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. ; 2007.
140. Creative Commons. [Online]. [cited 2012 1 3. Available from: <http://creativecommons.org/licenses/by-sa/3.0/>.
141. DBpedia. [Online]. [cited 2012 1 4 [[Αναφορά στα Συνδεδεμένα δεδομένα]]. Available from: <http://wiki.dbpedia.org/Interlinking>.
142. Wikipedia. [Online]. [cited 2012 1 7 [[Αναφορά στη GNU Free Documentation License]]. Available from: http://en.wikipedia.org/wiki/GNU_Free_Documentation_License.
143. Wikipedia. [Online]. [cited 2012 1 9 [[Αναφορά στα Συστήματα Διαχείρισης Περιεχομένου]]. Available from: http://en.wikipedia.org/wiki/Content_management_systems.
144. W3C. [Online]. [cited 2012 1 10 [[Πληροφορίες για την Xpath 2.0]]. Available from: <http://www.w3.org/TR/xpath20/>.
145. cs.umd.edu. [Online]. [cited 2012 1 11 [[Αναφορά στη SHOE]]. Available from: <http://www.cs.umd.edu/projects/plus/SHOE/>.
146. Fensel D, van Harmelen F, Horrocks I, McGuinness D, Patel-Schneider P. OIL: An Ontology Infrastructure for the Semantic Web. ; 2001.
147. w3.org. [Online]. [cited 2012 1 11 [[Αναφορά στους Κανόνες Συνδεδεμένων Δεδομένων]]. Available from: <http://www.w3.org/DesignIssues/LinkedData>.
148. creativecommons.org. [Online]. [cited 2012 1 11 [[Αναφορά στην άδεια Creative Commons]]. Available from: <http://creativecommons.org/choose/>.
149. Open Definition. [Online]. [cited 2012 1 11 [[Αναφορά στην άδεια TALIS]]. Available from: <http://opendefinition.org/licenses/tcl/>.

150. Wikipedia. [Online]. [cited 2012 1 25 [[Αναφορά στο Situation Awareness]]. Available from: http://en.wikipedia.org/wiki/Situation_awareness.
151. MARC Standards. [Online]. [cited 2012 1 29 [[Αναφορά στο Πρωτόκολλο MARC]]. Available from: <http://www.loc.gov/marc/faq.html>.
152. semanticweb.org. [Online]. [cited 2012 2 5 [[Αναφορά στο SPARQL endpoint]]. Available from: http://semanticweb.org/wiki/SPARQL_endpoint.
153. plato.stanford.edu. [Online]. [cited 2012 2 5 [[Αναφορά στη Μηχανή Turing]]. Available from: <http://plato.stanford.edu/entries/turing-machine/>.
154. Wikipedia. [Online]. [cited 2012 2 7 [[Αναφορά στην άρνηση αποτυχίας]]. Available from: http://en.wikipedia.org/wiki/Negation_as_failure.
155. Wikipedia. [Online]. [cited 2012 2 7 [[Αναφορά στην ισχυρή άρνηση]]. Available from: http://en.wikipedia.org/wiki/Stable_model_semantics#Strong_negation.
156. Goranko V. maths.wits.ac.za. [Online]. [cited 2012 2 6 [[Αναφορά στο Skolemization]]. Available from: <http://www.maths.wits.ac.za/~goranko/FMAI2008/FirstOrderLogicSkolemizationClausalFormUnificationResolutionSld.pdf>.
157. Βικιεπιστήμιο. [Online]. [cited 2012 3 14 [[Αναφορά στο Λογικό Προγραμματισμό]]. Available from: http://el.wikiversity.org/wiki/Λογικός_Προγραμματισμός.
158. Wikipedia. [Online]. [cited 2012 2 4 [[Αναφορά στο ASP]]. Available from: http://en.wikipedia.org/wiki/Application_service_provider.
159. Βικιπαίδεια. [Online]. [cited 2012 2 7 [[Αναφορά στη στήλη της Ροζέττας]]. Available from: http://el.wikipedia.org/wiki/Στήλη_της_Ροζέττας.
160. ypeka.gr. [Online]. [cited 2012 2 8 [[Αναφορά στη ΕΥΓΕΠ]]. Available from: <http://www.ypeka.gr/Default.aspx?tabid=649&language=en-US>.
161. Wikipedia. [Online]. [cited 2012 2 2 [[Αναφορά στο Modus ponens]]. Available from: http://en.wikipedia.org/wiki/Modus_ponens.
162. Wikipedia. [Online]. [cited 2012 1 21 [[Αναφορά στο Modus tollens]]. Available from: http://en.wikipedia.org/wiki/Modus_tollens.
163. W3C RIF-WG Wiki. [Online]. [cited 2012 2 1 [[Αναφορά στη λογική Horn]].

Available from: http://www.w3.org/2005/rules/wg/wiki/Horn_Logic.

164. Antoniou G. Nonmonotonic Reasoning; 1997.
165. w3.org. [Online]. [cited 2012 3 1 [[Αναφορά στα IRIs]]. Available from: <http://www.w3.org/International/O-URL-and-ident.html>.
166. Βικιπαίδεια. [Online]. [cited 2012 1 6 [[Αναφορά στη Datalog]]. Available from: <http://el.wikipedia.org/wiki/Datalog>.
167. Wikipedia. [Online]. [cited 2012 2 8 [[Αναφορά στη Χρονική Λογική]]. Available from: http://en.wikipedia.org/wiki/Temporal_logic.
168. Βικιπαίδεια. [Online]. [cited 2012 2 5 [[Αναφορά στη Lisp]]. Available from: <http://el.wikipedia.org/wiki/Lisp>.
169. Wikipedia. [Online]. [cited 2012 2 15 [[Αναφορά στην F-λογική]]. Available from: <http://en.wikipedia.org/wiki/F-logic>.
170. diavgeia.gov.gr. [Online]. [cited 2012 3 3 [[Αναφορά στη Διαύγεια]]. Available from: <http://diavgeia.gov.gr>.
171. cyber.law.harvard.edu. [Online]. [cited 2012 3 3 [[Αναφορά στο RSS 2.0]]. Available from: <http://cyber.law.harvard.edu/rss/rss.html>.
172. mpeg.chiariglione.org. [Online]. [cited 2012 3 3 [[Αναφορά στο MPEG7]]. Available from: <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>.
173. cyc.com. [Online]. [cited 2012 3 15 [[Πληροφορίες για την κατακευματισμένη Τεχνητή Νοημοσύνη]]. Available from: http://cyc.com/cyc/cycrandd/areasofrandd_dir/distributedai.